

NASA SPACE SCIENCE DATA COORDINATED ARCHIVE

ARCHIVE PLAN FOR 2014 – 2015

Ed Grayzeck
NASA Space Science Data Coordinated Archive
Greenbelt, Maryland 20771
2014-03-31

ABSTRACT

This archive plan shows that the NASA Space Science Data Coordinated Archive (NSSDCA) expects to accept ~400 TB of data into the archive in 2014 and ~600 TB in 2015.

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 2014-2015. These are estimates for planning purposes, not exact data projections.

This is the successor to earlier plans covering 3-4 years each and updated at that same interval. With this version we have chosen to plan for two years at a time and update the plan annually, so the estimates should be more accurate and more relevant for planning.

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.

Given the prevalence of incoming data from the PDS nodes and subnodes for 2014-15, we have reorganized Table 2 by data contributors rather than by missions as in previous plans.

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 as AIPs.

An Archive Information Package (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 preferred delivery and storage means. In the long-term most of the non-AIP data in the permanent archive is planned to be converted 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 2014-2015 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 ~400 TB of data arriving at the archive in 2014 and ~600 TB in 2015.. 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. Clearly, planetary missions dominate; their coming contribution to the NSSDCA is estimated to be over 800 TB in 2014-15.

TABLE 2. Summary of data expected at NSSDCA, 2014-2015.

Project	Service Level* & Discipline+		Expected Data Volume (GB)		Totals (GB)
			2014	2015	
<i>PDS Nodes</i>					
PDS_ATM	A	P	3	4	7
PDS_GEO	A	P	14	47	61
PDS_IMG	A	P	230	400	630
PDS_NAI	A	P	0.5	0.6	1.1
PDS_PPI	A	P	4	4	8
PDS_PSI	A	P	60	40	100
PDS_RINGS	A	P	0.6	0.4	1.0
PDS_SBN	A	P	1	2	3
<i>Missions</i>					
FERMI	B	A	7	7	14
RHESSI	B	H	1	1	2
WIND/WAVES	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	<1	<1	<1
SPDF	B	H	0	0	0
TOTALS			411	596	1008

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

+Discipline: A = Astrophysics; H = Heliophysics; P = Planetary & Lunar.

TABLE 3a	
Service Level	TB (2014-2015)
Permanent Archive	811
Backup	196

TABLE 3b	
Discipline	TB (2014-2015)
Astrophysics	194
Heliophysics	2
Planetary & Lunar	811

Glossary

AIP	Archive Information Package
GB	Gigabyte
HEASARC	High Energy Astrophysics Science Archive Research Center
IRSA	Infrared Science Archive
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 Facility
PDS_PPI	PDS Planetary Plasma Interactions Node
PDS_PSI	PDS Planetary Science Institute (sub-node of Small Bodies)
PDS_RINGS	PDS Rings Node
PDS_SBN	PDS Small Bodies Node
RHESSI	Reuven Ramaty High Energy Solar Spectroscopic Imager
SPDF	Space Physics Data Facility
TB	Terabyte
WIND	NASA Spacecraft to study solar Wind (not an acronym)
WAVES	Plasma Waves instrument on WIND (not an acronym)
WISE	Wide-field Infrared Survey Explorer