ATTACH. K

NAS5-98156
MOD 21

NASA

Space Science Data Operations Office

Archive Management Plan

NAS5-98156
Version: February 3, 2000

Prepared by

Raytheon Information Technology and Scientific Services

Raytheon

Approval Signatures

Contractor Managers

Program Manager	2/7/00 Date
, Performance Based Contracting Manager	2/10/00 Date
, Data Center Manager	2/10/00 Date

Government Managers

Dr. Joseph H. King, Head NSSDC and WDC-A	2/8/00 Date
George Fleming, SSDOO PBC Work Element Manager	2/10/00 Date
Mary Stevens, SSDOO Contracting Officer	2/14/00 Date

Table of Contents

Section	Page
1.0 Introduction	4
2.0 Scope	5
3.0 Applicable Documents / References	6
4.0 NSSDC Archive Management	7
4.1 Performance Measurement	7
4.2 Archive Management Procedures	7
5.0 Plan / Schedule	13
5.1 Milestone List	13
5.2 Milestone Explanations	13
Appendix A: The Film Archive	15

1.0 Introduction

The purpose of this document is to plan for the long-term handling, preservation and availability of space science data sets at the NASA Space Science Data Operations Office (SSDOO). It establishes a set of approaches that have been coordinated with, reviewed by, and approved by both Contractor and Government management that will result in an archive that is able to meet the needs of the scientific community for data and information at the SSDOO.

As a jointly approved document, this plan witnesses the united commitment of both Contractor and Government. The plan does not distinguish between contractor and government roles, however, the work will be performed primarily by the contractor with regular consultation with the government Head of the NSSDC. Strategic, evolutionary planning regarding all aspects of the NSSDC will officially come through LOE task 318. The data center contractor personnel will support such efforts as requested and will also submit unsolicited improvement ideas from time to time for consideration. But, evolution planning along with requirements definition is primarily a Government matter.

As a "plan" this document provides a top-level overview of our approach and a high-level plan for maintaining the present high quality archive management operations at the SSDOO, and for evolving and improving operations in the future. Other documentation, that is referenced herein, contains the detailed procedures and schedules that will be used on a day-to-day basis for management and training purposes. This document will be updated no less frequently than once per year and approved by the Government in order to ensure that it remains consistent with current SSDOO policies and the Government's evolution planning.

2.0 Scope

The SSDOO acquires, processes, organizes, documents, archives and disseminates data from NASA's space science missions. It supports the world community of scientists and the public. The SSDOO's archive management charter is comprised of the work of the Astrophysics Data Facility (ADF), the Space Physics Data Facility (SPDF), the National Space Science Data Center (NSSDC), and the Orbiting Satellites Project (OSP).

This document addresses only the SSDOO archive management activities performed under the auspices of the NSSDC. The NSSDC's development and data / information handling, processing, and storage occur primarily at the NASA Goddard Space Flight Center's Building B28 in Greenbelt MD with additional storage at the National Archives' Federal Records Center in Suitland, MD, and the Iron Mountain off-site facility near Jessup, MD.

This archive management plan covers NSSDC activities performed by the Raytheon ITSS Performance Based Contracting (PBC) staff in fulfillment of the requirements in the SSDOO / Raytheon Contract (see Applicable Documents/References below) section 1.2.2 - Archive Management.

3.0 Applicable Documents/References

1) "Managing Electronic Records," National Archives and Records Administration (NARA), Office of Records Administration, Washington D. C., 1990.
2) "Care and Handling of Computer Magnetic Storage Media" (NBS Special Publication 500-101), National Bureau of Standards, Institute for Computer Sciences and Technology, 1983.
3) NARA Code of Federal Regulations, Subchapter B "Records Management", Part 1220 "Federal Records, General," see gopher://gopher.nara.gov/00/about/cfr/records/1220.txt
4) NARA Code of Federal Regulations, Subchapter B "Records Management", Part 1234 "Electronic Records Management," see gopher://gopher.nara.gov/00/about/cfr/records/1234.txt.
5) NASA Procedures and Guidelines (NPG) 1441.1C, "NASA Records Retention Schedules," see http://www.hq.nasa.gov/office/cio/records.htm.
6) National Media Lab (NML) Magnetic Tape Storage and Handling "A Guide for Libraries and Archives."
7) SSDOO / Raytheon Contract (NAS5-98156, 12/1/98)
8) SSDOO Standard Operating Procedures (NSSDC/WDC-A-R & S 95-05).
9) "Reference Model for an Open Archival Information System (OAIS)," Consultative Committee for Space Data Systems, (CCSDS 650.0-W-4.0), see http://ssdoo.gsfc.nasa.gov/nost/isoas/.
10) Reference material for checking the viability of CDs: see http://www.mscience.com/; http://www.eyecote.com/clover.htm; and http://www.kochdigi.com/PAGE/TECHNO/PRODUCT/CS42_LOC.HTML.

4.0 NSSDC Archive Management

4.1 Performance Measurement

The performance of the Archive Management Staff is measured against success criteria specified in the SSDOO Raytheon Statement of Work. The PBC manager will ensure that accurate and complete performance measurement records are collected and maintained in a Metrics Tracking System database.

4.2 Archive Management Procedures

Detailed procedures for managing the NSSDC archive are documented in the NSSDC Standard Operating Procedures (SOP) document. These procedures are established to make the NSSDC fully compliant with NARA and NASA guidelines for archive management (within the budgetary and physical restrictions imposed by the NSSDC's sponsors). In this section, we outline the kinds of procedures contained in the SOP, the key features they address, and plans for implementing and enhancing the procedures.

4.2.1 Inventory

The NSSDC has thousands of volumes of electronic and other media from many data providers. NSSDC's efficient operation relies on the utility, completeness, and accuracy of its detailed information management system of metadata associated with these data and information so that we know at all times what we have and where it is, and can determine what actions need to be taken to properly maintain the archive.

Detailed descriptions of the NSSDC databases can be found in Product Baseline documentation maintained by the SSDOO CCB. Procedures for maintaining and reporting on the contents of the databases can be found in the SOP.

4.2.2 Media

As a data center that supports many and varied space science programs, the NSSDC has procedures, hardware, and software to handle a variety of media types. The NSSDC has media standards for data and information stored in its permanent archive and its data dissemination environment.

Supported media types and standards, and procedures for replenishing the media stock are specified in the SOP.

4.2.3 Acquisition

A key NSSDC function is to acquire data from NASA-funded and NASA-relevant missions (including data generating programs and NASA/OSS active archives) in such a way that they can be archived and independently used by future requesters who need not be familiar with the particular data set or the mission from which it came. In practice, an acquisition scientist, supported by the Archive Management Staff, acts as the interface between the NSSDC and the data provider.

Procedures associated with this activity, from identification of data for acquisition, through creation of a project data management plan, and delivery of Submission Information Packages to the Archive Management Staff for ingest are specified in the SSDOO Acquisition Scientist Handbook.

4.2.4 Ingest

The ingesting of data and information into the NSSDC archive begins when a Submission Information Package is physically or electronically received in the NSSDC data operations area, an archive identifier is assigned, and a database entry is established for the data. Ingest concludes with the data or information item becoming part of the NSSDC's permanent archive and/or being placed in the NSSDC's data dissemination environment. During the ingest process, metadata is extracted and validated, and information systems databases are populated to support proper life cycle data management, response to user requests, and creation of accurate management reports.

Procedures associated with Ingest activities are specified in the SOP.

4.2.5 Retrieval

The requesting and retrieval of data and information from the NSSDC archive begins with a user request and concludes with data being delivered to or electronically downloaded by the user / customer. Two general categories of data retrieval processes are handled by the NSSDC:

1. Automated electronic data retrieval from the NSSDC's data dissemination environment via FTP, a direct e-mail request from the user, or via e-mail sent by one of several NSSDC user interface software packages.
2. Telephonic, mail, or other physical request mechanisms for electronic, document, or other NSSDC data products.

Automated electronic data retrieval procedures are specified at Internet site:
http://nssdca.gsfc.nasa.gov/ndads/archives_homepage.html.

All other data and information retrieval procedures are specified in the SOP.

4.2.6 Preservation

The most critical function of the NSSDC archive is to preserve data and information so that it is available to the scientific community over the long term. The importance of this activity is shown in the SSDOO Raytheon contract that severely penalizes the contractor when data is lost. Data preservation begins when the Government clearly identifies the data sets that must be protected and maintained indefinitely by the NSSDC. All other data are treated as requiring less attention. The identified data sets are marked in the NSSDC's databases for appropriate action. Through data backups, data monitoring, and media transitioning processes, specified data and information continue to be viable and available to users and customers into the indefinite future.

4.2.6.1 Data Reliability Standards

At the NSSDC, reliable storage entails three principles as follows:

1. There is a backup copy of the data.
2. The data is regularly monitored.
3. There is no media in the library, identified by the Government for long-term protection and maintenance, that is more than 10 years old. (The 10-year standard is important for the older 9-track tapes presently in the archive. Other media like CDs do not require this restrictive standard. We will refine this standard in subsequent releases of this document.)

The archive keeps this information in the NSSDC database for each item in the permanent archive (including the film archive). We create reports that show which media do not conform with our reliability standards.

4.2.6.2 Backups

As the most basic step toward ensuring the long-term viability of all data managed by the NSSDC, except in unusual circumstances that are accompanied by signed official Government direction, duplicate copies of the data on all ingested media are created as part of the data ingest process. It is our goal to store the backups in a different facility from the original from the original. Additionally, we regularly backup all operational databases stored on NSSDC systems. Procedures associated with the NSSDC data and database backups are specified in the SOP.

BACKUPS GOAL: The NSSDC will *"backup"* data to ensure conformance with reliability standards. During this year, space physics data sets not having a backup copy and not identified as being exempt from the backup requirement will be copied to backup media. By May 31, 2000, we will backup, as a minimum, at least 1/3 of this media set. We will provide the Government with a report of our progress in connection with the June 6-month metrics report. The remainder will be completed by November 30, 2000.

We will provide the Government with a report of our progress in connection with the December 6-month metrics report.

4.2.6.3 Monitoring

A statistical sample of the media/datasets identified for protection and maintenance will be "read" each year to ensure that the NSSDC can say with 90% confidence that no data have been corrupted. In the event that data are ever found to be corrupted (where "corrupt" means as little as one bit in a file is bad), the backup copy of the media will be used to restore the original data and information to the permanent archive. The goal is that no data identified for long-term maintenance will ever be lost. This approach applies to the "PBC contract" as follows: the contractor does not incur a "data loss" penalty if a file is found to be corrupted, only if the data is "lost;" i.e., the file is corrupt and the contractor is unable to restore the file from a backup. Procedures associated with the NSSDC media monitoring are specified in the SOP.

Because it is an impracticably large task to monitor the entire NSSDC archives in a year, we will apply statistical sampling methods to a portion of the archive that is most susceptible to "data loss," e.g., 9-track tapes. Based on lessons learned from initial monitoring efforts we will set more aggressive monitoring goals and plan to monitor other types of media, even media that do not pose a great loss risk such as CDs.

MONITORING GOAL: The NSSDC will *"monitor"* data that have been properly ingested to ensure conformance with reliability standards. By May 31, 2000, we will monitor, as a minimum, the first 2000 tapes from the Data Restoration Project set by randomly selecting fifty tapes and verifying their readability. This will demonstrate with 90% confidence that the entire set of 2000 tapes are viable. These tapes have been in a controlled environment since their creation approximately 8 years ago and should provide us with a very good initial measurement of the viability of the entire set. Additional monitoring, as negotiated by the Government and Contractor will be completed by November 30, 2000.

The following describes our statistical approach to monitoring media with 90% confidence.

Assumptions:

- 0% of the tapes in the NSSDC archive are not usable (i.e., less than one in two thousand).
- The results of our tape monitoring should prove that a 90% confidence interval for the mean percentage of bad tapes in a selected tape population is within .5% of zero.

Analysis:

- The required tape random sample size is determined by the equation:

$$.005 = 1.645 \times SQRT(((1/2000)x(1999/2000))/\text{sample size})$$

- Random sample size is about 50 tapes.
- If the fifty randomly sampled tapes are all good, then we can assume that the entire population is good.
- If one of the fifty tapes is bad then a total of about 500 tapes must be sampled as shown by the following calculation of a 90% confidence interval.

$$1/500 + 1.645 \times SQRT(((1/500)x(499/500))/500) = .005$$

- If two of the 500 tapes are bad then a total of about 800 tapes must be sampled as shown by the following calculation of a 90% confidence interval.

$$1/800 + 1.645 \times SQRT(((2/800)x(798/800))/800) = .005$$

4.2.6.4 Transitioning

As certain media near the end of their technological or physical life, and, as technology advances result in media options that hold larger volumes of data, have faster data transfer rates, and promise greater reliability, longevity, etc., new permanent archive media standards will be established, and, data and information will be copied from older to newer media types.

Fields in the IDA database are maintained by the contractor to indicate transitioning plans and status. The contractor is fully responsible for the goodness of all data ingested since 12/1/98. The contractor is also responsible for transitioning data in accordance with this signed AMP. This approach applies to the "PBC contract" as follows: a whole media volume gets marked as good in IDA, then any files that go irretrievably bad on that volume incur a penalty under the PBC contract. General procedures associated with the NSSDC media transitioning are specified in the SOP.

TRANSITIONING GOAL: The NSSDC will *"transition"* data to ensure conformance with reliability standards. We will begin to transition the Nimbus-7 THIR CLDT data set. This task involves copying 4900 tapes from older 9-track tapes to newer media. These tapes are ready for data transitioning in terms of age and there is a need to preserve this data for a current user. By May 31, 2000, we will transition, as a minimum, 500 tapes. Additional transitioning, as negotiated by the Government and Contractor will be completed by November 30, 2000.

4.2.6.5 Environmental Controls

The temperature and humidity of the SSDOO tape library is continually monitored to ensure that media do not degrade as a result of environmental conditions. Within the limitations of the physical plant in which the NSSDC archive resides, we strive to meet NARA guidelines. Temperature and humidity chart recorders are calibrated and weekly

paper recording charts are kept. Rolls of plastic are maintained in the library in case of a sprinkler leak. We inform the Government when environmental conditions are out of range.

General procedures associated with the NSSDC environmental controls are specified in the SOP.

4.2.6.6 Archive Security

We operate the media library during regular business hours to ensure that media are routinely handled by members of the Archive Management Staff. At the end of each workday the data in the library are secured -- internally locking the moveable archive racks and locking the outer door to the media library. In order to maintain "ownership" of archive tapes, sign-out/sign-in procedures are implemented. Tape librarians are the only authorized source for obtaining media from the library.

The NSSDC uses off-site facilities for storing backup copies of data and, in some cases, the only copy of inactive data sets. One facility is Iron Mountain located near Jessup, Maryland and the second facility is the Federal Records Center in Suitland, Maryland. These facilities are provided by the Government. Relocation of facilities is considered periodically, taking into account costs and care for the data.

General procedures associated with the NSSDC archive security are specified in the SOP.

5.0 Plan / Schedule

This section of the Archive Management Plan contains the schedule by which the NSSDC plans to implement improvements. It is a living section of the document that will be modified as needed. As such, it may be modified and published separately from the document as a whole. Any changes to the schedule shall be approved jointly by the contractor and the National Space Sciences Data Center & World Data Center A, Head Code 633 prior to implementation.

5.1 Milestone List

Following is a list of the milestones for efforts covered under this plan. More detailed explanations of the meanings of the milestones are in section 5.2.

Milestone Date YYMMDD	Reference AMP Paragraph	Milestone Description
000104	5.2.1	January metrics report
000107	5.2.3	Publish new SOP
000202	5.2.1	February metrics report
000302	5.2.1	March metrics report
000404	5.2.1	April metrics report
000502	5.2.1	May metrics report
000531	5.2.4	Complete Backups Part 1
000531	5.2.5	Complete Monitoring Part 1
000531	5.2.6	Complete Transitioning Part 1
000602	5.2.1	June metrics report
000602	5.2.1	Six month metrics report
000705	5.2.1	July metrics report
000802	5.2.1	August metrics report
000904	5.2.1	September metrics report
001003	5.2.1	October metrics report
001102	5.2.1	November metrics report
001130	5.2.4	Complete Backups Part 2
001130	5.2.5	Complete Monitoring Part 2
001130	5.2.6	Complete Transitioning Part 2
001204	5.2.1	December metrics report
001204	5.2.1	Six month metrics report
010101	5.2.2	Republish AMP
010101	5.2.3	Republish SOP

5.2 Milestone Explanations

5.2.1. Performance against contractual archive management metric objectives will be reported each month by the second work day of the following month. A six month

metrics report will be delivered to the Government for Incentive Fee award calculations by the second work day of the following months June 2000 and December 2000.

5.2.2. A new version of the Archive Management Plan will be published by January 1, 2001.

5.2.3. The NSSDC Standard Operating Procedures (SOP) (version 1995)will be updated as needed and published by January 7, 2000, republished by January 1, 2001.

5.2.4. The NSSDC will begin immediately to *"backup"* data that does not conform to the reliability standards in section 4.2.6.1. By May 31, 2000, we will backup, as a minimum, the media set indicated in section 4.2.6.2. Additional backups, as indicated in section 4.2.6.2, will be completed by November 30, 2000.

5.2.5. The NSSDC will begin immediately to *"monitor"* data that does not conform to the reliability standards in section 4.2.6.1. By May 31, 2000, we will monitor, as a minimum, the media set indicated in section 4.2.6.3. Additional monitoring, as indicated in section 4.2.6.3, will be completed by November 30, 2000.

5.2.6. The NSSDC will begin immediately to *"transition"* data that does not conform to the reliability standards in section 4.2.6.1. By May 31, 2000, we will transition, as a minimum, the media set indicated in section 4.2.6.4. Additional transitioning, as indicated in section 4.2.6.4, will be completed by November 30, 2000.

5.2.7. We will develop procedures and a plan for ingesting film and other image data, and for storing / transitioning film data on computer compatible media for long-term storage by May 31, 2000. When procedures and a plan are approved by the Government, we will begin immediately to convert 5% of the applicable film archive per year to digital format.

Appendix A: The Film Archive

The film and image component of the NSSDC archive has a very large volume of film products that are stored in many different formats. The film archive presently does not have an automated database system. However, it operates very effectively based the expertise of the staff and the application of our four focus areas that support the ISO Archive Reference Model. This appendix addresses archive management procedures that apply to the film and image component of the NSSDC archive. The NSSDC electronic data and document archive, including the permanent archive and data distribution environment component, is addressed in the main body of this document (as prescribed in the SOW).

A.1 Performance Measurement

The performance of the Archive Management Staff is measured against technical, cost, and schedule success criteria specified in the SSDOO Raytheon Statement of Work. The PBC manager will ensure that accurate and complete performance measurement records are collected and maintained in a Metrics Tracking System database.

A.2 Archive Management Procedures

We have detailed procedures for managing the NSSDC archive that are documented in the NSSDC Standard Operating Procedures (SOP) document. These procedures are established to make the NSSDC fully compliant with NARA and NASA guidelines for archive management (within the budgetary and physical restrictions imposed by the NSSDC's sponsors). In this section, we outline the kinds of procedures contained in the SOP and the key features they address, and our plans for implementing and enhancing the procedures in the future.

A.2.1 Inventory

With the present redesign of the IDA database, the current flat file database will be imported into the new relational database. It is our plan, over time, to complete the population of the NSSDC film archive inventory. Procedures for entering film items into the inventory database and for generating film archive inventory reports will be documented in the SOP.

A.2.2 Media

It is our plan to develop NSSDC film archive media standards. Procedures for generating film archive inventory reports will be documented in the SOP.

A.2.3 Acquisition

We presently encourage our data providers to submit film and image data on electronic media. Procedures associated with this film and image acquisition, from identification of data for acquisition, through creation of a project data management plan, and delivery of Submission Information Packages to the Archive Management Staff for ingest are specified in the SSDOO Acquisition Scientist Handbook.

A.2.4 Ingest

We will develop procedures for ingesting film and other image data, and for storing / transitioning film data on computer compatible media for long-term storage. When this plan is approved, we will begin immediately to convert 5% of the applicable film archive per year to digital format. Our approach is as follows:

- Prioritize - We will assess the interest in / importance of the various film datasets and prioritize them for digitization. We plan to gain official Government approval of our prioritization list.
- Migrate - After obtaining Government approval of our prioritization list, we plan to migrate to digital form at least 5% of the Government-directed non-perishable archive per year. A few images are presently being digitized to support customer requests.

Film archive ingest procedures are located in the SOP.

A.2.5 Retrieval

Film archive items are retrieved using the same manual retrieval procedures that can be applied to obtain items from the permanent data archive. There are presently no mechanisms for anonymous electronic download of film data.

It is our plan to develop NSSDC film archive retrieval procedures and document them in the SOP.

A.2.6 Preservation

We have multiple copies of some film and image data. It is our plan to develop NSSDC film archive preservation procedures for backups, monitoring and transitioning, and document them in the SOP. With the transitioning of the film products to digital media preservation of the film archive will be much easier and more efficient in the future.