National Space Science Data Center
Senior Review 2009

Cognizant Personnel: E. Grayzeck, B. Jacobs, N. James, J. Thieman, D. Williams, E. Bell, J. Garrett, H.K. Hills, L. Martin, P. McCaslin, C. Ng, and J. Russell

Contents

Executive Summary

The National Space Science Data Center (NSSDC) provides a vital service to NASA and the scientific community as its long-term, multi-disciplinary Space Science archive. From its founding as the archive for all NASA science data, NSSDC has had to maintain a broader view than subsequently established specialized archives, both in data diversity and in data longevity. Sustaining and improving its curation activities are essential to ensuring that data gathered at public expense will continue to be available and usable into the indefinite future.

All portions of the Senior Review Call are answered. Specifically, Section 1 addresses the science merits and specific contributions of various activities, and how NSSDC will continue to discover and communicate new scientific knowledge in line with NASA's Strategic Goals, Objectives and Research focus areas. The relationship of NSSDC to other Heliophysics data services is discussed in Section 2 along with how NSSDC is working with these services to ensure conformity of data and metadata to accepted standards (*i.e.*, how each project providing data services is implementing the Heliophysics Science Data Management Policy).

In this context NSSDC recognizes the following six challenges to be addressed in the next few years:

1. provide researchers a means for searching many widely-distributed archives in order to retrieve relevant data of interest;
2. make data delivery and internal processing simple and seamless;
3. facilitate the migration of archive data collections, especially in Heliophysics, to more usable, modern formats;
4. streamline the process for capturing metadata for data being archived;
5. convert selected parts of NSSDC's vast analog holdings to digital form when feasible for increased science return; and,
6. monitor and evaluate changing technology to ensure the long-term preservation of digital data.

NSSDC's solutions (initiatives) to these challenges are discussed throughout this proposal. Each initiative is designated in the table of contents, *e.g.*, Section 1.5.1 contains C-1 in its title, indicating that it discusses initiative 1 related to challenge 1 in the above list.

NSSDC has and will continue to work with: (a) flight projects in planning their data products and formats (*i.e.*, Project Data Management Plans); (b) organizations sponsoring Heliophysics Resident Archives (RAs), to assist their formation; (c) other data facilities (*e.g.*, Heliophysics Final Archives) to help provide data to scientists to solve their science problems; and, (d) flight projects and data facilities to store, maintain, and distribute their data after they have closed shop. NSSDC remains strongly active in data standards methodologies both inside and outside of NASA and also provides four levels of archive storage: permanent archive, second archive, backup, and analog archive, which are discussed in the proposal.

The NSSDC has archived data in multiple science areas (heliophysics, planetary, and astrophysics) and its planning for data preservation and usability focuses on the indefinite future, 100-200 years, not just the life of a mission nor the life of a mission team. That longer-term focus imposes different standards for data and metadata than that for other archives. The future plans, initiatives, and expectations are based on our expertise associated with efficiently running a multi-discipline, long-term archive, entrusted with the indefinite preservation of irreplaceable science data obtained at great public expense.

1. Enabling Science

A main goal for NSSDC is to preserve data over the long term for scientific investigation. As described later, NSSDC follows the ISO standard "Reference Model for an Open Archival Information System (OAIS)". By adhering to this model, NSSDC enables a broad range of science as will be described below. First, we relate the NSSDC operations to the Strategic Goals of NASA as identified in its Agency-wide Strategic Plan. Second, we assert a set of definitions that guide the value we provide to science as well as how we implement the curation role for NASA. Finally, we give instances where these principles have enabled science.

1.1 Relationship to NASA Strategic Plan

Based on the NASA strategic plan, the Heliophysics, Planetary, and Astrophysics roadmaps, and the 2006 Senior Review discussions with headquarters, NSSDC activities are aligned with the following Level I requirements:

1. Function as the space science data/metadata repository -- to obtain data generated from missions, hold it safely and securely for as long as it has significant value, and make it available to the research community when it is not available elsewhere or when specified by other agreements. This is accomplished by working with discipline data systems, their repositories, missions, and investigators.
2. Provide the space science community with data stewardship guidance and support so that data made available by various repositories is well documented in order to support independent usability via, for example, Virtual Observatory access. Independent usability is also critical in recognition of the reality that a large fraction of NASA and support contractor staff are likely to retire in the near future.
3. Make unique data and metadata available, as well as participate in Virtual Observatory development efforts to assist in the practical evolution of those concepts.
4. Pursue Education and Public Outreach efforts when it can cost-effectively leverage its repository activity, to also encourage an adequate Science, Technology, Engineering and Math (STEM) workforce to meet NASA's needs.

These requirements are supported by a combination of implicit and explicit text from the space science strategic plan and roadmaps. NASA strategic plan goal 3, "*Develop a balanced overall program of science, exploration, and aeronautics consistent with re-direction of human spaceflight program to focus on exploration*", and its sub-goal 3b "*Understand the Sun and its effects on Earth and solar system*", clearly motivate the missions and roadmaps of the heliophysics, planetary, and astrophysics disciplines. These new understandings would be severely impacted unless the data from these programs are available to the broad research community long enough to cost-effectively extract relevant information (Level I-1).

Specifically, the 2006 NASA Science Plan chapter 6 on Heliophysics notes that "*Enabling the return and synthesis of vast datasets from anywhere in the solar system [6.4.3]*" is part of the Technology investment. In addition, section 6.4.4 states that *"harnessing the full benefit of heliophysics science over the coming decade requires a move from the single topic analysis techniques to the implementation of well-managed archives, virtual observatory systems …"* This is not possible in practice without reliable repositories of data (Level I-1). It also states that "*Examples currently under development include virtual observatory tools, which are providing integrated access to the distributed data bases built from the observations of multiple platforms…*" The existence of a 'data repository', clearly distributed as part of the Great Observatory, is assumed (Level I-1). Comparisons of data from different missions, as through use of Virtual Observatories, is also expected and supports the need for Level I requirements '2' and '3'. Under the Heliophysics Data Environment Policy, it states that a variety of internal and external organizations want access to NASA's knowledge and data in this area, and clearly this is supported by requirements C-1, 2 and 3. In addition, NSSDC is identified in that policy as an option for "… long-term preservation. In this context, the NSSDC has offered and will continue to offer one means of providing safe, long-term backup of data from various mission phases, including the level-zero data of an active mission through to final archive products."

The Planetary and Astrophysics roadmaps stay focused on the mission elements and the unsolved technical and environmental problems. They do not discuss specific implementations supporting research and data analysis, apart from noting the need for 'ground truth' data to support some missions. Clearly the capture and use of data from missions, along with some ground truth data, is central to accomplishing the scientific and exploration objectives outlined (Level I-1). In addition the latter notes "*The large data volumes may be accommodated in part by distributing data sets and analysis. However, the software tools as well as the connectivity for such data systems will require new approaches and architectures for synthesizing these data streams (e.g., National Virtual Observatory (NVO)).*" The need for data archives/repositories and virtual observatory access is given explicitly (Level I-1,2 and 3). Under Appendices it notes "*Information gathered from NASA web sites serves, in many cases, as the primary scientific archive for students of all ages*" and "*Finally, the general public, as our ultimate customer, is a crucial external constituency. Results from NASA space missions, as covered by popular media, have the demonstrated ability to inspire citizens of all ages, and they engender good will in support of the overall vision and goals of the agency*." Requirement C-4 supports this.

From the Level I requirements and an understanding of the current roles of repositories within NASA, NSSDC's Level II functional requirements are derived. NSSDC provides long-term, cost-effective curation to NASA's data and information resources within a multi-discipline environment, based on MOUs established with data providers. NSSDC organizes its functions to align with the international ISO standard OAIS Reference Model for archives and repositories and simultaneously support Level I requirements as follows:

1. Ingest data and metadata from data providers with adequate documentation for long term independent usage (OAIS Ingest)[Level I-1, I-2]
2. Store data safely, securely and cost-effectively to maintain its viability as long as it is deemed to be of sufficient value (OAIS Archival Storage)[Level I-1]
3. Store metadata supporting finding and access via the user community, as well as supporting internal administration needs. (OAIS Data Management)[Level I-1, I-2]
4. The Interactions of NSSDC encompass the following: provide an access service to the NSSDC holdings, including support for Web services, Virtual Observatory access to NSSDC unique data, and selected data of general interest to the public. (OAIS Access) [Level I-3]. Provide advice from NSSDC on standards and templates supporting cost-effective data and metadata management, and on standards and guidelines for data provider supplied metadata to help ensure long-term independent usability. (OAIS Preservation Planning) [Level I-1, I-2]. Coordinate NSSDC functions including personnel, budgets, space, MOUs with data providers, and the migration of data to ensure continued cost-effective curation. (OAIS Administration) [Level I-1, I-2, I-3]

1.2 Definitions: Science Productivity, Science Merit, Science/$

In the Senior Review call we are asked to express our archiving functions in terms of science per dollar. In order to do that NSSDC uses the following definitions:

- *Science Productivity*: the ability to provide many years of accumulated data
- *Science Merit:* NSSDC houses many unique data that can't be duplicated.
- *Enabling science*: NSSDC seeks to provide data with enough metadata to be understandable and useful.

"Science/$" is difficult for NSSDC to measure as its goals are to provide a multi-discipline archive based on international standards, and to be an insurance policy to the specific science discipline active archives or virtual observatories.

NSSDC's digital archive is 77 TB. A too simplistic definition of "Science/$" would be to take the current budget ($4.2M) and divide by that total volume of data. To narrow this discussion, we must realize that about 50% of the effort is devoted to Heliophysics based on the number of data collections. In 2008 in Heliophysics NSSDC accrued 1.2 TB of data. These data may be added to existing collections (e.g., RHESSI) or be new additions. NSSDC has 25.7 TB in heliophysics digital data and over 1000 analog data collections as well. Half ($2.1M) of the NSSDC budget is devoted to heliophysics activity, so it is possible to calculate cost per TB. A more realistic example is to multiply the full set of 1,062+ digital data collections by an average number of products in each collection. From our work in creating Archival Information Packets (AIPs), 2,000 products (files) are estimated per digital data collection so it is possible to calculate the cost per product. An alternative approach for "Science/$" is to provide reports on cases when NSSDC supplied data or expert advice that required extra effort. In Section 1.4 we give such examples by science discipline.

1.3 Terminology: Curation, Completeness, Standards, Access

- *Curation* is the preservation and maintenance of data. The preservation of data means to ensure that it remains safe from harm, that the data which were received remain intact. The maintenance of data means to ensure that it is kept in an appropriate condition so that it can be accessed and used into the indefinite future. In order to ensure accessibility it is often necessary to not simply preserve the "bits", but also to transform or modernize the data so that it continues to be of use.
- *Completeness of data*, for our purposes, is defined as that part of the data and metadata generated by an instrument or investigation which has been determined by a group of discipline scientists (which might be limited to the investigators involved in capturing/processing the data) to encapsulate the salient information captured thereby.
- *Standards* are generally accepted principles for the best/most appropriate way to preserve and/or exchange data and/or metadata. These may be community *ad hoc* standards (such as SPASE) or ones based on organizational formalism (as with ISO or CCSDS).
- *Data access* is the ability to discover and subsequently obtain data. This includes the metadata infrastructure that is necessary to support such functions.

1.4 Working in the SMD Data Environment

Curating data involves interacting with missions and other archives, often on special projects designed to enhance the science return of archived data or future missions. Specifically this can mean reading and reformatting data to make them usable by mission planners, researchers, or other archives, or working with instrument teams to ensure the proper metadata are fully captured. Below are some recent examples from space physics, solar physics, planetary science, and astrophysics.

1.4.1 Heliophysics: Space Physics

NSSDC has participated in the Resident Archive (RA) process from the start, laying out the general outlines to RAs through a white paper (http://nssdc.gsfc.nasa.gov/nssdc/RAXWhite_Paperu.doc). A Resident Archive is tasked with providing access to data and expert assistance for some period of time after missions have ended. The first RA, the SAMPEX Data Center, has now moved into the maintenance phase, and NSSDC staff acted as advisors in the execution of the initial proposal to revamp

SAMPEX data and then stage it at the Ace Science Center.

NSSDC is involved in converting a complicated Apollo sub-satellite magnetometer data collection for a requester. Peter Chi (P.I., EM Wave Resident Archive) of UCLA asked for help on this same conversion, on which he was working as part of a funded LASER proposal. NSSDC staff met with him and provided help on the various steps involved, and the converted version of the data will be obtained for the archive and for our requester, rather than expending our limited resources in duplicating the effort.

Various ISEE data were staged for retrieval for Dr. Theodore Fritz of Boston University, including: ISEE-1 and ISEE-2 4-Second Averaged Electron and Proton Flux data, ISEE-1 Proton Fluid Parameters 6 Earth Radii - Bow Shock, ISEE-1 LEPEDEA 3-D Proton & Electron Distribution Functions; 128/512-Seconds, ISEE-1 0.1-Second Electron and Proton Flux (IEEE binary), ISEE-2 Three Dimensional Fast Plasma Data.

A recent request was received from Ryan Olds, a graduate student at the University of Colorado, for Explorer 35 data "in a modern format". He was directed to our anonymous FTP site to a tar file of data from the MIT plasma instrument and was sent a format description. He also was interested in magnetic field data, so a staff scientist adapted an existing routine to unblock the IBM binary blocked data collection, then generated a new customized program to convert the contents to an ASCII format and sent the data file as an email attachment, together with the documentation describing the data collection. The additional files of this data collection will be similarly converted and the complete set will become a new archived data collection. An additional request for energetic particle data from the same spacecraft will be addressed next; if the data are as described, it will be a relatively simple conversion.

To the right is a note from Dr. William Reupke at NASA GSFC following his request for a document titled *Data Processing System Used for Radiation Data of Explorer VII*. The document was scanned from microfiche and e-mailed to him.

All,

I just want to thank all of you for assisting me in locating the report, cited below, by Suomi, Parent and Swift. The copy was found at the National Space Sciences Data Center located near my office!

This report contains a data flow diagram for the Wisconsin earth radiation experiment launched aboard Explorer VII in 1959. The diagram will prove very useful in preparing my talk Satellite Meteorological Product Generation Systems: Past, Present and Future.

Again, thanks !

Bill Reupke
Dr. Bill Reupke
GOES-R Ground Segment
Goddard Space Flight Center

1.4.2 Heliophysics: Solar Physics

Yohkoh data is in four different NSSDC data collections spanning 1991-09-01 to 1997-06-07. The Yohkoh Legacy data Archive (YLA) at Montana State University is one of the first group of the Resident Archives (RAs) selected for funding for NASA's Virtual Observatories for Heliophysics Data program. YLA provides the best corrected data set of solar X-ray images and spectra from the Yohkoh satellite with a user-friendly web interface. As an RA they promise to: (1) provide a corrected data set with improved image quality for SXT data; (2) provide FITS files as well as a specific data format called XDA; (3) develop a user-friendly web interface for data access; and (4) accumulate value-added data and documents. They are promoting the use of the YLA data in the hope to extend their two-year grant. Dr.

Aki Takeda of MSU has promised to notify NSSDC when they will no longer be able to maintain the YLA archive. Details of how the Yohkoh archive might be transferred to NSSDC have yet to be determined.

1.4.3 Planetary Science

A research team at Arizona State University's School of Earth and Space Exploration (ASU SESE), under the direction of Mark Robinson, is using a high-resolution scanner to digitize all the original Apollo 15, 16, and 17 metric (mapping) and panoramic camera film. These represent some of the best photography of the lunar surface taken to date. In addition to their obvious historical importance, they are also scientifically important, covering over 20% of the Moon's surface at resolutions up to 1-2 m. The images will be used for targeting the Lunar Reconnaissance Orbiter Camera (LROC), for which Robinson is principal investigator. To use these images for camera targeting, rigorous navigation and pointing information is required for each frame. These data only exist on microfilm indexes archived at NSSDC, so we digitized these ten microfilm reels (roughly 20,000 frames) and delivered the digital copies to ASU SESE. There they are run through optical character recognition (OCR) software and converted into Planetary Data System (PDS) SPICE files archived at the PDS NAIF (Navigation and Ancillary Information Facility) Node and will be used to make the images scientifically usable for the MESSENGER LROC and future researchers. The ASU SESE project is documented in *The Apollo Digital Image Archive*, M.S. Robinson *et al.*, Lunar and Planetary Science Conf. 39, 1515, 2008 and *Targeting the Lunar Reconnaissance Orbiter Narrow Angle Cameras*, B.L. Joliff *et al.*, Lunar and Planetary Science Conf. 40, 2343, 2009.

NSSDC responded to a request by Wendell Mendell (JSC) to read a set of tapes containing data from the Apollo 17 Infrared Radiometer. His collection consisted of 7- and 9-track tapes which were sent to NSSDC. Dr. Mendell did not know exactly what data were contained on the tapes, only that it was Apollo radiometer data. The capability existed at NSSDC to read the 9-track tapes, which was done and the results sent back to Dr. Mendell. The 7-track tapes are awaiting a decision by Dr. Mendell as to whether the data they contain justifies the effort to read them. If so, they will be sent to John Bordynuik Inc., as NSSDC no longer has the capability to read 7-track tapes. NSSDC plans to archive these data upon completion of this effort.

1.4.4 Astrophysics

James Adams at the National Space Science and Technology Center at NASA's Marshall Space Flight Center requested data from Gravity Probe-B (GPB), specifically the Level 1 and Level 2 proton monitor, magnetometer, and orientation data. He and a colleague, J. Kolodziejczak (Project Scientist for Gravity Probe-B), are working on the next generation of astrophysics missions (OASIS). They want to characterize the environment as sampled by GPB in its orbit. Their analysis centers around the raw proton monitor data, included in Level 1. Ancillary data to the analysis is the magnetometer data, as well as the roll phase and spacecraft coordinates from the Level 2 data. The latter is important to understand the “roll-polhode resonance” used to estimate the “frame-dragging” effect.

The Gravity Probe-B archive has not been fully delivered at this time, so the requester was provided with a preliminary version of the data submitted, including all available documentation but without the star tracker data. The archive in this preliminary form is written as CDF files (designed with assistance from NSSDC) so that the proton monitor data are easily used in this type of space physics investigation.

The quote to the right refers to an NMC request which originated from a university customer requesting analog IMP data on microfilm. After talking with one of our scientists who was familiar with the data, he learned that the data had already been digitized and, with the proper coordinate system, could be found using our NMC system.

> *His [curation scientist's] comments are very helpful. The true-equator issue is not critical for what I'm trying to do, but I do encourage you to capture that level of metadata for future datasets - I'm wearing my Virtual Astronomical Observatory hat here, we've been putting a lot of work into standards for interoperable use of archival astronomy datasets, and the same must apply to space physics: being VERY explicit about coordinate systems is important.*
> *I do realize, however, how severely understaffed and underfunded you are there and am just glad that you are making good progress on getting stuff online.*
> *- Jonathan*
> *Jonathan McDowell [*jcm@head.cfa.harvard.edu*]*

1.5 Science-Based Initiatives and Major Themes

1.5.1 SPASE Support in Heliophysics (C-1)

The Space Physics Archive Search and Extract (SPASE) project is intended to provide researchers in the heliophysics science community a way to search many widely-distributed heliophysics archives for specified types of data and retrieving those which are of interest. A SPASE Data Model has been developed by international representatives within heliophysics to serve as a common schema for describing the available data. With a common metadata language, data collection descriptions since have been created that use standardized terminology, enabling queries among many heliophysics data centers that have widely-varying approaches for their internal representations of the data they hold.

SPASE-based user interfaces rely on this approach for rapid data search and retrieval. The Virtual Space Physics Observatory Product Finder (http://vspo.gsfc.nasa.gov/) has "SPASE inside" and works directly from the SPASE data descriptions that it stores. The Virtual Energetic Particle Observatory (VEPO; http://vepo.gsfc.nasa.gov/), Virtual Magnetospheric Observatory (VMO; http://vmo.nasa.gov/), and Virtual Heliospheric Observatory (VHO: http://vho.nasa.gov/) all have similar hierarchical data query interfaces based on the SPASE metadata. Further refinement continues for the SPASE Data Model to improve the data descriptions and provide more detailed and accurate requests to be constructed for whole data collections or subsets of the data.

1.5.1.1 Scenario

Consider the following scenario posed by Prof. Ray Walker in 2003 as an example of the research problem he wanted to investigate and posed as a type of problem that SPASE would help to solve.

> "I want to study the flow of energy through the solar wind magnetosphere and ionosphere system during magnetospheric substorms. Observations and a simulation of a substorm on December 22, 1996 revealed a previously unknown state of magnetospheric convection (1). Earthward flow from a tail neutral line reversed direction in the inner magnetosphere and formed a large scale vortical nightside convection pattern. The dawnside vortex formed a flux rope just prior to the substorm onset but this was not the cause of the substorm breakup. In the dusk side vortex the flow created strong field aligned currents directed away from the Earth that connected to the region of maximum auroral luminosity in the ionosphere. We believe these currents were related to the westward traveling surge. The flow reversal and formation of the tail vortices may result from the simulation's earthward convection being limited by high near-Earth plasma pressure or line-tying caused by high ionospheric conductance. This is a very different picture of inner tail convection than that traditionally associated with substorms. Now I would like to know how common this complex scenario is during substorms. To do that I will need observations in the near Earth tail, observations in the auroral ionosphere and observations in the solar wind. Selecting appropriate observations is a major task. In our combined data and simulation studies we use solar wind data to drive our simulation of the magnetosphere. In addition, for this study I

will need observations of plasmas and fields from spacecraft appropriately located in the near-Earth plasma sheet near the onset of a substorm and auroral observations to determine the time and location of onset. We typically use Polar or IMAGE UV images to select candidate events, and trajectory plots from NSSDC to figure out which spacecraft were in the tail. Then we go to ACE or Wind and check for appropriate solar wind data. Finally we check the browse data from Geotail, Interball and/or Cluster to figure out just where they are in the tail. It is a lot of work and setting up one study can take weeks or months. Just to select events we need to visit as many as 20 or 30 distributed data sites. I believe that the science return in this type of multi-instrument, multi-spacecraft study is the future of space physics and that we desperately need to make it easier. A simple tool that allows me to access data from all of these distributed sites from one stop would greatly facilitate my task."

This is clearly an elaborate and difficult study, but SPASE is now the unifying tool for any scientist interested in quickly locating and retrieving heliophysics data. It has taken a while to reach this point, but the establishment of the VxOs and the Space Physics Data Environment is now reaching the fruition envisioned many years ago. Data descriptions following a unified metadata model are now enabling the search and extraction of data needed for solution of complex multidisciplinary research.

Can SPASE be extended to more than just the heliophysics domain? Probably the closest discipline that overlaps with Heliophysics is Planetary Science. Many planetary missions carry instruments yielding types of data (fields and particles) that are useful in the study of comparative magnetospheres. It is possible to extend SPASE to describe such data. Planetary imaging data collections are not too far from the solar imaging data collections which are presently within the domain of SPASE. Can SPASE also be used to describe atmospheric, geological, gravitational, etc. data collections? The intent now is to extend SPASE into space physics types of planetary data. Extending beyond that realm will depend upon need and the success with space physics applications.

For the last few years funding for SPASE developments has come from the NSSDC core budget and this is expected to continue into the foreseeable future. The funding includes:

- Supporting the part time work of the SPASE manager;
- Hosting the SPASE web site on an NSSDC machine;
- Editing and updating the Data Model as well as website maintenance and development by UCLA staff;
- Programming development by both NSSDC and UCLA staff; and
- Working on data descriptions of NSSDC holdings by staff and consultants.

1.5.1.2 Future Initiatives for SPASE

The applications of SPASE to the Heliophysics Data Environment are presently being tested. It is anticipated that extending SPASE to other Space Physics domains will occur in 2010–2011 and going beyond those boundaries will be tried in 2012–2014. PDS personnel at UCLA already are applying SPASE to Voyager data within the planetary area. It is expected that it will be expanded further from Heliophysics into other associated disciplines.

Support for SPASE at NSSDC has included the generation of data collection descriptions by both staff and consultants using only a small amount of internal funding. The success of SPASE will depend greatly on having up-to-date information about the most popular data collections. Consequently, it is important to continue/increase funds for the SPASE descriptions of the highest priority data collections. The information contained in NSSDC's NIMS database of metadata is more comprehensive than the information represented in all of the SPASE data descriptions. New SPASE data collection descriptions can be generated from NIMS, but this requires frequent manual intervention to properly transfer/translate the information. The process can be made more automated if the database is enhanced. The enhancements would also benefit NSSDC, allowing the reverse process -- taking information from SPASE data collection descriptions and adding it to the NIMS database as well.

1.5.2 Conversion of High Priority Heliophysics Data During the Migration Process (C-3)

Many of the Space Physics data collections archived at NSSDC are written in the formats of now-outmoded computers and therefore are not easily useable by current researchers, who often have limited, if any, computer programmer support. Even if they have support, the programmer may be unfamiliar with the many problems encountered in converting from one machine representation to another or from 7- to 9-track tape. Also it is wasteful of limited research resources that each requester must do their own conversion. It is much more efficient for the archive to make the conversion once and archive, available to all, the resulting version, whether it is in IEEE binary format or ASCII character mode, or a generic format (e.g., CDF, FITS, HDF, etc.). NSSDC has provided many conversions of data collections to modern formats in the past upon request, but has not made a concerted effort to always offer this service as a possibility.

The Apollo 15 and 17 Heat Flow Experiment (HFE) data were converted for a requester who couldn't read and recover the data. Those readily-usable data collections have since been distributed to other requesters of the originals. In this case not only was it necessary to convert from an obsolete binary representation, it was also necessary to read the data in reverse order (because the provider had written it that way from his database), and recognize the records (all written together into a single file) that belonged to each of five different record types, and separate them into the five logical files, each with its own record type and record length. This represents one of the more difficult cases of converting an old data collection. An example from the other end of the spectrum is data collections that were written by an IBM 360 in fixed length records, either binary or Extended Binary Coded Decimal Interchange Code (EBCDIC) character mode, which can be read directly on NSSDC's DEC Alpha with an appropriate VMS compiler command to make that conversion. Many intermediate cases occur, for example, when a previous copy step onto higher-density media utilized an inappropriate copy mode from 7-track to 9-track tape and inserted fill bits. This fill operation then must be recognized and undone in the conversion process.

When an old data collection is restored, generating the output in plain American Standard Code for Information Interchange (ASCII) character mode ensures that it can be easily read by any user. However, considering that a 32-bit binary word can carry information that requires eight decimal figures to convey its original precision, a 4-byte binary floating-point number can expand (by a factor of 3.5) into 14 characters when rendered as ASCII carrying the original level of precision. The extra cost of storing in ASCII is readily justified by the ease of access for the whole science community. In practice the expansion factor frequently is not as large; because data sometimes carry a specific lesser degree of precision, or a science judgment is made that for a certain kind of measurement, a lesser degree of precision is sufficient.

NSSDC staff are presently involved in converting several problematic data collections into ASCII or Institute of Electrical and Electronics Engineers (IEEE) binary. The PDS Lunar Data Node is also currently restoring several old Apollo datasets, including one where microfilm number tables are digitized and (via OCR) made into a digital version of the data. This is a labor-intensive operation – to scan, OCR and verify the accuracy of the conversion. Consequently, such extensive efforts are expended only after judicious consideration and involvement of the community in the evaluation.

NSSDC held a recent data review (Sec 2.6.4.2) to categorize the 1,062 Solar and Space Physics data collections considered for conversion. The results, plus those from an earlier review, will be the basis used to assemble a representative set of 50 data collections of high priority that will be converted first. Our initial step will be the ASCII conversion in which the level of effort for conversion and the difficulty will be determined. From current experience it is expected that a half dozen of the fifty will take considerable resources (a man-month each), while the majority will take from 1-5 days effort each. During the initial period, community experts drawn from HDMC and SPDF will be assembled to assess whether a further conversion, for example to CDF, is in order for the prototype set of 50 data collections. This level of effort is within the in-guide budget constraints and staff. From NSSDC's current priority listings of the legacy tape data collections, it is desired to extend this conversion effort to all high level data (rankings of 4-6), so the project is multi-year and would convert the estimated 200 data collections.

1.5.3 Prototype Analog Archive Modernization: Review and Digitization (C-5)

NSSDC has begun to review the lunar portion of NSSDC's analog holdings by matrixing a GSFC lunar scientist, Paul Lowman, to help. The data include the Hasselblad photography from the Apollo missions, the Surveyor and Lunar Orbiter images, the collection of Soviet imagery, our microfilm lunar data collections, and, if time permits, our lunar map collection. The purpose of the review is to:

- prioritize images and data collections for digitization,
- review the metadata (database information and ancillary documentation) for completeness and usefulness,
- help prioritize microfilm data collections for scanning, restoration, and digitization,
- review and enhance our indexes of photography,
- assess our lunar map collection,
- aid us in determining the usefulness and completeness of our collection of Soviet analog data.

He has reviewed the Apollo 17 photography and has produced a progress report, describing and prioritizing the film magazines from the surface, orbital, and trans-Earth portions of the mission. At the end of this effort, Lowman will put together a report on all these aspects of our lunar analog collection and deliver it to NSSDC.

Currently NSSDC holds roughly 2000 reels of microfilm with planetary and lunar data. The vast majority of these are lunar-related. If the non-lunar data are excluded, as are the reels that contain photographic indexes and those that contain SIDE, LACE, seismic and other data for which digital versions already exist and have been restored or are being restored as part of an existing LASER program, only 120 reels of microfilm covering 40 data collections remain. As part of the NSSDC lunar analog review, Lowman will evaluate these data collections for scientific merit, determine if digital versions of the data already exist, and prioritize them for scanning and possible OCR for conversion to digital data in a prototype effort. Some of these data, if appropriate, could also be proposed for restoration through LASER or similar programs. Digitization of these data, strongly recommended by the NSSDC User Group in 2009, will ensure preservation of the data and also greatly enhance usability.

NSSDC holds a large number of microfiche and microfilm copies of Solar/Space Physics spacecraft data. Many of these are actually duplicated in the form of older digitized tapes. NSSDC has also started the review of these holdings to determine which are unique and if the analog versions should be digitized. An experienced member of the staff, Joe King, has taken on this task and will assign a priority (1-5, low to high) for digitization in the area of cosmic ray and magnetospheric physics. Since NSSDC now has the capability to OCR scanned images to provide tabular data, the uniqueness and usability of applying such a technique will also be ranked to the sample of data he is reviewing. Once the results of this effort are compiled in a written report, NSSDC management will decide, based on a sample of 50 data collections, whether to expand the effort. Any larger effort will be reviewed by community experts as the demand on resources and personnel may be significant.

One immediate finding has been that there are a large number (14,500 or 2/3 of microfilm holdings) of analog ionograms from the ISIS and Alouette spacecraft. NSSDC has subsequently consulted with Bob Benson, provider of the archived ionograms, to outline a process to select a set of ionograms based on time periods and satellite/station combinations for immediate testing. NSSDC has described the technique used in previous work to digitize lunar microfilm. In some cases the scanning of microfilm and microfiche can be outsourced to vendors using automated machinery. We are engaged in a small prototype effort to determine if this is feasible. Good quality examples as well as others with poor clarity and contrast were chosen and sent to John Bordynuik Inc. (JBI) to be scanned. The results will be used to determine how well bulk scanning will work for digitizing our microfilm and microfiche, and what quality of products can be reliably outsourced, which is more cost effective. The more difficult images will be handled internally, where we have more extensive control. So the current effort is considered In-Guideline but large scale digitization of heliophysics (or planetary) data is described in the narrative (3.2) as a Requested/Enhanced budget.

2. Technical

2.1. NSSDC Background and Relationship to the Heliophysics Data Environment

The National Space Science Data Center was established to provide the preservation and dissemination of scientific data from NASA missions. It now manages an archive of data from roughly 1500 investigations carried on roughly 550 spacecraft over the past 49 years. The digital component of the archive amounts to over 77 TB distributed across 4500+ distinct data collections. In recent years the rate of growth of the data volume was increasing – the archive grew by 3 TB in 2005; 4 TB in 2006; 13 TB in 2007, and 24TB in 2008. In 2009 and beyond, growth is expected to be about 20 TB/yr, due largely to planetary missions. There is also a legacy, non-digital, component of the archive. With regard to heliophysics for 2008 there was 1.2 TB of data, and through May of this year it is 0.6 TB of data accrued to the archive.

In a new approach to data dissemination in recent years, distributed, active archives were established in the Planetary, Astrophysics and Solar sub-disciplines. Those Active Archives (AAs) assumed 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 space physics. NSSDC continues as the permanent archive for all space science data, interacting with active archives per Memoranda of Understanding (MOUs) that clarify relationships and responsibilities. NSSDC also uniquely provides access to the taxpaying public for certain data and information of interest to them.

As a general policy NSSDC acquires data from active archives for long term preservation and provides it back to them if requested, but that is only the first of many archive services provided. NSSDC acquires data directly from projects and researchers when those data are not of interest to AAs, and makes such data available to researchers and to the general public. NSSDC similarly serves legacy data that pre-dated the formation of active archives. In accordance with particular MOUs, NSSDC handles data dissemination requests that are beyond the mission of particular active archives, such as responding to large data transfer requests or supporting access by the general public. In addition NSSDC 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 (WDC) for Satellite Information and is the only NASA part of the WDC system for the past 40+ years.

The NSSDC formerly included the SPDF as a component. NSSDC now serves as an “Archive for Archives” that provides backup and media-handling/refresh capabilities that the active archives do not have the resources to maintain. The NSSDC serves this function not only for Final Archives, but also for active missions such as RHESSI and Gravity Probe-B, at their request. NSSDC is the home of the SPASE group, and funding for core SPASE services – coordination; maintaining documents, schema and registries; and tool support -- is through NSSDC.

SPDF is the Heliophysics Final Archive for Space Physics data, which is served through CDAWeb and other Web portals and ftp sites. It also develops and maintains a variety of services, *e.g.*, SSCWeb, and actively supports the CDF format that is in very common use by missions, VxOs, and the CCMC. The SPDF and NSSDC are completing an MOU to coordinate the backup archiving of its space physics data.

SDAC plays a role similar to SPDF but for the solar component of Heliophysics data. It works with missions to serve large amounts of solar data (e.g., STEREO and SOHO), and will play an essential role in the serving of SDO data. SDAC is responsible for such essential tools as the SolarSoft analysis package. NSSDC provides backup as needed.

The Heliophysics Data and Model Consortium, serves the Heliophysics community by providing middleware services and model data; the HDMC is directly linked to the VxOs (including the Virtual Model Repository). The HDMC also works closely with the CCMC, exploiting tools and providing a coordinating function to the recently selected RAs.

2.2 Supporting Science Data

NSSDC archives data from heliophysics (space physics, solar physics), planetary, and astrophysics missions. Each discipline, and often each individual project, is different in terms of the archiving agreement (secondary archive, backup archive, etc.), method of data delivery (electronic, tape, hard drive), and data format (Flexible Image Transport System (FITS), Planetary Data System (PDS), Common Data Format (CDF), etc.). The holdings for each discipline include older legacy data and data still arriving from currently active or recently completed missions.

2.2.1 Heliophysics: Space Physics

In the heliophysics discipline, NSSDC's data providers are projects, project data centers, active archives,

Data Provider	Mission	Comments
Science Data Centers (SDC) and Active Archives (AA)	ACE*,FAST*, IMAGE*,SAMPEX*	Permanent archive. SAMPEX was first RA.
	TIMED*	Also in CDAWeb, 584 GB/year
	IBEX*	20 GB for nominal two-year mission. Also available in CDAWeb and via VHO and VEPO
	THEMIS*	220-290 GB/year
	TWINS*+	~1684 GB for 2 two-year mission
	AIM	1.5 TB total for mission
SPDF	CDAWeb data	Ingest into permanent archive
	Multiple spacecraft in FTP	Ingest into permanent archive
Older missions	Geotail, IMP8*, ISEE*, Polar, Ulysses*, Voyager*, Wind	Relatively small volumes
Legacy	145 missions	250 GB

Table 2.2.1. Summary of space physics missions and interactions. NSSDC acts as a stable archive to ensure the usability of heliophysics data by future researchers. MOUs with active archives and PDMPs with projects guarantee data are preserved which may also be available elsewhere, for a great many legacy missions NSSDC is the primary or sole data source. *PDMP or MOU exists. **Data held in SDC, RA, or AA until distant future, then will come to NSSDC for permanent archive. +Data to NSSDC for archive within about one month after reception.

and legacy projects and PIs. Past interactions with data providers were on a more informal, *ad hoc* basis, while the interactions with newer projects are governed by a Project Data Management Plan (PDMP) or equivalent. For example, help was provided to the Interstellar Boundary Explorer (IBEX) project in preparing their draft PDMP. A summary of missions and interactions for space physics is shown in Table 2.2.1.

NSSDC receives data as files transferred electronically, on media (currently CD, DVD, DLT, and disks or "data bricks"). The heliophysics curation scientist reviews the available information and creates (*cf* Table. 2.5.1) the records describing the data collection and additional necessary information. The curation scientist also reviews the documentation provided for the data collection, to ensure that it is sufficient for a researcher knowledgeable in the field, and may request additional information or clarification if needed.

2.2.2 Heliophysics: Solar Physics

In Solar Physics most data reside either in the individual mission science centers or in the Solar Data Analysis Center (SDAC). Recently several resident archive centers began to offer access to mission data and consider NSSDC as a deep archive or permanent repository. Past interactions with solar data providers were on a more informal, *ad hoc* basis, while interactions with newer projects are governed by a PDMP or equivalent. NSSDC can provide a long-term preservation function for solar mission data. A summary of missions and interactions for solar physics is shown in Table 2.2.2.

Data Provider	Mission	Comments
GSFC RHESSI Data Center	RHESSI*	Current mission; Level 0 permanent deep archive is provided through NSSDC
STEREO Science Center	STEREO*	No data at NSSDC yet.
SDAC	Yohkoh	Data permanently archived at NSSDC. Data and software available from SDAC.
	Hinode	Current JAXA mission
	SOHO EIT	Current mission, archival plan shared with NSSDC
SDAC and/or Stanford	SDO	Future mission, large data volume
Stanford Solar Center	SOHO MDI	Current mission
TRACE Data Center	TRACE	Current mission, soon to begin RA
Legacy/Inactive	OSO 7 and others	Permanent archive. Data restoration at NSSDC.

Table 2.2.2. Summary of solar missions and interactions. Critical long-term preservation of data and easy access to the data by researchers and the general public are two of NSSDCs primary functions. Solar mission data are held at NSSDC as a backup to SDAC and Archives. It also acts as a primary archive for some solar mission data collections.

Here is a summary of the solar missions and their data:

- SDO will be launched in a few months. At 15 PB per year, the huge volume of data will be handled by the Joint Science Operations Center at Stanford.
- STEREO has a PDMP signed in 2002; it needs updating to reflect modifications in mission archival plans. Data are available through SDAC and directly through the STEREO Science Center.
- Hinode is primarily a JAXA mission, and therefore, does not have a PDMP. SOT and XRT data are available from the LMSAL Hinode Data Center and through SDAC.
- RHESSI has a PDMP which hasn't been updated since before launch, however, there is a Mission Archive Plan attached to the 2008 Senior Review. Since 2003 about 600 GB/year of RHESSI data have been ingested to the NSSDC near-line permanent archive. This copy is considered a "deep archive" copy, which stays within four months of the current date. When RHESSI is no longer an active mission, the Solar Data Analysis Center will be its resident archive, and then NSSDC as a

permanent archive. NSSDC will work with the RHESSI PI on an MOU for a clear transfer plan.

- TRACE has generated 150 GB/year of data for more than 10 years. There was no PDMP. Once the mission ends, the PI plans to embed the data into the SDO AIA archive at the Joint Science Operations Center at Stanford. NSSDC will work with the PI on an MOU.
- SOHO, a joint mission between NASA and ESA, has no requirements for a PDMP. However, the mission made available an archival plan that was included in the 2008 Senior Review proposal. The bulk of MDI (and therefore SOHO) data resides at the Stanford MDI data center. Before they are migrated into the SDO Joint Science Operations Center, NSSDC will work with the Stanford Solar Center on an MOU regarding our roles as a permanent archive for all SOHO data.
- Yohkoh data from 1991-1997 have been archived at NSSDC. The entire data set and software are available from SDAC, and mostly recently, from Montana State University, which is funded as a Resident Archive Center. NSSDC will work with the YLA PI at MSU on a clear transfer plan.

2.2.3 Planetary Science

Planetary science data comes to NSSDC primarily through the Planetary Data System (PDS). NSSDC acts as the second archive for the Planetary Data System, serving as a repository for PDS data sets on hard media such as CDs, DVDs, and tapes. Part of the holdings are data from older missions, such as Viking, Voyager, and Pioneer Venus. These collections are largely complete at PDS and NSSDC, but in most cases are being redelivered electronically from PDS to ensure the archive is complete and up-to-date, and, since future data deliveries will be made electronically, to have an established baseline for all data deliveries. Other more recent spacecraft, such as Galileo and Mars Global Surveyor, have completed their missions, but not all data have been archived. Many continuing missions, such as Cassini, have a

Data Provider	Mission	Comments
Planetary Data System (PDS)	Permanent archive: 17 missions – Many data collections	For example: NEAR, Magellan, Viking, Voyager. Many currently being redelivered electronically
	Cassini-Huygens, MER, New Horizons	Continuing mission, data flow to NSSDC established
	Stardust, Deep Impact, Dawn, Phoenix, NEXT, EPOXI	Small data volume expected, to NSSDC via PDS
	Chandrayaan 1	Current ISRO mission with limited PDS role
	Hayabusa, Kaguya	Current and recent JAXA missions with limited PDS role
	SMART-1, Mars Express, Rosetta, Venus Express	Current and recent ESA missions with limited PDS role
	Chang'e 1	Current Chinese mission, no present NSSDC role
	Mars Reconnaissance Orbiter. Mars 2001 Odyssey	Current missions, large data volume (70 TB) to NSSDC via PDS
	Lunar Reconnaissance Orbiter	Future mission, large data volume (200 TB) to NSSDC via PDS
	GRAIL, LADEE, Mars Science Laboratory	Future missions, discussions about data volume with PDS, NSSDC
Legacy	>95 missions	For example, Apollo ALSEP and orbital data being restored as part of the PDS Lunar Data Node at NSSDC

Table 2.2.3. Summary of planetary missions and interactions. NSSDC fulfills an important role as the primary distributer of legacy planetary data to researchers and the general public. Through a unique MOU with PDS, NSSDC also provides long-term preservation of data from recent planetary missions and from the restoration of older data.

well-established, scheduled data flow. A summary of missions and interactions for planetary science is shown in Table 2.2.3. Plans are in place or being readied for future missions with both large and small expected data flows. Agreements with non-NASA missions for archiving data are not systematic and are largely under negotiation.

2.2.4 Astrophysics

In astrophysics, NSSDC receives and ingests 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. Table 2.2.4 summarizes the various data provider categories, including the SARCs: high-energy astrophysics (HEASARC), which now includes the microwave and submillimeter data previously associated with LAMBDA; UV-Optical (MAST); and infrared data (IRSA). NSSDC received approximately 22.4TB of new astrophysics data in 2008, of which 15.5TB came from HEASARC. Most data products are delivered and ingested in astronomical standard FITS format. In some cases, such as Gravity Probe B (GPB), the project team delivered the mission document set along with the observation attribute database exported into Common Data Format (CDF).

Data Provider	Mission	Comments
Science Archive Research Centers (SARCS)	9 active missions across HEASARC*, MAST*, IRSA+, MSC	Permanent archive for 20 completed mission. Backup archive for SARCs per MOU. Data ingest ongoing, *e.g.*, GALEX, WISE
Fermi Science Support Center	Fermi+	LOU signed, data arriving at NSSDC
Other	Genesis	MOU in development
	Gravity Probe-B*	Permanent archive per MOU, data delivered
	~10 missions	Future archive for JPL Navigator
Inactive	70 missions	Permanent archive, *e.g.*, Uhuru

Table 2.2.4. Summary of astrophysics missions and interactions. NSSDC performs a crucial backup role for astrophysics data through MOUs with the SARCs as well as acting as the primary archive for a number of new and legacy missions (+MOU exists. *PDMP exists. #LOU exists.)

2.3 Functional Organization

NSSDC's current and proposed future activities will be described within the framework of the functional model shown in Figure 2.3.1. This conceptual model is adapted from the ISO standard "Reference Model for an Open Archival Information System (OAIS)" that is now used worldwide as a starting framework in which to address digital preservation of archives/repositories. This model, whose standardization was led by NSSDC staff, has facilitated all types of archives and/or repositories in describing their activities and data using its concepts. The major OAIS functions shown and described below, are: Ingest, Archival Storage, Data Management, Administration, Access, and Preservation Planning.

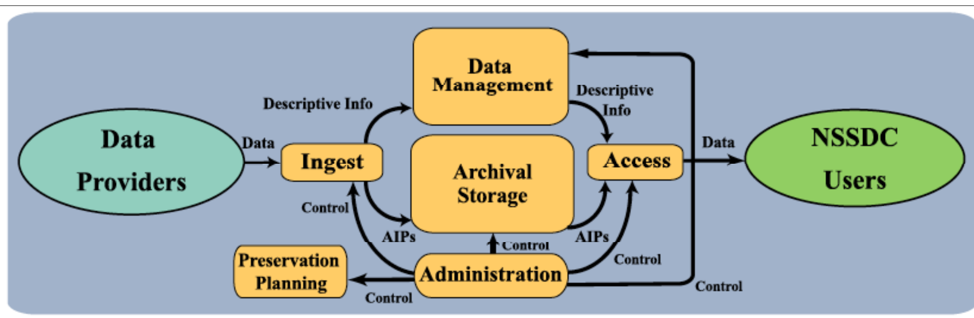

Figure 2.3.1. The internationally standardized OAIS model for data centers has been followed in the evolution of the NSSDC's architecture to ensure the most effective use of data and resources.

Three NSSDC Activities – Ingest, Data Management, and Archival Storage – map directly to the OAIS functions. For ease of contracting, NSSDC has combined the other three OAIS functions: Administration, Access, and Preservation Planning - into a single NSSDC activity called Interactions.

Data Providers (AAs, RAs, missions, scientists) provide data and supporting information to the OAIS Ingest function of the archive. They arrive via Submission Information Packages (SIPs), which for NSSDC may be any set of data and documentation files. The OAIS Ingest function turns these into one or more Archival Information Packages (AIPs), using either the legacy Multi-file Packager and Analyzer (MPGA) capabilities or the new XML-based Xware capabilities, for preservation within the OAIS Archival Storage function. To facilitate low-cost management, NSSDC's formalized AIP is a single file container that holds one or many science data files, a number of attributes about each file that help NSSDC manage its AIPs, and pointers to all of the supporting documentation, including calibration and format 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 NSSDC defined attributes.

The OAIS Archival Storage function is responsible for managing and preserving the AIPs. It also holds and manages any data that is not in an AIP form, i.e. much of the data. As a result of NSSDC's long history it has received and used for storage many types of media. The OAIS Data Management function receives attributes from the OAIS Ingest function and stores these in an NSSDC database known as the NSSDC Information Management System (NIMS). These attributes are those typically used to aid in locating data of interest to NSSDC's users. Most of these attributes have been entered manually over the years.

The OAIS Access function provides services and functions to support users in obtaining data or other information from NSSDC. The interface services are supported by the information stored in the OAIS Data Management function. For example, a user may use a web form to access the NASA Master Directory, which in turns queries NIMS under the OAIS Data Management function. For the delivery of data, the OAIS Access function is used to query the OAIS Archival Storage function and to perform any processing on the AIPs (or non-AIPs) needed to meet the request. Most data deliveries are of the basic science files and documentation files, whether from an AIP or not.

The OAIS Administration function has many tasks, but basically oversees the day-to-day operation of the archive and manages the migration of data from one form or media to another. The OAIS Preservation Planning function tracks changes in technology, standards, and the communities NSSDC serves in order to recommend appropriate preservation strategies and standards to the OAIS Administration function. For example, the current AIP implementation was developed by this function.

2.4 Ingest

Ingest provides the services and functions to accept data and documentation from data providers and to prepare the contents for storage and management within the archive. Ingest functions include receiving data submissions to a staging area, performing quality assurance, generating descriptive information for inclusion in the NSSDC's Data Management database, and coordinating inputs to Archival Storage and Data Management.

2.4.1 Archive Information Packages

NSSDC has adopted the Archival Information Package (AIP) as a key component of data preservation. The AIP is an information object composed of the data content to be preserved along with the reference, context, provenance, and fixity information required for long-term preservation and understanding of the data. The NSSDC implementation of the AIP is a single file, independent of the underlying media type, which can hold single data files, multiple files, or directory hierarchies. Since AIP generation began in 2000, NSSDC has ingested 2.4 million AIPs containing 8.1TB of data.

2.4.2 Xware and Ingest Process (C-2)

NSSDC's original suite of AIP generator and AIP analyzer software has been largely supplanted by a simpler set of 'Xware' tools that use XML for the collection and transmission of metadata during the ingest process and preservation of Preservation Description Information (PDI) within the AIP. Xware for generation of data submission manifests is supplied to data providers. Other Xware tools are used by NSSDC to verify integrity of data deliveries and to package data into AIPs.

The Planetary Data System is currently using Xware for data submissions to NSSDC. SIP manifests are created with Xman software. Small data submissions may be delivered to NSSDC electronically. Large submissions are written to USB disks up to 2TB in size and are then validated using Xcheck before being shipped to NSSDC. Data submissions are validated at NSSDC with Xcheck before being packaged into AIPs with the Xchunk software. The overall process is shown in Figure 2.4.1.

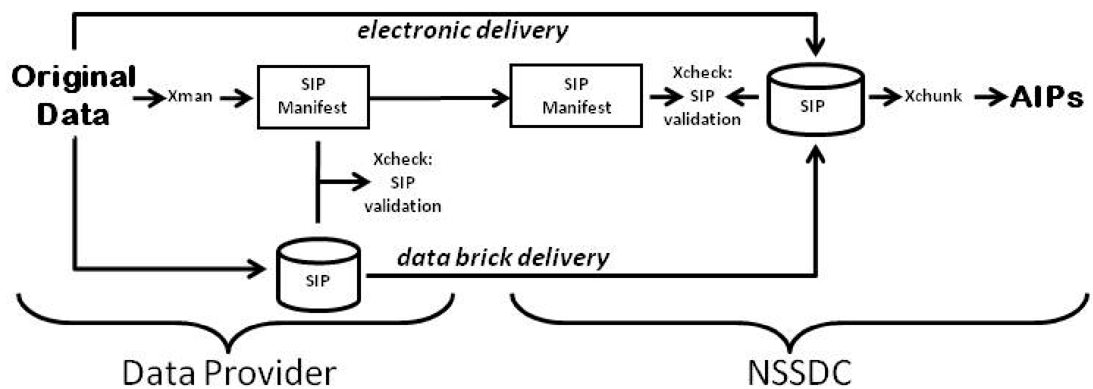

Figure 2.4.1 NSSDC continues to automate and enhance the efficiency of its submission process using Xware tools. This overview shows the flow of PDS submissions to NSSDC at this time.

NSSDC can ingest and archive digital files and directory structures. NSSDC does not ingest non-exported databases or data that can only be realized using proprietary or specialized products. There is an increased use of databases to represent data products rather than using traditional data files. NSSDC will work with providers to find ways to save the content of databases in preservable formats. This was done with the Gravity Probe B mission who stored their data products in a Sybase relational database. Ultimately, Gravity Probe B delivered their L0, L1, and L2 data to NSSDC in CDF files.

NSSDC still accepts data submissions on media, currently including CDs, DVDs, several varieties of DLT and SuperDLT magnetic tape, and transportable magnetic disks (data bricks). Network deliveries are supported as well, although the existing 10Mb network restricts the practical size of network deliveries.

2.5 Information System Management

The ability to search for NSSDC's data holdings based on spacecraft name and discipline, has been in

Partition	External Interface	Internal Interface	Number of Records
Spacecraft	NMC, SPASE registry	JEDS	6,500
Experiments	NMC, SPASE registry	JEDS	5,400
Data Collections	NMC	JEDS	5,400
Personnel	NMC, SPASE registry	JRAND	61,000
Publications	NMC	JEDS	48,000
Media (digital, off-line)	---	JIN	75,000*

Table 2.5.1. NSSDC uses its NIMS database to track information. External interfaces are those which are available to the general public. Internal interfaces are those used by staff via named user accounts. The number of records is approximate. * indicates the number includes only those records for which NSSDC has current media.

place since the mid-1980s and web-accessible since the mid-1990s. This is supported by the metadata maintained in the NSSDC Information Management System (NIMS) database. It contains information primarily for describing and searching for data holdings. In addition, it can also handle an administrative capability to track the archiving status of various flight projects. Table 2.5.1 shows some of the information that is tracked in NIMS.

The database and the interfaces described herein support both the Data Management and Data Access portions of the OAIS model. The external interfaces (NSSDC Master Catalog and the SPASE registry) support data access and data management. The associated internal software is referred to as Jware, derived from its latest version in Java.

2.5.1 Master Catalog

External access to the information about NSSDC's data holdings is primarily via the NSSDC Master Catalog (NMC). It queries NIMS to produce dynamically generated web pages. In 2007 this interface was completely re-engineered. In addition to a web form which supports queries by spacecraft name, there are now forms to query for experiments and data collections directly, eliminating the need to drill down to data information via the spacecraft.

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 data collections, but the number of collections that are accessible in this manner has continued to grow at a steady pace. The main portal to the NMC is via http://nssdc.gsfc.nasa.gov/nmc/.

2.5.2 SPASE Registry

Another means of external access to NSSDC information is the SPASE Registry. This registry currently contains access to information about the spacecraft (observatories), experiments (instruments), and people (PIs, team members) tracked in NIMS in a SPASE-compliant (XML) format. Version 2.0.0 of this interface was released shortly following the release of the V2.0.0 SPASE Data Model in April 2009. The portal to the NSSDC SPASE registry is via http://nssdc.gsfc.nasa.gov/spase/.

2.5.3 Jware (JEDS, JRAND, JIN)

Jware are Java-based software that are the main access interfaces for manipulation of the information about NSSDC data. They provide capability to insert, find, and modify the underlying metadata stored in the NIMS database tables. The existing components of Jware include:

- JEDS for spacecraft, experiments, data collections, and publications;
- JRAND for people, both requestors and others involved with projects and/or data collections,; and

- JIN for off-line media, tape and disk holdings.

Access to these interfaces is on a named-user basis and is available only within the GSFC computing environment.

2.5.4 Streamlining Metadata Flow (C-4)

NSSDC has begun to streamline its process for ingesting data and significant progress has been made. A big part of the change is the adoption of a single user interface for the collection of metadata required to ingest, archive, preserve, and access data. The initial version of this new mechanism was implemented using a spreadsheet-based form, which has allowed proof-of-concept and immediate adoption of the new process. Collected metadata is subsequently manually transferred (via copy-paste) to JEDS and JRAND. (See previous section.)

NSSDC plans to implement a web interface to replace the spreadsheet-based form. Backing up the web interface is an eXtensible Markup Language (XML) schema, already created, to encapsulate the information into an XML file. Collected metadata will be "shredded" (i.e., mapped) into appropriate databases, eliminating the current manual operation. Data collection descriptive and search information will be shredded into NIMS. Information supporting the data ingest and data storage processes will be shredded into a new ingest database.

The new ingest database will manage the information needed by data providers to create XML (SIP) manifests and the information required by NSSDC to create Archive Information Packages (AIPs). A web interface will supply a view of the ingest process to data providers, from the identification of a new NSSDC collection through the storage of AIPs on archive media.

2.6. NSSDC Interactions Activity

2.6.1 Internal Projects

Figure 2.6.1 shows data migration for earlier, legacy data products (including digitized materials) into media-independent forms for archival storage. NSSDC extracts files and related attributes for legacy

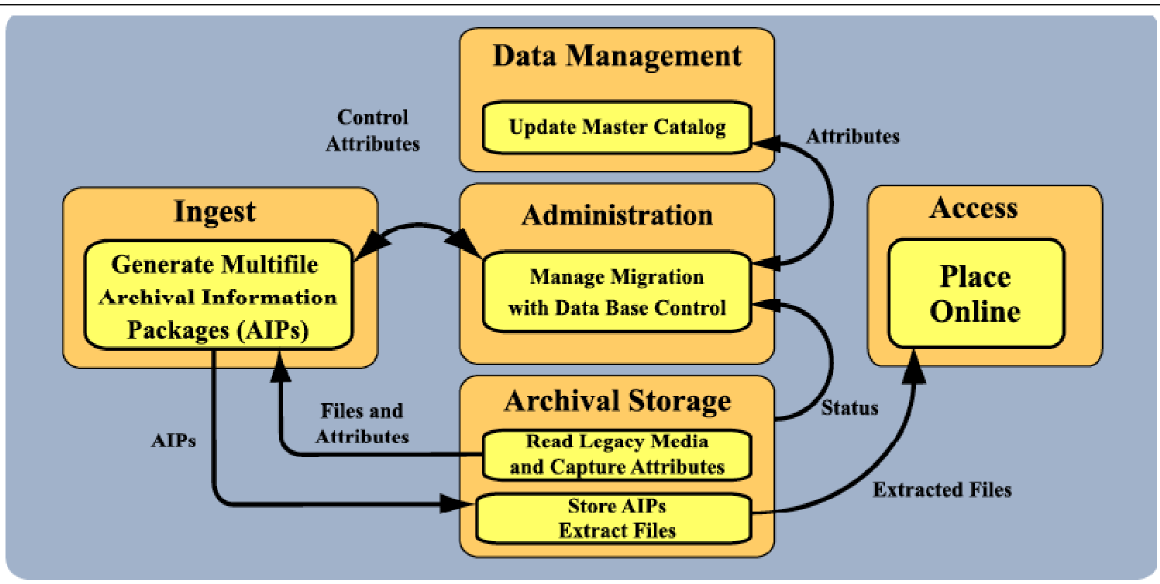

Figure 2.6.1 Data migration, controlled by administration, ensures that data remain available into the future.

magnetic tapes and these are provided to the OAIS Ingest function staging area. Then 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 OAIS Access function.

Processing legacy data includes capturing provenance information and checking documentation to make the data more reliable and independently useable. Where practical, data will be updated to a more accessible format. NSSDC has validated the process on the first selection of 200 tapes (both 9 track and 3480 cartridges). This "legacy tape migration" is continuing as a background task that will take several years to complete. By then it will be necessary to start the media refreshment of the current SDLTs which have a planned eight-year life cycle.

2.6.2 External Projects

External administration includes developing and evolving the current curation policy for NSSDC as a facilitator for data ingest from discipline-specific active archives, missions, and PIs. NSSDC assesses the needs of these groups to define the services that NSSDC can provide as the permanent archive. In addition to existing MOUs, NSSDC has recently concluded an MOU with the Infrared Science Archive (IRSA) and separate Letters of Understanding (LOU) with the Fermi and Wise missions. When a mission does not work directly through an active archive (e.g., Gravity Probe-B), a separate agreement may be sought to define roles and responsibilities.

The PDS MOU has gone through an extensive update and now covers large (2 TB) “data brick” submissions. The delivery process of data from PDS to the NSSDC addresses the need to transfer the contents of PDS volume content that exceeds the capacity of the archive media currently in use at NSSDC (300 GB) but can be applied to the transfer of data from smaller volumes as well. For data retrieval the providing PDS node contacts the NSSDC curation scientist or NSSDC Coordinated Request and User Support Office (CRUSO). PDS volumes to be retrieved are identified by their transfer object IDs. For PDS, where each SIP contains only one transfer object, supplying the SIP ID will provide sufficient information for NSSDC to retrieve the data.

In the heliophysics discipline, NSSDC is usually identified in each mission’s Project Data Management Plan as the final archive. 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). NSSDC also ingests a limited amount of data from recent missions, for example, data from TWINS.

To support the submissions of, and access to, data, NSSDC maintains web pages that locate data, provide fact sheets, identify white papers (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, NSSDC users include active archives, researchers, educators, and the public. As the permanent archive for NASA, NSSDC also acts 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, science data are also provided to international 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 data collections archived at NSSDC (and for a limited number of data collections available elsewhere that are of interest to space scientists).

NSSDC generates a monthly summary and more extensive bi-monthly progress report for NASA HQ. In addition, interactions with NASA HQ include periodic discussions on the budget and presentations to the NSSDC sponsor. NSSDC monitors customer satisfaction through review of user comments and direct interactions with active archives. The NSSDC Annual Report, posted on-line in March of the following year, summarizes the NSSDC status and statistics. Similar to recent annual reports, Table 2.6.2 totals digital data arriving at NSSDC from January 2008 through May 2009.

Astrophysics	**Data volume (GB)**
FUSE	803.08
GALEX	5,090.84
HEASARC	15,498.42
Fermi	4,800.02
Planetary	
PDS	289.73
Heliophysics	
ACE	27.95
Alouette	9.42
FAST	688.07
Geotail	3.77
ISIS	49.85
Polar	9.28
RHESSI	817.52
Twins	151.75
Ulysses	13.45
Wind	37.53
Total	**28,290.68**

Table 2.6.2. The unprecedented quantity of data arriving at NSSDC January 2008 through May 2009, shown here by mission/project, will be dwarfed by future data deliveries.

2.6.3 Preservation Planning: SCaN/CCSDS, TRAC, NARA

The responsibility of NSSDC is to preserve data, thus it needs to be aware of and apply new, more effective methods of data archiving and to participate in the development of these methods to assure their utility for NASA needs. NSSDC is also involved in researching archiving and standards not only for NASA, but for international space agencies and for all institutions in general that need archiving. NSSDC-associated personnel work with the Consultative Committee of Space Data Systems (CCSDS) to develop standards for effective archiving that can be applied throughout the world's space agencies. CCSDS has also received funds from the U.S. National Archives and Records Administration (NARA) to investigate archiving approaches that would be of benefit to both NASA and NARA.

The SPASE activity mentioned previously also plays a major role in preservation planning for the future. NSSDC has long maintained databases of metadata information about NASA and other space agency satellites and spacecraft, important information sources for searching and inventorying worldwide data holdings. NSSDC is upgrading access to the metadata information so that users in the SPASE community and throughout the world can easily access NSSDC metadata, update it, and use it to create comprehensive registries of spacecraft, instruments, and the personnel associated with them.

2.6.3.1 CCSDS/SCaN Standards Development

The GSFC Space Communications and Networks (SCaN) group, under the auspices of the CCSDS, funds some of the NSSDC standards development. Due to the long history of digital archiving at NSSDC, the space and archiving communities have benefited from NSSDC participation in CCSDS. NSSDC has also benefited in this relationship from the unique experiences and implementation efforts of the other space agencies. In recent years CCSDS has attracted participation and contributions from many affiliated organizations including national libraries, archives, museums, and library associations.

NSSDC participation in CCSDS has advanced the state of the art in digital archiving and information packaging through development and maintenance of standards. Currently standards development work of interest to NSSDC is being done through several CCSDS Working Groups: Data Archive Ingest (DAI), Repository Audit and Certification (RAC), Information Packaging and Registries (IPR), and XML Standards and Guidelines (XSG). For many years SCaN has supplemented or funded initial NSSDC participation in these groups.

The most successful CCSDS standard ever developed is the Reference Model for Open Archival Information Systems (OAIS). Following its publication and almost universal acclaim the next step was to have independent criteria to objectively judge whether or not any of these archives have been adequately protecting their digital holdings and, in particular, whether the information in those holdings will remain understandable and usable by their Designated Communities in the future. Due to the wide acceptance of the OAIS Reference Model, it has been possible to leverage the resources of NSSDC to obtain support from other space agencies as well as the wider archiving community. For example, a project to develop an initial draft of these criteria was delegated to the joint leadership of NARA and the Research Library Group (RLG) with participation by NSSDC and other space agency representatives.

Trusted Repositories Audit and Certification (TRAC) provides tools for the audit, assessment, and potential certification of digital repositories, establishes the documentation requirements needed for audit, delineates a process for certification, and establishes appropriate methodologies for determining the soundness and sustainability of digital repositories. A Trustworthy (Digital) Repository is a repository capable of reliably storing, migrating, and providing access to digital collections. The repository's mission includes providing reliable, long-term access to managed resources to its designated community, now and into the future.

Several publications are available to provide further detailed criteria for Trusted Repositories. In 2007, NARA, the Center for Research Libraries (CRL), and the Online Computer Library Center (OCLC) jointly published "Trusted Repositories Audit and Certification: Checklist and Criteria" (TRAC) which articulated a framework of attributes and responsibilities for trusted, reliable, sustainable digital repositories capable of handling the range of materials held by large and small cultural heritage and research institutions. The Consultative Committee for Space Data Systems (CCSDS) Repository Audit and Certification (RAC) Working Group (WG) is using the TRAC document as a basis for creating a CCSDS and ISO Standard. Internal drafts of the proposed standards are available and a public review is expected this year. NSSDC has been viewed by the community as a leader in archiving of digital data and NSSDC personnel have been vital participants in both of these efforts. The TRAC document has been submitted to the Consultative Committee for Space Data Systems (CCSDS) for conversion into CCSDS and ISO Standards by which any science repository or archive could be formally, independently and objectively judged.

NSSDC has used the TRAC document to evaluate its internal structure and practices. NSSDC has also solicited input from the community through the NASA Astrophysics Data Centers Executive Council, Heliophysics Data and Model Consortium (HDMC), and Planetary Data System. Subsequently, NSSDC has modified that TRAC document for use with the Heliophysics Resident Archives. NSSDC continues to participate in the conversion of the TRAC document into an international standard. Once the draft of the international standard is available, NSSDC will cooperate with the standards developing group to conduct a test audit and then perhaps an actual certification audit of NSSDC. NSSDC would then share the experience gained during the audit process to aid RAs in their own self-certification or external certification audits. These certifications will do much to persuade both funder and customers that data being held in NASA archives are being properly curated.

NSSDC proposes sponsoring a workshop on the application of the TRAC document to science archives which would emphasize two aspects: bettering communication and sharing of common methods. The goals of the workshop are shown in Table 2.6.3.

Communication

- to establish the level of commonality of problems and best practices, as seen by the archives, and their interest in continuing to determine what is a Trustworthy Archive;
- to identify broadly based techniques and best practices that address how to perform the audit; and
- to determine who will perform the audit and with what results.

Common Methods

- help archivists locate Trustworthy Repositories required for a given study;
- provide archivists with access to those data;
- assure that those data are useable; and
- preserve the data forever.

Table 2.6.3. Goals for NSSDC-sponsored workshop on Trusted Repositories.

The proposed outline for NSSDC 2009 Scientific Archive Workshop follows that implemented for the first Scientific Archives Workshop (in the 21st Century)" but would have a specific focus and smaller attendance. We anticipate presenting the result of the workshop at a session associated with MSST-IEEE in 2010.

2.6.3.2 NASA/NARA Interagency Agreement

The Interagency Agreement (IA) establishes a working relationship between NASA/GSFC Sciences and Exploration Directorate and the National Archives and Records Administration (NARA). The purpose of this IA is to allow NASA to conduct research that will specifically leverage and apply NSSDC's expertise and experience. The research will address NARA's critical requirements for preserving and providing the American public with access to the exponentially growing, extremely diverse and arbitrarily complex collections of historically valuable federal, presidential and congressional records.

Since NSSDC is embracing the use of XML for a variety of internal uses and in its interactions with the external community, it will become necessary to track instances of the XML Schemas that govern those uses. The XML Schemas will necessarily evolve and the various versions of the same XML Schema will need to be stored. XML Objects associated with XML Schemas may also need to be stored. The prototype XML Schema registry developed in conjunction with NARA and CCSDS research can be adapted for NSSDC use. Underlying the prototype registry is a freely available ebXML registry with a façade that has been developed to ease interactions with the underlying repository.

2.6.3.3 SPASE and Registries

The SPASE project is intended to facilitate the search and access to data throughout the heliophysics community by development of a data model for describing heliophysics data. The data model is implemented through XML schema. The data model and its use have been developed by the SPASE Working Group comprised of representatives of the international heliophysics community, especially the NASA-funded Virtual Observatories. The SPASE Data Model currently in use is Version 2.0. The document describing this model may be found at http://spase.gsfc.nasa.gov/ from the links in the area called "Data Model Document".

NSSDC personnel have led the SPASE effort through biweekly teleconferences and semiannual face-to-face meetings of the SPASE Working Group. Since the 2006 Senior Review NSSDC has supported SPASE from the regular NSSDC budget. The main SPASE website has been moved from the original UCLA location to NSSDC. Todd King of UCLA receives NSSDC funds to continue as a remote webmaster for this site and as the editor of the SPASE Data Model. NSSDC also supports the creation of data descriptions of key data holdings according to the Data Model.

The NIMS database information has been made available for the creation of SPASE registries. The registries provide basic information about (in SPASE terms) observatories, instruments, and the people, respectively, associated with these missions. Those making data descriptions can use this information for their descriptions.

At this time, several of the VxOs exchange the registry information between themselves by utilizing "git" repositories. Forthcoming work at NSSDC will involve implementing a local "git" repository which will enable updates by the VxOs to be incorporated to the NSSDC database and, similarly, enable updates to the registry information held locally at a VxO to be made as a consequence of updates to the NSSDC database.

2.6.4 Community Participation

NSSDC exchanges knowledge with the full spectrum of external entities, but in this section we will discuss interactions with only a few – specifically, other NASA entities and outside advisory groups, particularly the NSSDC Users Groups, its recommended specialty advisors and the standards community.

2.6.4.1 The NSSDC Users Group

The NSSDC Users Group (NUG) is a committee of experts in data archiving and related fields who meet bi-annually to review the NSSDC progress and plans, who suggest "course corrections" for NSSDC policies, procedures and plans, based on their perspective as outsiders and interested users.

The NUG met on 16 March 2009 at NASA/GSFC. The group members in attendance were Drs. Fred Bruhweiler (CUA), Ian Richardson (NASA/GSFC/UMD/CRESST), George Rossano (Aerospace Corp.), Robert Benson (NASA/GSFC), and Paul Lowman (NASA/GSFC). Presentations by members of the NSSDC staff covered a range of topics from the SPASE effort to preservation of the non-digital data in the possession of the NSSDC. Their recommendations are summarized as follows:

- safeguard and modernize NSSDC's Analog Archive;
- continue migration of legacy tape data, but provide a more useful format of copied data,
- develop a Trustworthy Archive document (TRAC)
- plan for a new science archive workshop (SAW II) centered around the TRAC document, and
- expand SPASE activities to include conversion of high priority NSSDC products to SPASE descriptions.

2.6.4.2 2008 NSSDC Data Review

An important milestone for NSSDC was the NSSDC Data Review Groups, which were formed in December 2007. Their mandate was to review and prioritize NSSDC data collections stored on older media (7-track, 9-track, 4mm, 8mm, cartridge) for migration to into the formal archive system in which AIPs are written. They did not re-evaluate the 495 data collection already chosen by an internal review in 2000-2002 and whose migration was already underway, but the remaining 1,062.

Ten members of the heliophysics community reviewed and prioritized 1,062 data collections for migration. Between December 2007 and August 2008 Bob Benson, Dieter Bilitza, Mike Collier, Shing Fung, H. Kent Hills, Joe King, Terry Kucera, Carolyn Ng, Ian Richardson, and Adolfo Vinas met five times as a panel and worked individually to determine data uniqueness and usability. Each member was assigned 50-100 data collections as the primary reviewer and about the same number as the secondary reviewer. A numerical score was recorded in a spreadsheet with 5 as the highest priority and 1 the lowest.

Most work was completed by email and in small groups. Of the 1,062 data collections, 24% were rated priority 4 and 5. Based on the community scientists' input, NSSDC folded the results into the migration list of data holdings.

The results of this data review are as follows:

- Heliophysics: 1,062 data collections reviewed with 24% as high priority.
- Planetary Science: 422 data collections reviewed with roughly 50% as high priority.
- Astrophysics: approximately 100 data collections reviewed, 30% priority 4-5
- Earth Science: approximately 100 data collections being reviewed by ES-DIS people
- Other: others such as ephemeris reviewed, approximately 50% high priority

2.7 Archival Storage

The archival storage function is responsible for ensuring that data entrusted to the NSSDC resides on appropriate archive media and that data content is preserved in usable form as long as they have appreciable science value.

2.7.1 NSSDC Archive Storage Services

NSSDC provides four levels of archival storage services as shown in Table 2.7.1.

Permanent Archive	Long-term curation of digital data. Data may be repackaged and/or transformed to maintain accessibility and usability. Content information of SIPs is preserved but not necessarily in the original format. AIPs are re-written to new media within eight years. Data that are not currently preserved in AIPs will eventually be migrated from their legacy media into AIPs. The MOU may specify that data access be restricted, e.g. proprietary data or data available through an active archive.
Second Archive	Storage of digital data that is also held by another archive. Data may be repackaged (e.g., put into AIPs) and/or reversibly transformed to maintain accessibility. Content information of SIPs is preserved and can be retrieved in the original format. AIPs are re-written to new media within eight years. Other media refreshment is performed per MOU with the primary archive. The MOU will specify whether NSSDC will restrict access to the data.
Backup	Storage of digital data to support another archive's contingency plan per MOU. Cognizance of data content and format is not required of the NSSDC. Data may be repackaged and/or reversibly transformed to maintain accessibility. Content information of SIPs is preserved and can be retrieved in the original format. Data access is restricted to the provider or other entities per the MOU.
Analog Archive	Preservation of analog data on a variety of media with external review to establish priorities for digitization. Selected retention of original analog data after digitization. Data are copied and disseminated by NSSDC.

Table 2.7.1. **NSSDC levels of archive storage services.** The level of archive service and the specific details of what is provided by NSSDC are determined by MOU on a case by case basis.

In the three digital archive storage levels multiple storage methods are used: some of the data are preserved in AIPs and some of the data reside on legacy media. It is NSSDC's intent to migrate all permanently archived data on legacy tape media into AIPs.

2.7.2 Media and Analog Storage

NSSDC uses several media types to support archival storage. Low-cost, high-capacity magnetic tape (SDLT, LTO) is used for long-term storage of AIPs. Data staging and short-term AIP storage is done on magnetic disk. Transportable magnetic disks or "data bricks" are used to deliver large data submissions to NSSDC and for short-term storage of data prior to packaging into AIPs. In addition NSSDC holds a large number of legacy magnetic tape volumes (including nine-track and 7-track) whose data content are being migrated to AIPs stored on modern magnetic tape. To handle cases where data recovery from old media is problematic, NSSDC has a contract with JBI, a leading commercial data recovery company. JBI uses sensitive state-of-the-art equipment that reads magnetic tapes with no physical contact between the equipment and the magnetic recording surface.

Historically, the media used to transfer digital data to the archive usually served as the storage media with a copy of each media volume being stored in an off-site location. An application called JIN (see section 2.5.3) is dedicated to the task of managing these digital media volumes. Analog media, when fully cataloged, will be integrated into the next JIN release.

Over the past 18 months the NSSDC has been working with the Earth Observing System (EOS) to transfer ownership of earth science data collections and is currently in the process of transferring twelve collections containing over 19,000 tapes. Ownership of tapes stored in Federal Records Center has been transferred to EOS. Tapes located at GSFC are being moved as EOS resources allow. In some cases the scanning of microfilm and microfiche can be outsourced to vendors using automated machinery. A small prototype effort is in progress to determine if this is feasible. Examples have been chosen with both good and poor clarity and contrast, and sent to the highly-regarded firm JBI to be scanned. Reels of microfilm have been previously scanned in-house (*e.g.*, the Apollo indexes for ASU and the Cold Cathode Ion Gage plots) and an expertise developed in evaluating a reel in terms of the quality of a digital scan depending on its contrast, sharpness, and blemishes in the film. This expertise will be used to determine how well bulk scanning will work for digitizing remaining microfilm and microfiche, and for which quality of products outsourcing is practical.

2.7.3 AIP Storage

AIPs are stored in the NSSDC AIP Storage System (NASS) first released in 2005. 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 and tracked within NASS. NSSDC currently extracts and stores the NSSDC defined attributes from all ingested AIPs. This is done to facilitate internal analysis of the AIP holdings. Cyclic redundancy checksums are used to safeguard against data corruption while creating and preserving AIPs. Validated copies of every AIP are written to multiple media in distributed locations. For any given AIP at least one copy is stored off-site.

The media employed for AIP storage are DLT 7000 and SDLT 600 magnetic tape with native capacities of 30 and 300 GB respectively. NSSDC policy calls for a eight year refreshment interval for magnetic media containing AIPs. Those AIPs residing on the older DLT 7000 tapes are now being moved to SDLT 600 tapes.

2.7.4 Digital Library

The NSSDC Digital Library is similar to an traditional library. Here the objects to be distributed are digital images, data, software, documentation, etc. The collection of digital items include stand-alone images, data, software and document objects (analogous to books). The collection of digital items also include serialized collections of images, data, software and document objects (analogous to journals/serials). Like a traditional library, there is a Master Catalog (analogous to a library catalog of all holdings), as well as sets of cross-indexes for serialized collections based on location and temporal parameters (analogous to journal cross-indexes). Also, like a traditional library, there are Curation Scientists and Discipline Specialists (analogous to reference librarians) to aide the researcher in finding images, data, software and document objects relevant to the particular science problem(s) the researcher is addressing.

2.7.5 Tech Upgrades: Tape or Not to Tape (C-6)

Magnetic tape is NSSDC's primary media for long-term preservation of digital data and is expected to remain so for the next few years. Although magnetic disk is now commonly used for data backup, tape is widely acknowledged to be superior for long-term archival applications. (See http://storageconference.org/2008/presentations/1.Monday-DAPS/9.Schechtman.pdf.)

The primary advantage of magnetic tape over disk is cost. Modern tape technologies, Digital Linear Tape (DLT) and Linear Tape-Open (or LTO), continue to hold a sizeable cost advantage over magnetic disk that is projected to continue into the foreseeable future. Data density of magnetic tape has doubled approximately every three years, in parallel with the density increases of magnetic disk. The inherent transportability of modern tapes coupled with their long shelf life facilitates the task of storing duplicate data off-site. Note also that maintenance costs of tape sitting on a shelf are much less than a continually spinning disk. The fact that magnetic tape has slower access time than magnetic disk is not significant in archival applications where the data is accesses relatively infrequently. NSSDC's objective is to maintain data in usable form as long as it has significant science value. Magnetic tape is the best medium to cost-effectively sustain that objective.

In addition to permanent archive support, magnetic tape is one of several data delivery mechanisms endorsed by NSSDC and is the preferred mechanism of some providers (e.g. HEASARC).

NSSDC currently uses SuperDLT-4 tapes (300 GB native capacity) for long-term archiving of digital data. Faced with an increased inflow of digital data, higher capacity drives are required. NSSDC plans to purchase a new tape library equipped with LTO-4 (800 GB native capacity) tape drives. The jukebox will be upgraded with the next generation LTO-5 tape drives (1.6 TB native capacity) when they are available. Upgrades to successive generations of tape media will occur as they become available. Historically, a new tape generation of magnetic tape with double the capacity of the preceding generation becomes available approximately every three years. A future expansion of the tape library is anticipated.

2009	LTO-4 tape jukebox	$12,000
	Network upgrade	$4,000
2010	Magnetic disks for staging/processing of very large data deliveries	$32,000
2011	LTO-5 tape upgrade	$4,000
2013	Additional tape library with next generation drives	$12,000

Table 2.7.2. In order to anticipate the trends of rapidly increasing data volumes and demand, NSSDC must invest in technology upgrades. Upgrades to higher-generation tape technology will enable the storage of more data in the same area, network upgrades will enable more efficient electronic delivery of data, and additional magnetic discs will allow larger data deliveries to be staged and processed.

Other planned technology augmentation includes an additional 100 TB of network attached disk storage to provide necessary space for data staging and processing of ever larger deliveries from providers, upgrading the NSSDC network hardware to gigabit capability for improved internal performance and better support for large network delivery. A summary of the upgrades and estimated cost per year are shown in Table 2.7.2.

2.8 Cost Efficiency

One helpful tool NSSDC uses to assess its performance is the Annual Report, which identifies the various types of data requests, volumes, and categories of customers who use our services. Reporting results from the last three years allows us to closely examine trends, anomalies, and processes for increased efficiency.

NSSDC continues to process offline science data requests that involve special handling. Out of the 64 requests received since 2006, over half are Planetary Science (51%), 31% are Heliophysics, and the remainder Astrophysics/other. In recent years NSSDC has encouraged electronic dissemination to all users whenever possible. In 2008 FTP distribution was a necessity for most requests due to budget limitations for most of the year, though some media was distributed. Since 2005 NSSDC has tracked data requests by "items" within four broad categories. Table 2.8.1 shows the distribution of data served within these categories from 2007 to present. The table shows the distribution of the items by discipline for the same years . The table also shows FTP as the dominant mode of distribution, even though these statistics include only FTP data that were newly posted in response to a request, not data previously made available.

	Items				Discipline			
	Disks	**Print**	**Analog**	**FTP**	**Astrophysics**	**Planetary**	**Heliophysics**	**Other**
2006	872	534	119	14,318	4,395	3,401	8,040	7
2007	441	177	9	20,907	277	2,311	18,946	0
2008	186	109	4	20,877	95	20,979	102	0

Table 2.8.1 Total items sent out monthly from NSSDC. Disks include CDs and DVDs. Print media include photos (prints, slides, negatives), maps, documents, and posters. Analog consists of microfilm, microfiche, and video. FTP is comprised of requested digital data files, photos and documents. The Other discipline category includes Earth Science and ephemeris data. As evidenced in this table, the requests do not follow a predictable pattern, but their frequency integrated over time appears to be consistent.

An example of an anomaly would be the large 2007 FTP offline request for Apollo Data on 10 reels of Microfilm. These reels were digitized and staged for electronic retrieval. Note the large request for Planetary data. The largest increase is due to having data digitized and staged.

A great many NSSDC data collections and other information services are held permanently on disk for FTP access. The reader is invited to review all these services from the FTP link on the NSSDC home page. Table 2.8.2 gives the annual counts of files downloaded, both overall and for selected directories with high activity. Downloading by researchers via ftp of data files from the spacecraft_data subdirectory remained dominant in 2008, showing the high interest in and great value of these services provided by NSSDC. The Photo Gallery also remains of high public interest.

NSSDC collected from its space science user community a list of requirements for science data archive usability. The results were presented at both the SAW I and the Web Fed conferences. From these interactions, NSSDC plans to use Web 2.0 technology to create a collaborative environment for web technologists and space scientists to work together with other archive communities with similar challenges.

	2006	**2007**	**2008**
Spacecraft Data (Heliophysics)	**802,438**	**689,961**	**1,856,362**
Photo Gallery	936,039	720,213	481,089
All others on nssdcftp	998,388	930,342	634,070
Total	**2,736,865**	**2,340,516**	**2,971,521**

Table 2.8.2 NSSDC provides much of its data at no cost through its online NSSDC FTP site. The number of files downloaded each year shows that the site provides a valuable service to data users.

Summarized, our reporting shows that even with a relatively flat budget, we continue to provide access to more data and enjoy increased usage of our data (more data being pulled back). Our headquarters sponsor can inspect the bi-monthlies submitted to headquarters which report activities in the six OAIS functions; the increase in data volume and AIPs written can be quantified. In addition, the bi-monthly contains a table of deliverables as well as response times that the NSSDC staff must meet for both data ingest and data distribution; contractor response is gathered through a Performance Evaluation Board.

3. Budget Narrative

To comply with the Senior Review call, the budgets are expressed in the 3-way form as both In-Guideline and Functional. For the latter case, four basic functions based on Level II requirements, are identified to describe our activities: Ingest, Interactions (with external customers and internal activities), Information Management (to track and retrieve data), and Archival Storage. In addition, sub-functions that emphasize expected major themes or have been specifically requested by our headquarters sponsor are called out. A separate Excel file is also provided with detailed budget breakouts in FTEs with sheets by function or “all functions” which is organized by year. Note, functions are shown as bold and can be expanded or contracted using the “group” option.

In section 3.1 below the In-Guideline and Functional budgets are given for sub-functions as described in one scenario. The same form has been used in section 3.2 to identify the Request/Enhanced and associated Functional budget.

3.1 In-Guideline

Large increases to the data volume were discussed under the Technical section, but now by streamlining the Ingest process, only a small resource increase is necessary for FY2010-2014 in the operational areas of this process. In that timeframe, a pipeline will be implemented for non-AIP data such as the on-line data collections for SPDF. An initiative to broaden our task to migrate legacy tape data by ensuring its usability through a format conversion is also outlined; this will be done for the highest priority Heliophysics data.

In the Interactions role, there are seven budgets items identified. The first item is our general function to interact with customers both internal and external to the project and stays relatively constant. There is an upturn in 2010 as some of the NSSDC staff will be moved to a new GSFC Science Exploration Building and reorganized. The legacy migration comprises a large segment of funding under internal Interactions; phase one should be completed within five years. The second item involves a new initiative to prototype the conversion of specific analog holdings discussed in section 3.2; this task is zero for the In-Guideline and Functional budgets. The third item is our support for the World Data Center for Satellite information; our resources provide the SPACEWARN Bulletin and data to international partners. Next, the activity listed for the In-Guideline budget is our effort to interact with the recent RAs and work with them to provide guidance such as the TRAC approach under the auspices of CCSDS. A related workshop to gather lessons learned is being planned, thus an increase in travel funds in FY2010 to bring in speakers is included. For Providing Data Access, web services will be developed in FY2010 as data distribution shifts more to electronic delivery; this will also support SPASE. The general request office will be continued as long as it can provide access to unique data holdings. Participating in appropriate data registries in space science will also allow users to locate this legacy data. The next two items are for efforts in the standards area. NASA currently has an Interagency agreement with NARA which allows NSSDC to conduct research into standards usage and this work will be continued. In Preservation Standards, there is an expected increase for FY2010 for the aforementioned workshop

In the Information Management item, FTE resources are expected to grow in FY2010 as the relevant databases (experiment spacecraft, request, personnel) are overhauled. Also included is a small effort for XML development and implementation in the NSSDC metadata systems to improve efficiency as well as to map SPASE numerical data. The SPASE item lists support separately as a major theme. Internal activities include web support, registry software maintenance, and model standardization. The current grant to the UCLA SPASE group will be continued for the FY2010-14 period to assist in model work as well as tool development.

For Archival Storage, many of the management storage functions will be automated to increase our reliability. Regarding infrastructure upgrade, projecting out through FY2011, the operations tape system (LTO4 drives) is expandable to accept the expected volumes (100 TB in FY2010 and FY2011). Our equipment expenditures includes purchasing a new tape jukebox ($12K each) which is planned in FY2013. Modernization is expected to continue by upgrading to low-cost Linux machines on a four-year cycle and purchasing a replacement RAID systems ($32K).

3.2 Requested/Enhanced (Usability of Analog Data)

NSSDC holds a large number of microfiche and microfilm copies of Solar/Space Physics archived data. Since NSSDC now has the capability to digitize analog holdings and OCR scanned images to provide tabular data, the uniqueness and usability of applying such a technique will be a new initiative. One immediate prototype activity will be to select and then digitize ionograms of interest to the community.

- FY10 provide digital copy of index file for full set of ionograms, evaluate reports on analog holdings by Lowman and King
- FY11 select samples that are unique in time or station/spacecraft combination, make inventory web accessible
- FY12 digitize samples and evaluate results of both in-house and contractor approach
- FY13 put ionograms on-line apply scanning process to lunar (Apollo) data and determine if OCR needed
- FY14 provide inventory with options for access to both scanned and tabular products

This process will be initiated with a small amount of curation scientist supervision (0.1 FTE) and on-site staff resource (0.5 FTE). It is expected that current review efforts will be reviewed in FY2010 and a prototype process started to identify low-contrast microfilm reels for scanning. From previous experience it is estimated that about 100 reels can be scanned during a year. The high-contrast data will go to our remote contractor, JBI, for bulk digitization using an existing NSSDC contract. From this source 400 more reels are expected to be scanned. Estimates indicate that there are roughly 15 frames per "observation" and 500 frames/reel. Once the process is developed, the task continues in the background with either on-site or contract support. As the second step the technique will be re-evaluated in FY13 to determine if other data should be scanned and subsequent curation scientist or staff time is decreased as work with the outside contractor is increased.

4. References

4.1 Bibliography List

http://nssdc.gsfc.nasa.gov/publication/bibliographies.html

4.2 Important URLs

NSSDC home page: http://nssdc.gsfc.nasa.gov/

NSSDC Master Catalog : http://nssdc.gsfc.nasa.gov/nmc/

SPACEWARN Bulletin: http://nssdc.gsfc.nasa.gov/spacewarn/

NSSDC Data Retention Policy: http://nssdc.gsfc.nasa.gov/nssdc/data_retention.html

NSSDC white paper on Resident Archives: http://nssdc.gsfc.nasa.gov/nssdc/RAXWhite_Paperu.doc

NSSDC response to questions about Resident Archives: http://nssdc.gsfc.nasa.gov/nssdc/ra7comb.doc

Reference Model for Open Archive Information Systems:
http://public.ccsds.org/publications/archive/650x0b1.pdf

Radio JOVE Project home page: http://radiojove.gsfc.nasa.gov/

Moon Trees Project: http://nssdc.gsfc.nasa.gov/planetary/lunar/moon_tree.html

HEASARC: http://heasarc.gsfc.nasa.gov/

IRSA: http://irsa.ipac.caltech.edu/

LAMBDA: http://lambda.gsfc.nasa.gov/

MAST: http://archive.stsci.edu/

PDS: http://pds.jpl.nasa.gov/

SDAC: http://umbra.nascom.nasa.gov/

SPDF (SECAA): http://spdf.gsfc.nasa.gov/

Heliophysics Data Policy: http://lwsde.gsfc.nasa.gov/Heliophysics_Data_Policy_2009Apr12.html

5. Acronyms

AA	Active Archive
ACE	Advanced Composition Explorer
ADEC	Astronomy Data Centers Executive Committee
AIM	Aeronomy of Ice in the Mesosphere
AIP	Archive Information Package
ALSEP	Apollo Lunar Surface Experiments Package
ASCII	American Standard Code for Information Interchange
CCSDS	The Consultative Committee for Space Data Systems
CD	Compact Disc
CDAWeb	Coordinated Data Analysis Workshop web interface
CDF	Common Data Format
CODATA	Committee on Data for Science and Technology
COSPAR	Committee on Space Research
CRA	Consortium of Resident Archives
CRC	Cyclic Redundancy Check
DCC	Digital Curation Centre
DLT	Digital Linear Tape
DPS	Division of Planetary Sciences
DVD	Digital Video Disc/Digital Versatile Disc
EBCDIC	Extended Binary Coded Decimal Interchange Code
EOS	Earth Observing System
ESA	European Space Agency
ESMD	Exploration Systems Mission Directorate
FAST	Fast Auroral Snapshot Explorer
FITS	Flexible Image Transport System
FTE	Full-time Equivalent
FTP	File Transfer Protocol
GALEX	Galaxy Evolution Explorer
GPB	Gravity Probe-B
GSCL	Genesis Sample Curation Laboratory
GSFC	Goddard Space Flight Center
HDCWG	Heliophysics Data and Computing Working Group
HDF	Hierarchical Data Format
HEASARC	High-Energy Astrophysics SARC
IBEX	Interstellar Boundary Explorer
ICSTI	International Council for Scientific and Technical Information
IEEE	Institute of Electrical and Electronics Engineers, Inc.
IMAGE	Imager for Magnetopause-to-Aurora Global Exploration
IMP	Interplanetary Monitoring Platform
IPDA	International Planetary Data Alliance
IRAS	Infrared Astronomical Satellite
IRSA	Infrared Science Archive
ISEE	International Sun-Earth Explorer
ISO	International Standards Organization
ISTP	International Solar Terrestrial Physics
JAXA	Japan Aerospace Exploration Agency
JBI	John Bordynuik, Inc.

JEDS	Java-interface to Experiment, Dataset, Spacecraft and publication information in NIMS
JIN	Java-interface to INventory information in NIMS
JPL	Jet Propulsion Laboratory
JRAND	Java-interface to Request And Name Directory information in NIMS
JSC	Johnson Space Center
LAMBDA	Legacy Archive for Microwave Background Data Analysis
LDP	Lunar Data Project
LPSC	Lunar and Planetary Science Conference
LROC	Lunar Reconnaissance Orbiter Camera
MAST	Multi-mission Archive at Space Telescope
MER	Mars Exploration Rovers
MESSENGER	Mercury Surface, Space Environment, Geochemistry and Ranging
MOU	Memorandum of Understanding
MPGA	Multi-file Packager and Analyzer
MSC	Michelson Science Center
NARA	National Archives and Records Administration
NASS	NSSDC AIP Storage System
NCITS	National Committee for Information Technology Standards
NEAR	Near Earth Asteroid Rendezvous
NIMS	NSSDC Information Management System
NMC	NSSDC Master Catalog (public view to NIMS)
NOST	NASA/Science Office of Standards and Technology
NSSDC	National Space Science Data Center
NUG	NSSDC Users Group
NVO	National Virtual Observatory
OAIS	Open Archival Information System
OCR	Optical Character Recognition
OASIS	Orbiting Astrophysical Spectrometer in Space
PDMP	Project Data Management Plan
PDI	Preservation Description Information
PDS	Planetary Data System
PDS NAIF	PDS Navigation and Ancillary Information Facility Node
PI	Principle Investigator
PV-2005 (PV 2005)	Ensuring Long-Term Preservation and Adding Value to Scientific and Technical Data
RA	Resident Archive
RADP	Resident Archive Data Plan
RAID	Redundant Array of Independent Disks
RAUG	Resident Archive User Group
RHESSI	Reuven Ramaty High Energy Solar Spectroscopic Imager
RLG	Research Libraries Group
ROM	Read Only Memory
SAMPEX	Solar Anomalous and Magnetospheric Particle Explorer
SARC	Science Archive Research Center
SDAC	Solar Data Archive Center
SDC	Science Data Center
SDLT	Super DLT
SDO	Solar Dynamics Observatory
SECAA	Sun Earth Connection Active Archive
SIP	Submission Information Package
SPASE	Space Physics Archive Search and Extract

SPDF	Space Physics Data Facility
SSSC	Sun-Solar System Connection
STEM	Science, Technology, Engineering, and Mathematical
STEREO	Solar Terrestrial Relations Observatory
SWAS	Submillimeter Wave Astronomy Satellite
TRAC	Trustworthy Repositories Audit and Certification
THEMIS	Time History of Events and Macroscale Interactions during Substorms
TIMED	Thermosphere Ionosphere Mesosphere Energetics and Dynamics
TWINS	Two Wide-angle Imaging Neutral-atom Spectrometers
WDC SI	World Data Center for Satellite Information
VEPO	Virtual Energetic Particle Observatory
VHO	Virtual Heliospheric Observatory
VMO	Virtual Magnetospheric Observatory
VO	Virtual Observatory
VSPO	Virtual Space Physics Observatory
VxO	Virtual Observatories in Heliophysics
XFDU	XML Formatted Data Unit
XML	eXtensible Markup Language
YLA	Yohkoh Legacy Data Archive