Planetary Data System Archive Preparation Guide (APG)

April 1, 2010
Version 1.4

Jet Propulsion Laboratory
Pasadena, California

JPL D-31224

Planetary Data System Archive Preparation Guide (APG)

April 1, 2010

Version 1.4

Approved:

______________________________ ____________
Edwin Grayzeck Date
PDS Program Manager

______________________________ ____________
Reta Beebe Date
PDS Chief Scientist

______________________________ ____________
William Knopf Date
PDS Program Executive

CHANGE LOG

Revision	Date	Description	Author
Start Draft	2004-06-01	Initial outline	Gordon/Simpson
Draft Release 1	2004-06-15	First Draft	R. Joyner
Draft New Outline	2004-07-09	Outline approved by MC	Working group
0.040712	2004-07-12	Second Draft	R. Joyner
0.040721ras	2004-07-31	Wholesale slash and burn	R. Simpson
0.040820	2004-08-20	Revised per telecon 2004-08-09	R. Simpson
0.040820a	2004-08-30	Small changes to 3.2.2; rewrite remainder of Section 3	R. Simpson
0.040906	2004-09-06	Rewrite Sections 4, 5	R. Simpson
0.040909	2004-09-09	Editorial improvements	R. Simpson
0.040916	2004-09-16	Per telecon 2004-09-15	Simpson, Joyner
0.041111	2004-11-13	Telecons and e-mail editing	R. Beebe MC & WG
0.041203	2004-12-03	Revised CH 1 & 2	R. Beebe & V. Smith
0.050107	2005-01-07	Revised CH 3-5	R. Beebe & L. Huber
0.050120	2005-01-20	Revised Preface & Introduction	R.Beebe & R. Simpson
0.050503	2005-05-03	Revised ICD & SIS relation	R.Beebe & S.Slavney
1.1	2006-29-08	Reconcile with PAG; update for CN/EN & media changes; add checklist, change example labels, various corrections and revision of Appendices	M. A'Hearn R. Beebe M. Gordon
1.2	2008-04-16	Remove mention of ICD, add concept of ADD, emphasize overall content requirements for documentation rather than listing specific document names (Preface paras 3-5; section 3.3.5; section 4.2.1.4; section 5.1 para 1; section 5.2 para 2; Appendix A; Appendix D.2; Appendix F)	Dick Simpson
1.2	2008-05-07	Change "Archive Description Document" to "Archive Description Documentation"	Dick Simpson
1.2	2008-06-04	Corrected typos and similar errors on cover pages; in Preface; Sections 3.2.1.2 (paragraph 3),3.2.2, 3.3.3, 4.1,	Dick Simpson

		4.1.2, and 5.2; captions for Figures 3.2.1.2aand 4.1; and Figure 3.2.1.2c	
1.3	2008-06-04	Revisions to Figure 3.2.1.2.c, correction to text at end of Section 3.2.2, and change of MAC to MAS in various places (for consistency with PAG).	Dick Simpson
1.4	2010-03-04	PREFACE, para 2: Added note that titles may vary	Dick Simpson
1.4	2010-03-04	Section 1.1: Correct pointer to Section 1.5	Dick Simpson
1.4	2010-03-04	Section 1.3, para 2: Minor change in last 3 words.	Dick Simpson
1.4	2010-03-04	Section 1.4: Updated URLs for [2] and [3]	Dick Simpson
1.4	2010-03-04	Section 2.0, next to last para: Changes “DN’s” to “DNs”	Dick Simpson
1.4	2010-03-04	Section 2.0, last para: corrected pointer to Section 1.4	Dick Simpson
1.4	2010-03-04	Section 2.2.2, keyword: Added sentence mentioning LDD.	Dick Simpson
1.4	2010-03-04	Section 2.2.2, value and standard value: adjusted line spacing.	Dick Simpson
1.4	2010-03-04	Page 12: restored figure 2.2.2 and caption	Dick Simpson
1.4	2010-03-04	Section 2.2.3: Reworded first sentence. Corrected reference to Fig. 2.2.2.	Dick Simpson
1.4	2010-03-04	Figures 3.2.1.2{b,c}: Converted to bordered text rather than pasted image to simplify pagination	Dick Simpson
1.4	2010-03-04	Page 20, 2nd full para: Updated to 40 characters maximum	Dick Simpson
1.4	2010-03-04	Page 20, Example directory and file names: Changed from W to E for alignment with current practice.	Dick Simpson
1.4	2010-03-04	Section 3.3.1, line 1: Minor wording change	Dick Simpson
1.4	2010-03-04	Section 3.3.1, next to last para, line 6: inserted “that”	Dick Simpson
1.4	2010-03-04	Section 3.3.3, para 2, line 5: substituted wording “may be”	Dick Simpson
1.4	2010-03-04	Section 3.3.4, para 2: Closed “key words” to “keywords” and inserted long dashes to improve sentence structure	Dick Simpson

1.4	2010-03-04	Section 3.3.5 and elsewhere: removed obsolete change bars	Dick Simpson
1.4	2010-03-04	Section 3.3.5, para 4: Revised wording to read: "…reproduction of such papers in …"	Dick Simpson
1.4	2010-03-04	Section 3.3.5, para 5: Corrected reference to Appendix E in last line	Dick Simpson
1.4	2010-03-04	Section 3.3.5: Appended paragraph on ITAR	Dick Simpson
1.4	2010-03-04	Section 4.2.1.4, para 2: Corrected reference to Appendix E	Dick Simpson
1.4	2010-03-04	Appendix A: Added RSA and UCD	Dick Simpson
1.4	2010-03-04	Appendix B: Removed empty row	Dick Simpson
1.4	2010-03-04	Appendix C: Updated URLs, moved RS to Support Nodes and functions section and rewrote, added UCD in same section, made Pluto a trans-Neptunian body.	Dick Simpson
1.4	2010-03-04	Appendix C, EN: Listed 'tools' URL and added note to contact PDS rep	Dick Simpson
1.4	2010-03-27	Appendix C: Revised org chart	Dick Simpson
1.4	2010-03-04	Appendix D: Updated URLs, spelled out acronyms, removed content-free section titled Calibration Files/Algorithms (D.5), renumbered following sections, added paragraph explaining data set citations.	Dick Simpson
1.4	2010-03-04	Appendix E: Changed heading from "System" to "Software", removed 3.4.3.3.2 (Mission Host Catalog file), renamed Host Instrument to Instrument Host (new 3.4.3.3.2), renumbered following sections, removed references to ESA files, changed index file names to upper case.	Dick Simpson
1.4	2010-03-04	Removed original Appendix F (Mission Archive Checklist), separated CODMAC Levels from Appendix E and made it the new Appendix F.	Dick Simpson
1.4	2010-03-04	Appendix A: Added ITAR	Dick Simpson

1.4	2010-03-27	Section 4.1.1: Removed list of example tools, which either don't exist or would be very hard to find and use.	Dick Simpson Anne Raugh
1.4	2010-03-27	Section 2.2.3: Restored paragraph describing volume series	Dick Simpson Ed Guinness
1.4	2010-04-01	Updated version date, Table of Contents	Dick Simpson

TABLE OF CONTENTS

"EXPECT TO ITERATE TO PERFECTION"

–Tom Stein (Geosciences Node, 2005)

PREFACE

Archives do not have to be perfect; but they do need to be *close enough* that future users, unfamiliar with how the data were acquired, can understand and apply them to new problems. This document defines the components of the archiving process that you will carry out and explains how you — a data or science professional recently assigned the task of archiving instrument data from a newly approved planetary mission — can produce a quality archive that meets the standards of NASA's Planetary Data System (PDS). If you do not fit the "new mission" profile, you may still find helpful information here that can be easily adapted to your situation.

PDS sets the standards for archiving; your mission provides the raw data and will establish the schedule for deliveries to PDS. Along the way, you can get help from both. Before (or soon after) you became involved, PDS identified one of its Discipline Nodes (DN) to lead its archiving effort for your mission and assigned a staff member to work with you and your instrument team. The mission designated one of its own staff to be the Mission Archive Scientist (MAS), providing a single point of contact for PDS. A Data Archiving Working Group (DAWG) will include representatives from each instrument team, PDS, and the mission (titles of groups and individuals may vary, but the positions/functions should exist in each mission). You should not be shy about requesting assistance from any and all of these people; iterating to perfection takes both time and attention, and you can benefit immensely from the experience of others.

In the definition stage of archive production, each instrument team, mission, and its assigned PDS representative(s) will work out an agreement that defines the archiving tasks that each party will perform, along with a schedule. Examples of tasks include writing of documentation, generation of PDS labels, definition and production of ancillary files, product delivery, and archive assembly. This agreement is documented in the Data Management and Archiving Plan (DMAP); it defines the scope of your task and is usually signed by the instrument P.I., the node manager, the mission project manager and the PDS program manager.

During the next stage, each instrument team will produce documents that describe data processing plans and the data products in sufficient detail for a scientist to use them. Your assigned PDS rep can provide templates for these documents as well as examples of completed documents from other missions. Since the names and contents of specific documents vary among missions and agencies we will call them collectively the Archive Description Documentation (ADD) here. You don't actually have to write an ADD; but you do need to provide enough documentation that your archive is usable. The ADD can often be assembled from documents written for other purposes.

The archiving task itself starts with an inventory of what you already have and what your instrument team has agreed to provide in the DMAP. The archival data products may have been clearly defined in a proposal; therefore, your team may already have a data processing pipeline — in concept, if not in fact. You can leverage these to advantage by integrating the archiving task with the primary data flow. The files distributed to your team members for analysis should be the same files that go to the archive, giving you immediate error detection and quality control. This also ensures that your archive is created during the main mission — not left as an inconvenient afterthought when interest, funding, and time are almost exhausted.

When you are ready, the "design" of your archive will be reviewed by a panel of peers. The design can then be converted into a working system; once real archival products have been generated, those will also be peer reviewed. The reviews will include your collection of documentation — is the information that accompanies the data sufficient for the future scientist to be able to both understand and use the data?

Do not underestimate the specificity of the task; iterating to perfection is not easy. But use the help that is available and be proud of the result when you have achieved success.

1.0 INTRODUCTION

Creating an archive of planetary science data that will be useful 10-50 years in the future is a tall order. But careful planning, adoption of models that have been successful in the past, integration of the archiving task with other activities, and conscientious production are within reach of even the most inexperienced archivist. Until now, there has been limited information available to new archivists. This document collects the most important concepts with good examples of current practice and guides the new archivist through the forest of Planetary Data System standards, mission requirements, and general good sense to an archive that is both achievable and of high quality.

1.1 Purpose

The *Archive Preparation Guide* (APG) provides a step-by-step, “cookbook” approach for preparing data for submission to NASA's Planetary Data System and links you to personnel who can assist you [See section 1.5 and Appendix C for a listing of Discipline Nodes (DNs) and associated staff members].

1.2 Audience

The data manager, or archivist, who is responsible for creating and delivering archival data from a planetary science instrument or investigation to the PDS will be the prime user of the APG document. Software professionals designing the archive and operators involved in production, as well as the principal investigator (PI) or team leader may also find this document to be a useful planning tool.

1.3 Scope

The APG focuses specifically on a flight project (mission) context but the principles apply to archiving results from a data analysis program or mission support activities. The approach taken in this document is to assume a typical large mission.

The archiving process can vary significantly from one mission to another for a variety of reasons. Of these, mission complexity or data volume are usually the two aspects of a mission carrying the most significant impact. Early identification of mission specific requirements, in addition to continued communication with your Planetary Data System representative (PDS rep), will allow you to identify adjustments that minimize the effort of designing your data set(s).

The APG is written to be consistent with the current PDS Standards Reference (See section 1.4, reference 1). This document provides examples of a small subset of the allowed possibilities. The examples illustrated are thought to be the most common and adaptable. In fact, we explicitly recommend that you do not exercise all of the options allowed in the PDS Standards Reference, which covers existing data sets as well as currently appropriate formats. Consult with your PDS rep early in your planning phase if you need to deviate from the APG recommendations in order to ensure that you will be able to make optimal use of validation tools and to assure ease of future access to the data.

1.4 Applicable Documents

The following documents may be obtained from the PDS Operator (pds_operator@jpl.nasa.gov), or downloaded from the indicated URL's

[1] Planetary Data System Standards Reference, PL D-7669, Part 2.

http://pds.nasa.gov/documents/sr/index.html

[2] Planetary Science Data Dictionary Document, JPL D-7116

http://pds.nasa.gov/tools/dictionary.shtml

[3] Planetary Data System (PDS) Proposer's Archiving Guide (PAG), JPL D-26359.

http://pds.nasa.gov/tools/proposing.shtml

1.5 External Resources and Contacts

Throughout this document, we repeatedly refer you to your PDS rep. Within the PDS, a distributed archive system, ingestion and validation of the data sets occur at the discipline nodes. The management structure of your mission may designate an individual as the Mission Archive Scientist (MAS) who will have responsibility for coordinating all archiving with the PDS. The MAS should direct you to the appropriate discipline node where you will be assigned an advisor. (If you are developing special products in a Data Analysis Program (DAP) or wishing to archive independently, review Appendix C in this document and pick a collaborating DN based on common research interests and/or areas of expertise. When you contact the DN, explain your situation and they will assign one of their people to work closely with you in creating your archive.) Your advisor should have answers to all of your PDS questions — or know how to find them. The MAS is the person within the mission most concerned with archiving—in particular, with getting high-quality data from each team into the public domain in a timely manner while meeting the schedule established by the mission and NASA headquarters. If you are not getting the response you need from your advisor, consult with your PI and/or contact your MAS to get remedial action.

2.0 OVERVIEW OF THE ARCHIVING PROCESS

Producing an archive generally proceeds in three steps:

- Planning and design
- Development and testing
- Data production, distribution, and maintenance.

Following comments on the structure of the PDS and archiving philosophy, each of these steps is described in separate sections of this document.

The PDS archives and distributes peer-reviewed data to the planetary science community. The structure of the PDS includes multiple discipline nodes, referred to as DNs, which specialize in certain scientific disciplines and/or technical skills. The DNs coordinate with the Mission Archive Scientists to help mission teams and science programs to plan and design individual archives. They assist in data validation, ingestion, and organization, as well as lead peer reviews of incoming data. DNs, their host institutions, and their areas of expertise are listed in Appendix C.

Data archived by the PDS nodes must follow organizational standards and the content must be comprehensible and submitted and stored in formats that the science community will find easy to use. The data format should facilitate machine-assisted searches, supporting correlative science across missions and science disciplines. The data must also meet published standards regarding format, content, and documentation (see section 1.4, references 1 and 2). PDS maintains an on-line archive and, once fully integrated into the PDS, archival data may be retrieved through electronic queries over the internet (transfer on physical media is possible if justified case-by-case). The PDS provides copies of all accepted data sets to the National Space Science Data Center (NSSDC), which serves as both the PDS 'deep archive' and the official distribution point for requests originating outside the NASA planetary research community; however, electronic access to the PDS is open to the public.

2.1 Elements of a Good Archive

A mission archive, as required by NASA NRA’s and AO's, should contain sufficient documentation of the mission, the instrument(s), and calibration procedures necessary for members of the current and future science community to effectively use and, if appropriate, recalibrate the data. An archive includes complete information about the geometry relevant to the observations (*e.g.*, spacecraft position and orientation relative to the target). It also includes catalog files that may be ingested into the PDS database along with the raw data and higher order data products. The specific products for your instrument have been negotiated within the mission and with NASA Headquarters and the individual instrument teams will be required to contribute to an integrated Archive Plan for Science Data (APSD) that identifies those products that each individual instrument is committed to deliver to the PDS, including higher order products, the most useful to the science and mission planning communities. Higher order products are more easily ingested into the PDS because documentation for lower order products from which these products have been derived has already been developed.

Finally, archive planners should design their pipelines to organize and store data in straightforward, widely recognized, non-proprietary formats. In preparation for a quality, streamlined archive, be sure to describe data as they exist in the archive files and include descriptions of algorithms and calibration used to generate the products in the delivery package. Specialized software (whether included in the archive or not) should not be required to display or manipulate the data. Consult with your PDS rep if special formats are needed.

2.2 PDS Concepts

This section introduces key terms and concepts as they are used within the PDS. The internal meanings of several key phrases and/or concepts are presented to facilitate a clear and concise communication between the PDS and the archivist. Appendix B provides more technical definitions.

2.2.1 Data - Logical Building Blocks

In this document, we use the terms 'primary data' or 'science data' to refer to the primary output from instruments making scientific observations or measurements. Results from the processing of instrument output—calibrated data, resampled or gridded data, maps, etc. are also included within the rubric of 'primary data'. Tables of summary output from many observations by several instruments may also be 'primary data'. The supporting material — also called 'ancillary data' — is needed to use and understand the primary data and may include calibrations, geometry information, documents, algorithms, indexes, and files used for cataloging. The individual terms are described in more detail below and the hierarchy is shown in Figure 2.2.1.

- **data object** - Data may be organized into PDS-recognized formats called 'objects', such as TABLE and IMAGE. A file can contain one or more objects, some of which may be nested. For example, a file may contain two TABLE objects, one containing instrument settings and the second containing measurements. By definition, each of these TABLE objects must include one or more COLUMN objects. All data objects have an associated label that describes the file (See 2.2.2).
- **data product** - The smallest locatable unit of data within the PDS contains a data object and its label and, within the data set, is identified by a unique alphanumeric identifier, the PRODUCT_ID.
- **data set** - A data set is a collection of related primary data products and the ancillary data products needed for their understanding and use. The primary data products typically (but not always) are a series of observations from an instrument that have been processed in the same manner. Each data set has a unique DATA_SET_ID and DATA_SET_NAME
- **data set collection** - Related data sets may be grouped into a data set collection. For example, all of the data sets (raw, calibrated, and processed) from a single instrument could form a data set collection.

2.2.2 Labels -- Keyword = value statements

The label is the basic element of the data set that allows data recovery and access. It describes both the content and format of its associated data product. The information in a label is used to catalog each data product so that it can be located directly and uniquely by its storage address or found as part of a search (i.e. Where are all the magnetometer data products collected near Ganymede?). Labels are composed of keyword=value pairs in text format.

- **keyword** - Keywords identify characteristics of data files and objects. For example, they may specify the version of the *PDS Standards Reference* under which the data product was defined, list physical characteristics of the file (*e.g.*, number of records and bytes per record), and identify and describe the objects (object types, file names, data set, *etc*.). Each keyword must be listed in the *Planetary Science Data Dictionary* (PSDD) [2]. If your data set has unique characteristics that require new keywords, your PDS rep can assist you in defining them and getting them included in the PSDD. You also have the option of using a mission-managed 'local data dictionary' (see the Locally-Defined Data Elements section of [1] and consult your PDS rep).

- **value (or keyword-value)** - A keyword-value can be a single value, an ordered sequence of values, or an unordered set of values depending on the established keyword that is being used. In the examples below, keywords are on the left, values on the right.

```
PDS_VERSION_ID          = PDS3
RECORD_TYPE             = FIXED_LENGTH
RECORD_BYTES            = 80
FILE_NAME               = "TESTDATA.TAB"
TARGET_NAME             = {"SATURN","TITAN","MIMAS"}
^IMAGE                  = ("DATAFILE.IMG",13)
```

- **standard value** – Some keywords may take on only a limited number of pre-defined values, known as "standard values". In the examples above, `RECORD_TYPE` is such a keyword; its values are limited to:

```
FIXED_LENGTH
STREAM
VARIABLE_LENGTH
UNDEFINED
```

The current set of PDS keywords and standard values can be found in the PSDD [2] or on-line using the PDS Data Dictionary Lookup tool. If no existing keyword meets your needs, contact your PDS rep.

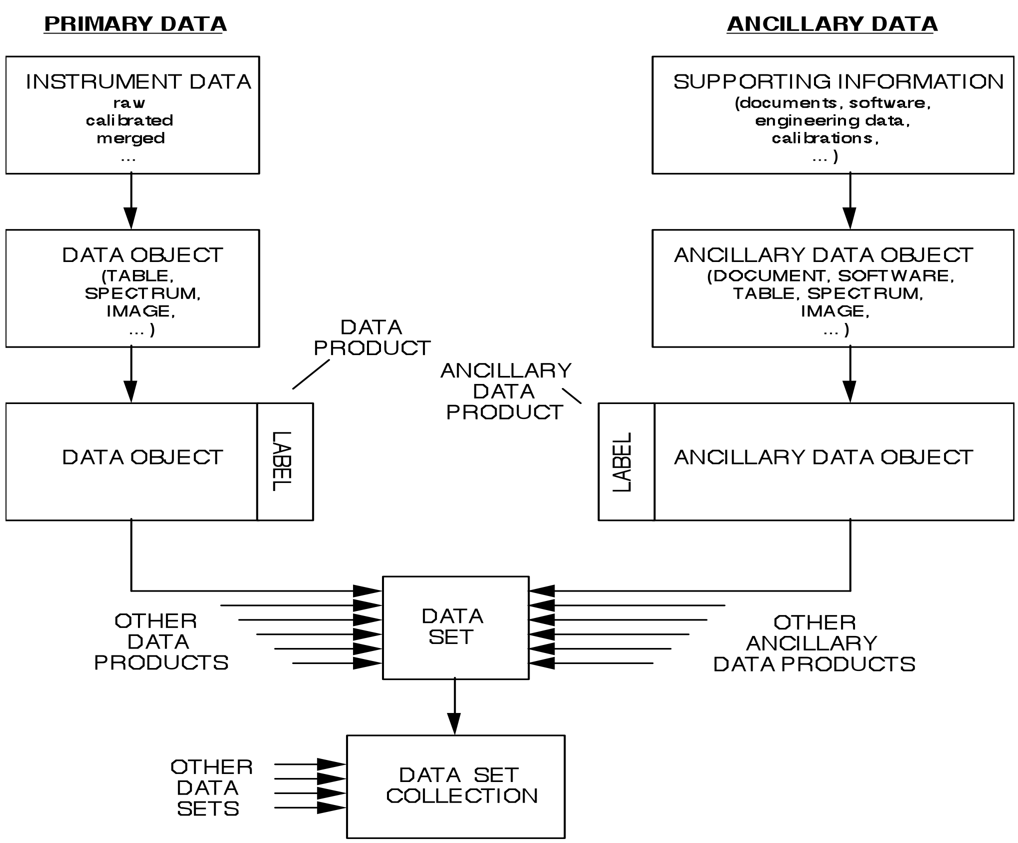


Figure 2.2.1. Hierarchy showing relationships among files, objects, data products, labels, data sets, and data set collections. Primary data originate with an instrument; ancillary data are everything else needed to use or understand the primary data.

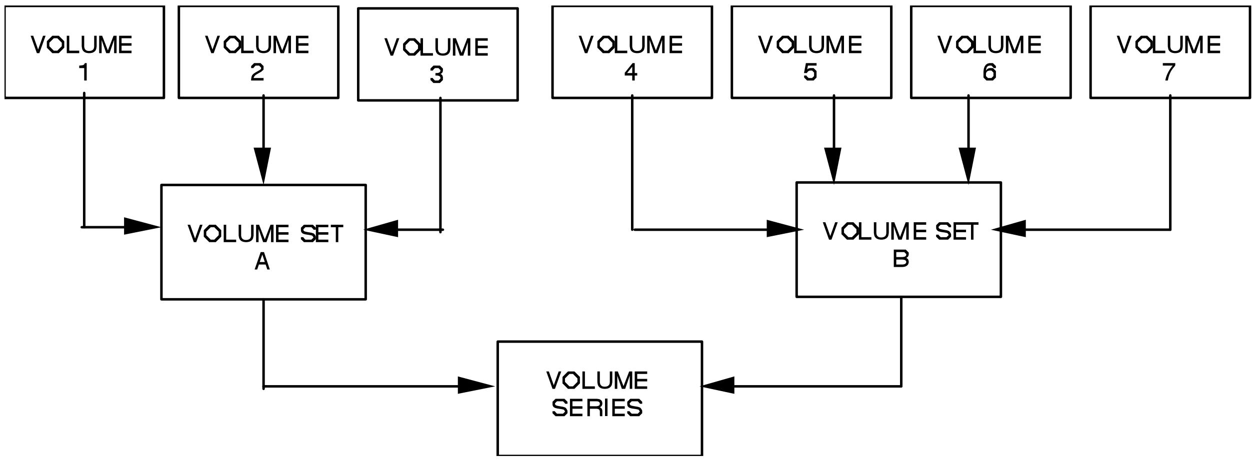


Figure 2.2.2. Relationships among volumes, volume sets, and volume series. Each volume is a unit of physical or logical storage within an archive; volume sets are aggregations of volumes, and volume series are aggregations of volume sets.

2.2.3 Data - Units of Storage

Data sets are organized into units of physical media for storage. Three hierarchal units are shown in Figure 2.2.2.

- **volume** - Data sets are stored on physical or logical media within an archive; one physical or logical unit of such storage is called a volume — for example, a single compact disk, magnetic tape, or disk partition. Each volume is identified by a unique VOLUME_ID and VOLUME_NAME. Depending on capacity of the volume relative to size of the data set(s), a single volume may hold several data sets or only part of one data set.

- **volume set** - a collection of related volumes is called a volume set and is identified by a unique VOLUME_SET_ID and VOLUME_SET_NAME. The volumes in a set often come from a single source. For example, the 150 CD-ROM volumes of raw data from the magnetometer on the hypothetical Geronimo mission might be identified in their labels by

```
VOLUME_SET_NAME = "GERONIMO MAGNETOMETER RAW DATA"
VOLUME_SET_ID = "GERMS_0xxx"
```

- **volume series** - a collection of related volume sets is called a volume series and is identified by a unique VOLUME_SERIES_NAME. The sets in a volume series usually come from different sources. For example, all of the Geronimo mission data (raw and reduced from all instruments) could be a volume series; or, all of the data from a single planet (from all missions) could form a series. For example the data label may contain

```
VOLUME_SERIES_NAME = "GERONIMO MISSION"
```

3.0 ARCHIVE PLANNING AND DESIGN

This chapter provides an overview of archive planning and design, which includes:

- Making an inventory of your data holdings (or scoping your future holdings)
- Understanding and applying PDS archiving concepts
- Designing the archive
- Prototyping the archive

Practical examples are provided to illustrate these steps. PDS often allows more than one implementation; in practice, some are favored over others. If you feel that you need to deviate significantly from the examples provided, contact your PDS rep to minimize false steps.

3.1 Inventory Data Holdings, Production Plans, and External Archiving Requirements

Taking an 'inventory' will size your archiving task and identify components of the process that are already in place. To get started, consider the following questions and ask your PDS rep for guidance as you assess the task. These questions will help you formulate queries to be addressed by instrument team members and to determine where to obtain useful input:

- What data have the team proposed to generate (i.e. data types, data rate, times of acquisition during the primary mission, and total data volume)?
    - The original instrument proposal will define the initial intent.
    - Preliminary mission data management plans may have modified it.
    - Work with the P.I. to obtain this information and to identify team members who can help you scope this.
- What documentation is (will be) available?
    - Mission and Spacecraft (Host) documentation is provided by the mission.
    - Frequently the team produces an instrument paper for a specified journal (Obtain the current version of this or seek help from specific team members).
    - Determine when and how the instrument data management plan will be developed.
- What calibration will be available?
    - Depending on the level of heritage of the instrument, PDS compatible files may exist that can be modified.
    - Consultation with your team members will clarify the extent of calibration problems and define associated time delays.
- What are the steps in processing the data and which levels of products are planned for delivery to the PDS?
    - This is defined in the mission and instrument data management plans and may have been specified in the initial Announcement of Opportunity (AO).

- Where does archiving fit within the data production pipeline?
    - Get a sense of the data flow.
    - Become familiar with what will go into your data processing pipeline.
    - Identify where you tap into the pipeline to pick off the archival files. Do NOT invent a separate process for archiving; that wastes time and effort and jeopardizes your quality control. You want the products that go into the archive to be the same products that have been thoroughly examined and tested by your team's scientists.
    - Consult with your PDS rep to determine what validation tools can be implemented to assure your products will be PDS compatible.
- How and when will the data be validated and peer reviewed?
    - Check with the P.I. to identify the Mission Archiving Scientist and ascertain if there is (will be) a mission-imposed schedule.
    - Expect a mission archiving working group to be set up.
    - Determine the schedule for generating required documents.
- Is there anything obviously needed for archiving but missing from the inventory?

All of these questions have been asked before — by both beginners and experienced data producers. Many answers fall directly out of your mission or instrument data management plan if they have been completed, but you may be leading this development. Your PDS rep can help you establish a solid work plan.

3.2 Preliminary Design

A common mistake of first-time archivists is to design the archive solely by reading PDS documentation. Use the experience of your PDS rep at this stage. Someone has probably produced an archive like yours previously; there is no harm in copying a good model. In fact, there may be advantages to a future scientist if your archive looks like others in the PDS holdings; it will be easier to compare data from different times, missions, and instruments. PDS personnel can point you toward examples of previous work.

Once you have a good sense of your own data flow, you need to define the products in PDS terms (objects, products, data sets, *etc.*, as introduced in Chapter 2). The principal activities are:

- Defining the data products
- Defining the data sets
- Developing file and directory naming conventions
- Identifying pipeline production issues
- Defining the data validation process

Some of these activities will be done in parallel; assigning file and directory names as you define data products is often convenient. A single naming convention that works on all of your files can pay handsome dividends. Expect several iterations and use the assistance of your PDS rep. Your team may consider several different file types before they establish the final data pipeline.

Additional description of these five activities is given below.

3.2.1 Define the Data Products

Data products should be defined based on science requirements — in many cases the file structures have been determined during the development of the instrument even before your instrument was selected for the mission. With raw data, the products are usually obvious and are determined by the manner in which the data are read out and transmitted. For partially or fully processed data, there are more choices; you can select parameters that will be more useful (*e.g.*, calibrated radiance) or drop others that seem unimportant (status flags of heaters that only operate during eclipses). If the products have already been defined, you may still have some options regarding file sizes and formats. Note, however, that the more you can use an existing conceptual pipeline to deliver archival products, the simpler and more successful your archiving task will be. Mainstream products will have been well-scrubbed by the time they go to archive, and you will have to implement fewer unique steps to get them there. If you are also designing the primary data pipeline, plan the simplest, most self-explanatory file structures possible (consider accessing the data 20 years later on an alien platform without the team software).

3.2.1.1 Data Product Usage

Typically, data products are defined based on factors such as their utility in supporting science investigations of the team, the probable way in which the data will be accessed, and the expected frequency of access. What does the team need in order to meet its science goals? What processing schemes have been employed in the past — either by your team or others? What types of information are presented at scientific and project meetings? Will any of the results be fed back into future planning? If so, what and how? Is your team centrally located? If not, are there concerns about delivering data to remote sites either because of the volume or poor communications links? Do your data appear in typically small or large blocks? If small, are there logical ways to aggregate several small blocks into a larger product? If large, does it make sense to subdivide the blocks into smaller pieces without affecting the science use?

3.2.1.2 Selecting Objects

Although PDS has defined a set of data formats or objects (see Appendix A in [1] for the complete list), 'tables' and 'images' account for the majority of products. Anticipated use of the science data will often dictate the archiving format — *e.g.*, TABLE, IMAGE, SPECTRUM, etc. For example, many spacecraft imaging data sets will be composed of single image data products, each defined as an IMAGE object. Once you have determined the 'logical' definition of your data products, you need to format them according to PDS Standards.

```
2004-05-12T12:00:00,  43200,"F9",  19.4
2004-05-12T13:00:00,  46800,"F9",  20.1
2004-05-12T14:00:00,  50400,"F9",  22.9
2004-05-12T15:00:01,  54001,"B4",  23.0
2004-05-12T16:00:00,  57600,"F9",  22.5
```

Figure 3.2.1.2a. Example TABLE object.

```
OBJECT                      = TABLE
  INTERCHANGE_FORMAT          = ASCII
  ROWS                        = 5
  COLUMNS                     = 4
  ROW_BYTES                   = 41
  DESCRIPTION                 = “Hourly temperature outside lab.
                                 Column delimited by ASCII
                                 carriage-return line-feed pairs.”

  OBJECT                      = COLUMN
    NAME                        = “UTC”
    COLUMN_NUMBER               = 1
    DATA_TYPE                   = TIME /* Data type TIME requires
                                        PDS format */
    START_BYTE                  = 1
    BYTES                       = 19
    DESCRIPTION                 = “Sample time in PDS standard month-
                                   -day time format with one minute
                                   precision implied.”
  END_OBJECT                  = COLUMN

  OBJECT                      = COLUMN
    NAME                        = TIME IN SECONDS”
    COLUMN_NUMBER               = 2
    DATA_TYPE                   = ASCII INTEGER
    START_BYTE                  = 21
    BYTES                       = 7
    FORMAT                      = “I7”
    UNIT                        = SECOND
    DESCRIPTION                 = “UTC converted to second of day.”
  END_OBJECT                  = COLUMN

  OBJECT                      = COLUMN
    NAME                        = “SENSOR ID”
    COLUMN_NUMBER               = 3
    DATA_TYPE                   = CHARACTER
    START_BYTE                  = 30
    BYTES                       = 2
```

```
    FORMAT                     = "A2"
    DESCRIPTION                = "Thermometer identifier. First
                                  character gives location: B, F, S =
                                  back, front, side door.
                                  Second indicates model number:
                                    4 = Temp-o-dat model 451;
                                    9 = Ice9 Fridge Sensor"
  END_OBJECT                 = COLUMN

  OBJECT                     = COLUMN
    NAME                       = TEMPERATURE
    COLUMN_NUMBER              = 4
    DATA_TYPE                  = ASCII_REAL
    START_BYTE                 = 34
    BYTES                      = 6
    FORMAT                     = "F6.1"
    UNIT                       = CELSIUS"
    DESCRIPTION                = "Temperature outside the lab."
  END_OBJECT                 = COLUMN

 END_OBJECT                = TABLE
```

Figure 3.2.1.2b. Example TABLE object definition using nested COLUMN objects.

```
OBJECT                           = IMAGE
   LINES                          = 244
   LINE_SAMPLES                   = 537
   SAMPLE_BITS                    = 32
   SAMPLE_TYPE                    = IEEE_REAL
   OFFSET                         = 0
   SCALING_FACTOR                 = 0.5
   LINE_DISPLAY_DIRECTION         = UP
   SAMPLE_DISPLAY_DIRECTION       = RIGHT
 END_OBJECT                       = IMAGE
```

Figure 3.2.1.2c. Example IMAGE object definition

TABLE is a natural choice for numerical data that can be easily visualized in rows and columns, the SPECTRUM being a special case of the TABLE that is defined for spectral data. The PDS TABLE object accommodates both binary and ASCII tables; the latter is encouraged for all but the most massive files. The SPREADSHEET object is similar to TABLE except that it uses FIELD objects with maximum byte counts instead of COLUMN's and may provide more efficient storage when fields are highly irregular and/or many values are missing. TABLE, SPECTRUM, and SPREADSHEET objects can be read into many database applications. It is recommended that you discuss data formats with your PDS rep in the initial steps of archive planning to assist you in choosing the best object for your data.

An example of a TABLE is given in Figures 3.2.1.2a,b -- Note: This TABLE contains the four typical data types (time, integer, character and real) that appear in ASCII tables. The data object (which we assume is a single, separate file) is given in Figure 3.2.1.2a; the object definition, which will be part of its detached label, follows in Figure 3.2.1.2b. Note that the value of ROW_BYTES includes delimiters but that the values of START_BYTE and BYTES do not and that the double quotes in Column 3 are considered to be the delimiters of a character string.

The IMAGE object is flexible in the sense that both ASCII and binary samples (pixels) can be accommodated. The image is specified by keywords that give the number of lines and samples and the properties of the latter. It is possible to assign part of the beginning and/or end of each line for storage of non-image data. Figure 3.2.1.2c shows an example of an IMAGE object definition.

3.2.1.3 Estimate File Sizes and Data Flow

Once data products have been defined, estimates of the file sizes and number of files need to be made. This is so you can predict the rate of data flowing into the archive, develop delivery modes and determine schedules. Estimate the number of archival products of each type that will be generated by your team within some time interval, multiply by the size of the respective products, and sum the results. Delivering data on CDs is not going to work if you plan to deliver terabytes of data per year. Electronic delivery, on the other hand, must be designed to be robust and allow for bandwidth between your institution and the relevant node of PDS. This will usually require consultation with your PDS rep and may require higher level discussions, e.g. between your PI and the Discipline Node manager.

It is common within missions to stage data deliveries at specific time intervals (quarterly for example). That is, you design your volume structure and validation approaches to deliver PDS compatible products on a schedule to which the mission and PDS have agreed. This schedule is usually specified early in mission planning in response to requirements levied by NASA Headquarters that non-mission scientists have timely access to the data.

3.2.2 Define Data Sets and Data Set Collections

The data set is a logical grouping of data products. Defining the scope of a data set is often done before, or concurrently with, defining the individual data products. All of the raw data from your instrument and the necessary supporting information could be one data set; the calibrated data could form the basis for a second data set, and processed data a third. Observation type, discipline, target, or time may also be used to discriminate among data sets.

If your mission includes several targets in succession — for example, two asteroid flybys followed by a comet rendezvous — you might want to define separate data sets for each body. If only a limited amount of data were collected at each asteroid, you might choose to combine the raw, calibrated, and processed data into a single data set for each asteroid. On the other hand, if the orbit around the comet were to last for several years, you might find it easier to manage those data by separating the files into data sets corresponding to mission phases (or years). You could have both raw and calibrated data sets for each of the primary and extended phases at the comet. Or, to facilitate future use, you could group all of the reduced data into a single data set. It is generally a bad idea to try to subdivide the data too finely. In other words, try to limit yourself to a small number of data sets with a large amount of data in each set to simplify things for an end user who needs to cross boundaries between data sets.

Data sets may be combined into data set collections. All of the data sets from one instrument from a given mission could be considered a data set collection.

Definition of both data sets and data set collections is, in many respects, arbitrary but should be logical. In a mission context, data sets and data set collections are often defined jointly by the group of all data producers, and in consultation with PDS reps. Together you select mission and instrument acronyms and build keyword-values for DATA_SET_ID and DATA_SET_NAME. For example

```
DATA_SET_ID      = "MER1-M-APXS-2-EDR-OPS-V1.0"
DATA_SET_NAME    = "MER 1 MARS ALPHA PARTICLE X-RAY SPECTROMETER
                    2 EDR OPS V1.0"
```

where MER1 is the instrument host, M signifies that the data set was collected at Mars, and APXS is the instrument. See the "Data Set/Data Set Collection Contents and Naming" chapter in [1] for details.

3.2.3 Develop Naming Conventions and Algorithms for Directories and Files

The directory should be self-explanatory (see the Directory Types and Naming chapter and the Volume Organization and Naming chapter in [1] for details). PDS recommends that directories contain no more than 256 files or subdirectories —approximately the number that would fit conveniently onto a single computer screen. If more exist, then adding a new level of subdirectories is preferred — up to a limit of eight total directories in depth. By using subdirectory names efficiently in conjunction with file names, both can be kept short.

Directory and file names should be intuitive — they should allow the user to navigate from one part of the data set to another. And, the naming convention adopted should result in names that uniquely identify the directories and files within an easily recognized pattern that can be used by both humans and machines. For example, if T810111.DAT contains data from January 11, 1981, then use of the wildcard T8101* selects all of the data from January.

It may be convenient to organize data (especially raw data) chronologically — i.e. directories named by year and month and files named by day, hour, minute, and second (note: Numeric dates in year-month-day order are preferred. In the case of alternative representations, FEB_2004 would appear before JAN_2003 and 02_01_2004.DAT would appear before 02_02_2003.DAT in most computer listings.) Names based on observation sequence numbers would also generally result in a time ordering of directories and/or files. Image or map data could be named efficiently by including target coordinates; spectral data could be organized using names that include wavelength.

The data product is stored under a file name within a directory and each data product must have a unique PRODUCT_ID within the data set in which it is contained. PDS allows a maximum of 40 characters in directory and file names to ensure that all file names are unique. For example, if an imaging sequence had been set up to map north-to-south as the planet rotated, the data might be stored in directories based on longitude bins, but would allow users to extract the files and group them in latitudinal bins. For example:

```
FILE_NAME  = "E340_350_N10_20.IMG"  /*IN DIRECTORY E340_350*/
PRODUCT_ID = "E340_350_N10_20.IMG"

FILE_NAME  = "E350_360_N10_20.IMG"  /*IN DIRECTORY E350_360*/
PRODUCT_ID = "E350_360_N10_20.IMG"
```

Although acceptable, a less desirable and non-unique approach could be to use shorter file names where

```
FILE_NAME  = "N10_20.IMG"  /*IN DIRECTORY E340_350*/
PRODUCT_ID = "E340_350_N10_20.IMG"

FILE_NAME =  "N10_20.IMG"  /*IN DIRECTORY E350_360*/
PRODUCT_ID = "E350_360_N10_20.IMG"
```

See the "File Specification and Naming" and "Directory Types and Naming" chapters in [1] for further information.

3.2.4 Identify Pipeline Production Issues

Whether they use the term or not, instrument teams will need to develop a 'pipeline' for handling mission data. The pipeline begins with data collection (as from a telemetry stream) and ends with generation of standard products, which may then be distributed to team members for analysis, fed back into planning for future observations, or both. In fact, 'standard products' may be data at any level of processing starting from 'scrubbed' telemetry (after transmission artifacts have been removed) to fully calibrated and registered products.

As the archivist, you should take advantage of this pipeline and assist in defining the level of human monitoring that will be incorporated with automation to handle the data flow and ensure quality control. Except for a few ancillary documents, the pipeline will provide most of the products you will need for your archive in the same (or very nearly the same) form that your team will see them; thus 'bugs' in the archive will be detected and reported by team members in the course of their regular science analysis.

Modifying the pipeline to include archiving steps *before production begins* can enable efficient archiving. For example, inclusion of label generation would force you to define objects and adopt a logical file-naming scheme that would be used by all team members. Properly constructed and populated with significant information, such as pointing geometry, the labels would automatically provide your team members with much of the identifying information they will need later when preparing publications or presentations and provide you additional validation checks of your archive.

As you integrate your archiving plan with the team's data pipeline, develop a checklist to determine whether you have optimized your plan. The questions originally posed in Section 3.1, or those listed below, which cover some of the same ground, may be helpful.

- How and where do you define data products?
- How and where do you create labels?
- What are the appropriate levels of processing for data going to the archive?
- How do you aggregate the data into data sets and archival volumes?
- How much data will be archived?
- What types of calibration files and algorithms are needed?
- What delivery time(s) are appropriate?

3.2.5 Identify Data Validation Issues

Even if you fully exploit the pipeline, there will be a few validation steps that will be unique to the archive. Only in unusual circumstances will your team members see the actual archive volumes, for example; in most circumstances, the data products will be distributed immediately for team use while creation of the volumes will lag behind. Although team members may be the most active long-term users of the archive, experience tells us that they do not have the time to provide formal validation during the mission.

PDS has tools that can assist you in data validation. (Note: Consult your PDS rep to obtain the latest versions of validation tools and for assistance in effective use of them.) For example, there may be a tool that scans each archive volume to check that you have all of the expected pieces in place. Other tools may provide convenient ways for you to check that individual products meet PDS Standards while you are debugging the software that creates them. After the initial volumes have been validated, you will need to validate the pipeline software after each software upgrade and make occasional spot checks during production.

The mission data management plan will define specific documents your team must generate and schedules you must meet. You will be asked to demonstrate in a review that you have developed a reasonable archive design; you will also be asked to circulate example products and volumes for review prior to production. Depending on the convening authority, your review will be conducted by groups that include others in the mission, external science reviewers, and PDS staff. Criteria will include adequacy of the archive in terms of content and documentation and adherence to PDS Standards. Once production begins, there should be a thorough validation of the first few volumes by external reviewers, then occasional spot checks to ensure that errors have not crept into the process.

Validation, testing, and review are discussed in more detail in Chapter 4.

3.3 Prototyping

Although prototyping may have seemed implicit in our discussions to this point, we list the task explicitly here to emphasize its importance. The chances of introducing random errors (spelling mistakes, missing elements, etc.) go down sharply and efficiency and ability to deliver on schedule goes up as you increase the amount of automation. Careful prototyping shows what each step in the pipeline must do and helps you identify the issues at the interfaces between steps and optimize the process.

We identify the key steps in prototyping here, and then expand on each in the following subsections. In most cases you will want software to generate files automatically, so we emphasize that aspect of prototyping. You can also assemble products by hand—simply to see what they look like.

- Primary Data
  - Instrument data objects
  - Labels

- Ancillary Data (and their labels)
  - Calibration data (including geometry data)
  - Catalog files
  - Index files
  - Documentation, software, etc

3.3.1 Primary Data

At this point, you should have identified the data objects upon which you will build your data products; if you have several different primary products, you need an object for each. The data files themselves should have PDS compatible formats. If not, you need to modify your pipeline software so that the data products you generate meet those format specifications.

Using advice from your PDS rep (and suggested examples from previous archives), you should produce labels for each product type. In addition to the object definitions already roughed out (*e.g.*, Figures 3.2.1.2b and 3.2.1.2c) you will need keywords (and values) to:

- identify the version of PDS Standards you are using (PDS_VERSION_ID)
- identify the data product (*e.g.*, PRODUCT_ID and DATA_SET_ID)
- characterize the file physically (*e.g.*, RECORD_TYPE and RECORD_BYTES)
- describe data object contents (*e.g.*, FILTER_NAME, OFFSET_MODE_ID, ...)
- point to the data object (*e.g.*, pointer ^IMAGE)

The chapter "Data Product Labels" in [1] has more information on label construction.

When you are developing software, try to keep it simple. (Label-making software that requires long lists of command line arguments will be more prone to operator errors.) Write software to find standard keyword-values within the data files or from other readily accessible sources. For instance, observation time tags often accompany even the most primitive raw data files. Because the time of observation is the most universal keyword in a mission, it is essential that all instruments use the same PDS compatible representation. Some desirable keywords, such as those associated with pointing geometry, could be deferred to a later step in processing after you have reconstructed the viewing geometry. When you have constructed your prototype labels, use the label verifier tool on your labels and then, to insure their completeness, run them past your PDS rep, your team members, and anyone else who will look.

In general, PDS encourages you to define a DATA directory within each archive volume – see the Volume Organization and Naming chapter in [1]. Subdirectories under DATA then hold the products themselves. For large data sets, you may wish to order the products chronologically, placing the earliest products on the first volumes and later products on subsequent volumes. Your PDS rep can provide guidance on this organization based on the number of files you expect, their sizes, and the rate at which they are generated.

3.3.2 Calibration Files and Algorithms

All calibration information needed for interpretation of primary data (including data needed to reconstruct the viewing geometry) must be produced, documented, and archived coincident with the science data that they describe. The calibration algorithms should be presented in a human readable form that would allow a future user to write software in a preferred language for a specified platform. Calibration software may be included in the archive, but such software is considered 'supplemental' and is not acceptable as a substitute for the required calibration algorithm. PDS does not commit to support the software.

It is particularly important that you address calibration data while prototyping the main archive (or earlier!) since some instruments have significant volumes of pre-launch calibration data. The suite of calibration data and algorithms will allow recovery of the full scientific value of the returned data and facilitate the correlation of data taken at different times, by different instruments, and by other spacecraft, observing platforms, and missions.

It is common to store calibration data within a CALIB directory on each archive volume, repeating the files on each volume if they represent only a small fraction of the volume capacity. Large volumes of pre-launch calibration data may be handled more conveniently by using one or more separate volumes. Your PDS rep can provide guidance on the appropriate way to archive calibration data.

If your mission (and your instrument) uses the SPICE system developed and supported by the PDS Navigation and Ancillary Information Facility (NAIF), you may opt to join in a common archive of SPICE files produced by NAIF. The NAIF archive has the advantage of including all SPICE software and documentation as well as the data files. Note, however, that the NAIF archive may be dynamic—as spacecraft trajectory and attitude reconstructions are improved, the set of files in the NAIF collection may be updated. Your archive needs to document the steps followed by *your* team that, at some point, may include archaic versions of these files. If your calibrated and/or reduced product labels point toward the SPICE files actually used and these can be found in the common archive, the requirements for archiving geometry information may be met. If the files your team used are not in the common archive, you may need to include a team-specific SPICE collection in your own archive. You may also choose to archive SPICE files simply for the convenience of having them locally. Check with your PDS rep or your MAS, who should be up-to-date on plans for the common SPICE archive, for guidance if you are uncertain about the geometry archiving requirements in your case. Note that there is a small set of SPICE-specific keywords that may be useful in constructing labels for SPICE products.

SPICE and other geometry data are usually stored in the GEOMETRY directory on archive volumes; this directory may be subdivided into other directories for spacecraft and planetary ephemerides, attitude files, clock conversions, etc. As with the CALIB directory, it is convenient if the geometry files needed to analyze primary data are included within the same volume; this may require duplication of files from volume to volume.

The calibration and geometry data must be formatted as PDS objects—just as primary data are—and each object must be part of a data product described by a label. Some calibration and geometry data naturally fit within the definition of TABLE, IMAGE, and other objects we have mentioned in connection with primary data. But you will also find a need for the DOCUMENT, TEXT, and (possibly) SOFTWARE objects. See Appendix A in [1] for further information on these objects.

3.3.3 Catalog Files

All data submitted to PDS must be accompanied by a set of catalog files which briefly describe the mission, instrument host (that is, the spacecraft or other facility within which the instrument operates), instrument, and data set. A fifth catalog file identifies key personnel associated with the instrument, data set, and archiving task. A sixth file lists references cited in the first four files or which might otherwise be useful to a scientist working with the data at some future time. These six files usually appear in the CATALOG directory of each archive volume. Some archives include additional catalog files, such as for targets of the instrument; these are usually considered optional but may be encouraged or required by some missions or by some disciplines.

Catalog files are written in a structured format so as to be easily read by humans. Your responsibility (at the instrument level) is limited to INST.CAT, DATASET.CAT, and PERSON.CAT. MISSION.CAT and INSTHOST.CAT should be written and maintained by the flight project; you only need to download the current versions and add them to each of your archive volumes. A baseline REF.CAT may be drafted by the project—including the references cited in MISSION.CAT and INSTHOST.CAT; you then need to add the references you cite in INST.CAT and DATASET.CAT. Some projects, in conjunction with PDS, maintain a full REF.CAT—to which you contribute; this has the advantage that all references have been checked against and entered into the PDS database before you deliver any volumes. Otherwise, you run the risk of submitting REF.CAT entries that conflict with entries already in the database. This most often occurs when you draft a REF.CAT entry with a REFERENCE_KEY_ID (see Appendix B in [1]) that has already been used. A less common problem, and one that is more difficult to detect, is when two or more archivists submit REF.CAT entries for the same paper but with different REFERENCE_KEY_ID values.

PDS maintains a set of catalog file templates online. These provide the outline structure expected. Or, with the help of your PDS rep, you may find it convenient to use an existing catalog file as a model.

Most catalog files evolve over a mission. For example, the flight project will update the MISSION.CAT file as new mission phases are defined. MISSION.CAT also documents important mission events, which may or may not have been predicted. A series of 'safing' incidents could affect data collection; the dates when these occurred will be of interest to future users of the data sets. Your own DATASET.CAT file has a section on "Data Coverage and Quality" which gives a very coarse summary of where and when you collected data and whether anomalous conditions existed. Some of this may be available in the SPICE E kernel that the mission is maintaining.

Rather than submit complete catalog files with your initial data delivery, you may choose to submit a set of partially completed skeleton files. Once the data become available to users through PDS, complete files should be entered in the PDS database. The evolutionary nature of these files argues for some flexibility in their completeness during the time when data are still being acquired. Rather than let this impede your progress, contact your PDS rep and mission archive coordinator for guidelines.

A seventh file (VOLDESC.CAT) appears in the root of each archive volume (see the Volume Organization and Naming chapter in [1]). Although technically a 'catalog' file, we have deliberately omitted it from the discussion here since its construction and use is somewhat different.

3.3.4 Index Files

Index files tell users where to find data products. There is an INDEX.TAB file in each volume. When there is more than one volume, there is a CUMINDEX.TAB for all volumes in the volume set, which is the result of appending each new INDEX.TAB to the previous CUMINDEX.TAB.

The mission goal will be to define keywords that are used consistently across the mission and incorporated into the appropriate instrument indices so that a master index will allow correlative studies spanning individual data sets. In each INDEX.TAB there is a separate entry for each product giving the VOLUME_ID, the path to the product, the PRODUCT_ID, the DATA_SET_ID, and the PRODUCT_CREATION_TIME. Inclusion of other keywords — for instance, associated latitude and longitude ranges or start and stop times — with each product will help team members and future users in searching for related files. You will need to include any keywords here that your own science team will find useful in searching for data. If the right information is included, CUMINDEX.TAB from the final volume in a volume set becomes an obvious place to search for specific data files.

PDS requires that all primary data objects be listed in INDEX.TAB (and CUMINDEX.TAB) and recommends that a wide variety of ancillary data products also be listed. The extent to which data producers include ancillary data is discipline dependent. A good rule-of-thumb is to include all data products that might be used in numerical calculations; this leaves things like DOCUMENT and SOFTWARE objects out of INDEX.TAB. For guidance on current practices within your discipline, consult your PDS rep.

Note that the INDEX_TABLE is a distinct PDS object; certain COLUMN objects and keywords are required. Each index file must have a label in which the INDEX_TABLE object is defined (see Appendix A in [1]).

3.3.5 Documentation

Compiling the Archive Description Documentation (ADD) is second in importance only to assembling the data files themselves. The ADD does not have to be a physical document; but its logical content must be sufficient that the archival data are useful.

The ADD can often be assembled from existing documents; the archivist need only provide a short description of what is available and where it can be found. The component documents are expected to meet not only the PDS label and format requirements, but also the structural, grammatical and lexical requirements of a refereed journal submission. Incomplete or poorly organized or edited documents submitted for archiving will require time-consuming revisions and may ultimately lead to rejection of the accompanying data. Your PDS rep can help you optimize your efforts by providing appropriate examples and early reviews and advice.

Documentation provided with the data set should be sufficient so that a future scientist could understand how the data were collected and interpret them. Many observational schemes and sequences are designed to fulfill science goals developed by the team. Preservation of this information will enhance the value of the data set. The role of the ancillary data should also be clear. If calibration data are included, a detailed algorithm describing the calibration procedure is required. If higher order products have been produced, the steps in processing need to be described.

In the mission context, it is common for science teams to write papers explaining their goals and the measurements that will be carried out to achieve those. Copyright protection may preclude reproduction of such papers in the archive, but some publishers have made exceptions. Check with your PDS rep regarding possibilities for gaining permission to include such documents.

The format of the data included in the archive must be described clearly. Full PDS labels should provide this information, but Software Interface Specifications (SISs) may additionally be required by the mission. There are two types of SISs: the Data Product SIS and the Archive Volume SIS. The Data Product SIS describes the individual products and their derivations and the Archive Volume SIS describes the contents and structure of an archive volume, including documentation, software, and other materials in addition to data products. In some cases, if it is more convenient, the Archive Volume SIS is combined with the Data Product SIS. Your assigned PDS rep can provide templates for all three documents as well as examples of completed documents from other missions. Appendix E provides a possible outline for a SIS document.

Documents usually appear in the DOCUMENT directory of a volume. Simple text files can appear as TEXT objects with attached labels; these are among the simplest products in an archive. Because ASCII text is expected to be viable over a longer time span than any of the specialized formats, a human-readable ASCII core file is *required* in all cases. More complicated documents can be built around the ASCII text core file, supporting files with graphics, and a label that wraps all of the components into a single DOCUMENT product. Because they may be more readable in the short term, inclusion of formats such as Microsoft Word and Adobe PDF is also encouraged.

ITAR Compliance: Many aspects of spacecraft design, construction, and operation are subject to International Trafficking in Arms Regulations (ITAR). Since PDS is an open, web-based system, all data (primary and ancillary) available through PDS must be free of ITAR restrictions. Past experience suggests that documentation written in 'marketing' terms — why you would want one of these, rather than how to build and operate one —

is most likely to pass ITAR review. The ITAR situation is fluid, so you should stay in close contact with your PDS rep while designing your archive — and especially while preparing documentation, including calibration reports. It is the responsibility of data providers (not PDS) to ensure ITAR compliance of submissions.

3.4 Archive Planning and Design Checklist

The following checklist is provided as a summary of the steps outlined in the Archive Planning / Design section. The checklist can be used as a means of ensuring the steps towards creating a successful archive are complete.

	SECTION	ACTIVITY
☐	3.1	Inventory - primary data - what are your data? - what are your processing steps? - what is appropriate to archive?
☐	3.1	Inventory - ancillary data - what are your calibrations? - what other supporting data are available/desirable for archive? - what documentation is (or should be) available? - should software be included; if so, what?
☐	3.1	Inventory - external requirements - what does your Data Management Archiving Plan (DMAP) say about archiving? - what does your mission require? - what does your team want?
☐	3.1	Inventory - resources and contacts - who is your Planetary Data System representative (PDS rep)? - who is your Mission Archive Scientist (MAS) or data archiving contact? - who, outside your team, provides data? - who, within your team, will be generating archival products (if any)?
☐	3.2	Define data products - associate a PDS object with each data type - adjust formats to be consistent with PDS object requirements - draft object definitions
☐	3.2.1.3	Estimate data flow - estimate size of each product - estimate production rate - estimate total data flow
☐	3.2.2	Define data sets - which products go into which data sets - organize products into directories - establish DATA_SET_ID and DATA_SET_NAME for each
☐	3.2.2	Define data set collections (optional) - which data sets group into which collections? - establish DATA_SET_COLLECTION_ID, DATA_SET_COLLECTION_NAME
☐	3.2.3	Develop naming conventions - file names - PRODUCT_ID's - directory names

	SECTION	ACTIVITY
☐	3.2.4	Sketch pipeline showing - where archival products appear - where labels are generated - aggregation into directories, data sets, and volumes
☐	3.2.5	Incorporate validation into pipeline - what validation is needed? - who does it and when?
☐	3.3	Prototype - primary data and labels - ancillary data and labels - index files and labels - pipeline - software for above - documentation - catalog files

4.0 DEVELOPMENT AND TESTING

'Development and testing' includes adaptation of your team's data processing pipeline to generate PDS-compliant archival data products and volumes. After you have designed and prototyped your products, constructing and validating a sample volume will ensure that the system you set up to produce your data volumes automatically will be PDS compliant and acceptable without revision.

4.1 Development

The main elements of a typical data processing pipeline are shown in Figure 4.1. The data are collected, inspected, cleaned up, and stored in the first stage; calibrated in the second; and processed to standard products in the third. If the flow is designed to support archiving, raw primary and ancillary data can be formatted as PDS objects, named, and labeled before entering the team database. As derivative files are created later in the pipeline, they are also named and labeled to PDS standards. The archiving step then becomes little more than bookkeeping—staging the new files until enough have been collected to assemble a new volume, creating and validating the volume, and delivering it to PDS.

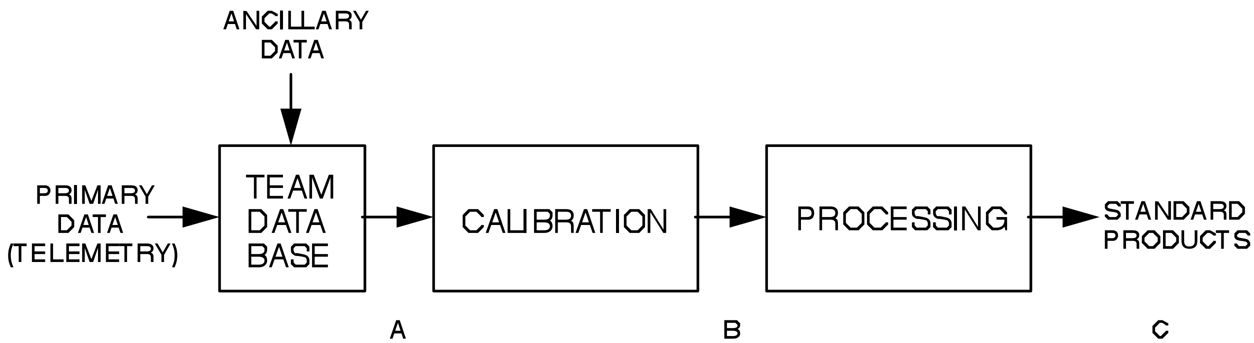


Figure 4.1. Simplified pipeline for mission processing. Data from the instrument and other sources are stored (with labels) in the team database, then used as input to a calibration stage. Later generic 'processing' yields standard data products. Files for the archive can be obtained at points A, B, and C; in NASA terminology, these correspond generally to data of Levels 0, 1, and 2 (see Appendix F).

Since PDS and future users of the data need to understand the steps leading to the products at steps A, B, and C (Figure 4.1), we encourage you to:

- keep your PDS rep in the loop on what the pipeline will do
- keep your PDS rep in the loop on what the pipeline will not do
- ensure that documentation is updated in parallel with changes to the pipeline and its software.

4.1.1 Label Generation

Your PDS rep can offer suggestions on automating the process for generating labels, including the use of PDS tools and commercial software.

In some cases, the label becomes comparable to or exceeds the length of the product itself. You can simplify the label by abstracting its repeated portions into 'format' (FMT extension) files which are stored only once; the individual labels then reference the FMT file(s) through pointers. Your PDS rep can help with this and other techniques to simplify your work.

Label verifier tools are an important adjunct to automating your label generation; use them frequently to make sure you have not introduced anomalies into your labels. Contact your PDS rep for information about the current tools available.

4.1.2 Volume Production

The "Volume Organization and Naming" chapter of [1] tells you what goes into a volume. Assume for the remainder of this section that you are part of a large mission and have one raw data set spread over many volumes. The following steps are needed to produce a volume of raw data. There may be second and third data sets for calibrated and reduced data, respectively; they can be handled in the same way.

Your database (Figure 4.1) should be set up to allow you to query and determine how much data have accumulated since you made the last volume. Efforts should be made to generate each volume when the amount of accumulated data fills a unit of your selected medium.

In addition to the new primary data files from your instrument and new ancillary files of geometry and other data, there will be a set of 'static' files that go into every volume—for example, your CALIB, DOCUMENT, and SOFTWARE directories. These are files that do not change—or change so infrequently that you can imagine hand-editing them when updates are needed. You should keep these in a convenient, but safe, location from which they can be copied. Keep a master file or devise a routine to generate skeleton volumes.

PDS also requires the CATALOG directory and its files. Check to see if you need to update the "Data Coverage and Quality" section in DATASET.CAT; otherwise, CATALOG should be essentially static across many volumes.

AAREADME.TXT, ERRATA.TXT, and VOLDESC.CAT are files in the root of each volume; they generally need small updates for each volume. You will probably be making continual updates to ERRATA.TXT as you discover minor errors and 'features' in the data set; these updates occur asynchronously with volume production. Users will be made aware that the most recent volume has the most recent ERRATA.TXT. The other two files will be customized for each new volume.

INDEX.TAB and its label need to be created for each new volume; the new INDEX.TAB is then appended to the previous CUMINDEX.TAB, and CUMINDEX.LBL is updated accordingly. Archivists have developed various pieces of software to assist you in generating index files; contact your PDS rep for suggestions and assistance.

4.1.3 Volume Validation

There are two parts to volume validation. First is confirmation that the volume meets archiving standards. This can be done automatically (albeit not perfectly) using tools available from PDS. These check that the volume has all required components, that all files are properly recorded in INDEX.TAB (and CUMINDEX.TAB), etc. They also call other PDS utilities to make sure that labels are correct, that all references cited in other CATALOG files actually appear in REF.CAT, and that line lengths and delimiters in text files meet PDS recommendations. Ask your PDS rep for help; then get the tools installed and running on your system.

The second part of volume validation is confirmation of the scientific integrity of the contents. This is more difficult and requires insight into the data. If your pipeline is designed so that team members are using the same products for their research that you include in the archive, the job is vastly simpler. If, unadvisedly, the path to the archive differs significantly—for example, you generate the products long after the science team has studied other versions—you and your team will have to develop a separate validation process. Scientific integrity of the volume also includes checking that documentation is sufficient to explain and use the data; if your DOCUMENT, SOFTWARE, and CALIB directories are complete for the first volume, all you need do is double check that the same files are appropriate, properly edited and have been incorporated into subsequent volumes. If your instrument or processing changes, you will, of course, need to keep the ancillary files current. Be careful, however, not to lose information important at the beginning of the mission by removing it from later versions of the files—unless you carry forward all versions of the documentation.

If your team has written SISs and/or a Data Management Plan, they may include sections on validation. Check that each step mentioned in each document has been implemented in your archiving pipeline. Make sure that each person with identified responsibilities in validation is ready to perform and understands those steps.

Implementing data validation is often an iterative process. Both the procedures themselves and the software written to perform certain steps will likely require a number of updates as your validation becomes both more comprehensive and also more automated.

4.2 Testing

Testing is done at two levels—the team level, where you assemble example products (based on simulated data, if possible) into sample volumes; and at the mission/PDS level, where those volumes are reviewed for compliance with PDS Standards and adequacy of documentation and other ancillary information. You are *strongly encouraged* to conduct at least the part of the PDS review concerned with archive design *before* you have made significant investments in software development; check with your PDS rep for guidance.

4.2.1 Team Testing

4.2.1.1 Data Product Simulation

Simulation is important because it gives you concrete examples of data products to test data flow and interfaces; your team may already be using simulated data to test its pipeline. Sample volumes of products generated from simulated data may be used to test archive interfaces with your mission and/or PDS. Real data (as from pre-launch bench tests) can also be helpful, but they may not be as robust in exercising processing steps that involve expected observing geometries. Tests with simulated data products may also provide insight into desirable enhancements of products (*e.g.*, additional keywords) that you may have missed in your initial design and prototyping.

4.2.1.2 Pipeline

As early as possible, you should get the data production pipeline into place. Once the pipeline is working, you can input simulated raw data (as from your instrument) and simulated files from other sources and test all of the steps, including generation of sample archive volumes. Check at each step that the pipeline operates as expected on the simulated data.

4.2.1.3 Sample Volumes of Archival Data

The sample volumes should be "complete" in the sense that they are representative of what a final "archive" volume will contain. They should contain examples of each data product using simulated or test data as available. It is important that the structure of the volume be accurately represented so that all automated validation steps can be tested. Once this step has been accomplished, you can focus on the second part of validation—checking the scientific integrity of the volume. This means getting real science data into the primary files and fleshing out the documentation and other ancillary information. If your mission has a "cruise" phase, you may be able to use those data for more realistic testing.

4.2.1.4 Documentation

You now have sample products (possibly with dummy values) stored in sample volumes; your labels should be mature and the core parts of your Archive Description Documentation (if not in final form) should at least be readable and understandable. If you are part of a typical mission, the production schedule for the SISs will be levied by the mission and the SISs will be an important component of your archive design review (see next section).

A common practice is to build the SISs around a set of example labels; each label describes the format and content of its data product; so all you need are bridges between labels to complete the SIS. The outline in Appendix E may look formidable, but you probably have most of the pieces already. Work closely with your PDS rep as you develop the product and volume SISs.

4.2.2 Testing at the Mission and/or PDS Level

Every archival data product (and volume) must meet PDS standards for format and content. Whether that has been accomplished is determined by a peer review committee. There are usually two levels of peer review in a mission context—an evaluation of the archive design, including sample products and volumes; and a review of mission data as volumes are delivered to PDS. Although the objectives are similar in both, the design review is often more thorough. Assuming that liens identified in the design review have been successfully resolved, the delivery review panel only spot checks that volumes continue to meet the design specifications. The remainder of this section is focused on the archive design review.

There are four general areas in which the archive is reviewed:

- Compliance with PDS Standards
- Compliance with the applicable Data Management Archive Plan (DMAP) and Software Interface Specifications (SISs),
- Scientific merit of data (are these the proper data to be archived?)
- Usability of the data (appropriate formats; completeness of data set, including ancillary data; comprehensive documentation)

Steps in the review process include:

- Establishing a review committee
- Conducting the review
- Completing the review (including documenting and resolving liens)

4.2.2.1 Establishing a Peer Review Committee

The mission and PDS will establish the archive design review schedule. Your PDS rep will recruit the review committee. The official reviewers will be planetary scientists with an interest in the data but who are not involved in preparing the archive and who are not connected with the discipline nodes with which you are working. The review panel will also usually include members from your advisory Discipline Node and the data preparer (probably you). Depending on the circumstances of your mission, it may also include the Principal Investigator (PI), one or more data suppliers (*e.g.*, someone from the Spacecraft Team or Ground Data System), and a PDS data engineer. A member from your advisory Discipline Node typically chairs the review committee.

4.2.2.2 Conducting the Review

You will be responsible for providing copies of the design documents (DMAP and SISs) and sample data products or volume(s) (*e.g.*, CD-WO or DVD media or via electronic access) to each member of the review committee. Normally this will be done through the lead Discipline Node for data reviews, but in some cases it may be directly from the mission to the reviewers. For pipeline reviews, it will usually be the mission's responsibility to distribute the data directly to the reviewers. As always, consultation with your PDS rep and/or review chair is essential.

The review committee chair will provide instructions on how the review will be conducted. The instructions may vary depending on the nature of the data set, but your PDS rep should provide adequate guidance to help you prepare. Sometimes you will be asked to walk the committee through the design, products, and volumes during a plenary session, which may be conducted as a face-to-face meeting, teleconference, or both. In other cases, the chair may ask committee members to provide written comments (*e.g.*, e-mail) on the design and sample products in the four areas mentioned above before the panel convenes. In unusual circumstances, the chair may ask the committee members to submit their reviews in advance, and the plenary will be your opportunity to respond to their comments.

Reviewers should evaluate documentation (including labels, calibration files, and CATALOG files): do they explain the data and how to use them, and will they be intelligible to a scientist 10, 20, or 50 years in the future? If your archive includes special software for displaying or manipulating the data, it is expected that committee members will evaluate the software both functionally and as a long-term component of the archive (i.e. how likely is it that the software will work in 20 years).

4.2.2.3 Completing the Review

At the end of the peer review session the chair will summarize the results of the review in a formal memo. The review committee is responsible for making a recommendation on whether:

- The design documents and sample data products/volumes passed peer review and development can proceed without additional review.
- The material did not pass peer review and must undergo an incremental (delta) review in certain specified areas.
- The material did not pass peer review and must undergo another full peer review.
- The material did not pass peer review and does not merit a second peer review.

In addition to an overall recommendation, the chair will compile a list of "liens" — questions about or requests for change in the archive design or sample products/volumes. The liens must be resolved before the archive can be accepted by PDS. Note, however, that you are not obligated to accept all requests for change. You may offer a rebuttal to one or more liens; it is then up to the chair and PDS as to how the lien should be "resolved." In extreme cases, where the peer review committee collectively agrees that a product does not meet archiving requirements and the team is unable or unwilling to resolve the liens, then PDS may elect to “safe store” the data (*i.e.*, the data are not deemed archive quality, but may be preserved and distributed). "Safing" data has been allowed only rarely and only long after the mission has been completed.

When the review committee has approved the design and the sample products/volumes, you should complete your development and testing at the team level and move to the Data Production stage.

4.3 Development and Testing Checklist

The following checklist is provided as a summary of the steps outlined in the Archive Development and Testing section. The checklist can be utilized as a means of ensuring the steps towards creating a successful archive are complete.

	SECTION	ACTIVITY
☐	4.1	Modify pipeline software to generate PDS objects
☐	4.1.1	Write or adapt software to generate PDS labels
☐	4.1.2	Add tracking software (or equivalent functionality) - what new files have been created? - how many have accumulated since the last volume was written?
☐	4.1.2	Write software to assemble archival volume - collect new primary data - add appropriate ancillary data calibration and geometry documentation software - generate new index files AAREADME.TXT ERRATA.TXT VOLDESC.CAT
☐	4.1.3	Develop procedures/tools to validate volumes - compare with PDS Standards - verify scientific integrity
☐	4.2.1.	Test the above steps using simulated (and/or bench test) data
☐	4.2.1.4	Document the above in one or more SISs
☐	4.2.2	Complete archive design review - prepare and distribute review materials - respond to review panel requests, questions, and comments - resolve liens to the satisfaction of the review panel and PDS

5.0 DATA PRODUCTION, DISTRIBUTION, AND MAINTENANCE

This chapter provides an overview of archive production, distribution, and maintenance. Maintenance includes reprocessing, which may be required for several reasons.

5.1 Data Production

Data Production entails running the data production pipeline, validating the archive volumes, and transferring them to PDS. The means of transfer and the schedule should be defined in the DMAP, with concurrence of the mission.

Data Production also includes peer review of products and volumes delivered to PDS. Rather than the comprehensive review of design and sample products/volumes described in the previous chapter, the review of delivered products is usually conducted using automated tools. It checks for format and PDS compliance on many, if not all, volumes. Often the first few volumes are examined very carefully to ensure that they follow the (previously approved) DMAP and SIS, then only spot checks are made to confirm science integrity on succeeding volumes. Only if significant problems are discovered or if the delivered products and volumes deviate from the designs approved in the earlier review is the flow of data into PDS interrupted.

5.2 Data Distribution

Once the data have been received, PDS is responsible for releasing them to the science community and the public. In PDS "data release" means adding data products to an online repository, ensuring that the data are accessible, and then announcing their availability. PDS operates a "subscription manager" service that allows interested scientists to sign up for automatic notifications of incremental data releases of a particular data set.

Data products are normally distributed by PDS through a DN. In special circumstances, distribution may be by a Data Node working in conjunction with a PDS/DN. The organization responsible for distributing each data product is identified in the DMAP. Data are officially distributed to the general public (anyone who is not a NASA-funded planetary scientist) through the National Space Science Data Center (NSSDC); NSSDC receives copies of the data from PDS as described in an MOU between PDS and the NSSDC.

Users may obtain products by downloading files from the archive. In cases of large downloads, users may also request that data be written onto physical media such as DVD.

5.3 Maintenance

Maintenance is the set of processes by which you respond to errors or updates in data products/volumes. These include:

- Reprocessing—if the team determines that there has been an error in previous processing or that refined algorithms will produce better results.
- Withdrawal—if the team determines that there has been an error in processing and that previously released products cannot be repaired.
- Update ERRATA.TXT—if an error of relatively little consequence has been identified and the cost of repair significantly outweighs any benefit, the team may simply note the problem in ERRATA.TXT.
- Update file—Some files can be updated with little impact on the archive as a whole. For example, corrected typographical errors in documentation or updates of "Data Coverage and Quality" in DATASET.CAT often do not warrant even the little attention provided by a note in ERRATA.TXT.

Any changes in the production and/or validation of the data products/volumes should be reflected in the design documents. For instance, extensive reprocessing may violate assumptions made by PDS when it agreed to your mission's (or instrument's) archive plan. If you expect significant fractions of your archive to be delivered more than once, you should make that clear in the archive plan (by amendment, if necessary).

5.4 Data Production, Distribution, and Maintenance Checklist

The following checklist is provided as a summary of the steps outlined in the Data Production / Distribution / Maintenance section. The checklist can be used as a means of ensuring the steps towards creating a successful archive are complete.

	SECTION	ACTIVITY
☐	5.1	Run the data production pipeline
☐	5.1	Validate each volume
☐	5.1	Submit early volumes for comprehensive "production review"
☐	5.1	Deliver volumes to PDS; expect automated and/or spot checks
☐	5.3	Respond to reports of anomalies in an appropriate manner - reprocessing for serious defects which can be corrected - withdrawal for serious anomalies which cannot be corrected - update ERRATA.TXT for minor anomalies - updated ancillary files for editorial changes

APPENDIX A Acronyms

The table below lists acronyms and abbreviations used in this document.

ADD	Archive Description Documentation
AFCR	(hypothetical) Asteroid Flyby Comet Rendezvous mission
ALC	*Archive Life Cycle* (PDS document, in preparation)
AO	Announcement of Opportunity
APG	*Archive Preparation Guide* (this document)
ASCII	American Standard Code for Information Interchange
CD-ROM	Compact Disc—Read-Only Memory
CODMAC	Committee On Data Management And Computation
DMAP	Data Management and Archiving Plan
DN	Discipline Node
DVD	Digital Versatile Disc
EDR	Experiment Data Record
ESA	European Space Agency
ISS	Imaging Subsystem
ITAR	International Traffic in Arms Regulations
JPL	Jet Propulsion Laboratory
lvtool	Label Verifier Tool (PDS software)
MAS	Mission Archive Scientist
MER	Mars Exploration Rover
NAIF	Navigation and Ancillary Information Facility
NASA	National Aeronautics and Space Administration
NASAVIEW	PDS display software
NRA	NASA Research Announcement
NSSDC	National Space Science Data Center
PAG	*Proposer's Archiving Guide* (PDS document)
PDF	Portable Document Format (Adobe software)
PDS	Planetary Data System
PI	Principal Investigator
PSDD	*Planetary Science Data Dictionary* (PDS document)
RDR	Reduced Data Record
RSA	Radio Science Advisor
SIS	Software Interface Specification
SPICE	Spacecraft ephemeris, Planet/satellite ephemeris, Instrument information, Camera orientation, Event information.
TBD	To be determined
tbtool	Table Browser Tool (PDS software)
UCD	User Centered Design
URL	Uniform Resource Locator
UTC	Coordinated Universal Time

Table A-1. Acronyms and abbreviations

APPENDIX B Definition of Terms

The table below lists terms used in this document.

Archive	(1) One or more data sets stored for long-term preservation; (2) the logical structure of such an archive without regard to its physical location or the media on which it is stored.
Volume	One physical or logical unit of storage within an archive identified by a unique VOLUME_ID and VOLUME_NAME—for example, a single compact disk, magnetic tape. or disk partition. Depending on capacity of the volume relative to size of the data set(s), a single volume may hold several data sets or only part of one data set.
Volume Set	One or more volumes containing data with a common origin, time span, and/or theme and identified by a unique VOLUME_SET_ID and VOLUME_SET_NAME.
Data Object	Data having a common origin or theme and organized into a recognized format. PDS recognizes about 30 objects including image, table, and text.
Label	A text file containing metadata (data about data) in a keyword=value format which describes the structure and content of one or more data objects.
Data Product	A label and the one or more data objects that it describes.
Data Set	A grouping of data products having a common origin, time span, and/or theme and documentation sufficient to permit use without recourse to other data. A data set is identified by a unique DATA_SET_ID and DATA_SET_NAME.
Data Set Collection	A grouping of data sets which are related within a specific scientific area, practice, or objective (i.e., by observation type, discipline, target, or time) and identified by a unique DATA_SET_COLLECTION_ID and DATA_SET_COLLECTION_NAME.
Standard data product	A data product that has been defined during the proposal and selection process and that is contractually promised by the PI as part of the investigation. Standard data products are generated in a predefined way, using well-understood procedures, and processed in "pipeline" fashion.
Primary Data	The scientific measurements made by an instrument
Ancillary Data	Other data needed to understand or use the primary data. Examples include instrument settings, calibration data, files specifying observation geometry, processing software, and documents.

APPENDIX C PDS Nodes and Contact Information

The PDS home page is the main entry portal to the PDS. http://pds.nasa.gov/

The structure of the PDS includes multiple discipline nodes (DNs) and a few Support Nodes, which provide valuable services that are not science discipline specific.

Discipline Nodes - each DN is responsible for data in a specific subject area or specialty; each maintains the PDS data holdings in that specialty. At least one discipline node will work closely with the mission team throughout archive development.

The DNs are briefly described below. The DNs provide your single most useful resource - contact them early and often. The most current list of Node contacts is at the following online address. http://pds.nasa.gov/contact/contact.shtml

- Planetary Atmospheres Node, located at New Mexico State University (Las Cruces, NM), concentrating on planetary atmospheres and including the gas giant planets. http://atmos.pds.nasa.gov/

- Geosciences Node, located at Washington University (St. Louis, MO), specializing in the surfaces and interiors of terrestrial planets. http://geo.pds.nasa.gov/

- Imaging Node, located at the U.S. Geological Survey (Flagstaff, AZ) and Jet Propulsion Laboratory (Pasadena, CA), specializing in digital imaging and image processing. Imaging also is responsible for large icy satellites. http://img.pds.nasa.gov/

- Planetary Plasma Interactions Node, located at the University of California, Los Angeles, concerned with fields and particles, plasmas, and interactions with the interplanetary medium. http://ppi.pds.nasa.gov/

- Planetary Rings Node, located at the SETI Institute (Mountain View, CA) and concentrating on planetary ring systems and dynamical interactions with inner moons. http://rings.pds.nasa.gov/

- Small Bodies Node, located at the University of Maryland (College Park, MD), and concerned with comets, asteroids, trans-Neptunian objects (including Pluto), and interplanetary dust. http://sbn.pds.nasa.gov/

There are two Support Nodes and two ‘functions’.

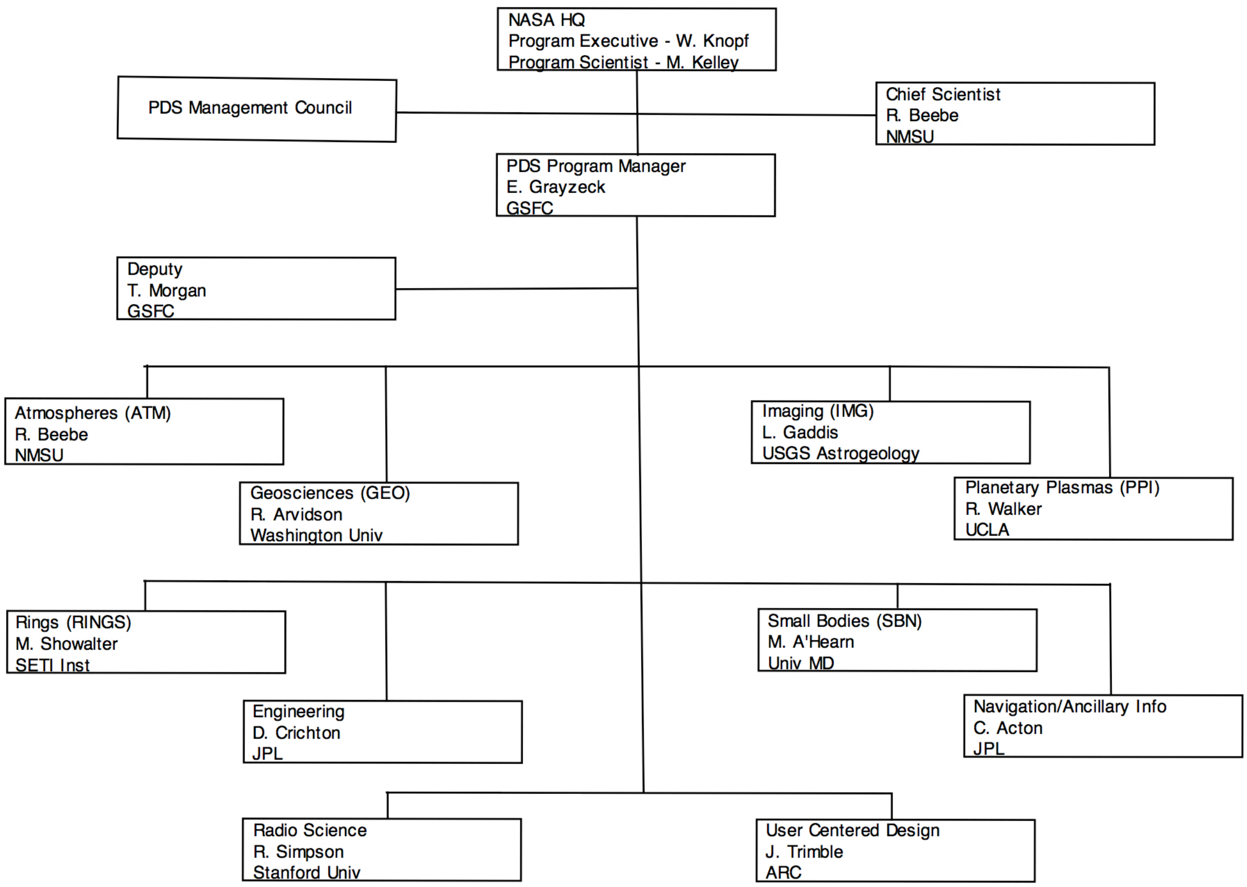
- PDS Engineering Node (EN), located at the NASA Jet Propulsion Laboratory (Pasadena, CA), maintains standards and provides tools and other infrastructure for PDS. http://pds.nasa.gov/tools/index.shtml/ or contact your PDS rep for additional assistance.

- Navigation and Ancillary Information Facility (NAIF), located at the Jet Propulsion Laboratory (Pasadena, CA), specializing in observational geometry (ephemerides and orientations, including the SPICE information system) and ancillary information (e.g., spacecraft engineering data). http://naif.pds.nasa.gov/

- The Radio Science Advisor (RSA), located at Stanford University (Stanford, CA), consults on archiving radio science data. The RSA does not maintain a web site or ingest or distribute PDS data except through other DNs.

- The User Centered Design (UCD) function is located at the NASA Ames Research Center (Moffett Field, CA) and consults on human-machine interfaces, data searches, and computer tools.

Planetary Data System Organization Chart



APPENDIX D On-Line Resources and Examples

The following resources are available electronically either at the designated URL or the *Archive Preparation Guide* (APG) web page: http://pds.nasa.gov/documents/apg

D.1 PDS Web Page

- PDS Home Page: http://pds.nasa.gov/
- Links to key PDS documents such as *PDS Standards Reference and PDS Data Dictionary* can be found at: http://pds.nasa.gov/documents/

D.2 Archiving Documents

See the APG web site for a set of example archiving documents; such as:

- MOU – Example Memorandum of Understanding for Mars Reconnaissance Observer (MRO)
- DMAP – Example Data Management and Archiving Plans
  - Messenger
  - Deep Impact
- SIS – Example Software Interface Specifications, which document volume and data product structure, content, and processing history
  - Draft Huygens – Surface Science Package (SSP)
  - Data Product SIS template
  - Archive Volume SIS template

D.3 Catalog Files

See the APG web site for a set of example catalog files; such as:

- MISSION.CAT
- INSTHOST.CAT
- INST.CAT
- DATASET.CAT
- PERSON.CAT
- REF.CAT
- TARGET.CAT

D.4 Product Files

See the APG web site for a set of example product files; such as:

- Example IMAGE data product and data product label
- Example TABLE data product and data product label
- Example Index files

D.5 Data Set Citation descriptions

A data set citation is the recommended listing when the author of a journal paper wishes to reference the data set. Formats and content vary among journals, but the basic information should be provided following these examples from Mars Exploration Rover (MER)

```
DATA_SET_ID      = "MER2-M-MTES-3-RDR-V1.0"
DATA_SET_NAME    = "MER 2 MARS MINI-TES RDR V1.0"
CITATION_DESC    = "Christensen, Phil, 'MER 2 Mars
                    Mini-TES RDR V1.0', NASA Planetary
                    Data System,
                    MER2-M-MTES-3-RDR-V1.0, 2004."


DATA_SET_ID      = "MER1-M-MTES-2-EDR-V1.0"
DATA_SET_NAME    = "MER 1 MARS MINI-TES EDR V1.0"
CITATION_DESC    = "Christensen, Phil, 'MER 1 Mars
                    Mini-Thermal Emission Spectrometer
                    EDR V1.0', NASA Planetary Data System,
                    MER1-M-MTES-2-EDR-V1.0, 2004."
```

D.6 PDS Tool Suite

The latest set of released PDS Tools may be downloaded from the PDS Software Library.

http://pds.nasa.gov/tools/index.shtml

D.7 Additional Resources

The following may also be of interest.

- PDS Interactive Phonebook – interactive search tool for PDS personnel.

  http://pds.nasa.gov/tools/phonebook/phonebook.cfm

- PDS Data Dictionary Lookup tool – interactive search tool for searching the keywords and objects that comprise a PDS label.

  http://pds.nasa.gov/tools/ddlookup/data_dictionary_lookup.cfm

- PDS Reference Lookup tool – interactive search for references used in the PDS catalog files.

  http://pds.nasa.gov/tools/rlookup/reference_lookup.cfm

APPENDIX E Software Interface Specification (SIS) Outline

The overall structure and content of the SIS should reflect the following:

1	INTRODUCTION
1.1	PURPOSE AND SCOPE
1.2	CONTENTS
1.3	INTENDED READERSHIP
1.4	APPLICABLE DOCUMENTS
1.5	RELATIONSHIPS TO OTHER INTERFACES
1.6	ACRONYMS AND ABBREVIATIONS
1.7	CONTACT NAMES AND ADDRESSES
2	OVERVIEW OF PROCESS AND PRODUCT GENERATION
3	ARCHIVE FORMAT AND CONTENT
3.1	FORMAT
3.1.1	Volume Format
3.1.2	Data Set Format
3.1.3	File Formats
3.2	STANDARDS USED IN DATA PRODUCT GENERATION
3.2.1	PDS Standards
3.2.2	Time Standards
3.3.3	Coordinate Systems
3.3	DATA VALIDATION
3.4	CONTENT
3.4.1	Volume Set
3.4.2	Data Set
3.4.3	Directories
3.4.3.1	Root Directory
3.4.3.2	Calibration Directory
3.4.3.3	Catalog Directory
3.4.3.3.1	Mission Catalog File
3.4.3.3.2	Instrument Host Instrument Catalog File
3.4.3.3.3	Instrument Catalog File
3.4.3.3.4	Personnel Catalog File
3.4.3.3.5	Reference Catalog File
3.4.3.3.6	Dataset Catalog File
3.4.3.3.7	Map Projections Catalog File
3.4.3.3.8	Target Catalog File
3.4.3.4	Index Directory
3.4.3.4.1	Dataset Index File, INDEX.LBL and INDEX.TAB
3.4.3.4.2	Geometric Index File, GEOINDEX.LBL and GEOINDEX.TAB
3.4.3.4.3	Other Index Files
3.4.3.5	Browse Directory and Browse Files
3.4.3.5	Geometry Directory
3.4.3.6	Software Directory
3.4.3.7	Gazetter Directory
3.4.3.8	Label Directory
3.4.3.9	Document Directory
3.4.3.10	Extras Directory
3.4.3.11	Data Directory
3.4.4	Data and Label Files
3.4.5	Pre-Flight Data Products
3.4.6	Sub-System Tests

APPENDIX F CODMAC Levels

NASA	CODMAC	Description
Packet data	Raw - Level 1	Telemetry data stream as received at the ground station, with science and engineering data embedded.
Level-0	Edited - Level 2	Instrument science data (e.g., raw voltages, counts) at full resolution, time ordered, with duplicates and transmission errors removed.
Level 1-A	Calibrated - Level 3	Level 0 data that have been located in space and may have been transformed (e.g., calibrated, rearranged) in a reversible manner and packaged with needed ancillary data (e.g., radiances with the calibration equations applied).
Level 1-B	Resampled - Level 4	Irreversibly transformed (e.g., resampled, remapped, calibrated) values of the instrument measurements (e.g., radiances, magnetic field strength).
Level 1-C	Derived - Level 5	Level 1A or 1B data that have been resampled and mapped onto uniform space-time grids. The data are calibrated (i.e., radiometrically corrected) and may have additional corrections applied (e.g., terrain correction).
Level 2	Derived - Level 5	Geophysical parameters, generally derived from Level 1 data, and located in space and time commensurate with instrument location, pointing, and sampling.
Level 3	Derived - Level 5	Geophysical parameters mapped onto uniform space-time grids.