

**NASA SPACE SCIENCE DATA COORDINATED ARCHIVE
ARCHIVE PLAN FOR 2015 – 2016**

Ed Grayzeck
National Space Science Data Center
Greenbelt, Maryland 20771
2015-03-18

ABSTRACT

This archive plan shows that the NASA Space Science Data Coordinated Archive (NSSDCA) expects to accept ~350 TB of data into the archive in 2015 and ~300 TB in 2016.

1. INTRODUCTION

NSSDCA provides a vital service as NASA's permanent multi-disciplinary Space Science archive. Its curation activities are essential to ensure that space science data will continue to be available and usable into the indefinite future. The need for long-term curation arises because in most cases the full value of any set of data cannot be known in advance. New science discoveries or changes in research and exploration priorities may make older data, seldom noticed before, suddenly highly relevant.

This archive plan summarizes the expected data inflow to NSSDCA (note the Acronym list at the end of this document) for the years 2015-2016. NSSDCA archive plans cover 2 years and are updated annually. Note these are estimates for planning purposes, not exact data projections.

1.1 Levels of Service

NSSDCA accepts and archives data under four levels of service, summarized in Table 1 below. The most familiar is the Permanent Archiving of data, but, as defined in MOUs with various data providers, it also provides Backup service, mostly for other Archives. The Analog Archive includes photos, maps, microfilm, microfiche, documents, etc, some analog copies of digital data and others supporting metadata; it is included in this list for completeness.

Table 1. NSSDCA Archival Storage Services	
Permanent Archive: AIPs	Preservation of digital data in Archival Information Packages delivered by a data producer or created at NSSDCA. AIPs are re-written to new media within six years. Data is disseminated by NSSDCA if not available through an active archive or per MOU.
Permanent Archive: non-AIP digital data	Preservation of non-packaged data on various media types. Data will eventually be migrated from legacy media to AIPs, though no media refresh will be made in the meantime. Data is disseminated by NSSDCA if not available through an active archive or per MOU.
Backup	Storage of digital data at climate-controlled off-site facility to support another archive's contingency plan per MOU. Data will not be disseminated by NSSDCA.
Analog Archive	Preservation of analog data on a variety of media with selected refreshment and selected digitization. Selected retention of original analog data after digitization. Data are copied and disseminated by NSSDCA.

1.2 Archive Information Packages (AIPs)

In Table 1 NSSDCA's permanent archive is digital data that is stored either as AIPs or not. The non-AIP digital data is stored on off-line media and tracked by the media on which it resides. The portion of the data stored near-line in LTO jukeboxes has been growing since 2000 and includes all new data inflows received via electronic transfer, plus some legacy data collections; it is notable not because of its media, but because those data are stored on LTOs in Archive Information Packages (AIPs).

An AIP is a single file container that holds one or many science data files, a number of attributes about each file that help NSSDCA manage its AIPs, and pointers to all of the supporting

documentation, including calibration information. Ideally this is enough information to allow a user to be able to utilize the data independently of the archive and the original producer of the data. No reformatting of the science data files is performed unless record boundaries need to be retained and are not already in the byte stream. Any files that are transformed may be returned to their original state using the NSSDCA defined attributes. Additionally, AIPs are media independent and platform independent, making AIPs the mechanism for long-term preservation. NSSDCA plans to convert most of the non-AIP data in the permanent archive to AIPs.

1.3 Active Archives

NASA has established a set of Active Archives, which receive data from missions and provide electronic access to the missions' data, along with documentation and tools for accessing and using the data. NSSDCA's mission is to accept data from the Active Archives or sometimes directly from missions, then provide long-term curation of the data. This is a critical service, since the full value of any set of data cannot be known in advance. New science discoveries or changes in research and exploration priorities may make older data, seldom requested, suddenly highly relevant.

2.0 ARCHIVE PLAN

The revised, detailed Archive Plan for NSSDCA for 2015-2016 is given below (next page) in Table 2. Table 2 lists the node/archive/mission and the estimated data volume to be delivered each year. Also included are the level of service (Permanent Archive - with or without AIPs - or Backup) defined by MOU for each data collection and the discipline (Astrophysics, Heliophysics, Planetary & Lunar) for each. For archives which require Backup service, the data volumes expected from individual missions are combined and listed in the table by the name of the archive, i.e. HEASARC, IRSA, MAST, PDS, and SPDF.

The totals in Table 2 show that NSSDCA is planning for ~348 TB of data arriving at the archive in 2015 and ~296 TB in 2016. By discipline, planetary missions will make the largest contributions with the PDS data nodes delivering nearly half the data. The single data provider expected to make the greatest data deliveries expected are those from the PDS Imaging Node, which is archiving data from the Lunar (LRO) and Mars (MRO) Reconnaissance Orbiters. The summary of the Table 2 entries by level of service and by discipline is given in Tables 3a and 3b, respectively.

TABLE 2. Summary of data expected at NSSDCA, 2015-2016.

Project	Service Level* & Discipline [†]		Expected Data Volume (TB)		Totals (TB)
			2015	2016	
<i>PDS Nodes</i>					
ATMOS	A	P	2	3	5
GEO	A	P	60.2	0.9	61.11
IMAGING	A	P	73	75	148
NAIF	A	P	0.6	0.7	1.3
PPI	A	P	10	15	25
RINGS	A	P	0.6	0.6	1.2
SBN-PSI [‡]	A	P	25	25	50
SBN-UMD	A	P	1.5	1	2.5
<i>Missions</i>					
FERMI	B	A	7	7	14
RHESSI	B	H	6.2	6.2	12.4
WINDWAV	B	H	<1	<1	<1
WISE	B	A	0	0	0
<i>Active Archives</i>					
HEASARC	B	A	90	90	180
IRSA	B	A	0	0	0
MAST	B	A	0	0	0
SPDF	B	H	72	72	144
TOTALS			348.1	296.4	644.5

*Service Levels: A = Permanent Archive (AIP or non-AIP); B = Backup.

[†]Discipline: A = Astrophysics; H = Heliophysics; P = Planetary & Lunar.

[‡]Subnode of SBN-UMD (PDS Small Bodies Node at the University of Maryland)

TABLE 3a	
Service Level	TB (2015-2016)
Permanent Archive	294
Backup	350

TABLE 3b	
Discipline	TB (2015-2016)
Astrophysics	194
Heliophysics	156
Planetary & Lunar	294

Glossary

AIP	Archive Information Package
GB	Gigabyte
HEASARC	High Energy Astrophysics Science Archive Research Center
MAST	Multi-mission Archive at Space Telescope Science Institute
NSSDC	National Space Science Data Center (now NSSDCA)
NSSDCA	NASA Space Science Data Coordinated Archive
PDS	Planetary Data System
PDS_ATM	PDS Atmospheres Node
PDS_GEO	PDS Geosciences Node
PDS_IMG	PDS Imaging Node
PDS_NAI	PDS Navigation and Ancillary Information Node
PDS_PPI	PDS Planetary Plasma Interactions Node
PDS_PSI	PDS Planetary Science Institute
PDS_RING	PDS Planetary Rings Node
PDS_SBN	PDS Small Bodies Node
RHESSI	Reuven Ramaty High Energy Solar Spectroscopic Imager
SPDF	Space Physics Data Facility
TB	Terabyte
WISE	Wide-field Infrared Survey Explorer