Introduction

Scope
This document provides the information required to package PDS volumes into Archive Information Packages (AIPs) and deliver them to NSSDC. That information includes:
- A description of the relationships between NSSDC data collections and PDS datasets and volumes
- Descriptions of various identifiers used during the AIP generation and delivery process
- A step-by-step description of the process

This document is intended as a startup guide and does not supply information about the design of the packaging software or provide a complete error code listing. It does list the most likely MPGA processing error messages and their causes.

PDS Archive Projects
Each installation of the MPGA software is uniquely identified by a four character Archive Project ID stored in an MPGA configuration file. Twelve PDS nodes have been identified as potential AIP producers and have been been assigned a single Archive Project ID. (See appendix A for a list of Archive Projects and their corresponding PDS Nodes.) If a PDS node desires to run MPGA on multiple machines, then additional Archive Project IDs must be assigned by NSSDC.

The AIP Archival Storage ID
Every AIP has a unique Archival Storage Identifier, calculated by MPGA during AIP generation and preserved in the AIP. The ID of the Archive Project that created an AIP is incorporated into the AIP's Archival Storage ID. This identifies the data provider of the AIP and ensures that the Archival Storage ID will not be duplicated elsewhere. The Archival Storage ID is in the form *AAAAnnnnnnnnnn* where *AAAA* is the Archive Project ID and *nnnnnnnnnn* is an 10 digit number derived from the current date and time by MPGA.

Relationships between NSSDC Collections, PDS Datasets, and PDS Volumes

An NSSDC collection contains one or more PDS datasets. A PDS dataset may only belong to one NSSDC collection. Any PDS volume containing PDS datasets (as listed in the VOLDESC.CAT file) belonging to an established NSSDC collection, or a subset of those datasets, is a member of that NSSDC collection. A PDS volume may belong to only one NSSDC collection and may contain only PDS datasets from a single defined NSSDC collection. Existing NSSDC collections may be modified. A PDS dataset that does not yet belong to an NSSDC collection may be added to an existing collection. Interaction with an NSSDC curation scientist is required to establish or modify an NSSDC collection.

Outline of the AIP Generation and Delivery Process

The process can be broken down into five steps:

1. Obtain an NSSDC FTP account for AIP delivery. This setup step will need to be performed only once for each PDS Archive Project.

2. Install the Multifile Package Generator and Analyzer (MPGA) software. A software installation must be performed for each PDS Archive Project.

3. Obtain NSSDC collection IDs. This step needs to be performed before a given PDS dataset can be packaged into an AIP for the first time.

4. Create AIPs with MPGA. This step is performed every time a PDS volume is packaged into an AIP.

5. Deliver AIPs to NSSDC. AIPs produced by the PDS nodes must be delivered individually.

Step 1: Obtain an NSSDC FTP account for AIP delivery

Delivery of AIPs to NSSDC will be via FTP. An individual FTP account has been created at NSSDC for each PDS Archive Project. The designated contact individual for each PDS Archive Project should contact the NSSDC system administrator to obtain FTP account information.

> Teresa Hall-Jackson
> NSSDC, Code 690.1
> NASA Goddard SFC
> Greenbelt, MD 20771
> +1 (301) 286-3299
> jacksont@pop600.gsfc.nasa.gov

Step 2:   Install the MPGA software

Software Delivery

NSSDC delivers MPGA software via FTP to the Engineering node in tar files, one tar file for each PDS Archive Project. The tar files are named according to the Archival Project Node for which it was intended, for example, the MPGA system configured for *Planetary Atmospheres* is the tar file named:
MPGA-OPSv6.0-PDS_ATM1-PLANETARY_ATMOSPHERES-1.tar (See appendix A for a list of Archive Projects and their corresponding PDS Nodes). The Engineering node distributes the MPGA tar files to other PDS nodes. The MPGA software is self-contained and the required executables, run scripts and configuration files are available within the delivered tar files.

System Requirements

The MPGA software requires the following environment:
- Linux-based operating system. (Tested with Red Hat and Fedora)
- An estimated 500 megabytes of disk space to hold the software and support operations during runtime for most data sets
- Disk space 2x the size of the input PDS data to support AIP generation

- Disk space 2x the size of the input AIP to support AIP unpacking

Installation & Verification
Unpack the tar file provided by the Engineering node (the tar ball filename will be named to identify *your* Archival Project Node and will begin with the prefix *MPGA-OPSv6.0-*, replacing the tar filename in the example below):

tar –xvf MPGA-OPSv6.0-PDS_ATM1-PLANETARY_ATMOSPHERES-1.tar

The directory *MPGA/* is extracted. This directory contains the system components, all arranged across the directories *sys/*, *bvers/*, *data/* and *run/*.

The additional directory, *stage_out/,* is where the AIPs and output log and diagnostic files will be created.

Step 3: Obtain NSSDC Collection IDs

Interactions with the NSSDC Curation Scientist
NSSDC collection IDs for PDS data are issued by the NSSDC curation scientist for PDS, Dave Williams. His contact information is:

> Dave Williams
> NSSDC, Code 690.1
> NASA Goddard SFC
> Greenbelt, MD 20771
> +1 (301) 286-1258
> dave.williams@gsfc.nasa.gov

The submitting PDS node will need to supply the NSSDC curation scientist with the VOLDESC.CAT file and the AAREADME file from a representative PDS volume. The DATASET.CAT file would be useful but is not required. The curation scientist will verify that the information represents a new collection or a modification to an existing collection. A new collection ID or confirmation that an existing NSSDC collection has been modified will be provided to the submitting PDS node. The process will usually take three to four days.

When to Establish or Modify an NSSDC Collection
A valid NSSDC collection ID is required to generate an AIP with MPGA. The NSSDC collection ID is one of the parameters supplied to MPGA and is preserved in the resulting AIP. For any PDS volume being packaged, one of the following conditions will apply:
- None of the PDS datasets on the PDS volume belong to an NSSDC collection. A new NSSDC collection must be established before packaging the volume. The new collection provided by NSSDC should be used when generating the AIP.
- The PDS volume contains PDS datasets that belong to a single defined NSSDC collection. The existing NSSDC collection ID should be used when generating the AIP.
- The PDS volume contains PDS datasets that belong to a defined NSSDC collection and also contains datasets that do not belong to an NSSDC collection ID. The additional PDS datasets must be added to the existing NSSDC collection before packaging the volume.

Note on NSSDC Collection ID Validation

AIPs containing PDS volumes will be checked at the NSSDC to verify that the PDS datasets contained in the volume belong to a single defined NSSDC collection. An AIP that does not contain members of single NSSDC collection will not be ingested. The PDS provider will be notified in the event of an AIP failing this check. A planned release of the MPGA software will eliminate the need to perform the manual dataset check at NSSDC by using a curation scientist-provided dataset matrix to determine the correct NSSDC collection ID when AIPs are generated. This will also eliminate the need for manual entry of the NSSDC collection ID in the MPGA input list file by the PDS provider.

Step 4: Creating AIPs with MPGA

Packager Setup

The MPGA packager enables packaging of a volume directory into an Archive Information Package (AIP). The run script is named *START-JOB--MPGA-packager_OP-VERS_v6_0.com* and its configuration file is *stage_in_glst_directory.lst*. Both files are located in the *MPGA/run/* directory.

The normal usage is to run the packager from the *MPGA/run/* directory. In the simplest configuration the *stage_in_glst_directory.lst* file contains one line, with a single <tab> as the delimiter between the fields and terminated with a new line character. The field values are:

DATA_SET_DIRECTORY_PATH/	OUTPUT_AIP_DIRECTORY_PATH/	NSSDC_COLLECTION_ID

and each field should not be bound by quotes (i.e: do not use quotes – " "). A correct example:
  /data/dmgsc_1001/mgsc_1122/   ../stage_out/   PSPG-00999

Each field is described below:
- DATA_SET_DIRECTORY_PATH/: (in example: /data/dmgsc_1001/mgsc_1122/) the data directory where the data set resides. This can be relative from the current working directory or can be an absolute directory from root "/", such as */data/volume_name/*. The last slash "/" at the end is required
- OUTPUT_AIP_DIRECTORY_PATH/: The output data directory where the AIP is to be created. AIP filenames are generated by the packager and AIPs are put here. The last slash "/" at the end is required. To be consistent with the script used to call the MPGA AIP reconstituter, the *stage_in_glst_directory.lst* file as delivered specifies the directory *../stage_out/* for this variable. You can use the directory *MPGA/stage_out* as is or you can optionally create a soft link with the same name as in the following example:
  # cd MPGA/   (go into the *./MPGA/* directory)
  # dir /mystage_area/   (an example of some output directory of your choice)
  # rm –R ./stage_out   (the *stage_out/* directory will be removed!)
  # ln -s /mystage_area ./stage_out   (now, the AIPs will actually be put in your directory of choice and you can keep the OUTPUT_AIP_DIRECTORY_PATH/ value equal to ../stage_out/)
- NSSDC_COLLECTION_ID: (in example: PSPG-00999) NSSDC supplied collection identifier, stored in the AIP as an attribute called "PRIMARY_COLLECTION_ID"

Running the Packager

Package an AIP by running the script *START-JOB--MPGA-packager_OP-VERS_v6_0.com*. It must be run from inside the *MPGA/run/* directory as follows:

cd MPGA/run

./START-JOB--MPGA-packager_OP-VERS_v6_0.com

The packager runs as a background process. The elapsed time varies depending on the size of a data volume. Three output files are generated by this script:

- AIP file - located at *MPGA/stage_out* as defined in *MPGA/run/stage_in_glst_directory.lst*. Every AIP produced will have a unique file name in the form *AAAAnnnnnnnnnn_*package.aip where *AAAAnnnnnnnnnn* is the Archival Storage ID of the AIP.
- Diagnostic file - located at *MPGA/stage_out* as defined in *MPGA/run/START-JOB--MPGA-packager_OP-VERS_v6_0.com* as –output. Here, it is named *out-packager.diag*. This file may be renamed to one of your choice. You must remember to send it with the AIPs created to NSSDC for Ingest. There is one diagnostic file per run, so if it is a fixed name be careful to store it so it is not overwritten during the next packaging run
- Log file located at *MPGA/stage_out* as defined in *START-JOB--MPGA-packager_OP-VERS_v6_0.com* as –log. Here, it is named *out-packager.log*, and you should treat it just like the Diagnostic file described above

Packager Operation

Once the packager has completed it will output the words *SUCCESS* or *FAILURE* to the screen. To verify that an AIP was created look in the output directory and make sure there is an AIP file with the suffix *.aip* plus the *.log* and *.diag* files. This is important because in some error conditions the software can issue a *SUCCESS* to the screen, but the AIP is not created due to some of its more advanced modes of operation. If the AIP is not created, you can look in the *.diag* file to get a description of what happened. The most common errors are the following (not a complete list):

- could not write a char to file – this means the software ran out of disk space
- could not create log/output file – output directories may be wrong or have insufficient privileges
- bad directory – input or output directories may be wrong or have insufficient privileges
- expected ascii but found binary – a data file in the input PDS volume, after its type was identified, contained invalid character ranges
- zero byte file found – a data file in the input PDS volume contained no data
- voldesc & dataset catalog errors – some kind of problem while parsing the voldesc or dataset catalogs for the required attributes, including the following:

VOLUME_ID	VOLUME_VERSION_ID	START_TIME	STOP_TIME

Step 5: Delivering AIPs to NSSDC

This step is performed every time a PDS volume is packaged into an AIP.

Delivery Products

AIPs are delivered to NSSDC in tar files via FTP. After an AIP is created, the submitting PDS node must create a tar file containing the following files:

- a single AIP
- the MPGA log file generated when creating the AIP
- the MPGA diagnostic file generated the AIP

If the AIP being delivered is a replacement for an earlier AIP, then include:

- a text file named replacement.txt containing the Archival Storage ID of the AIP being replaced.

An AIP may replace, in whole, a single previously delivered SIP. Partial AIP replacements or replacements of multiple AIPs by a single AIP are not currently supported.

Tar file names should be in the form *AAAAnnnnnnnnnn*.tar where *AAAAnnnnnnnnnn* is the Archival Storage ID of the AIP. The completed tar file should be delivered to NSSDC using the FTP account provided to the submitting PDS Archive Project.

Ingest Notification

NSSDC will notify the submitting PDS Archive Project when the AIP has been successfully ingested. An email message will be sent to the PDS Archive Project identified in Appendix A.

Note on the AIP delivery process
The AIP delivery process outlined above is an interim process that will change as NSSDC develops a standard approach, with reduced manual intervention, for transfer of Submission Information Packages from data providers to NSSDC.

Appendix A: PDS Archival Project List

Archive Project ID (FTP account name)	Node Name	Contact Name	Contact Email
ATM1	Planetary Atmospheres	Lyle Huber	lhuber@nmsu.edu
ENX1	Engineering	Emily Law	Emily.Law@jpl.nasa.gov
GEO1	Geosciences	Tom Stein - Thomas C. Stein	stein@rsmail.wustl.edu
IMA1	Imaging – USGS	Chris Isbell - Christopher E Isbell	cisbell@usgs.gov
IMA2	Imaging - JPL	Mike McAuley - Michael McAuley	myche@jpl.nasa.gov
NAI1	Navigation Ancillary Information Facility	Boris Semenov	Boris.V.Semenov@jpl.nasa.gov
PPI1	Planetary Plasma Interactions – UCLA	Todd King	tking@igpp.ucla.edu
PPI2	Planetary Plasma Interactions - Iowa	Todd King	tking@igpp.ucla.edu
RIN1	Planetary Rings	Mitch Gordon	mgordon@seti.org
RSX1	Radio Science	Dick Simpson	rsimpson@magellan.stanford.edu
SBN1	Small Bodies - Maryland	Anne Raugh	raugh@astro.umd.edu
SBN2	Small Bodies - Planetary Science Institute	Carol Neese	neese@psi.edu

Appendix B: Instructions for Reconstituting AIPs

Reconstitution of AIPs is not required for delivery to the NSSDC.

The MPGA reconstituter, *START-JOB--MPGA-reconstitute_aips_OP-VERS_v6_0.com* unpacks AIPs. It is also run from the *MPGA/run/* directory as follows:
```
# cd MPGA/run
# ./START-JOB--MPGA-reconstitute_aips_OP-VERS_v6_0.com
```

No configuration file is required to run the reconstituter. The reconstituter can unpack multiple AIPs as long as they are at the same location.

Note: It's recommended to examine the reconstituter script: *START-JOB--MPGA-reconstitute_aips_OP-VERS_v6_0.com* prior to the execution to ensure the definition of the *–aip_wildcard* option points to the correct source location of AIP files to be unpacked.

Two outputs are generated as a result of running the reconstituter:
• An unpacked directory with the same content of the data volume before it was packaged into an AIP. The name of the directory is the Archive Storage ID (ASID) string that was calculated by the packager, such as TEST0000000037.dat.dir.
• An attribute file that corresponds with the unpacked data volume. It has the same name as the above directory, with ".att" as the extension.