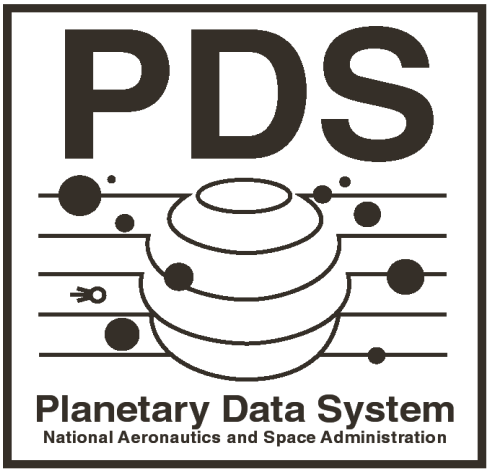
Data Nodes: Principles and Practices

NASA/PDS

White Paper



version 060301

Contents

SECTION 1 4

INTRODUCTION 4

SECTION 2 6

BACKGROUND 6

Data Nodes for Recovery 6

Data Nodes for Mission Interface 7

SECTION 3 8

ISSUES 8

Advantages 8

Disadvantages 8

Cautionary Note 8

SECTION 4 9

GUIDELINES 9

Purpose 9

Longevity 9

Interface 9

Dissolution 9

SECTION 5 10

APPENDIX A1 - EXAMPLE MISSION INTERFACE DATA NODE AGREEMENT 10

Preamble 11

General 11

XXX Mission/Instrument Responsibilities 11

XXX Data Node Responsibilities 11

YYY Discipline Node Responsibilities 12

APPENDIX A2 - EXAMPLE RECOVERY DATA NODE AGREEMENT 13

Preamble 13

General 14

XXX Data Node Responsibilities 14

YYY Discipline Node Responsibilities 14

SECTION 6 15

APPENDIX B – SUGGESTED CHRONOLOGY FOR MISSION INTERFACE DATA NODES 15

Section 1

Introduction

The Planetary Data System (PDS) operates as a federation of Discipline Nodes (DN's), established in response to NASA NRA 03-OSS-04. Each Discipline Node receives, curates, and distributes data of interest to a community of planetary scientists. Some DN's sponsor permanent "sub-nodes" to maintain expertise in special areas, (Figure 1).

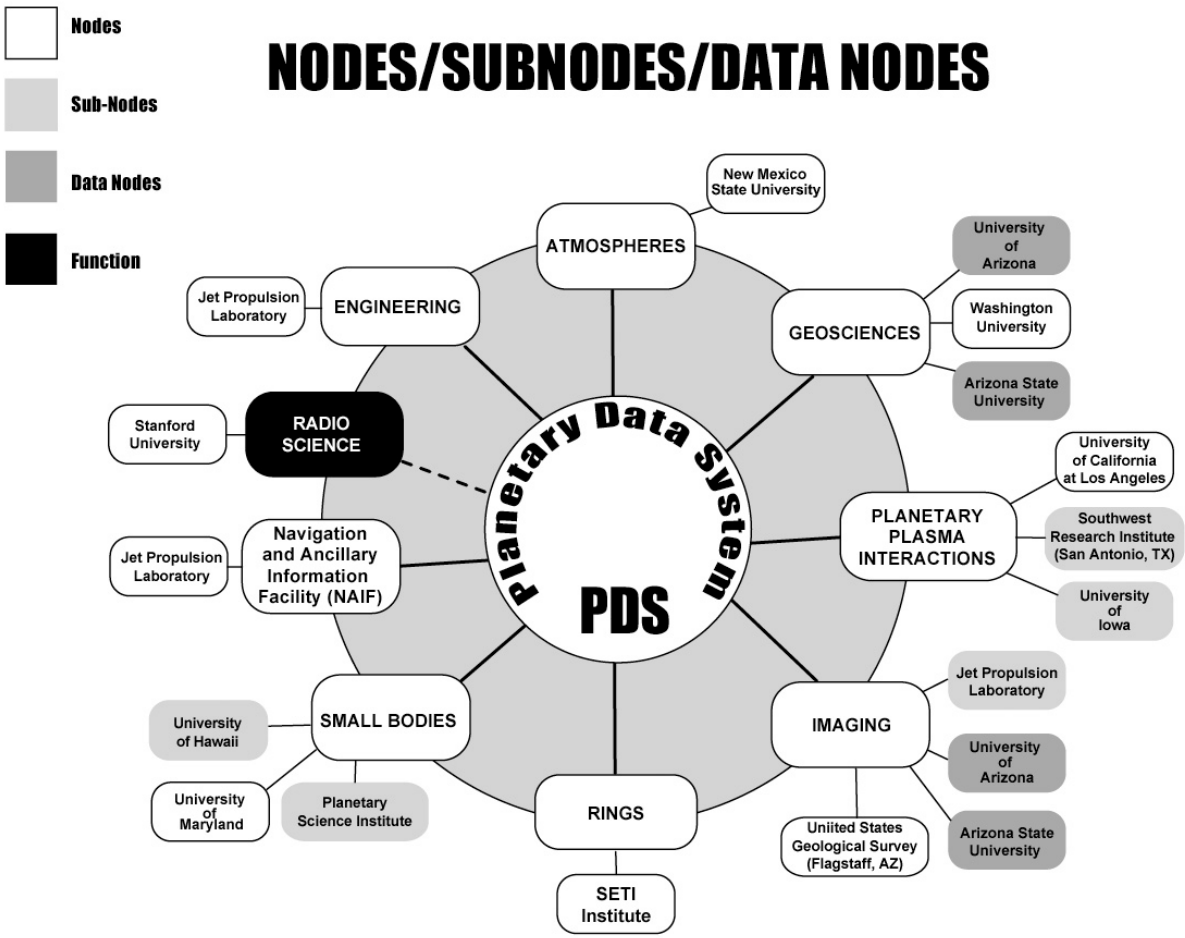


Figure 1. Planetary Data System structure. Primary user interface is through the central hub, which contains catalog information and provides connections to Discipline Nodes (DN's). Each DN has a home institution (*e.g.*, University of Maryland for Small Bodies); some have sub-nodes for special expertise (*e.g.*, the Small Bodies Dust Sub-Node at the Planetary Science Institute).

To facilitate cross-discipline investigations, archival formats and day-to-day operations within PDS are governed by a common set of documents including the *PDS Standards Reference* and the *Planetary Science Data Dictionary Document*.

From time to time it has been desirable to establish temporary adjuncts to the PDS structure. Historically the reasons for setting up such "data nodes" (dN's) have ranged from interest in rehabilitating existing data sets so they could be used more widely or more intensively for science or future mission planning to the need for buffering between a flight project and PDS. In all cases, however, there has been an implicit assumption that the dN was temporary and that the data holdings would eventually be absorbed by the parent DN.

Recently there has been increased interest in establishment of data nodes, including at least one proposal to NASA to set up a data node which had not previously been planned by PDS. This document describes the framework within which PDS expects data nodes to operate, so that future dN proposals (both solicited and unsolicited) may be more easily integrated into the system.

Section 2

Background

Data nodes fall broadly into two categories — those for upgrading existing data holdings, which have been incompletely processed and/or are poorly documented, so that they are suitable for general use; and operations (usually much larger) that provide PDS with an interface to an active flight project. Each is described in more detail below. Table 1 summarizes recent experience with data nodes.

Table 1 - PDS Data Nodes

Discipline Node (DN)	Number of Data Nodes (approx.)		Notes
	Past	Active	
Atmospheres (ATM)	4	1	Completion of mission data (past); mapping Galileo data (active)
Geosciences (GEO)	2	2	telescopic and laboratory data (past); MGS TES and Odyssey GRS (active)
Imaging (IMG)	0	2	Odyssey THEMIS and MRO HiRISE (active)
Planetary Plasma Interactions (PPI)	3	0	restorations; late focus on mission data
Small Bodies (SBN)	5	1	restoration of some Lowell Observatory holdings, restoration of some International Ultraviolet Explorer data, others (past); Arecibo radar data (active)

Data Nodes for Recovery

The majority of data nodes (Table 1) have been established for restoration or special handling of existing data; the Planetary Plasma Interactions Node (PPI) has also used "recovery" data nodes to provide additional focus on data archiving near the end of a spacecraft mission. The contractual arrangements have varied, usually chosen on grounds of administrative simplicity. A short statement of work and budget described the tasks; the typical data node term and funding level have been one year and $30K, respectively. Within this class, only the Atmospheres data node for mapping Galileo atmospheric data and the Small Bodies Node (SBN) data node for archiving asteroid

radar data at Arecibo Observatory are currently active. Aside from administrative delays in funding, the general reaction to establishing and working with this type of data node has been positive, though Geosciences (GEO) reported one experience in which the provider failed to adhere to PDS standards and significant portions of the work had to be redone by GEO staff.

Data Nodes for Mission Interface

The Imaging Discipline Node (IMG) and GEO each have two data nodes, operated by active mission instrument teams (Table 1). Each dN is co-located with its instrument team, leverages expertise of the team and capabilities developed to support the mission, and has scheduled data deliveries to PDS. At termination, each dN will transfer its hardware and software, as well as its data, to the parent DN. The TES dN has been operational since 1998; it is funded by the TES Team through the Mars Global Surveyor Project, which will determine its lifetime. The GRS and THEMIS data nodes have been operational since 2001; they are funded by the Odyssey mission, which will determine their lifetimes. The HiRISE data node will begin delivering data after MRO reaches Mars; it is partially funded by PDS as part of the most recent USGS/JPL proposal to continue operating the IMG Discipline Node.

Section 3

Issues

The reasons for establishing a data node vary depending on both the organizations involved and the data to be handled.

Advantages

The advantages of establishing a Data Node include:

- Improving quality of the archive by leveraging expertise, such as offering user support for the duration of a mission from the team actively working with the data
- Reducing overall costs by avoiding unnecessary duplication—also by tapping into the data system of an active mission team
- Ensuring more timely public access to new data
- Capturing data which might otherwise be lost (aging investigators) or missed (lab or observatory data not considered within the normal scope of PDS)
- On a temporary basis, providing higher levels of support, products, and/or services than is customary for the parent DN

Disadvantages

The disadvantages of establishing a Data Node include:

- The requirement to provide additional supervision over a (potentially) remote facility
- Dealing with data delivery and services that are beyond normal PDS scope (in the sense that the user community may continue to expect extended services after the data node has been dissolved).
- Inheriting potentially old hardware/software and the need to maintain or replace both

Cautionary Note

It should be noted here that establishment of a data node is intended to benefit the PDS and its community of users. It is expected that the archive which results from existence of the dN will be better than the archive which would have resulted from "normal" archiving with equivalent funding. A data node is not intended to provide either additional or extended funding to an investigation beyond that which would have been allocated by the mission itself.

Section 4

Guidelines

Every data node requires a Discipline Node parent. PDS requires a formal agreement between the interested parties covering establishment, operation, and dissolution of the data node. A proposal to NASA will not lead to establishment of a data node until such an agreement has been negotiated. Appendix A1 is an example agreement for a mission interface data node; agreements for "recovery" dN's can follow the same model, but are likely to be simpler (*e.g.*, Appendix A2)

Purpose

Define clearly why the data node is needed. List the benefits to PDS from establishing this arrangement. If the data node will restore old data, explain why the data node is needed rather than simply delivering through existing channels If the data node will provide a buffer between PDS and an active flight project, explain what is missing from (or inadequate in) PDS that forces establishment of the data node.

Longevity

Data nodes are *temporary* adjuncts to PDS. Define the lifetime of the data node. For an active flight project with possible extended missions, set firm dates based on current expectations; these can be amended later, if mutually agreeable. Include the schedule for data delivery and milestones , showing whether the schedule is being met. If the data node is to provide an interface to an active mission, consider whether the data node can be phased out before the mission is completed. If not, include provisions to ensure that the job will be completed satisfactorily as the mission (or instrument) winds down.

Interface

Clearly and completely, list the tasks to be performed, the schedule to be followed, and the roles and responsibilities of each party in achieving them. In the case of a data node, which handles data from an active mission, the agreement must state which tasks will be performed as mission responsibilities and which will be performed for PDS as the data node (see Appendix B for further information). The management and support responsibilities of the parent DN should also be listed. If not clear from other provisions in the agreement, the funding expectations should be included.

Dissolution

Dissolving a data node requires transfer of data node holdings to the DN on or before the negotiated completion date. Dissolution and transfer must be completed with minimal interruption of services to users. The services (if any) which the DN will continue after dissolution of the data node will be described. The data transferred will meet PDS standards. There may be hardware and software issues in some transfers in addition to the data.

Section 5

Appendix A1 - Example Mission Interface Data Node Agreement

MEMORANDUM OF AGREEMENT
FOR ESTABLISHMENT OF THE XXX DATA NODE

BETWEEN

PDS YYY DISCIPLINE NODE
Institution Name
Location

AND

XXX INSTRUMENT TEAM
Institution Name
Location

Date

_____ _____
PI Name Date
Principal Investigator, XXX Instrument Team

_____ _____
DN Manager Name Date
Manager, PDS YYY Node

_____ _____
PDS Manager Name Date
Program Manager, Planetary Data System

_____ _____
Mission Manager Name Date
Manager, ZZZ Mission

Preamble

To facilitate transfer of archival data from the XXX instrument of the ZZZ mission to the Planetary Data System and to PDS users, the responsible parties agree to establishment of the XXX data node under the conditions listed below. Terms of the agreement may be amended by unanimous agreement among the parties.

General

1. The XXX data node is established for the purpose of archiving and distributing standard and special data products from the XXX instrument. For definitions of terms, such as "standard" and "special" data products, see the current version of the *PDS Archive Preparation Guide*.

2. Explain here why establishment of the data node is preferable to using the normal archiving path to PDS.

3. The data node will be funded by and meet requirements for data delivery established by the XXX instrument and the ZZZ mission.

4. The data node will report to the YYY Discipline Node of the PDS. During the term of this agreement it will meet archiving, distribution, and performance standards established by the PDS and the YYY Discipline Node. At the termination of this agreement its holdings will be transferred in an orderly manner to the YYY Discipline Node.

5. The data node is established for a time period not to exceed the primary phase of the ZZZ mission plus six months.

XXX Mission/Instrument Responsibilities

During the term of this agreement the data node will assume the following responsibilities, which are expected of all entities delivering data to PDS.

1. Design and implement a PDS-compliant science data archive of Standard Data Products.

2. At the discretion of the XXX Team, augment the archive to include Special Data Products.

3. Work with PDS to ensure that the archive will be of maximum scientific use to the community consistent with the resources available.

4. Describe the archive structure and contents in a set of documents including data product Interface Control Documents (ICD's), Software Interface Specifications (SIS's), INST.CAT files, DATASET.CAT files, and other documents.

5. Generate Standard and Special products as defined by the documentation.

6. Organize the data products with appropriate documentation, software, and other information for delivery to PDS on a regular basis.

7. Validate the deliveries based on standards set by the mission, the team, and PDS.

8. Deliver the archives to PDS.

XXX Data Node Responsibilities

The following are unique responsibilities for the XXX data node within PDS.

1. Distribute archival products from the XXX Team electronically to planetary scientists via

web and/or ftp interfaces. As resources allow, distribute data on hard media (*e.g.*, CD-R or DVD-R) to users without adequate network access.

2. Provide support to users, such as:

a. answering questions by phone or e-mail about the XXX data

b. assisting users to identify, locate, and download data of interest

c. generating special products upon request, as resources allow

3. Maintain electronic interfaces so that the XXX catalog is visible to PDS and the XXX holdings can be downloaded or transferred within the system. Specifically, this includes installing, activating, and maintaining PDS-D software on the XXX data node system.

4. Report monthly on activity to the YYY DN Manager

5. Plan and implement (with the YYY DN) an orderly transfer of data node holdings and services before the end of this agreement.

YYY Discipline Node Responsibilities

During the period of this agreement, the YYY Discipline Node will

1. Supervise and oversee operation of the XXX data node

2. Provide assistance during integration of the data node with PDS, especially regarding installation of PDS-D

3. Ensure that the data node has access to other PDS data management tools and that its installation is properly maintained.

4. Provide guidance in responding to user requests for information and assistance, particularly where significant resources may be required to satisfy a request.

5. Receive and validate data deliveries and monthly activity reports

6. Plan and implement (with the XXX data node) an orderly transfer of data node holdings and services before the end of this agreement.

Appendix A2 - Example Recovery Data Node Agreement

MEMORANDUM OF AGREEMENT
FOR ESTABLISHMENT OF THE XXX DATA NODE

BETWEEN

PDS YYY DISCIPLINE NODE
Institution Name
Location

AND

XXX DATA NODE TEAM
Institution Name
Location

Date

_____ _____
PI Name Date
Principal Investigator, XXX Data Node Team

_____ _____
DN Manager Name Date
Manager, PDS YYY Node

_____ _____
PDS Manager Name Date
Program Manager, Planetary Data System

Preamble

To facilitate transfer of archival data from the XXX data node team to the Planetary Data System and to PDS users, the responsible parties agree to establishment of the XXX data node under the conditions listed below. Terms of the agreement may be amended by unanimous agreement among the parties.

General

1. The XXX data node is established for the purpose of archiving standard and special data products from the XXX instrument. For definitions of terms, such as "standard" and "special" data products, see the current version of the *PDS Archive Preparation Guide*.

2. Explain here why establishment of the data node is preferable to using the normal archiving path to PDS.

3. The data node will report to the YYY Discipline Node of the PDS. During the term of this agreement it will meet archiving, distribution, and performance standards established by the PDS and the YYY Discipline Node. At the termination of this agreement its holdings will be transferred in an orderly manner to the YYY Discipline Node.

4. The data node is established for a time period not to exceed twelve months.

XXX Data Node Responsibilities

During the term of this agreement the data node will assume the following responsibilities, which are expected of all entities delivering data to PDS.

1. Design and implement a PDS-compliant science data archive of Standard Data Products. The Standard Data Products shall be ... (list) ...with an estimated total volume of ... bytes.

2. At the discretion of the XXX Team, augment the archive to include Special Data Products.

3. Work with PDS to ensure that the archive will be of maximum scientific use to the community consistent with the resources available.

4. Describe the archive structure and contents in a set of documents including data product Interface Control Documents (ICD's), Software Interface Specifications (SIS's), INST.CAT files, DATASET.CAT files, and other documents as needed and/or appropriate.

5. Generate Standard and Special products as defined by the documentation.

6. Organize the data products with appropriate documentation, software, and other information for delivery to PDS.

7. Validate the delivery or deliveries based on standards set by the XXX data node team and the YYY PDS Discipline Node.

8. Deliver the archives to the YYY Discipline Node in such a way that the final delivery is complete before the end of this agreement.

9. Report quarterly on activity to the YYY DN Manager

YYY Discipline Node Responsibilities

During the period of this agreement, the YYY Discipline Node will

1. Monitor operation of the XXX data node

2. Ensure that the data node has access to PDS data management and other tools

3. Receive and validate data deliveries and quarterly activity reports

4. Plan and implement (with the XXX data node) an orderly transfer of final data node holdings before the end of this agreement.

Section 6

Appendix B – Suggested Chronology for Mission Interface Data Nodes

Early in Phase B of the mission, a Memorandum of Understanding (MOU) defining/delineating roles and mutual responsibilities regarding archiving and distribution of mission data should be negotiated between the mission and the PDS with a Discipline Node leading the way. Also, the PDS will work with the mission to complete the archive portion of the Data Management and Archiving Plan (DMAP). The DMAP will contain sufficient detail about the data flow, the archival products, and the procedures for generating and distributing the archive products. Additionally the DMAP will provide sufficient detail to enable both the mission and the PDS to develop reliable estimates of the personnel and costs for their respective activities.

It is at this stage, that the Discipline Node will ascertain the feasibility of establishing a Data Node. If the Discipline Node decides in favor of establishing a Data Node, then the Discipline Node and the Instrument Team (IT), should establish a written agreement, a Data Node Memorandum of Understanding (DN MOU), which defines/delineates roles and mutual responsibilities regarding archiving and distribution of specific data sets.