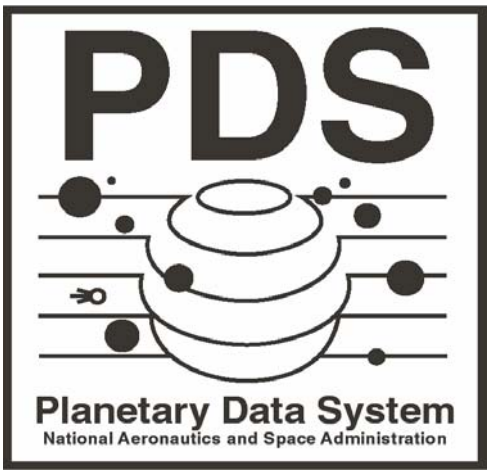
Planetary Data System

Archive and Delivery Tracking Level 4, 5, and 6 Requirements

DRAFT

March 20, 2007

Version 0.10070320



Jet Propulsion Laboratory
Pasadena, California

JPL D-xxxxx

CHANGE LOG

Revision	Date	Description	Author
Start Draft	2007–02-04	First Draft	S. Hughes, D. Crichton
	2007-03-07	Updated after telecom of 3/6/07	D. Crichton, S. Hughes
	2007-03-20	Updated after telecom of 3/20/07	S. Hughes, D. Crichton

Table of Contents

1 Introduction

The purpose of this document is to capture requirements for tracking the status of data from its delivery by the data providers through the PDS to the deep archive, NSSDC. Tracking is differentiated into delivery tracking, primarily tracking data from the data provider to the PDS and archive tracking, the tracking of data within the PDS archive. The archive tracking function includes making an inventory of all files and ensuring that all files delivered are accounted for and available to the community. Within the PDS, files are components of collections. These collections include data products, data deliveries and data sets.

A key concept introduced in these tracking use cases is the delivery escrow. This concept allows the provider to start a delivery however the receiver does not have to accept the delivery until certain delivery criteria have been met. The receiver can choose to reject a delivery.

Tracking files is one part of maintaining the integrity of the archive and understanding the status of deliveries both to PDS and the NSSDC. Recently, PDS used the Cassini Archive Tracking System (CATS) as a tool for tracking product deliveries from Cassini instrument teams to the PDS nodes. As a result of this experience, PDS decided to repurpose the CATS software for the Phoenix mission to tracking deliveries from Phoenix instrument teams in a manner similar to Cassini. In addition, the PDS Management Council approved the level 3 requirement 2.2.2 for tracking which states, "PDS will track the status of data deliveries from data providers through the PDS to the deep archive."

PDS has been also developing a set of related use cases for *Data Integrity.* While the Data Integrity use cases focused on *file corruption,* the use cases defined for tracking will focus on tracking of files. However the *Data Integrity* use cases will be referenced from this document when appropriate.

2 Related Documents

[1] Planetary Data System Data Integrity Use Cases Document, October, 2006,

[2] Planetary Data System Data Integrity Requirements Document, November, 2006,

[3] Planetary Data System Level 123 Requirements, August, 2006,

3 Definitions

The following definitions are used in the requirements.

1. **Acceptance Status** – A status indicating whether a data delivery has been accepted.
2. **Actors**. An actor is a person, organization, or external system that plays a role in one or more interactions with your system. For tracking system notification, actors are those entities (individuals) that have been identified to be informed of actions by the system.
3. **Archive Manifest** – An archive manifest is a list of files and should include at least filenames, file checksums, and the checksum type.
4. **Authorized Actor** – An actor who has been approved to be informed of actions by the system. For tracking system notification, authorized actors are a subset of tracking system actors that have been specifically authorized to be informed of specific actions by the system.
5. **Data Product** – A data product label and one or more data objects. A data product consists of not only the product label file, but all files that comprise each data object.
6. **Data Product Label** – One or more data object descriptions.
7. **Data Provider** - Entities that submit data to PDS.
8. **Data Set** – A data set is a collection of product together with ancillary data and documentation. A data set is organized as one or more archive volumes.
9. **File** – A collection of bytes stored as a single unit within a file system.
10. **Delivery Manifest** – The delivery manifest is a list of files that should include at least the file names, the file checksums, and the checksum type.
11. **Delivery Status –** A status indicating whether a data delivery has been successfully transferred.
12. **Manifest** - A manifest is a list of items. For the purpose of tracking, the manifest is a list of files and will include at a minimum the filenames, checksum, and checksum type. (See Archive Manifest and Delivery Manifest).
13. **Product Collection** – A product collection is a set of data products collected for a specific purpose.
14. **PDS Node** – Any PDS node including Discipline Nodes, Data Nodes, and the Engineering Node. The Discipline Nodes include both science and support nodes.
15. **Physical Media –** Any computer system device used for short or long term storage of data including but not limited to optical media, tape, and magnetic disk.
16. **Use cases**. A use case describes a sequence of actions that provide something of measurable value to an actor.
17. **Volume** – Any organized collection of files that reside on physical media for the purpose of near term storage, online access, data submission, electronic distribution, or long-term archive. Note that this definition includes the PDS archive volume.

4 Notation

The numbering of the requirements in this document will be formatted as **Ln.xT.AA.N**, where:

- **Ln** indicates the requirement level number.
- **xT** is an acronym representing either Delivery or Archive Tracking requirements.
- **AA** is a two letter acronym for the requirement subcategory. (Optional)
- **N** is a unique number for the requirement.

Following the text of a requirement may be a reference to the requirement from which it was derived. The reference will be in parenthesis.

A paragraph following a requirement, which is indented and has a reduced font size, represents a comment that provides additional insight for the requirement. This comment should not be considered part of the requirement for development or testing purposes.

5 Delivery Tracking Requirements

The following requirements are for data delivery tracking. In this document, "delivery tracking" encompasses the process of tracking files associated with data deliveries from any data provider through to the acceptance of the data by a PDS Node. These requirements are on the data tracking system, a subsystem that tracks data delivery events and their statuses. An important feature suggested by these requirements allows for data deliveries to be successfully transferred to but not necessarily accepted by a PDS node until reviewed. This “escrow” feature ensures that only data deliveries that meet certain criteria are actually accepted. These requirements are derived from level one, two, and three PDS requirements, PDS Tracking Use Cases, and PDS Policies.

5.1 General

L4.DT.GR.1 – The Delivery Tracking System shall track the contents of a data delivery identifying each file associated with a delivery (UC-1.2, 2.2.2)

L4.DT.GR.2 – The Delivery Tracking System shall allow verification that the contents of a data delivery package matches its delivery manifest by checking that every file listed in the manifest is in the package and vice versa. (2.2.2)

L5.DT.GR.2.1 – The PDS shall define a standard delivery manifest.(2.2.2)

The delivery manifest lists the files to be delivered (see proposed definition). The definition will specify the information content of the delivery manifest but not its format or structure.

L4.DT.GR.3 – The Delivery Tracking System shall notify actors about events including scheduled events associated with deliveries and any change in delivery or acceptance statuses. (3.3.6)

The PDS Notification/Subscription system could be extended to fulfill this requirement. The notification of an actor will be triggered by an event. Various requirements in the following sections will signal events. The delivery plan will provide a list of actors to be notified for specific events. The subset of “authorized” actors will also be provided in the delivery plan.

L5.DT.GR.3.1 – The Delivery Tracking System shall trigger notification of an actor for the following types of events.

Delivery Plan Modification

Pending Data Delivery Event
Data Delivery Event
Tracking System Status Changes (i.e. Delivery Status and Acceptance Status)

L4.DT.GR.4 – The Delivery Tracking System shall allow authorized actors to set the tracking system statuses. (UC-1.2)

L4.DT.GR.5 – The Delivery Tracking System shall allow the generation of reports detailing the statuses of past and projected future data delivery events. (UC-4, UC-5, 2.2.2)

L5.DT.GR.5.1 – The Delivery Tracking System shall provide a GUI and a log-in capability for access to information about data deliveries and their events.

L5.DT.GR.5.2 – The Delivery Tracking System shall allow a authorized actor to query interactively for information about data deliveries and their events.

L5.DT.GR.5.3 – The Delivery Tracking System shall be able to generate reports on data deliveries and their events.

5.2 Data Provider Delivers Data

L4.DT.DP.1 – The Delivery Tracking System shall track data delivery events and their statuses based on a negotiated delivery plan between a data provider and a PDS Node. (UC-1.3, UC-6, 2.2.2, 2.6.2, 2.9.1)

Possible data delivery events include Delivery to PDS, and Acceptance by PDS. Each delivery event can have one or more event statuses which are captured by respectively, Delivery_Status and Acceptance_Status. On integration with the PDS Notification system, these delivery events will be added to the already existing Release and Archive events.

L4.DT.DP.2 – The Delivery Tracking System shall be configured based on a data delivery plan that has been negotiated between a Data Provider and a PDS Node. (UC-1.1, 2.2.2)

L5.DT.DP.2.1 – The Delivery Tracking System shall maintain provisional data delivery event schedules.

L5.DT.DP.2.2 – The Delivery Tracking System shall maintain attributes that characterize the data delivery (e.g. product type, orbit number, etc).

L5.DT.DP.2.3 – The Delivery Tracking System shall maintain the predicted number of products and the predicted size of the data delivery

L5.DT.DP.2.4 – The Delivery Tracking System shall maintain specifications as to what constitutes an acceptable data delivery

L5.DT.DP.2.5 – The Delivery Tracking System shall maintain the version of PDS data dictionary and local data dictionaries associated with the data delivery.

L5.DT.DP.2.6 – The Delivery Tracking System shall maintain a list of authorized actors who should be informed of data delivery events associated with the delivery plan.

L5.DT.DP.2.7 – The Delivery Tracking System shall allow a PDS Node to configure the tracking system with the delivery plan and to signal an event to notify the authorized actors that a configuration is complete.

L5.DT.DP.2.8 – The Delivery Tracking System shall allow modification of a data delivery plan. (2.2.2)

Modifying the delivery plan includes allowing the permanent removal of a data delivery from the plan.

L6.DT.DP.2.8.1 – The Delivery Tracking System shall allow Data Provider to remove a data delivery from the delivery plan and set the delivery status to "REMOVED".

L6.DT.DP.2.8.2 – The Delivery Tracking System shall be able to signal an event to notify authorized actors of a change of delivery status to "REMOVED".

L6.DT.DP.2.8.3 – The Delivery Tracking System shall allow a Data Provider to change the delivery plan in general.

L4.DT.DP.3 – The Delivery Tracking System shall maintain the delivery status of a data delivery from a Data Provider to a PDS Node. (UC-1.2, 2.2.2)

L5.DT.DP.3.1 – The Delivery Tracking System shall allow a Data Provider to signal a pending delivery event to notify the PDS Node that the data delivery is ready for transfer to begin.

L5.DT.DP.3.2 – The Delivery Tracking System shall allow a PDS Node to update the delivery status to "RECEIVED" if the transfer was successful. (UC-1.2)

L5.DT.DP.3.3 – The Delivery Tracking System shall allow a PDS Node to update the delivery status to "UNSUCCESSFUL" if the transfer was not successful. (UC-1.2)

L4.DT.DP.4 – The Delivery Tracking System shall be able to signal an event to notify authorized actors of a change in the delivery status. (UC-6, 2.2.2)

L4.DT.DP.5 – The Delivery Tracking System shall allow a PDS Node to accept, reject or mark as incomplete a data delivery depending on whether the deliver is acceptable according to negotiated delivery parameters. (2.2.2)

> **L5.DT.DP.5.1 –** The Delivery Tracking System shall maintain an acceptance status that can be set by a PDS Node to ACCEPTED, INCOMPLETE or REJECTED. (UC-1.3, UC-6, 2.2.2, 2.9.1)
>
> > **The possible reasons for rejection include the delivery not being PDS compliant, ITAR restricted, etc**
>
> **L5.DT.DP.5.2 –** The Delivery Tracking System shall maintain an acceptance status note that a PDS Node can use to provide additional information about the acceptance status.

L4.DT.DP.6 – The Delivery Tracking System shall be able to signal an event to notify authorized actors of a change in a data delivery's acceptance status. (UC-6, 2.2.2)

L4.DT.DP.7 – The Delivery Tracking System shall allow the withdrawal of a data delivery. (2.2.2)

> **The assumption is that the data provider will re-submit the data delivery later.**
>
> **L5.DT.DP.7.1 –** The Delivery Tracking System shall allow a PDS Node to update the delivery status to "WITHDRAWN" if a data delivery has been withdrawn by the data provider.

5.3 Data Node Termination

L4.DT. NT.1 – The Delivery Tracking System shall support the dissolution of a Data Node and the transfer of the Data Note's holdings to a permanent PDS Node. (UC-3, 2.2.2)

> **The requirements in section 5.2 apply after the Data Node assumes the role of Data Provider.**

6 Archive Tracking Requirements

The following requirements are for archive tracking. In this document, "archive tracking" encompasses the tracking of data holdings in the PDS archive and are levied on the PDS to track data files, data products, data volumes, and data sets from the acceptance of the data at a PDS Node through to the submission of the data to the NSSDC. These requirements are on the PDS to track the files in the archive as a whole, spanning all nodes. An important feature suggested by these requirements allows for data deliveries to be successfully transferred to but not necessarily accepted by the NSSDC until reviewed. This “escrow” feature is implemented implicitly by differentiating between the transfer to and acceptance of a package by the NSSDC. In other words, a delivery can be successfully transferred to but not necessarily accepted until the NSSDC has ensured that the delivery meets certain criteria. These requirements are derived from level one, two, and three PDS requirements, PDS Tracking Use Cases, and PDS Policies.

6.1 General

L4.AT.GR.1 – The PDS shall be able to provide an accounting of all files in the archive (PDS Policy, 2.2.2 - see Policies Section)

L5.AT.GR.1.1 – The PDS shall track all files and their membership in data deliveries, data sets, volumes, data products from their acceptance into a node repository through to their delivery to the NSSDC. (UC-1, UC-2)

L6.AT.GR.1.1.1 – The PDS shall be able to track all files in a data delivery that have been ACCEPTED by a PDS Node.

L6.AT.GR.1.1.2 – The PDS shall be able to associate a file to a data product.

L6.AT.GR.1.1.3 – The PDS shall be able to associate a file to a volume.

L6.AT.GR.1.1.4 – The PDS shall be able to associate a file to a data set.

L6.AT.GR.1.1.5 – The PDS shall be able to track a collection of files prepared for distribution including both files from a data delivery as well as additional ancillary files.

L4.AT.GR.2 – The PDS shall be able to verify that all files in the archive are accessible. (4.1.1)

L4.AT.GR.3 – PDS shall preserve information required to verify the original file delivered to PDS node has not been corrupted. (4.1.1, L4.DI.3, L4.DI.4, L4.DI.5, L4.DI.6, L4.DI.7, L4.DI.9, L4.DI.10, L4.DI.11, L4.DI.12, L4.DI.13, L4.DI.16)

> **A possible implementation would be maintaining the original checksum value for each file and using that value throughout the end-to-end system.**

L4.AT.GR.4 – The PDS shall maintain status and location information for each data set and volume in the archive in a common catalog. (2.2.2, 2.6.3).

> **The PDS Catalog tracks data collections including data sets, volumes, and data deliveries. Specific information beyond data collection information to be tracked includes the curating_node_id, distribution_node_id, backup_node_id, archive_status, acceptance_status, delivery_status, nssdc_id, etc.**

L4.AT.GR.5 – The PDS shall notify actors about events including scheduled events associated with deliveries within or from the PDS archive and any change in delivery or acceptance statuses. (3.3.6)

> **This requirement focuses primarily on the NSSDC interface. The PDS Notification/Subscription system could be extended to fulfill this requirement. The notification of an actor will be triggered by an event. Various requirements in the following sections will signal events. The delivery plan will provide a list of actors to be notified for specific events. The subset of "authorized" actors will also be provided in the delivery plan.**

L5.DT.GR.5.1 – The PDS shall notify an actor for the following types of events.
Delivery Plan Modification
Pending Delivery Event
Delivery Event
Status Changes (i.e. Delivery Status and Acceptance Status)

6.2 Data Transfer between PDS Nodes

L4.AT.TR.1 – The PDS shall be able to track the transfer of files from one PDS Node archive repository to another (2.2.2, 2.6.3).

L5.AT.TR.1.1 – The PDS shall be able to create a delivery manifest for files that have been prepared for transfer to another node archive repository.

L5.AT.TR.1.2 – The PDS shall be able to validate that all files have been transferred from one PDS Node archive repository to another using a delivery manifest.

L5.AT.TR.1.3 – The PDS shall provide the information necessary to update a common catalog after a data collection has been transferred from one PDS Node archive repository to another. (UC-2, 2.6.3)

Possible information includes the curating_node_id, distribution_node_id, backup_node_id, archive_status, acceptance_status, delivery_status, nssdc_id, etc. Data collections include data sets, volumes, and deliveries.

6.3 Preservation of Data at NSSDC

L4.AT.NS.1 – The PDS shall be able to track the transfer of a data set from a PDS Node to the NSSDC for long term preservation. (UC-3, L4.DI.14, 2.2.2, 2.9.1)

Possible data delivery events include Transfer to NSSDC. Each delivery event can have one or more event statuses which are captured by the NSSDC_Transfer_Status. On integration with the PDS Notification system, these delivery events will be added to the already existing Release and Archive events.

L5.AT.NS.1.1 – The PDS shall maintain information about the proposed contents of a data set delivery and the associated transfer schedule.

L5.AT.NS.1.2 – The PDS shall maintain the delivery status of a data set transferred from a PDS Node to the NSSDC for preservation.

L6.AT.NS.1.2.1 – The PDS shall be able to partition large data sets for delivery to the NSSDC.

Some large PDS Data Sets might require more than one AIP.

L6.AT.NS.1.2.2 – The PDS shall notify the NSSDC that a data delivery is ready for transfer to begin.

L6.AT.NS.1.2.3 – The PDS shall set the delivery status set to "RECEIVED" if the transfer to the NSSDC was a success.

The NSSDC notifies the PDS Node of the delivery status.

L6.AT.NS.1.2.4 – The PDS shall set the delivery status to "UNSUCCESSFUL" if the transfer to the NSSDC was deemed to not be successful.

The NSSDC notifies the PDS Node of the delivery status.

L6.AT.NS.1.2.5 – The PDS shall notify all authorized actors of the result of the delivery to NSSDC, including a request for retransmission if necessary.

L5.AT.NS.1.3 – The PDS shall maintain the acceptance status of a data set transferred from a PDS Node to the NSSDC

L6.AT.NS.1.3.1 – The PDS shall set the acceptance status to "ACCEPTED" if the NSSDC determines that the data set delivery is acceptable according to the negotiated delivery parameters.

The NSSDC notifies the PDS Node of the status.

L6.AT.NS.1.3.2 – The PDS shall obtain the NSSDC_ID of the data delivery for updating the PDS Catalog. (2.6.3)

L6.AT.NS.1.3.3 – The PDS shall set the acceptance status to "REJECTED" if the NSSDC determines that the data set delivery is unacceptable according to the negotiated delivery parameters.

The NSSDC notifies the PDS Node of the status.

L6.AT.NS.1.3.4 – The PDS shall notify all authorized actors of the acceptance status of the delivery and a description of any problems if necessary.

6.4 Data Inventory Reporting

L4.AT.IR.1 – The PDS shall allow the generation of reports on its data holdings including the location, status, and integrity of data deliveries, data volumes, data sets, data products and their associated files. (DI UC-6, 2.2.2, 2.6.3, PDS Policy, 2.2.2 - see Policies Section)

6.5 PDS Node Delivery of a Data Set

PDS Nodes need to schedule events and notify actors of the events and status changes associated with the delivery of a data set to the community. For example, a node might want to modify and then deliver a new version of a data set. The archive tracking system might want to consider a light-weight delivery plan for this use case. The subscription and notification system must have requirements for this use case.

7 Level 3 Requirements

The following existing level three requirements relate to tracking PDS data and are referenced in the requirements defined in Section 5.

2.2.2 PDS will track the status of data deliveries from data providers through the PDS to the deep archive

2.4.5 PDS will track the status of each peer review

2.5.2 PDS will implement procedures for accepting archival data

2.6.3 PDS will integrate the catalog with the system for tracking data throughout the PDS

2.8.2 PDS will maintain a distributed catalog system which describes the holdings of the archive

2.9.1 PDS will accept and distribute only those items which are not restricted by the International Traffic in Arms Regulations (ITAR)

8 Policies

1) Each node is responsible for periodically verifying the integrity of its archival holdings based on a schedule approved by the Management Council. Verification includes confirming that all files are accounted for, are not corrupted, and can be accessed regardless of the medium on which they are stored. Each node will report on its verification to the PDS Program Manager, who will report the results to the Management Council. - Planetary Data System Management Council Meeting Minutes, 29-30 November 2006.