Version 37 NUG

# NSSDC Activities, Accomplishments, and Future Directions

# 1 Activities and Accomplishments, FY 2004-2005

## *1.1 NSSDC Overview*

The NSSDC was established by NASA in the mid-60s to provide for the preservation and dissemination of scientific data from NASA missions. It now manages an archive of data from 1,546 investigations carried on 545 spacecraft over the past 47 years. The digital component of the archive amounts to over 36 TB distributed across 2,300 distinct "datasets." In recent years the data volume is growing at about 3 TB/yr but is expected to accelerate substantially to 10-15 TB/yr due largely to planetary missions. There is also a legacy, non-digital, component of the archive.

In a new approach to dissemination of data in the last few years, distributed, active archives began to be established in the Planetary, Astrophysics and Solar sub-disciplines. They took over responsibility for working with new projects to acquire and distribute data for community researchers while NSSDC remained vital in acquiring and disseminating space plasma physics data. A few mission-specific archives also were established in the space physics area. NSSDC continues as the permanent archive for all space science data, interacting with active archives per MOUs that clarify relationships and responsibilities. NSSDC also uniquely provides access to the taxpaying public for certain data and information of interest to them.

Beginning in fiscal year 2003, the NSSDC budget was formally split to provide separate funding and a more distinctive identity for the Space Physics active-archive component known as the Space Physics Data Facility (SPDF). More recently, a Goddard reorganization assigned the NSSDC and SPDF to different parent organizations. It should be recognized by readers of this document that many systems and activities long identified with NSSDC, such as OMNI data and its OMNIWeb interface and empirical modeling now fall into the SPDF domain and are not further addressed herein.

As a general policy, NSSDC acquires data from active archives for long term preservation and provides it back to them if requested. NSSDC acquires data from projects and researchers, for long term preservation, when those data are not of interest to other active archives, and it makes such data available to researchers and the general public. This also applies to much legacy data that pre-dated the formation of active archives. In accordance with particular MOUs, NSSDC also handles data dissemination requests that are beyond the mission of particular active archives, such as responding to large data transfer requests and supporting access by the general public. In a similar vein, NSSDC also provides a remote backup capability for some active archives. NSSDC also provides data to foreign requesters (scientific researchers, students, and the general public) through the mechanism of the World Data Center for Satellite Information and is the only NASA part of the WDC system for the past 30+ years.

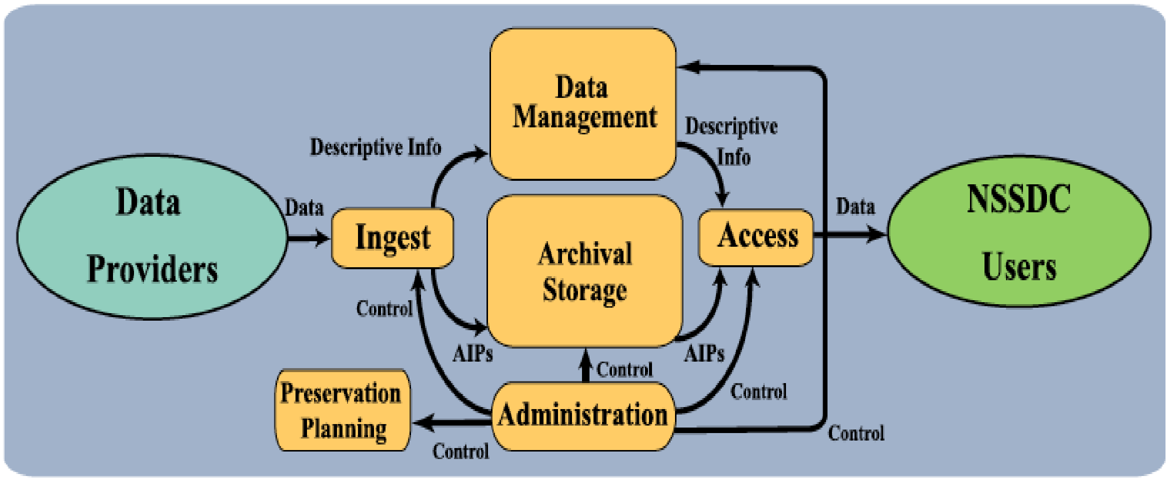


Figure 1: The internationally-standardized OAIS model for data centers has been followed in the evolution of the NSSDC's architecture to ensure the most effective use of data and resources.

NSSDC operates within the context of the conceptual framework established by the internationally-standardized Reference Model for an Open Archival Information System (OAIS) as illustrated in the functional model overview shown in Figure 1. In this model, Data Providers provide data and supporting information to the Ingest function of the archive. They arrive, either conceptually or actually, via Submission Information Packages (SIPs). The Ingest function turns these into one or more Archival Information Packages (AIPs) for preservation within the Archival Storage function.. The AIPs contain all of the science data and contain, or points to, all of the supporting documentation, including calibration information, that is required to allow a user to be able to use the data independently of the archive and of the original producer of the data. In some cases, the AIP might be generated directly by the data provider, as is the case with the IMAGE spacecraft. This is the preferred approach as it allows automation of the ingest process.

The ingest function also extracts, and/or generates, the descriptive information that will be needed to support finding of the data by external customers (NSSDC Users). It provides this to the Data Management function. NSSDC users interact with the Access function, which uses the descriptive information in Data Management, to find the data of interest. Access translates this into requests to the Archival Storage function for the AIPs needed, and then it processes the AIPs to satisfy the user's request. The resulting data and documentation returned to the NSSDC user is provided, either conceptually or actually, in the form of Dissemination Information Packages (DIPs).

This is managed, on a day-to-day basis, by an Administrative function. This function has many responsibilities, both internally and externally, as described later. It is supported by the Preservation Planning function with recommendation on standards and processes to be used, and with information on the state of technology used by the NSSDC's data providers and users.

The OAIS functional model provides a convenient way to discuss the activities of an archive and is used to organize discussion of NSSDC's activities and accomplishments in the following sections.

In addition to acquiring, preserving, and disseminating data, NSSDC is playing a growing role in supporting Space Science's (NASA Headquarters and community) needs for multi-archive guidance, standards, and services. These are discussed in more detail in subsequent sections.

## *1.2 Data ingest*

Ingest, as shown in Figure 2, provides the services and functions to accept data and documentation, often in the form of SIPs, from a variety of data providers and to prepare the contents for storage and management within the archive. Ingest functions include receiving SIPs to a staging area, performing quality assurance on SIPs, generating (from SIPs and from other resources) Descriptive Information for inclusion in the archive's Data Management database, generating AIPs (if not provided directly by the data provider), and coordinating updates to Archival Storage and Data Management.

While receiving SIPs and generating AIPs for permanent preservation is the primary ingest activity, there are two other types of ingest activity that are currently in use to support NSSDC's data producers and related user community. The first of these is referred to as a 'second archive' service in which digital data are received on distributable media that is also held by a primary archive. The data are not transferred to AIPs, however the media are kept under environmental control but no media refreshment is performed. NSSDC may disseminate copies of the distributable media if authorized to do so by the primary archive as per MOU. The second of these ingest activities is referred to as a 'backup' service. Digital data are stored, typically offsite, to support another archive's contingency plan per MOU. Backup data are not disseminated by NSSDC. NSSDC may also receive analog data., but this has been very rare in the last few years.

For all but the backup service, information about the data is inserted into the Data Management database. These include such things as identification of the spacecraft and experiment(s) that contributed to the data; the identity of the PI and other investigators; brief description of the data, written not to describe how to use/interpret the data, but to allow prospective data users to determine if this data will be of use for their purposes; time span covered by the data; pertinent discipline of the data; reference to published journal article(s) describing the instrument and its operation and calibrations, etc.

The appropriate evaluation of a dataset and its associated supporting information, and the generation of most of the related entries for Data Management (i.e., NSSDC Information and Management System (NIMS)) , requires some specialized knowledge of the particular scientific discipline involved. To this end, NSSDC maintains a small staff of acquisition scientists, with different backgrounds and areas of expertise, allowing them to handle datasets in any of the disciplines in which NSSDC archives data.. They also interact knowledgeably with the data providers and help data users locate and use data. The detailed nature of these interactions and the particular tasks and efforts required for ingest differ with the disciplines (Sun-Solar System Connection, Solar System, and Exploration of the Universe).

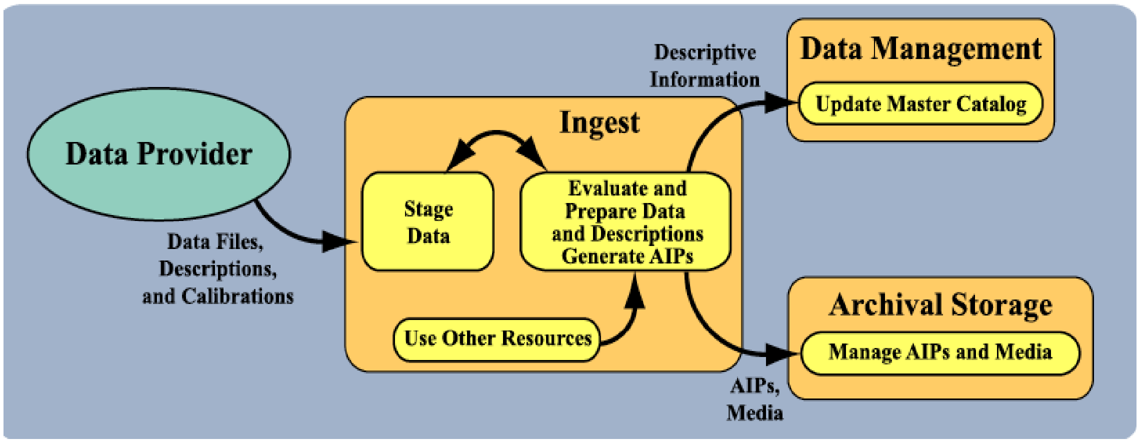


Figure 2: The OAIS model provides a clean separation between the ingest activities and the long-term management of data and metadata, thereby promoting independence and facilitating cost-effective evolution of system components.

Over much of NSSDC's history, datasets were submitted using a wide variety of media: computer tapes, punch cards, chart rolls, hard copy, diskettes, photographic film, microfilm, and microfiche. Hard copy was microfilmed and the microfilm archived. Currently, data typically arrive via CD, DVD, DLT, and the Internet.

The NSSDC developed the Multifile Package Generator and Analyzer (MPGA) software to create AIPs for the preservation of science files and related metadata. The NSSDC uses MPGA during ingest to package SIPs into AIPs. In cases where the data provider delivers SIPs in the form of AIPs, the NSSDC uses MPGA to identify the AIPs and extract from them the metadata required to support the ingest process. Both of these MPGA processing scenarios are highly automated with MPGA being invoked by the NSSDC's DIOnAS (Data Ingest and Online Access System).

Currently one data provider, the IMAGE project, is directly generating AIPs for delivery to the NSSDC. These AIPs are delivered to DIOnAS for archival storage. The NSSDC anticipates an expansion of this practice and has produced an upgraded version of MPGA to supply to data providers that will run under multiple operating systems. The NSSDC is currently working with PDS to develop an electronic AIP delivery method.

In addition to the location of AIP generation, other ingest details may also vary, depending on the science discipline involved, as described in the following three sections.

### 1.2.1 Heliophysics

In the Heliophysics discipline, NSSDC's data providers are Projects, Project Data Centers, Active Archive Centers, and legacy projects and PIs. Past interactions with data providers were on a more informal ad hoc basis, while the interactions with the newer projects are governed by a PDMP or equivalent; for example, we provided help to the future IBEX project in preparing their draft PDMP. A summary of missions and interactions for SSSC is shown in Table 1. Note that we continue to receive new data from the ISTP spacecraft and make it available thru the CDAWeb interface, which offers display and download services. SSSC data volumes ingested to the near-line permanent archive totaled ~764 GB in 2004, with ~678 GB being backup for RHESSI. These are only for data archived in AIPs; in Space Physics there are other data archived but not in AIP format. The corresponding volumes for 2003 were ~1004 GB and ~856 GB. For the year 2005, these volumes are 1487 GB and 796 GB. The total inputs to the archive, including both AIP and non-AIP data, were 2758 GB in 2003, 2181 GB in 2004, and 3803 GB in 2005. We receive data as files transferred over the Internet, on CD-ROM, DVD, or DLT media. The SSSC acquisition scientist reviews the available information and creates (in Data Management, Fig. 1) the records describing the dataset and additional necessary information, as detailed in section 1.2 above. The acquisition scientist also reviews the documentation provided for the dataset, to ensure that it is sufficient for a researcher knowledgeable in the field, and may request additional information or clarification of some elements of the provided documentation material.

Table 1: Through MOUs with active archives and through project PDMPs, NSSDC acquires data for preservation that is often available elsewhere, but for a great many legacy missions it is now the primary data source, thus ensuring its usability into the future.

| **Data Provider** | **Missions** | **Comments** |
|---|---|---|
| Science Data Centers (SDC) & Active Archives (AA) | ACE*, FAST*, IMAGE*, RHESSI*, SAMPEX*, TIMED* | Permanent Archive<br>SAMPEX is first Resident Archive Center |
| Space Physics Data Facility | Multiple ISTP era s/c in CDAWeb | Began ingest of data into archive.<br>Ongoing receipt of Level 0 from Polar-Wind-Geotail processing center |
| Solar Data Analysis Center (SDAC) | Yohkoh+ | Yohkoh data through SDAC and separate MOU |
| Other | IBEX,THEMIS,AIM | Future missions. NSSDC consulted on draft PDMP |
| Older missions; ongoing data flow | Geotail, IMP8*, ISEE*, Polar, Ulysses, Voyager**, Wind | No Science Data Center |
| Legacy | 684 missions | Permanent archive. |

+ Memorandum of Understanding (MOU) exists. *Project Data Management Plan (PDMP) exists.
** PDMP for planetary encounters exists.

### 1.2.2 Planetary Science

In the Planetary Science discipline, NSSDC's current data providers are the Planetary Data System (PDS) for whom NSSDC provides a second archive service. The NSSDC also acts as the permanent archive for lunar and planetary science data from older missions that are not otherwise available. PDS negotiates PDMP's with the individual projects and the interactions between PDS and NSSDC are specified in an MOU. Our clients are scientists, educators, and the general public. The total volume of planetary data received in 2005 is 338 GB. The corresponding volume for 2004 was 104 GB, and for 2003 was 723 GB. Table 2 provides a summary of the solar system mission ingest activities and their status. The data from PDS typically arrive on CD-ROM or DVD-ROM, as specified in the MOU. This is feasible for such missions as Mars Global Surveyor (1.2 TB total data expected), the Mars Exploration Rovers (2.0 TB), Cassini (3.5 TB), and Mars Odyssey (4.5 TB). However, large amounts of data expected from the Mars Reconnaissance Orbiter (~70 TB) and the Lunar Reconnaissance Orbiter (400 TB) will make archive and delivery on DVD-ROMs impractical. We are negotiating a process for electronic delivery with PDS. This will involve PDS transmitting the data electronically as Archival Information Packages (AIPs) with identifying wrappers. The details of this are still being worked out. AIP software created at NSSDC has been sent to PDS for testing and evaluation. We are also in the process of revising the MOU between NSSDC and PDS to encompass electronic delivery.

The data and metadata received at NSSDC has been inspected and validated by PDS. The planetary acquisition scientist reviews the information provided and creates dataset records in Data Management (Figure 1). The dataset record comprises the metadata necessary to identify, describe, and evaluate these data. The dataset record and all associated information such as descriptions of the instrument and spacecraft mission created by the acquisition scientist are available online through the Access function. Where possible, this metadata includes a network link which points to online data, typically housed on a PDS server.

Table 2: Through a unique MOU with PDS, NSSDC provides a long-term preservation function for planetary mission data, supports access by the general public, and is also the primary archive for many legacy planetary/lunar mission datasets not under PDS management.

| **Data Provider** | **Missions** | **Comments** |
|---|---|---|
| Planetary Data System (PDS)+ | Permanent Archive: 16 missions | For example: NEAR, Magellan, Viking, Voyager |
| | Mars Global Surveyor | Continuing mission, data flow established |
| | Cassini-Huygens, MER | Continuing mission, PDMP reviewed |
| | Stardust, Deep Impact | Small data volume expected |
| | Mars 2001 Odyssey | Radio Science level 0 data received, other pending |
| | Hayabusa, Selene, Lunar-A | Current and future JAXA missions |
| | SMART-1, Mars Express, Rosetta, Venus Express | Current and future ESA missions |
| | MESSENGER | Mercury orbiter, expect moderate data volume |
| | Mars Reconnaissance Orbiter, Lunar Reconnaissance Orbiter | Future missions, large data volume, expect electronic delivery via PDS |
| | New Horizons, Phoenix | Future missions, discussions about data volume |
| Legacy | >95 missions | For example, Apollo ALSEP and orbital data being restored as NSSDC's "Lunar Data Project" |

+ MOU (Memorandum of Understanding) exists.

### 1.2.3 Astrophysical Science

In the Astrophysical Science discipline, NSSDC receives and ingests astrophysics data primarily from NASA's distributed network of astrophysics "active archives." The activities of these SARCs (Science Archive Research Centers) are coordinated by the Astrophysics Data Centers Executive Council (ADEC), of which NSSDC is a member. NSSDC ingests astrophysics data initially as backup copies for the active archives, and later as permanently archived stable datasets upon agreement with the active archives. A stable dataset is one for which the mission has ended and no further reprocessing will take place. The separate roles and responsibilities of the NSSDC and the active archives are established and documented through MOUs and PDMPs. PDMPs are typically established between missions and the active archives, though occasionally a PDMP has signatory approval by the NSSDC. PDMPs are used to establish data formats, volumes, expected documentation, and delivery schedules.

Table 3 summarizes the various data provider categories, including the SARCs: high-energy astrophysics data are ingested from the HEASARC; UV-Optical data are ingested from MAST; infrared data are ingested from IRSA; and long-wavelength (microwave and sub-mm) data are ingested from the Legacy Archive for Microwave Background Data Analysis (LAMBDA). Table 3 distinguishes SARC-related data transfers from three other categories of data provider: large missions (e.g., HST, Spitzer Space Telescope); active missions that have a PDMP agreement directly with NSSDC (e.g., GPB); and NASA's legacy astrophysics mission datasets (e.g., IRAS, COBE, UHURU).

NSSDC has received approximately 4.75 TB of new astrophysics data in 2005, of which 3.9 TB came from the HEASARC. For comparison, NSSDC received 2.3 TB of new astrophysics data in 2002. These data come to NSSDC via a combination of network transfers and hard media. The modes of data delivery are also varied. For example, the full GPB mission data archive is to be delivered all at once as an Oracle database. In contrast to this, the FUSE data are shipped to NSSDC on DVDs throughout the year on a continuing basis. Most data products are delivered and ingested in astronomical standard FITS format. In some cases, such as GPB, the project team will deliver the entire mission document set and will deliver the mission observation attribute database exported in CDF format. This is a common practice –for a mission to convert their project-specific data formats into a standard format prior to delivery to the NSSDC.

An astrophysics acquisition scientist at NSSDC provides the interface to the active archives and astrophysics missions, while the routine acceptance, logging, and ingest of data and shipment manifests are handled by NSSDC ingest operations staff. The acquisition scientist reviews available information and creates dataset records, describing the dataset and additional necessary information. The acquisition scientist is the principal reviewer of mission documentation for accuracy and adequacy as the sole future means by which the mission dataset can be independently and correctly used by non-mission scientists.

Table 3: Through MOUs with the SARCs, NSSDC typically provides primarily a backup function, however there are a number of new and legacy missions for which NSSDC is the primary archive.

| Data Provider | Missions | Comments |
|---|---|---|
| Science Archive Research Center (SARC) and Science Center (SC) | 9 active missions across HEASARC+, MAST+, LAMBDA+, IRSA, MSC | Permanent Archive for 20 completed missions across four entities. Backup archive for SARCs per MOU. Data ingest ongoing, e.g., GALEX. |
| Other | 10 missions | Future archive for Swift, Spitzer, JPL Navigator |
| Other | Gravity Probe-B (GPB)* | Future permanent archive per PDMP, draft MOU |
| Legacy | 70 missions. | Permanent archive. For example, Uhuru |

+MOU (Memorandum of Understanding) exists. *PDMP (Project Data Management Plan) exists.

## *1.3 Administration*

At NSSDC, we divide the administrative functions into externally and internally focused activities .

### 1.3.1 External Administration

External administration includes developing and evolving the current acquisition policy for NSSDC as a facilitator for data ingest from discipline-specific active archives, missions, and PIs. In this role, NSSDC assesses the needs of these groups to define the services that NSSDC can provide as the permanent archive for all NASA space science data. In addition to existing MOUs with the PDS and HEASARC, NSSDC has recently concluded MOUs with two other astrophysics groups: the Multi-mission Archive at Space Telescope (MAST), and the Legacy Archive for Microwave Background Data Analysis (LAMBDA). We continue to negotiate an MOU with the Infrared Science Archive (IRSA). In the space physics discipline, NSSDC is usually identified in each mission's Project Data Management Plan as the final archive and agreements are reached either with the individual missions or the active archives such as Space Physics Data Facility (SPDF) and Solar Data Archive Center (SDAC). In certain cases when a mission does not work directly through an active archive (e.g., Gravity Probe-B) a separate agreement may also be sought to define the roles and responsibilities. For new sample return missions, NSSDC is working with PDS to define interactions with the Astromaterial Curation Facility (ACF) at the Johnson Space Center (JSC) for missions such as Genesis, Stardust and Hayabusa. NSSDC also ingests a limited amount of data from foreign sources; for example, Cluster and Interball data.

To support the submissions of, and access to, data, NSSDC maintains web pages that locate data, provide fact sheets, identify white papers (e.g., Data Retention Policy http://nssdc.gsfc.nasa.gov/nssdc/data_retention.html), and general information on the archive process and widely used formats. Since a large fraction of the NSSDC holdings are analog or legacy data archived before the institution of the active archives, customers include active archives, researchers, educators, and the public. As the permanent archive for NASA, NSSDC also functions as the World Data Center for Satellite Information (WDC SI), providing email announcements of satellite launches around the world, posting a monthly SpaceWarn bulletin on the WDC SI website, and providing trajectory information. Through the WDC, we also provide data to foreign requesters. enabling worldwide data access.

On behalf of COSPAR (Committee On SPace Research), NSSDC is privileged to receive notices of spacecraft launches and it assigns and disseminates unique international satellite designators for all spacecraft that attain orbit . NSSDC generates and maintains brief descriptions and limited other information about those spacecraft. These brief descriptions and other information are also maintained for scientific experiments on board satellites, and for their resulting datasets that are archived at NSSDC (and for a limited number of datasets available elsewhere that are of interest to space scientists).

NSSDC generates a bi-monthly report for NASA HQ to identify the progress of archiving for the various NASA missions and active archives. In addition, interactions include periodic updates on the budget and presentation to the SSSCDCWG advisory group; the latter has resulted in a white paper (see http://nssdc.gsfc.nasa.gov/nssdc/RAXWhite_Paperu.doc) on how to form Resident Archives at a mission's end to facilitate data being submitted to the permanent archive. NSSDC is tasked with responding to special projects for headquarters. An example of one such project is the development of a self-registry web site to track all NASA space science data.

NSSDC monitors customer satisfaction under the administration function through review of comments and direct interactions with active archives.

### 1.3.2 Internal Administration

NSSDC also performs an internal administration function. One example is to track the levels of service provided to data submitters and identify whether the accompanying documentation makes the data independently useable, NSSDC interacts with the active archives to plan infrastructure upgrades so as to handle new and voluminous data products. This planning is necessary to schedule tasks efficiently for the staff and allow for restoration of interesting or useful data products. An example is the restoration of Apollo *in situ* data by NSSDC, using separate funding, that will be ingested into the PDS in support of the lunar exploration initiative. This internal administration function forecasts the permanent archive needs to make effective use of market driven changes in technology and software for data storage. An ongoing task is to modernize the hardware/software system that we monitor monthly to look for cost effective approaches and changes.

Administration also oversees migrations. Figure 3 shows data migration for earlier data products (including digitizing relevant material) into media-independent forms for archival storage. NSSDC extracts files and related attributes from legacy magnetic tapes and the tape inventory database, respectively, and these are then provided to the Ingest function staging area. Ingest creates multifile AIPs on hard disk for pickup by the Archival Storage function. Updates to support discovery of the AIPs are sent to the database in Data Management. We control the entire migration process by a temporary database to ensure quality control. As the process continues, AIPs are stored both on robotic SDLT media and on hard disk. In some cases, unique or high demand data may be extracted from the AIPs and made electronically accessible via an FTP server under the Access function.

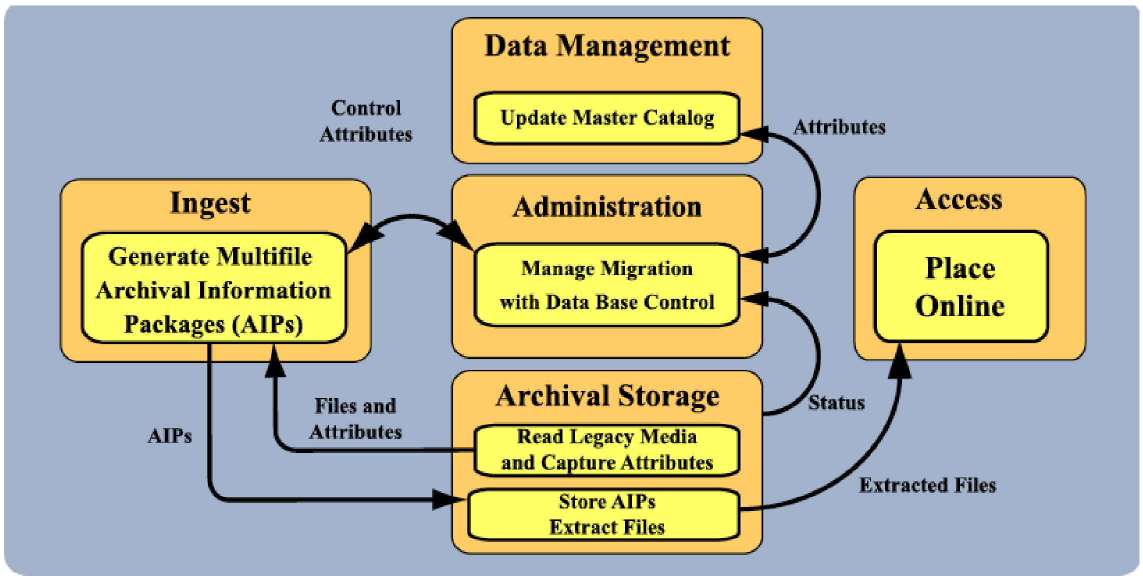


Figure 3: Data migration, controlled by administration, ensures that data remain available into the future.

This process includes capturing provenance information and checking documentation to make the data more reliably and independently useable. Where practical, some updates to a more accessible format will be made. NSSDC has migrated a selection of 200 tapes (both 9 track and 3480 cartridges) to initiate and validate the process. The effort, referred to as the 'legacy tape migration', is considered a background task that will take several years to complete. At that time, it will be necessary to start the refreshing of the current SDLT media on the planned 6-year life cycle.

## *1.4 Archival storage*

NSSDC is responsible for the long term archiving and preservation of space science data and maintains a multidisciplinary archive encompassing nearly five decades of space science digital data. These data include over 2,300 distinct digital datasets in excess of 36 Terabytes preserved on a variety of magnetic and optical media. As of January 2006 NSSDC held approximately 39,000 magnetic tape volumes (not counting copies) and 27,000 optical disk volumes. Also maintained by NSSDC is over 2000 analog datasets including over 700,000 still photographs, over 150,000 feet of motion picture film, and an extensive set of microfilm and microfiche. However virtually all new data are received in digital form.

In addition to holding data for which NSSDC has the primary preservation responsibility, NSSDC also provides "second archive" and "backup" services for data held by other archives as noted earlier. Table 4 illustrates the five levels of archive service provided by NSSDC.

Table 4: NSSDC provides a range of archival services as needed to meet the evolving needs of the larger Space Science data management and distribution environment

| **NSSDC Archival Storage Service** | |
|---|---|
| **Permanent Archive: AIPs** | Preservation of digital data in Archival Information Packages delivered by a data producer or created at NSSDC. AIPs are re-written to new media within six years. Data is disseminated by NSSDC 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. Data is disseminated by NSSDC if not available through an active archive or per MOU. |
| **Second Archive** | Storage of digital data on distributable media that is also held by another archive. No media refreshment is performed. NSSDC may disseminate the data if authorized to do so by the primary archive as per MOU. |
| **Backup** | Storage of digital data to support another archive's contingency plan per MOU. Data will not be disseminated by NSSDC. |
| **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 NSSDC. |

The 'permanent archive: AIP' service is the primary focus for new digital data arriving at NSSDC. The 'permanent archive: non-AIP' service reflects the historical buildup of data on a variety of media. These data will eventually be put into AIPs as well. The 'second archive' service reflects historical agreements for distributing copies of media. In the future, NSSDC will attempt to negotiate transfers of this information to AIPs in order to improve preservation efficiency. Some external archives have asked NSSDC to maintain backup copies of their archival storage, so a 'backup' service has been defined. As there is a substantial amount of analog data in the archive, much of it unique, NSSDC maintains an 'analog' service as well.

The AIP is an information package containing data and documentation needed to make the resulting information content understandable over a long-term period. In 2000, NSSDC adopted the approach of preserving data in AIPs as well as other concepts from the CCSDS/ISO "Reference Model for an Open Archival Information System (OAIS)". Use of the AIP to archive data dissociates digital data content from specific media volumes and provides increased efficiency when performing the digital data migrations that are required to preserve data into the indefinite future.

Historically, the media used to transfer digital data to NSSDC usually served as NSSDC's storage media with a copy of each media volume being stored in an off-site location. NSSDC's digital data holdings have been managed as a collection of digital media volumes. The Java Inventory (JIN), an application running within the NSSDC Information Management System (NIMS) is dedicated to this task. JIN is a recently released product, replacing an older less capable inventory system. Future releases of JIN will provide for the management of additional media. The release of JIN version 1.1 is expected later in this year, providing support for an improved process for releasing media and upgraded reports. Work on JIN upgrades to support the photoproducts inventory will commence this year with an expected deployment in 2007.

At this time, only those data delivered electronically to NSSDC are permanently archived in AIPs. Data delivered on transfer media (second archive) are still archived on media volumes in the traditional manner. Ultimately, all NSSDC permanently archived digital data will be migrated to AIPs.

To create and manage AIPs, NSSDC developed the Data Ingest and Online Access System (DIOnAS), deployed in the summer of 2000. DIOnAS generates AIPs at NSSDC or accepts pre-generated AIPs (e.g., from AIPs created by a data provider external to NSSDC), writes AIPs to permanent storage media, and can populate an FTP data distribution site with data extracted from AIPs. Thus, DIOnAS is a system whose role extends into multiple archive functions : Ingest, Archival Storage, and Access. The media types currently used for permanent storage of AIPs are DLT type IV cartridges (35 GB capacity) and SuperDLT type II (300 GB capacity). DIOnAS consists of Java software and a database containing ingest processing information and archive inventory information. The Multifile Packager and Analyzer utility (MPGA) can be called by DIOnAS to create AIPs and extract information from AIPs. Cyclic redundancy checksums (CRCs) are used extensively to safeguard against data corruption while creating, ingesting, and preserving AIPs. Validated copies of every AIP are written on multiple media in distributed locations. NSSDC policy calls for a six year refreshment interval for media containing AIPs.

The first version of the new NSSDC AIP Storage System (NASS) was released in 2005. NASS is a Java application and a database for managing the storage of AIPs at NSSDC. NASS is dedicated to the Archival Storage function, with no responsibility for Ingest or Access. The NASS architecture is not media specific any digital media may be used for AIP storage.. NASS development activities in 2006 are: queued media read and write requests providing concurrent processing capability, a graphical user interface for NSSDC staff to facilitate access to existing NASS management functions.

NSSDC currently extracts and stores as text objects the preservation description information from all ingested AIPs. This is done to facilitate internal analysis of the AIP holdings. A conversion of the database holding the AIP preservation description to XML is planned for this year.

Provenance information for NSSDC data is captured in multiple locations and is not easily accessed by the user community. NSSDC plans to develop a system for capture and preservation of provenance information, thus making the data more usable to future researchers. Development of an initial provenance information management system that addresses the legacy tape data migration is planned for this year.

NSSDC has multiple independently administered collections of digital documents. To provide a more efficient control mechanism and supply a uniform access method to the user community the NSSDC will develop an integrated document management system. Development of the initial system will start in summer of 2006 with an initial deployment late in the year or early in 2007.

## *1.5 Data management*

The ability to search for NSSDC's data holdings based on spacecraft name and discipline has been in place since the mid-1980s, with the current web-accessible incarnation being available since the mid-1990s. This is supported by the metadata maintained in a database under the Data Management function. This function also provides maintenance and upgrades to existing database related software and works to identify needed software to meet new operational requirements. This includes running specialized reports based on staff and external requests.

The database infrastructure is referred to as the NSSDC Information Management System (NIMS). NIMS contains information primarily of use for describing and searching for data holdings. In addition, NIMS also has the capability of handling many administrative functions that pertain to the tracking of archiving status of various flight projects.

A core piece of NIMS is the information about the datasets archived at the NSSDC. This partition describes the data so that interested parties can determine what would be of most use to their particular needs. Ancillary information about the data, including descriptions of the experiment(s) that generated the data and the spacecraft on which the experiment(s) flew, are also available. Other ancillary information in the database is the publications generated from or describing additional details about the datasets, experiments, and spacecraft. These parts of NIMS are primarily for use as discriminators and for searching.

Other partitions in NIMS (as well as some parts in the core set of partitions mentioned above) are used primarily for administrative purposes. This includes (for example) information on the number of requests for a given dataset, information about requestors, total volume of data in a dataset, etc.

The primary means by which information is entered into the NIMS database is via a pair of Java-based interfaces: JEDS (which is used to enter information about spacecraft, experiments, datasets, and publications) and JRAND (which is used to track information about people, both requestors and those involved with flight projects and/or datasets, and about requests for data). Access to these interfaces is on a named-user basis and is only provided to those within the GSFC computing environment.

The NIMS database is moderately sized in terms of the number of records it holds, roughly 6,000 spacecraft and about 5,000 each of experiments and datasets with about ten times that of publications and people. Some parts of the database are quite simple in nature, with people information spanning only five tables and a few dozen fields, while others are very complex. Spacecraft information is among the most complex in the database, comprised of 24 main tables with roughly 50 supporting (validation) tables and dozens of accompanying triggers and stored procedures to ensure consistent and reliable data. This difference in complexity is further reflected by the interfaces, with JRAND comprising about 13,000 lines of code while JEDS comprises about 40,000.

The primary means of external access to the NIMS information is the NSSDC Master Catalog (NMC) software. This activity more correctly falls under the Access function and will be discussed in the next section.

## *1.6 Access (Distribution of data)*

The Access function serves to identify and provide all of NSSDC's users with the data for which they are searching, as shown in general in Figure 4. The Access function depends on input from the Data Management and Archival Storage functions for additional support needed to fulfill the wide range of requests made to the data center. Note our primary mission is to be the Permanent, not an Active, Archive, with emphasis on preservation more than access, but access remains an important function. For many datasets NSSDC is the unique holder and for many more NSSDC provides distribution service along with the mission archive.

NSSDC users, who cover the full spectrum from scientific researchers to the general public, both domestic and international, generally obtain access to data through a collection of web pages and/or web-enabled interfaces maintained by NSSDC staff. These access information about data holdings and allow direct access to selected datasets on-line. Nearly 1 TB of the most requested datasets are maintained on-line at ftp://nssdcftp.gsfc.nasa.gov/. Most datasets are not directly accessible via the internet and options are provided that allow interested parties to request further information through the NSSDC Request Office. Other users come to NSSDC through personal contracts, such as acquisition scientists, who also provide access support.

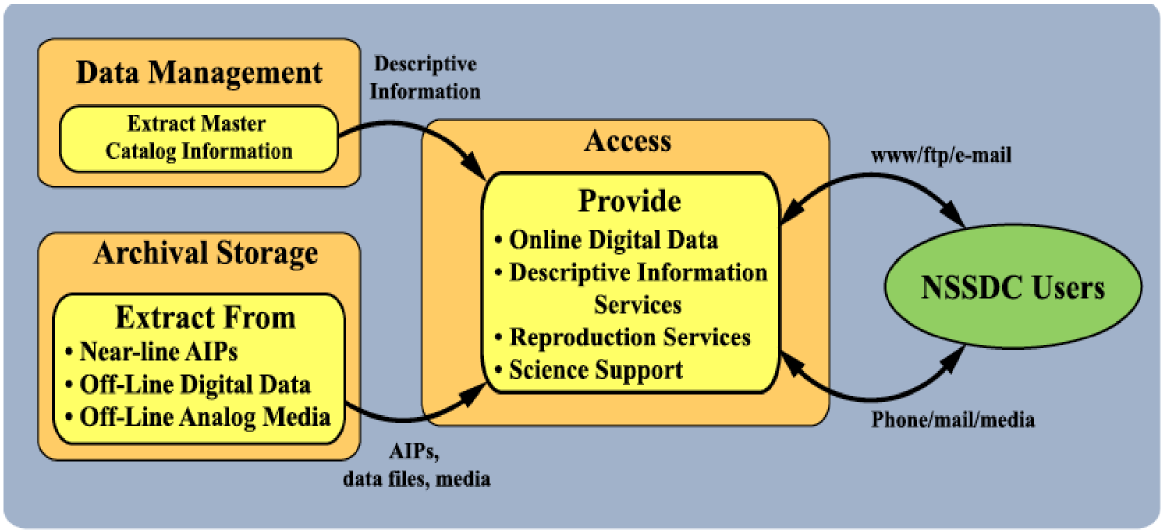


Figure 4: NSSDC users access data efficiently and cost-effectively since the function is modeled after OAIS and is cleanly separated from the long-term management of data and metadata.

Access to the information about NSSDC's data holdings is through the NSSDC Master Catalog (NMC), the web interface to the NIMS database, which can be linked to from the NSSDC homepage. It consists of an extensive set of Java code (primarily servlets) that queries the database in various ways to produce dynamically generated web pages. The database is organized to link data to an experiment and experiments to a spacecraft, and this is reflected in the Master Catalog software. Users can search though the NSSDC Master Catalog and locate data of interest, in many cases finding links to on-line sources of data holdings, both at NSSDC and elsewhere. This linking ability is supported mainly for highly-sought-after datasets currently, but the number of datasets that are accessible in this manner has continued to grow at a steady pace. The on-line query form allows users to query for spacecraft which meet several criteria (name, discipline, and/or launch date). From a spacecraft one can explore information about its experiments and further to the datasets generated, either by spacecraft or by experiment (as shown in Fig 4). Approximately 5,063 datasets are catalogs in the NMC, 4400 of these datasets are archived at NSSDC.

Two other available, although less publicized, query interfaces can also be accessed on the NSSDC public web site. Both can be reached via the URL http://nssdc.gsfc.nasa.gov/nmc/. One allows a query of the publication information tracked in NIMS (from which one can then go further into related spacecraft, experiment, dataset information). The other generates a listing based on data that have recently arrived at NSSDC within a specified time span.

The total number of direct requests to NSSDC has evolved as expected in recent years with the growth of the internet. The percentage of core users from universities or government labs remains about 25%, which is mirrored in the FTP usage statistics. Scientific data requests have been less frequent as active NASA datasets are being offered by research teams directly from their own web servers and often-requested datasets are posted on the NSSDC anonymous FTP site or on other NASA archive sites. In many cases the scientific request directed to NSSDC is for unique data archived before the advent of the active archives. Many general public requests for photos and videos have been replaced with web views and downloads. When requests are answered with hard copy or media, then a small handling fee may be charged to recoup the duplication costs.

For our scientific colleagues searching for space science data that are not easily accessed on-line, the Request Office coordinates inquiries with our in-house scientists to identify desired data, locate it and/or its associated metadata, discuss options for its distribution, and disseminate the data. The preferred dissemination method for digital data is FTP, but large datasets may be provided on CDs or DVDs. Special requests for data on microfiche, microfilm or paper printout are accommodated on a case by case basis. Photos are provided digitally via FTP or as prints or negatives. Documents are generally provided as digital or paper copies.

NSSDC also receives and answers requests from the general public. These requests are mostly for NASA photographs, videos or posters, such as the wildly popular Multi-wavelength View of the Milky Way. The experienced staff in the NSSDC Request Office are excellent at discerning the requester's requirements, e.g. which set of lunar photographs best show an Apollo landing site. The general users include commercial entities (e.g. calendar makers, book illustrators) and persons of all ages looking for their general interest. Requested videos, CDs, DVDs, and posters are shipped for a fee which may be waived by the Head of NSSDC.

Careful records are kept on electronic accesses, searches and downloads, as well as detailed logs of requests on physical media. These are reviewed periodically to see where we can improve services, e.g. identify newly popular datasets which will be posted on our ftp site for easier access.

One of the lesser known, continuing activities of NSSDC is to serve as the World Data Center (WDC) for Satellite Information. The WDC captures and disseminates information about spacecraft launches and other information. It also serves as a conduit fro request for NSSDC data from the international community.

The core activities for access to NSSDC's holdings, as described so far in this section, are planned to continue into the foreseeable future. Requests for individual datasets will wax and wane, but researchers and general users will always seek NASA data. We continue to refine and improve our Access Services as much as resources will allow. Currently we have begun or planned several improvements.

NSSDC is increasing its web services. Those that support SPASE, the international effort to define a dictionary for Space Physics terms as the basis for sharing data among Space Physics data centers, will be developed and posted this year. SPASE v 1.0.1 has just been posted, so the service associated with it will likely grow in coming years (see sec. 1.7 Preservation Planning).

Another web-related project in planning is to organize and make available via a web interface the NSSDC digital documents. Many supporting documents for datasets are received electronically, while others have been digitized internally. The digital documents will provide additional information about NSSDC datasets to aid our users in their selection and use of various datasets.

Two other development projects in progress will create a much more robust basis for the NSSDC archives and for future development of its systems. The first is to create an Access/Storage interface, a link between the NIMS database of supporting information and the DIOnAS database tracking data stored as AIPs. This link should be completed later this calendar year. The second is to begin the conversion of the NSSDC Master Catalog to web services. Rather than the current query system described above, where the database is queried and then data retrieved from the designated directories, the documents themselves would become database entries and thus retrieved directly via the query.

The largest improvement for data center access is becoming part of a Virtual Observatory, a specialized form of data mining across data centers in the same disciplines. Various Virtual Observatory (VO) projects exist or are planned for astronomy (NVO), planetary (IPDA), solar physics, and various space physics subdisciplines (VxO). The VOs are currently developing standard protocols for registering data, discovering data, and accessing data. These protocols include data/metadata standards, XML standards, and OAI standards. Because these VO projects are implementing such protocols, standards, tools, and services in a manner that allows cost-effective adoption by many other data centers, the NSSDC plans to adopt these same protocols in its interfaces. These interfaces will follow the Web Services paradigm in which machine-to-machine interaction occurs with minimal human intervention. Thereby, data available on-line are delivered to analysis and visualization packages semi-autonomously. Data archived near line will return a message that they are available via individual request. As a first step in our participation, we will provide a registry of NSSDC data products and services, and we will follow the emerging XML-based VO protocols for metadata, semantic markup, and database access. *As NASA's Permanent Archive, NSSDC is the unique holder of many datasets, from early balloon flights to the soon-to-arrive Gravity Probe-B, so has a significant contribution to make.* The value of the national and international VO projects is primarily in shared development and rapid low-cost buy-in from any data provider. We will leverage those developments and participate in this new virtual environment of seamless access to distributed data.

## 1.7 Preservation planning

The purpose of Preservation Planning is to provide advice to the NSSDC Administration regarding preservation strategies, design of packaging templates, and monitoring of technology as it affects archive services for NSSDC users and data providers. This function also supports NSSDC's role in providing advice to, and promoting synergism among, Space Science's active archives. An overview of the Preservation Planning function at NSSDC is given in Figure 5.

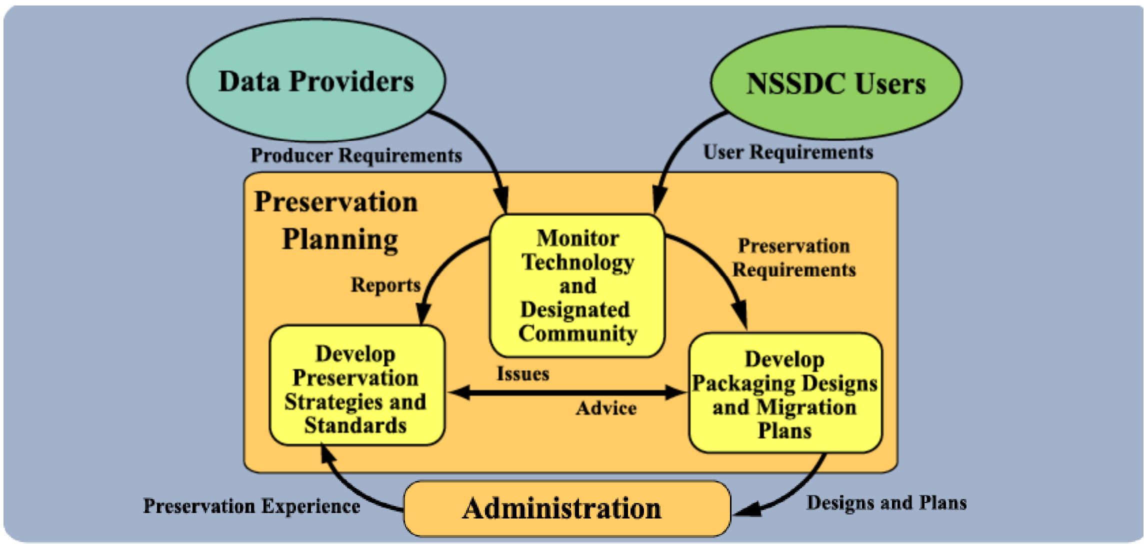


Figure 5: By carefully monitoring user requirements and technology capabilities, preservation planning allows maximal efficiency of the entire archives.

"Develop Preservation Strategies and Standards" identifies and evolves strategies and related standards that facilitate the long-term preservation of the scientific information. To facilitate a community focus for standards, NSSDC created a virtual organization, the NASA/Science Office of Standards and Technology (NOST). NSSDC/NOST is engaged in several related activities internal and external to the organization. It is important for a long term archive to stay abreast of the best thinking on preservation strategies for science data. This is facilitated by participation in relevant standards development activities that leverage the expertise of space and non-space organizations participating in the Consultative Committee for Space Data Systems. NSSDC/NOST leads the working group that developed the CCSDS/ISO "Reference Model for an Open Archival Information System (OAIS)." It has become the fundamental framework for communicating about digital archive functions including migration and interactions with other archives and the modeling of information for long term understanding. This framework, adopted by all types of archives all over the world, is instrumental in helping NSSDC evolve its systems to improve both efficiency and effectiveness. This includes the Archival Information Package (AIP) approach that supports more automation on ingest and low-cost archival storage and migrations. Most NSSDC/NOST participation in CCSDS has been funded externally which has greatly leveraged our contribution.

A recently completed standards activity that is beginning to influence NSSDC approaches to ingest and preservation is the CCSDS "Producer-Archive Interface

Methodology Abstract Standard."  It provides a checklist of activities and approaches to negotiating data submission from data providers and will help NSSDC evolve guidelines for data submission.  NSSDC is also an invited participant in the international Archive Certification Task Force that has recently released a draft certification checklist for public review.

NSSDC is leading and supporting an international effort, called "SPASE", to create a common Space Physics data model. It will facilitate access to distributed data services under the Virtual Observatory (VO) concepts.  Although some external funding supports this effort, NSSDC has also contributed its resources.  A first version of the data model has been released and is undergoing public comment and testing.

NSSDC will be developing draft guidelines in FY 06 for the Heliophysics discipline addressing documentation needs for datasets to be submitted to the NSSDC.

"Develop Packaging Designs and Migration Plans" evolves internal data structuring approaches and proposes plans for accomplishing specific data migration efforts.  The Archival Information Package (AIP) concept has been adopted to isolate data files from dependencies on underlying media types and the VAX/VMS file system, and it provides pointers to associated descriptive information so data bits may be understood.  Recent accomplishments include development of an AIP (version 3) that can hold multiple files and their related metadata. This is necessary to cost-effectively capture legacy data from 9-track and 3480 magnetic tape.  This function has also been leading a working group that is developing a detailed design for the migration of this data that is well documented not only to ensure consistent operations but to provide the provenance (processing history) information that will be needed in the future.  Our experience over 40 years greatly attests to the need to capture  processing history to address future use and migration issues.

Updates to the AIP functionality, based on experience with the legacy tape migration and the interactions with the PDS, will be proposed later in FY 06.  A prototype for permanent identifier functionality for internal NSSDC document management will be developed in FY 06 to improve our ability to manage our electronic documentation without breaking links and spending extra resources to fix links.

We will design and prototype an Acquisition Scientist ingest support interface to facilitate more automated ingest of dataset documentation, population of NSSDC's management system, and better quality control with less human resource expenditure.

"Monitor Technology and Designated Community" is used to stay abreast of changes in technology impacting NSSDC internal systems and NSSDC's producer and user communities.  NSSDC is now recognized world-wide for its expertise in digital preservation.  We accomplished this through participation and liaison with various standards groups, participation in selected conferences and workshops, and serving on review committees.  These activities provide a broad view that spans multiple agencies (including non-Space agencies) and technologies.  Examples include: participation in the international Consultative Committee for Space Data Systems, RLG/NARA Archive Certification Task Force, US CODATA/ICSTI Working Group on Permanent Access to Scientific Data and Information, UK Digital Curation Centre (DCC), NASA workshops on VO concepts, international SPASE working group, and liaisons with NCITS L8 working group, INTERPARES working group, GSFC Ontology Working Group, NASA XML Working Group, and the GSFC Library Metadata Review Group. We are also co-sponsoring the GSFC/IEEE conference on mass storage and archiving. We follow discussions within the PDS on standards and approaches, are a member of the Astrophysics ADEC consortium, and have close ties with the SPDF. Recent accomplishments include presentation of a paper at the PV-2005 conference on lessons learned from migrations in the context of the OAIS reference model, participation on an NSF review committee addressing digital preservation, and participation on the DCC review panel addressing a "Manual for Digital Curation." NSSDC is developing Web pages to facilitate synergism and good preservation practices among the Space Science archives. It makes current and evolving standards and guidelines, including relevant international material, readily visible, and it provides a view of standards that have been adopted by the major Space Science disciplines.

## 1.8 NSSDC Education and Public Outreach

The NSSDC continues to serve not just the science research community, but also the general public which is well aware of the valuable database of information on satellites, instruments, and general space science datasets and services. Just recently the NSSDC web server passed the 1 billion accesses mark with the statistics that have been kept since 1996. This makes it one of the ten most popular sites within NASA. The main customers of NSSDC are researchers, but the general public forms a large fraction of those accessing the web sites as well and education-related activities address this **customer focus**. The NSSDC information sites serve both the formal education community (students gaining knowledge for classroom purposes) and the informal education individuals and groups (learning taking place outside of the classroom). This follows the recommendation from the Sun-Solar System Connection Roadmap that says: "NASA Education and Public Outreach should take advantage of the existing connection and overlap between the formal and informal education arenas." Thus, the site continues to provide information for those interested in space data, possibly inspiring their involvement in Science, Technology, Engineering, and Mathematical (STEM) subjects.

As many NASA astronauts, scientists, and other workers will tell you, it is often the first encounter with pictures of other worlds or the exposure to results from space exploration experiments that stimulates the interest to go deeper into this subject area and perhaps follow it as a career. To quote from NASA's Strategic Objective for Education and Public Outreach, "Use NASA missions and other activities to inspire and motivate the nations' students and teachers, to engage and educate the public, and to advance the scientific and technological capabilities of the nation." In this case we use the NASA mission data for this purpose. Also, NSSDC staff lead the Radio JOVE educational radio astronomy project (described below) which is being adapted for use by the visually impaired, an underrepresented group needed to increase **diversity** in STEM careers. These activities are what is needed to fill the "**pipeline**" with high-quality students who will take up STEM professions and possibly join NASA in the future.

The NSSDC is a prime location for having educational information available since ready access to a wide variety of research data can be inexpensively made available to the public as well. This **leveraging** of funding is most efficiently done within the NSSDC. This is essentially a **partnership** between NSSDC and the education community.

The value of the educational activities done at NSSDC are indirectly **evaluated** by the popularity of the associated web pages. As indicated by the statistics above, the existing sites are apparently considered to be of value. Periodically we propose to add online surveys of NSSDC users to directly query their reaction to web pages of interest and then update the pages accordingl

### 1.8.1 NSSDC Educational Web Pages

The NSSDC photo archive is a very popular service for the general public. The overall NSSDC analog photo archive had been prioritized for digitization. When there is time available for the photo archivist, Mr. Robert Tice, or for summer students, additional digitization of the archive has been done according to the prioritization and according to indications of recent public interest based on requests.

The NSSDC continues to digitize and send out copies of images from the analog photo archive upon request from researchers and for small requests from the general public. This leads to a gradually increasing set of digital imagery that will be available to the public on the educational web pages. Thus, the archive will automatically include those images most requested and of greatest interest to the public.

The NSSDC education web pages point to many other pages and web sites of interest. Some of the pages maintained by Mr. Nathan James, the NSSDC webmaster, have proven to be of particular interest to the public are the general public page (http://nssdc.gsfc.nasa.gov/nssdc/gen_public.html) and the links from that page to imagery, especially planetary imagery. There are also special pages devoted to specific topics of interest that are often being sought by general public users, such as information about the Apollo project. NSSDC also hosts a number of educational activity sites on its computer infrastructure. All of the educational activities will be advertised through notices on the main NSSDC pages and on the NSSDC education home page and NSSDC advertising brochures will be added to the offline data orders.

### 1.8.2 Radio JOVE Project

The Radio JOVE Project is a hands-on educational program aimed at high school level students, but suitable for middle schools, colleges or any other interested groups or individuals. The groups or individuals build a radio telescope from an inexpensive kit and use it to observe the Sun, Jupiter, our galaxy, and Earth-based radio emissions. They also learn about space physics of solar system magnetospheres, the solar wind, and solar storms. Students can also participate by monitoring remote radio telescopes through the internet. The project's leader, Dr. James Thieman is an NSSDC senior scientist with a Ph.D. in planetary radio astronomy. Radio JOVE has been selling/distributing radio astronomy kits for eight years at a rate of more than 100 per year on a non-profit basis around the world. The project emphasizes the inspirational approach of hands-on inquiry-based learning of astronomy and, more generally, the scientific method. The archive of data records contributed by amateur radio astronomy observers now has several thousand records and this archive is maintained at the NSSDC. Many NASA Explorer schools have been trained to use the equipment as part of the Student Observation Network (SON) module called "Tracking a Solar Storm". Radio JOVE has received funding to help broaden the diversity of students doing astronomy by adapting the project for use by blind and otherwise disabled people.

A new capability has recently been added to Radio JOVE with the development of multi-frequency spectrograph monitoring software that can be freely downloaded and used by observers to monitor two professional radio spectrographs located in Hawaii and Florida. The data are of professional quality and could be used by both amateur and professional radio astronomers. The data from the spectrographs are being archived at NSSDC as well.

### 1.8.3 Moon Trees Project

Dr. David Williams, a contract staff member for the NSSDC has been involved in the Moon Trees project since 1996. The project involves tree seeds that were flown to the Moon on Apollo 14 in 1971 and brought back to Earth. The seedlings were widely distributed to schools and organizations for planting as part of the nation's bicentennial celebration. The trees were planted as a "tribute to our national space program" and "a living symbol of our ... scientific achievements" and still serve to inspire people, especially children, about the role they might play in future space exploration. Most of the trees are not prominently marked as "Moon Trees" and systematic records were not kept, so the project is attempting to locate these trees and find out some of the history behind them. The NSSDC staff member is coordinating this effort which includes a set of web pages on the Moon trees. These pages link back to information about the Apollo missions, science experiments, and information on the Moon which will help to motivate people to learn more about lunar exploration. In recognition of this work, Dr. Willliams was invited to deliver the commencement address at the French International School in San Francisco in June 2004. More recently he participated in a Moon Tree planting ceremony in Arlington National Cemetery.

# 2 Future Directions FY 2006-2010

## *2.1 Major themes*

Over the 40+ years of NSSDC’s existence, we have seen substantial evolution in the Space Science’s information systems, including archives. As noted, NSSDC’s role has also evolved substantially while retaining a focus on long-term preservation.. As we look to the future, it is clear that NSSDC must retain this focus while promoting the best values for Space Science’s information system needs. It must be pro-active in promoting cost-effective approaches, and it must be flexible and nimble in responding to the changing organizational, implementation, technical and policy environments.

To this end, NSSDC has identified a number of significant thrusts that it intends to pursue given the current environment. It must, of course, periodically evaluate them for any needed changes in direction and prioritization.

### 2.1.1 Preserving information efficiently

Approach: Work with the discipline archives and broader community to ensure its space science data is adequately documented for understandability and is in a form that is cost-effectively preservable for the long-term.

The most basic function of the NSSDC is to ensure the long term preservation of space science data in a form that is understandable and usable to the broad space science community. This means that space science data must be in a form that can be readily migrated across different media types and must have adequate accompanying documentation that can also be readily migrated with it. The techniques and criteria used to reach this state, or some approximation to it, vary widely across the space science disciplines and often from one set of data to another. NSSDC will more aggressively promote categories of documentation that should be available, and it will work with the disciplines and active archives to reach consensus on these requirements and on mechanisms to improve conformance, as needed. It will also address the issue of the cost-effective long term preservation of information currently residing in databases.

### 2.1.2 Doing more cost effectively (more NSSDC for the buck)

Approach: Benefit preservation operations by improved processes with an increased reliance on low-maintenance system capabilities (more work for less cost).

In the long and short term plans for NSSDC, automation plays a key role. We certainly will benefit from increased reliance on lower maintenance system capabilities, i.e. more work for lower cost. Optimistically, this is a way to lower costs. Realistically, with the projected growth in the volume of datasets, more-automated and lower-maintenance systems are a necessary survival skill for archives. With nearly static budgets, more efficient archive systems are the primary way to prevent our archive drowning and sinking under the massive volumes of data predicted in our near future.

The menu of possibilities to increase efficiency include those improvements already in progress, areas we have targeted for future improvement, and longer term issues we need to address.

The legacy tape migration project is our largest, ongoing, improvement in efficiency. Not only is it essential for preservation of the data's information content, but the transformation into AIPs and storage in the near-line DLT jukeboxes will greatly improve our monitoring and access capabilities. Future migrations of this content to newer media will be far more cost-effective. Our long-term plan is to move most content, both permanent and second archive, into AIPs to take advantage of these efficiencies. Magnetic tape currently remains the most cost effective storage method.

We DO NOT plan to put the entire, or even most of, the permanent archive on the web in the near term. We will, however, continue to make more data web accessible as per MOUs with other archives and projects, and as the sole provider of data of significant interest to the research community.

We plan to improve web access to information about the digital datasets in the permanent archive, such as with improved search capabilities. We may also post sample data from various datasets, which may be especially useful for some of the "ancient" (epoch < 1970) datasets in the permanent archive. We need to similarly expand web information about our analog datasets – i.e. images, documents and maps.

We plan to make more of our analog archive electronically visible and accessible. Long-term plans include digitizing most of the NSSDC's archive of analog datasets, primarily images, documents and maps. We would likely preserve the original analog data off site, post the most requested data on line and keep all digital copies at least near line.

Historically the archive has been a very human-intensive effort, with scientists or highly-skilled technical staff required to perform extensive work to document and permanently archive arriving data. With increasing volumes of data arriving from flight projects, such a labor-intensive effort can no longer be supported. Large volumes of data will need to be ingested quickly, along with ancillary information and data. NSSDC is working with active archives and other data providers to ensure that the ingest process is as streamlined as possible, with automation and programming solutions handling the bulk of ingest, freeing staff to work on providing solutions to future demands. Specifically, an internal assessment of the entire process of preparing a dataset for ingest was completed. It showed that the current process relies primarily on the acquisition scientist to provide supporting information in a series of four steps. Each subsequent step was with a different member of the acquisitions staff. The improved system will rely on the acquisition scientist to provide all the necessary information about the dataset in one consolidated submission early in the process. The data and supporting information then will be marshaled through the internal NSSDC processes by the ingest manager, who will confirm that all the required steps have been completed successfully and then authorize the actual ingest of the data. We will further review which of these internal steps can be completed in parallel rather than in series and assess which steps might be automated, combined or eliminated.

NSSDC's hardware systems are under constant review, as necessitated by the continuing evolution of technology. We are counting on increased storage per magnetic medium and on larger jukeboxes with shorter access times for data from magnetic media. For example, the ongoing legacy tape migration from 9-track and other equally obsolete tape formats to DLTs and Super DLTs is a major step forward. To continue to hold onto our gains and keep a viable archive requires continuing media refreshes.

### 2.1.3 Providing better solutions for the future while preserving the past

Approach: Increase the integration of NSSDC's metadata/data repositories through improved architectural designs and implementation that improve consistency and cost-effectiveness.

NSSDC has a long history of archiving space data and of documenting not only the data, but ancillary information about its sources. While this has led to a rich information environment, it has also led to disparity as additional systems or architectures were adopted to take advantage of new technologies. The result has been that some parts of the NSSDC's environment have evolved more quickly than others and some have remained relatively static for decades.

In today's world of demanding consumers, where information and data are required quickly, it is necessary to be responsive to the higher expectations that researcher and other interested parties have. While great strides have been taken to better integrate our various data and metadata systems, there is still more that needs to be done to stay competitive.

NSSDC's data and information systems must take advantage of new capabilities and new architectures as they evolve to address these demands. Several disparate metadata systems have been re-integrated over the last decade and that process is expected to continue. Plans are in place to better integrate NSSDC's off-line and near-line inventory systems as well as to make its extensive photographic inventory more accessible to researchers and the general public.

### 2.1.4 Resident Archives – bridge to quality mission data

Approach: Work with space physics discipline scientists to evolve the Resident Archive concept for topical or scientific groupings.

That is a goal of the Resident Archive (RA) effort. NSSDC has been involved since the inception of the RA concept and is a key player in the effective management of RA resources that enhance the quality of data delivered to the permanent archive. A vital aspect of NSSDC's interactions with Resident Archives is to provide the management needed to bridge the responsibilities of the missions with the requirements of the larger solar and space physics communities.

Following the specific requirements of its MOUs with projects and RAs, NSSDC provides the model needed to formulate a successful RA Data Plan (RADP), assembles the standards and processes for providing quality data on a continuing basis, and incorporates the guidance provided by review committees and interactions. Each RA will be a unique case but it can be built on the experience transmitted through an RA User Group established under NSSDC policies. A future evolution involves the grouping of RAs into natural coalitions where economy of scale or expertise comes into play. The resulting loose confederation can then provide NASA with a cost effective means to provide quality data based on the mission experts, in a continual manner for the benefit of the community and with an eye toward the long-term preservation for future generations.

### 2.1.5 Finding the right data, right now

Approach: work with VO middleware developers, active archives, project archives, including RAs, so as to improve research effectiveness of space science services (right data, right now).

Finding the right data right now is the goal of virtual observatory (VO) efforts, including registries and Web Services for transparent distributed data discovery and access. VOs enable interoperability, and interoperable systems enable powerful capabilities for correlative science. This is one of the primary scientific justifications for interoperability. Some of the greatest scientific discoveries of the past decade have come at the intersection of disciplines or sub-disciplines, where data from multiple sources are combined, correlated, and inter-compared for new insights into the nature of physical and astrophysical systems. But, such interoperability delivers an even broader set of value-added functions. These include: (a) data quality verification and data validation (from complementary data); (b) increased usefulness and usability of data when examined in concert with other datasets; and (c) ease of data access from multiple distributed systems. This is essentially the motivation for VO efforts, and VOs are in fact the instantiation of interoperable systems for space science data services.

NSSDC will work with VO middleware developers to take advantage of the latest developments for registering datasets and data ordering interfaces. This is a vital aspect of NSSDC's interactions with active archives and Resident Archives – in such cases, the vision of one-stop shopping for all of a user's data needs requires active VO participation of all archive sources (permanent, active, and Resident, plus project archives). The research effectiveness of the community is thereby significantly enhanced, as data searches, requests, and accesses are brought to a single focus at whatever NASA space science information service entry point the user chooses.

NSSDC will provide the legacy data component of the full spectrum of space science data resources. In this VO-enabled environment, cost-effectiveness is achieved through reuse, common development, and resource-sharing. Cost savings in these areas translate into greater end-user services, capabilities, and space science research support functions elsewhere within NASA's permanent, active, Resident, and project archives. Consequently, NSSDC will be able to provide a richer and more accessible data environment for space science archival research, long-term studies, historical data anchor points, reference data, trend analyses, and new mission planning. VO-style data discovery and access services will augment NSSDC's strength in fulfilling these obligations, thus enabling greater research capacity within the community and greater return on NASA's archival systems investments.

NSSDC is also the logical home for some data and service (resource) registries, particularly for the space physics community. Where appropriate, NSSDC will implement these information and middleware services in XML form, so that the content can be cost-effectively reused in many applications through the use of XSLT stylesheet transformations. XSLT allows the same meta-tagged information content to generate forms, to populate databases, to generate web pages or other documentation, to drive applications or services, to populate registries, or even to transform data and metadata from one format to another, tailored to the specific end-user needs of those data and services.

### 2.1.6 Working Smarter

Approach: Work with other archives to identify the problems and solution for handling complex and voluminous data through evolution of technology and standards.

It is no longer (if ever it was) appropriate for stove-pipe development activities among NASA's distributed science archives. Smarter processes begin with effective communication, cross-fertilization, resource-sharing, reuse, and shared lessons learned. NASA supports a number of archives in the space science domain, and many of them are already older than 10 years. These archives also have a number of common, largely discipline independent, issues associated with preserving and managing their holdings. These include cost-effective solutions for handling large volumes of data; ensuring the information content is well identified and preserved against the ravages of technology evolution, security breaches, and operator error; maintaining awareness of growing consensus on good digital preservation practices in the broader digital preservation community; and maintaining awareness of evolving technologies that can benefit the archives operations.

NSSDC plans to play a growing role in facilitating information discovery and sharing about these issues through a number of approaches, including continued participation in relevant standards developments, publicizing resources that come to its attention by virtue of its participation in broadly based standards development, working more closely with various archives on a one-to-one basis to understand and publicize their approaches and to offer NSSDC perspectives, and facilitating multi-archive workshops to address issues as appropriate.

NSSDC will participate and contribute to these interoperable archive efforts through participation in such cross-archive working groups and steering committees as the Resident Archives Users Group, PDS Management Council, and ADEC (Astrophysics Data Centers Executive Council). Within these contexts, NSSDC and the active, Resident, and project archives identify common problems, reusable and cost-effective solutions, opportunities for resource-sharing, and opportunities for interoperable data services. As NASA's space science data enterprise encounters skyrocketing data volumes and increasing complexity of data types, data organization, and database technologies, it is imperative for NSSDC and the other archives to work more closely together in addressing these challenges and in providing new powerful functionalities to meet them.

Through its active participation in standards development, long-term preservation, and OAIS reference implementation, NSSDC is therefore poised to contribute significantly to NASA's interoperable cross-archive improvement and enhancement efforts. In the end, the research capabilities of NASA's space science community will grow and the potential for exciting new scientific discoveries from multi-datasets will be greatly magnified. Correlative science is the strong end-user benefit derived from interoperability and closer cross-archive activities.

# 3 Appendix A: Acronym List

| | |
|---|---|
| AA | Active Archive |
| ACE | Advanced Composition Explorer |
| ADC | Astronomical Data Center |

| | |
|---|---|
| ADEC | Astronomy Data Centers Executive Committee |
| AIM | Aeronomy of Ice in the Mesosphere |
| AIP | Archive Information Package |
| ALSEP | Apollo Lunar Surface Experiments Package |
| CCSDS | The Consultative Committee for Space Data Systems |
| CD | Compact Disc |
| CDAWeb | Coordinated Data Analysis Workshop web interface |
| COBE | Cosmic Background Explorer |
| CODATA | Committee on Data for Science and Technology |
| COSPAR | Committee on Space Research |
| CRC | Cyclic Redundancy Check |
| DCC | Digital Curation Centre |
| DIOnAS | Data Ingest and Online Access System |
| DLT | Digital Linear Tape |
| DVD | Digital Video Disc |
| ESA | European Space Agency |
| FAST | Fast Auroral Snapshot Explorer |
| GALEX | Galaxy Evolution Explorer |
| HEASARC | High-Energy Astrophysics SARC |
| IBEX | Interstellar Boundary Explorer |
| ICSTI | International Council for Scientific and Technical Information |
| IDA | Interactive Digital Archive |
| IMAGE | Imager for Magnetopause-to-Aurora Global Exploration |
| IMP | Interplanetary Monitoring Platform |
| InterPARES | The International Research on Permanent Authentic Records in Electronic Systems |
| IRAS | Infrared Astronomical Satellite |
| ISEE | International Sun-Earth Explorer |
| ISO | International Standards Organization |
| JAXA | Japan Aerospace Exploration Agency |
| JEDS | Java-interface to Experiment, Dataset, and Spacecraft (and publication) information in NIMS |
| JIN | Java-interface to INventory information in NIMS |
| JRAND | Java-interface to Request And Name Directory information in NIMS |
| LAMBDA | Legacy Archive for Microwave Background Data Analysis |
| MAST | Multi-mission Archive at Space Telescope |
| NARA | National Archives and Records Administration |
| NCITS | National Committee for Information Technology Standards |
| MER | Mars Exploration Rovers |
| MOU | Memorandum of Understanding |
| MPGA | Multifile Packager and Analyzer |
| NASS | NSSDC AIP Storage System |
| NEAR | Near Earth Asteroid Rendezvous |
| NIMS | NSSDC Information Management system |
| NMC | NSSDC Master Catalog (public view to NIMS) |
| NSF | National Science Foundation |

| | |
|---|---|
| NSSDC | National Space Science Data Center |
| OAIS | Open Archival Information System |
| PDMP | Project Data Management Plan |
| PDS | Planetary Data system |
| PI | Principle Investigator |
| PV-2005 | Ensuring Long-Term **P**reservation and Adding **V**alue to Scientific and Technical Data (PV 2005) |
| RHESSI | Reuven Ramaty High Energy Solar Spectroscopic Imager |
| RLG | Research Libraries Group |
| SAMPEX | Solar Anomalous and Magnetospheric Particle Explorer |
| SARC | Science Archive Research Center |
| SDAC | Solar Data Archive Center |
| SDC | Science Data Center |
| SDLT | Super DLT |
| SIP | Submission Information Package |
| SON | Student Observation Netword |
| SPASE | Space Physics Search and Exchange |
| SPDF | Space Physics Data Facility |
| STEM | Science, Technology, Engineering, and Mathematical |
| SWAS | Submillimeter Wave Astronomy Satellite |
| THEMIS | Time History of Events and Macroscale Interactions during Substorms |
| TIMED | Thermosphere Ionosphere Mesosphere Energetics and Dynamics |
| WDC SI | World Data Center for Satellite Information |
| XML | eXtensible Markup Language |

=======