

GAO

Report to the Chairman, Subcommittee  
on Investigations and Oversight,  
Committee on Science, Space, and  
Technology, House of Representatives

December 1993

# SPACE OPERATIONS

## Archiving Space Science Data Needs Further Management Improvements



**RESTRICTED--Not to be released outside the  
General Accounting Office unless specifically  
approved by the Office of Congressional  
Relations.**

**RELEASED**

---

1

---



United States  
General Accounting Office  
Washington, D.C. 20548

National Security and  
International Affairs Division

B-254957

December 9, 1993

The Honorable James A. Hayes  
Chairman, Subcommittee on  
Investigations and Oversight  
Committee on Science, Space,  
and Technology  
House of Representatives

Dear Mr. Chairman:

The National Aeronautics and Space Administration (NASA) creates long-term archives of scientific data from its space missions because of the potential importance of the data in future studies such as long-term analyses of the earth's environment. This report responds to the former Chairman's request that we review changes to NASA's data archiving program since we last reported on the program in 1990.<sup>1</sup> Our reports criticized the conditions of NASA's data archives and pointed out that the archives were incomplete for many important missions and held no data for others. In summary, we recommended that NASA improve the physical condition of its archives, obtain and restore all appropriate data from past missions, and revise its data management policies to ensure that all valuable data are adequately archived in the future.

Our objectives in this review were to determine (1) if NASA had improved its data archiving policies and practices in response to our recommendations and (2) whether there are other opportunities to improve that program.

---

## Results in Brief

NASA has addressed recommendations in our prior reports by improving its data archiving policies and practices, but further improvements are needed to correct some significant, long-standing problems.

Among other things, NASA has improved data storage conditions, created subject area archives, and revised its data management policies and procedures. Those actions should help ensure that important science data are available for future use. Improvements were accomplished through the agency's data management initiative and related activities. The data

---

<sup>1</sup>Space Operations: NASA Is Not Properly Safeguarding Valuable Data From Past Missions (GAO/IMTEC-90-1, Mar. 2, 1990) and Space Operations: NASA Is Not Archiving All Potentially Valuable Data (GAO/IMTEC-91-3, Nov. 2, 1990). A listing of related GAO products is provided at the end of this report.

---

management initiative is an effort to ensure the archiving of appropriate data from past missions and to create an infrastructure that will facilitate the orderly archiving and use of data from future missions.

While we were encouraged by these improvements, some significant data archiving problems remain. Specifically:

- Efforts to locate and archive all available data from past missions are incomplete, in part, because NASA did not review data held at temporary storage locations. Potentially valuable science data have been stored for extended periods—some more than 10 years—at these temporary facilities, which do not always meet archive requirements. The amount of data in “temporary” storage exceeds what is stored in permanent archives. As a result, potentially valuable science data could be lost and NASA may not realize possible savings in storage costs that could result from discarding data tapes when they are no longer needed.
- Budgeting for future archiving is inadequate because data management plans are not available for most science missions. According to NASA’s January 1993 State of the Data Union Report, many of the agency’s ongoing or planned science missions did not have approved project data management plans. These plans help define future archiving requirements and their absence helps mask true program costs and contributes to NASA’s overall affordability problems.
- Potentially valuable data could be lost because procedures have not been implemented to ensure that original data are preserved until permanent archives are created.
- Data could be damaged or destroyed because the agency still does not have a program for periodically inspecting data archive facilities and sampling data quality to ensure that physical conditions are adequate to protect data quality.
- NASA’s data management policies have not been updated to reflect changes in responsibilities resulting from a March 1993 reorganization of the agency’s space science division, and the policies have not been incorporated into NASA’s contract for services from the Jet Propulsion Laboratory, where large volumes of NASA data are stored.

---

## Background

NASA is responsible for space exploration and for managing, archiving, and disseminating space science data. The agency has spent billions of dollars on its space science programs and, through these efforts, has collected massive volumes of data on magnetic tape and other media such as optical disks for immediate and long-term scientific use. These science data can

be used for retrospective analyses and correlative studies that may not have been envisioned when the data were first collected. For example, one of NASA's most important current projects—the Earth Observing System—is making use of large data sets from several past missions because they span enough years for scientists to analyze climate trends. The older data will be used to support new theoretical models about, among other important research, global warming. NASA scientists have found that new developments in software and computer technology have made it possible for the first time to fully exploit the potential of older data sets.

When NASA receives data from its space missions, data processing centers create various types of data records. These include original data records, which contain all of the raw data acquired by spacecraft instruments and sensors; master data records that contain complete data from mission experiments combined with supporting information such as the spacecraft's orbital position and attitude; and experiment data records, which contain the data from each experiment. The data processing centers send the experiment data records to the principal investigators who designed and conducted the respective experiments. At the same time, the centers send copies of all data records to temporary storage facilities. As appropriate, copies of these records are also sent to permanent archives. After the principal investigators analyze the data and report the results of their research, they provide the reduced and analyzed data to NASA for permanent archiving. The intermediate data in temporary storage can then be destroyed and tapes or other media released for reuse. Data in permanent archives are available for use by other researchers. According to data management officials, in some more recent missions, such as astrophysics missions, NASA uses principal investigators and guest observers rather than principal investigators only. In those cases, data are sent directly to archives and are available to the general scientific community almost immediately. In commenting on a draft of this report, NASA officials stated that we did not adequately account for "the extent, impact, and future implications" of the new approach. We focused on NASA's more traditional environment because, at the time of our review, that approach was NASA's primary approach to archiving data, and magnetic tape was NASA's primary storage medium for digitized data. We note, however, that future configurations of data archives are subject to the same data storage requirements that apply to NASA's current facilities, and that non-tape storage media require appropriate storage practices to ensure that the data are safeguarded.

Because archiving activities of individual science missions are not separately identified in the agency's budget, agency officials could not provide comprehensive information about the amount of resources dedicated to archiving. Funding to create mission archives is typically included in each of the science divisions' mission operations and data analysis budget account.

NASA expects that the annual volume of its space science data will increase from one-half terabyte<sup>2</sup> in 1989 to over 2,500 terabytes by the late 1990s. NASA's primary archive facility—the National Space Science Data Center—expects to receive over 8.8 terabytes of data from 1993 through 1997. Some missions will likely generate larger separate archives. The Earth Observing System, for example, will generate about 1 terabyte of data per day and is expected to generate several thousand terabytes of data over the operational life of the program. These data are expected to be archived at mission-specific facilities.

## NASA Has Improved Archive Management and Policies

Through its data management initiative and related activities, NASA has made or is making improvements in its data archiving policies and practices to address most of our prior recommendations. The principal activities of the initiative, which was begun in 1991, were to improve the archive facilities used by the former Office of Space Science and Applications, create subject area archives in the office's science divisions, and locate, restore, and archive appropriate data sets from past missions. Through this initiative, NASA has eliminated many of the deficiencies in its data archiving practices that we identified in the prior reports. Our recommendations, and NASA actions in response to them, are summarized below:

- NASA should archive its valuable scientific data in facilities that meet federal standards. The agency has addressed this recommendation by improving physical storage conditions at its major data archiving facilities. Improvements such as adding smoke and fire detectors and card key access systems will better enable the facilities to meet federal requirements<sup>3</sup> regarding security, environmental, and maintenance controls for magnetic media. Also, NASA has moved data from a location

<sup>2</sup>A terabyte is one trillion bytes of data. About 5,000 high-density magnetic tapes would be required to store one terabyte of data.

<sup>3</sup>Federal requirements are established by the National Archives and Records Administration, which is responsible for setting standards for records retention, maintenance, and storage, and for evaluating agency records management programs.

where conditions were inadequate to an improved facility, and the agency plans to move additional data to improved facilities in the future.

- NASA should reconstruct data from aging magnetic tapes to archival quality tapes or other media. NASA has addressed this recommendation by copying data to new tapes and other media such as optical disks. The National Space Science Data Center has restored data from over 24,000 magnetic tapes to high-capacity data storage media. Also, in August 1993 Jet Propulsion Laboratory personnel began restoring data from about 50,000 tapes.
- NASA should improve its archive resource allocations. NASA created the data management initiative through which it is funding the creation of discipline data systems within each of the science divisions. These data systems will be used to store important project data near the personnel who are most familiar with the data so that they can provide expertise about it to users such as scientists who did not participate in the experiments and observations that generated the data.
- NASA should develop and implement agencywide tape management and maintenance standards that include all federal archive standards. NASA issued a records management guide in July 1992 that contains standards for the management and maintenance of data on magnetic tapes and other media. The guide also incorporates National Archives requirements for the storage of magnetic media.
- NASA should thoroughly inventory, assess, and, where feasible, archive all appropriate data from past missions. In August 1989, NASA began efforts to locate and inventory data from past missions. Agency officials assessed data holdings at two of NASA's major storage facilities and surveyed principal investigators from prior missions in an attempt to locate data for archives. As discussed below, these efforts were not fully effective because NASA did not assess data in its temporary storage facilities and its survey of data held by principal investigators from past missions received a very low response rate. However, officials located some additional data sets through this inventory, and NASA science divisions have been tasked with continuing to search for additional data sets that should be archived.
- NASA should identify offices and officials responsible for managing science archives, recognize the need to archive selected data of potential long-term scientific value from all types of missions, and ensure that all missions develop and submit approved data management plans. NASA issued a revised data management policy in March 1992. The revised policy defines the roles and responsibilities for data management officials, recognizes that NASA should archive all valuable science data, and requires the preparation and approval of data management plans for each space mission.

- NASA should improve the participation of the scientific community in managing science archives. NASA has taken several steps to address this recommendation. Improved participation by the science community has been accomplished by discussing data management issues in NASA conferences, improving the rate of preparation of mission project data management plans, and obtaining scientists' participation in the development of data archiving systems through science group advisory committees. The science community has also helped NASA locate data from older missions and to prioritize data for restoration.

---

## Further Improvements Are Needed in Data Archiving Program

Although NASA made certain improvements to its data archiving program, some long-standing problems remain. For example, the inventory of data from past missions is incomplete and, as a result, NASA could be incurring unnecessary storage costs, could risk losing data stored in inadequate facilities, and lacks a clear understanding of the amount of prior mission data that needs to be restored and archived. Also, the agency cannot adequately budget for data archiving because approved data management plans are not available for most ongoing and planned science missions. Further, NASA lacks procedures to ensure that original data records are preserved until permanent archives are created, and has no program for inspecting the physical condition of data storage facilities or periodically sampling the quality of data stored in the facilities. Also, NASA's data management policies do not reflect a March 1993 reorganization, and they are not specifically incorporated into NASA's contract for operation of the Jet Propulsion Laboratory, where large volumes of NASA data are stored.

---

## Efforts to Locate Archival Data From Past Missions Are Incomplete

In August 1989, NASA began efforts to locate and inventory data from prior missions so that it could permanently archive appropriate data. However, these efforts were not fully effective because NASA did not assess data held in temporary storage facilities and because the survey of principal investigators had a low response rate. Because NASA has not assessed data held in temporary storage facilities, the agency may be incurring unnecessary data storage costs and some potentially valuable data may be stored under inadequate conditions. Because the inventory is not complete, NASA does not know how much data needs to be restored or the resources that will be needed to accomplish the restoration.

NASA inventoried data stored at the Jet Propulsion Laboratory and National Space Science Data Center and, later, surveyed principal investigators from past missions. Three science divisions also have initiated efforts to



assess some additional data sets at other locations to determine the need for restoring older mission tapes. On the basis of those efforts, NASA officials concluded that most surviving data that should have been archived were already archived.

However, in inventorying data from prior missions NASA did not assess the archival value of data sets held in what it considers to be temporary storage. Although these facilities are intended as temporary storage locations, science data are frequently retained there for long periods. For example, in November 1992 the Tape Staging and Storage Facility at Landover, Maryland, was holding about 153,000 magnetic tapes beyond the 2-year retention period established for that facility.<sup>4</sup> Over 52,000 of the tapes had been held for more than 10 years; over 55,000 had been held from 5 to 10 years; and about 45,000 had been stored from 2 to 5 years.

According to NASA officials, during the inventory, they obtained an identification of the numbers of tapes by spacecraft mission held at the Tape Staging and Storage Facility. However, the officials did not assess the archival value of the data. By reviewing those tapes that have been held longer than the normal 2-year retention period for this facility, archiving appropriate data, and releasing the remaining tapes, NASA may be able to reduce data storage costs at this facility. According to facility managers, storage costs for those tapes that have been held for more than 2 years amount to about \$175,000 annually.

NASA also did not review data at several other temporary locations, such as university and other contractor facilities, where data sets of undetermined scientific value are being stored. While NASA did send questionnaires to former principal investigators, in most cases, the investigators did not respond. Storage conditions at some of these facilities are inadequate and may result in damage or destruction to data that should be archived. For example, several thousand tapes are stored at the Center for Astrophysics and Space Sciences. Our review of conditions at this facility showed that some tapes are still stored under inadequate conditions. For example, we observed one tape storage area with steam and sewage pipes running overhead, trash on the floor, water on the floor in some areas, and generally poor conditions for tape storage. Figure 1 shows conditions in this storage area.

<sup>4</sup>By comparison, NASA's primary archive—the National Space Science Data Center—held about 84,000 tapes in January 1992, and that inventory was being reduced by NASA's restoration program.

**Figure 1: Data Storage Conditions at the Center for Astrophysics and Space Sciences**



While we did not inspect other temporary non-NASA facilities, minutes from NASA science division meetings show that additional data sets considered to be of scientific value are held on aging magnetic tapes in various temporary facilities around the nation.

NASA still has not located data from many of its past missions, in part because its survey of principal investigators from prior missions was not effective. NASA contacted scientists involved in 549 investigations from prior missions in an attempt to locate additional data for archiving. However, NASA received no response about data for 345, or 63 percent, of the experiments, and officials were unsuccessful in most follow-up attempts to determine why principal investigators did not respond.<sup>5</sup>

Efforts to locate additional data for archives are continuing but NASA cannot effectively estimate the amount of funds needed to restore data for archives because it lacks information about the availability of data from the past missions. For example, after completing its survey of principal investigators, NASA officials identified 100,000 tapes of data at the

<sup>5</sup>A non-response does not necessarily mean that data that should be archived are being held by a principal investigator.

---

University of Iowa. According to science officials, the data are valuable and should be archived. However, costs to prepare these data for archiving are estimated to be \$2 million to \$3 million—which could exceed the \$2 million or more that NASA officials said the agency currently spends a year on data restoration activities. As a result, much of these data will probably be abandoned, according to one science division official.

In commenting on a draft of this report, NASA officials expressed a concern that we were citing its Tape Staging and Storage Facility for inadequate storage practices. We inspected the facility during this review and found that its storage areas had been improved. At our request, officials at this facility tested 384 randomly selected tapes for readability. All of the tapes were successfully read. We also note that, according to agency officials, the facility has reduced its old tape inventory; however, it continues to hold 76,000 tapes beyond their 2-year storage requirement.

NASA officials also stated that NASA has completed its inventory efforts and that the portion of unaccounted data is small, marginally useful, and not important enough to justify resources to restore and archive the data. NASA officials stated that the low response rate from former principal investigators supported this opinion. We disagree with this view and believe that there are opportunities to locate and archive more data. For example, in our November 1990 report,<sup>6</sup> we identified 18 missions for which NASA held no digitized data; at the time a National Space Science Data Center official stated that it was likely that data for 12 of these missions were held by the principal investigators or co-investigators. However, NASA identified additional data for only one of these missions in its census of principal investigators.

---

**NASA Cannot Adequately Budget for Archiving Because Many Projects Lack Project Data Management Plans**

Individual project data management plans are intended, in part, to provide the consistent documentation needed to facilitate planning for data archives. NASA's data management policy requires that all data gathered by science missions be addressed in a data management plan. The plan is to address the total flow of data throughout a project's life and is to be prepared and approved after new start and budget approval. Thereafter, the plan should be reviewed periodically to determine if significant changes have occurred that would require project officials to change it.

---

<sup>6</sup>Space Operations: NASA Is Not Archiving All Potentially Valuable Data (GAO/IMTEC-91-3, Nov. 2, 1990).

---

Although required by policy, at the time of completion of our review in September 1993, approved project data management plans were not available for most ongoing or planned science missions. According to NASA's January 1993 State of the Data Union Report, of the 83 operational and planned missions, only 18, or about 22 percent, had approved plans. Sixty-five lacked completed data management plans although, according to NASA, 12 had draft data management plans, and 13 had an alternative document for archiving and data management.

Some projects that lack final plans include 21 of the 22 current or approved Earth Science missions. This includes missions associated with the Earth Observing System—one of NASA's most expensive and highest priority programs.<sup>7</sup> According to an Earth Observing System official, the Earth Observing System Data and Information System will rely on agreements to establish archives with the eight distributed active archive centers instead of project plans. Through these agreements the Earth Observing System program will fund universities' and NASA facilities' programs to generate data products and manage and archive data. However, NASA had completed agreements to establish archives with only one of the eight organizations at the time of our review. According to data management officials, "straw-man" versions of agreements had been prepared for the other archive centers, but the agreements had not been finalized. According to these officials, a long-term archiving plan for the total Earth Observing System program is also being prepared. This document will gather all existing information about the plans for long-term archiving of Earth Observing System data into one short document.

Also, large volumes of Earth Observing System data will be turned over to the U.S. Geological Survey and the National Oceanic and Atmospheric Administration for permanent archiving; however, NASA has not reached final agreement with either agency on the details of how data will be archived or how the archiving activities will be funded. NASA has established memorandums of agreement with both agencies, but the agreements do not contain details about data archiving. An addendum to the agreement with the National Oceanic and Atmospheric Administration was drafted in May 1990 to provide more specifics related to long-term data archiving, but this addendum has not yet been finalized. In commenting on the importance of a long-term archiving plan in an

---

<sup>7</sup>The Earth Observing System is projected to cost billions of dollars and generate a several thousand terabyte data base over a 15-year span.

---

April 1992 report, the National Research Council's Panel to Review the Earth Observing System Data and Information System Plans<sup>8</sup> stated that:

"Long-term archiving of [Earth Observing System] data is an issue that has not been addressed. Long-term commitment to maintaining data collected as part of [the Earth Observing System Data and Information System] is a critical component of the U.S. Global Change Research Program. NASA, in its response to questions from the panel, correctly pointed out that the issue of maintaining long-term archives is one that must be addressed by all participating federal agencies. Without a concrete plan and agency coordination for establishing permanent data archives, however, the overall objectives of [the Earth Observing System], and, therefore, of the U.S. Global Change Research Program, are jeopardized."

Data management and archive budget plans are essential documents for preparing and budgeting for the archiving of mission data. We have reported previously that NASA could provide better program cost estimates to account for the full range of mission costs, and that its projections of funding needs were too short range to use for making forward-looking decisions.<sup>9</sup> One of the most important functions of a data management plan should be to support planning for missions' archive budgets. However, project archive activities typically lack dedicated budgets. NASA archive managers told us that mission budgets rarely fund archiving activities and, because archiving is typically completed at the end of a mission, adequate resources can be difficult to obtain. According to NASA data management officials, the real issue is not so much the existence of data management plans, but rather a commitment to follow through with priority and funding to actually do the job.

In commenting on a draft of this report, NASA officials stated that they believed that simply counting the number of approved project data management plans could be misleading because some missions were already operational and others would not be operational until after 1996. The officials also said that they believe only one data management plan is needed for the 22 instruments planned for the Earth Observing System. We note, however, that the NASA data management directive requires a project data management plan for all data captured by NASA science projects, and that a preliminary project data management plan is required for review before a project receives new start approval. Also, the statistics on the

---

<sup>8</sup>The National Research Council created the Panel to review the system and ensure that it makes data from past, current, and future earth observation missions readily available.

<sup>9</sup>See our report, NASA: Large Programs May Consume Increasing Share of Limited Future Budgets (GAO/NSIAD-92-278, Sept. 4, 1992) and our testimony NASA Budget: Potential Shortfalls in Funding NASA's 5-Year Plan (GAO/T-NSIAD-92-18, Mar. 17, 1992).

---

data management plans were derived from a NASA document that, according to an agency official, was reviewed and approved by NASA's science divisions. In earlier comments, officials stated that there are some missions that should have completed and approved data management plans but do not, which we believe is the important point.

---

**NASA Has Not Established Controls Over Original Data**

To prevent the possible loss of important science data, our November 1990 report concluded that NASA needed to establish controls to ensure that original data records are not destroyed before all appropriate data are stored in permanent archives. We recommended that NASA establish and enforce a system of internal controls to ensure that original data are retained until NASA's archives have received all appropriate data.

In response, NASA officials stated that, as of September 1992, the agency would require that original data records from space flights be held until archival data sets were created. However, no actual controls were established to retain original data records. To the contrary, NASA's records disposition schedules still provide for the destruction of original records not later than 2 years after experiment data records are created. There are no provisions for verifying whether archives had been created and delivered to NASA before the original records are destroyed.

---

**NASA Officials Do Not Periodically Inspect Data Archive Facilities**

In our March 1990 report, we pointed out that NASA's Information Resources Management Office was not providing adequate oversight of data management activities at NASA field centers. We recommended that the Information Resources Management officials periodically review data management and archiving activities to ensure compliance with National Archives and Records Administration regulations.

In February 1993 NASA revised its information resources management inspection program to focus on self-inspection by agency personnel, who are to evaluate the effectiveness of their management of information resources, including an assessment of facilities' data management practices. However, according to NASA records management and review officials, the self-assessments address data management policies and controls. NASA officials stated that individual NASA installations are responsible for deciding whether to conduct reviews of data storage conditions and whether to sample data to determine its quality.

---

In commenting on a draft of this report, NASA officials stated that future self-assessment surveys will be revised to place added emphasis on inventory and data storage practices, and that the surveys will be augmented by periodic visits and inspections by NASA headquarters personnel.

---

### Data Management Policies Need Updating and Uniform Application

NASA's former Office of Space Science and Applications revised its science data management policy and issued the revised policy in the form of a program directive in March 1992. The revised directive was responsive to our prior recommendation that offices and officials responsible for managing space science data be identified and their responsibilities clearly defined. However, in March 1993 NASA reorganized its space science activities, abolishing the Office of Space Science and Applications and grouping that office's six science divisions into three separate offices.<sup>10</sup> The restructuring changed the organizational location of personnel responsible for data archiving and raised questions about how these new organizations would continue to manage and fund data archives and projects under the data management initiative. The Chief of Information Systems within the new Office of Space Science stated that within the next few weeks the policy directive will be established as an Office of Space Science Policy Directive with no substantive change. The document will then be coordinated with other NASA organizations to be included in their policy directives.

Also, NASA's data management policies are not specifically included in its contract for services from the Jet Propulsion Laboratory, where large volumes of NASA data are stored. The government-owned laboratory is staffed by personnel from the California Institute of Technology under a contract with NASA. NASA policies must be specifically incorporated into the contract with the Institute to be legally binding. Currently, the laboratory is not under contractual obligation to follow the data management directive. Officials told us that although the laboratory voluntarily complies with accepted data management practices, that compliance could be unilaterally ended at any time. NASA's current contract with the Institute was scheduled to expire at the end of September 1993, and NASA is currently preparing a new contract. According to a data management official, the status of contract negotiations is such that the data management policies will not be put into the new contract. According to

---

<sup>10</sup>The new offices are Mission to Planet Earth (the former Earth Science and Applications); Space Science (combining the former Space Physics, Solar System Exploration, and Astrophysics); and Life & Microgravity Sciences and Applications (combining the former Life Sciences and Microgravity Science and Applications).

---

this official, however, NASA intends to require the Jet Propulsion Laboratory to comply with appropriate data management policies in all subsequent task orders for data management related work.

---

## Recommendations

We recommend that the NASA Administrator instruct officials responsible for data archiving and management to (1) expand NASA's science data inventory to include sites that have not been surveyed to date and archive or release data as appropriate, (2) ensure that mission project data management plans are prepared at the beginning of a project so that mission archives can be adequately planned and budgeted for in the context of overall program affordability, (3) establish milestones for completing plans and agreements with other agencies on Earth Observing System permanent archives, (4) require confirmation that permanent archives have been created before original data records are destroyed, (5) establish procedures requiring periodic review of the adequacy of physical conditions at data storage locations and sampling of the quality of data, (6) update NASA's science data management policy directive to reflect roles and responsibilities consistent with the current organization of NASA science divisions, and (7) include appropriate elements of NASA data management policies in future contracts with the California Institute of Technology for operation of the Jet Propulsion Laboratory.

---

## Agency Comments and Our Evaluation

NASA generally agreed with the findings of this report, commenting that it found the report to be constructive and useful in addressing some valid issues concerning science data management and archiving. Officials provided specific comments addressing the report's accuracy, relative emphasis, and perspective. These comments are presented and evaluated in appropriate sections of the report. NASA commented that the recommendations in this report are reasonable and NASA will pursue them. The NASA comments are reprinted in full in appendix I.

---

## Scope and Methodology

We analyzed NASA reports, briefings, and budget submissions, and reports prepared by external scientific groups and committees. We discussed NASA's data management program and the agency's actions to address recommendations in our prior reports with officials at NASA Headquarters, the Goddard Space Flight Center, and the Jet Propulsion Laboratory. We inspected data storage facilities at Goddard, the Tape Staging and Storage Facility, the Washington National Records Center, the Jet Propulsion Laboratory, and the Center for Astrophysics and Space Sciences. We also



---

discussed NASA's records management program with General Services Administration and National Archives and Records Administration officials.

We performed our work from November 1992 to September 1993 in accordance with generally accepted government auditing standards.

---

Unless you publicly announce its contents earlier, we plan no further distribution of this report until 30 days after its issue date. At that time, we will send copies of this report to other appropriate congressional committees, the Administrator of NASA, and the Director of the Office of Management and Budget. We will also provide copies to others upon request.

Please contact me on (202) 512-8412 if you or your staff have any questions concerning this report. Major contributors to this report are listed in appendix II.

Sincerely yours,



Donna M. Heivilin, Director  
Defense Management and NASA Issues

---

# Contents

---

|   |    |
|---|----|
| Letter  | 1  |
| Appendix I<br>Agency Comments<br>From the National<br>Aeronautics and<br>Space Administration | 18 |
| Appendix II<br>Major Contributors to<br>This Report   | 23 |
| Related GAO Products  | 24 |
| Figure  | 8  |

Figure 1: Data Storage Conditions at the Center for Astrophysics and Space Sciences

---

## Abbreviations

NASA National Aeronautics and Space Administration

---

---

# Agency Comments From the National Aeronautics and Space Administration



National Aeronautics and  
Space Administration

Washington, D.C.  
20546

Office of the Administrator

NOV 10 1993

Mr. Frank C. Conahan  
Assistant Comptroller General  
National Security and International Affairs Divisions  
General Accounting Office  
Washington, DC 20548

Dear Mr. Conahan:


Thank you for the opportunity to review the GAO draft report entitled "SPACE OPERATIONS: Archiving Space Science Data Needs Further Management Improvements" (GAO/NSTAD-93-294) dated October 1993.

In general, we find the draft report to be constructive and useful in addressing some valid issues concerning science data management and archiving. This is an important topic to which NASA has been devoting considerable management attention, initiative, and resources over the past several years to improve the stewardship of science data assets acquired from past missions and to assure appropriate priority for data management in planning upcoming missions.

Our principal issue with the draft report is one of relative emphasis and perspective. The enclosure provides more specific comments, factual corrections, and other suggestions that we believe will strengthen the report, improve its overall balance, and reduce possible misinterpretations.

We appreciate the cooperative approach taken by your staff throughout the process of the review. We also believe that the commitment of time and effort on the part of personnel at NASA Headquarters, Goddard Space Flight Center, and the Jet Propulsion Laboratory to this review bears further testimony to the importance and priority of the topic within NASA and hope that, too, can be reflected in the final report.

Sincerely,



G. F. Dailey  
Associate Deputy Administrator

Enclosure

**NASA Response To GAO Draft Report**  
**SPACE OPERATIONS: Archiving Space Science Data**  
**Needs Further Management Improvements**  
**GAO/NSIAD-93-294**

**I. General Comments**

1. The report emphasizes tape media, a single central archive site at the National Space Science Data Center (NSSDC), and the flow of data from individual Principal Investigators (PI's) back into archives, all of which represent former modes as opposed to current approaches. The report fails to capture the extent, impact, and future implications of distributed science discipline systems for archiving data, where the timely flow of data into archives is planned and managed directly through projects. Examples include the Planetary Data System (PDS) and the Astrophysics Data Centers which actively support project archive planning and implementation. Also, note that there is a significant migration to optical media for storage instead of magnetic media. In the case of PDS, they have established nationally recognized leadership in the area of new media technology options and have produced 190 Compact Disk-Read Only Memory (CD-ROM) titles for missions ranging from Voyager and Viking, to current high visibility missions such as Magellan and Galileo.

Perhaps there is no more striking example of how early and continuing attention to data issues, along with the change in the data flow model above, is in establishing the foundation to deal with the future onslaught of data than the Earth Observing System (EOS). Science data product generation and archival will be accomplished through the Distributed Active Archive Centers (DAAC's), and interim versions of these will operate for ongoing missions even prior to the launch of the first EOS platform.

2. The report acknowledges progress and improvements in the NASA archiving environment and attributes much of the progress to the data management initiative undertaken to accelerate the restoration of data from past missions and to complete science discipline infrastructure for the relatively new space science disciplines of Microgravity and Life Sciences.

It is important to note that this initiative was undertaken despite the inability to obtain a budget augmentation for it. This demonstrates the importance and priority within the science programs to get our archiving program in order and point to the future.

As we look to the future, however, the funding issue is even more crucial, especially regarding continued restoration of old data from past missions. This must be weighed in the context of tightly constrained funding resources. The cost of restoring older datasets with potential scientific value must be traded against return on investment of the limited resources on more current efforts. The scientific community needs to participate in establishing this balance, and there are panels in virtually all of the science disciplines dealing with this very issue at this time.

**Appendix I  
Agency Comments From the National  
Aeronautics and Space Administration**

3. The report devotes a disproportionate amount of discussion concerning the incomplete inventory of data from past missions. It is GAO's judgment that there is a substantial amount of unidentified data held at PI sites because of the relatively low response rate to surveys.

We believe the inventory is complete for which the vast majority of important data holdings is accounted. This is based on analysis of detailed databases at NSSDC and at JPL which represent the core holdings of large, critical science datasets. This was augmented by a census effort in 1990-91 to identify potentially important community-held data. This has been further augmented by solicitation of proposals to restore datasets as part of astrophysics research announcements.

The results have been reviewed by science disciplines for accuracy, completeness, and priority for restoration and archiving. The focus now is to work in conjunction with the science disciplines to establish priorities for actually archiving datasets from the list.

While the response rate of individual PI's to the survey was relatively low, this does not necessarily correlate to the volume of potentially valuable archivable data. We believe the portion of unaccounted data residing at PI sites is small and marginally useful and not important enough to justify resources to restore and archive permanently. In fact, this determination by PI's in not responding could provide some insight into the overall low response.

**II. Clarifications, Omissions, and Corrections**

1. Throughout the report, references are made to "temporary storage facilities". Clarification is needed to determine which sites fall into this category, especially since these form the basis for many of the shortcomings cited in the report.

It is assumed that the Tape Staging and Storage Facility (TSSF) at the Goddard Space Flight Center is included in the category of "temporary storage facility" and that statements such as those on pages 2 and 9 concerning "... potentially valuable science data could be lost", and "... agency may be incurring unnecessary storage costs and some potentially valuable data may be stored under inadequate conditions", apply to the TSSF.

The bottom line in data storage facilities is always the readability of the data. The GAO auditor challenged TSSF to prove that the old data tapes were readable. TSSF agreed to conduct this time-consuming, expensive test with the understanding that the results would be reflected in the report. GAO statisticians used a random number generator and selected 400 tapes to be read. Every tape was read successfully, using GAO criteria. The test tapes, many of which are 20 years old, are indeed readable.

We believe this should be explicitly acknowledged in the report, especially since this facility was singled out for criticism in the previous report.

Now on pp. 2 and 6.

Appendix I  
Agency Comments From the National  
Aeronautics and Space Administration

The other concern about incurring unnecessary storage costs by holding tapes too long also needs to be clarified. Through extensive reviews, the tapes that have been held for more than 2 years in the TSSF have been reduced from 154,000 in September of 1992 to 76,000 in October 1993, a 50-percent reduction in 1 year.

2. Simply counting the number of approved Project Data Management Plans (PDMP) can be misleading. Of the 83 missions cited in the total count, 48 are currently operating missions. While many of these do not have formal PDMP's, there is nonetheless systematic, effective archiving of data by virtually all of them.

Of the 35 future missions included in the total of 83, even the relatively far-future (beyond 1996), which understandably do not have formally approved PDMP's, virtually all including AXAF, Cassini, EOS, and TRMM, are actively addressing data management as part of overall project planning. Formal PDMP's and related agreements are also in progress.

The case of EOS warrants further comment. First, it is not appropriate to consider the need for "22 plans" by instrument but rather one basic plan which all PI's are required to follow. At this stage of the program the EOS Science Data Plan is the appropriate document to reference. The DAAC's have been written into EOSDIS plans, and interface agreements have even been written by the DAAC's with non-EOS projects such as TOPEX, NSCAT, and UARS.

3. Page 4 - The discussion of the manner in which NASA processes the data it receives is too general and does not capture the data flow model discussed above, whereby data products flow directly from projects to archive centers for distribution to the community. Also, the statement that NASA uses guest observers rather than principal investigators is not true. Most disciplines have opportunities for both guest observers and principal investigators
4. Page 5 - The apparent inconsistency between the annual rate of data expected in the late 1990's and the aggregate expected to be received by the NSSDC should be clarified.
5. Page 12 - NASA spends much more than \$400k per year on data restoration. Between the science-wide data management initiative and science-discipline-specific efforts, there is more than \$2 million per year put into data restoration.
6. Page 17 - Correct the first paragraph to read:

In February of 1993, NASA revised its information resources management review program to focus on self-assessment by agency personnel, who are to evaluate the overall effectiveness of their management of information resources, including an assessment of facilities' data management practices. *However, according to the NASA records management and review officials, the self-assessments address policies and management controls as they pertain to archiving and data management. It is the responsibility of NASA component installations to determine the scope of their reviews conducted of physical data storage conditions and sampling of data.*

Now on p. 3.

Now on p. 4.

Now on p. 9.

Now on p 12.

---

**Appendix I  
Agency Comments From the National  
Aeronautics and Space Administration**

---

**III. Recommendations**

The recommendations are reasonable, and NASA will pursue them. In particular, regarding recommendation (5), "establish procedures requiring periodic review of physical conditions at data storage locations":

The first agencywide IRM self-assessment did not include review of records management activities. Self-assessment surveys, which will be used to review records management in 1994, will be revised to place added emphasis on inventory and data storage practices. The self-assessment activity will be augmented by periodic staff assistance visits and spot checks by NASA Headquarters personnel.



---

# Major Contributors to This Report

---

**National Security and  
International Affairs  
Division, Washington,  
D.C.**

David R. Warren, Associate Director  
B. Scott Pettis, Evaluator-in-Charge  
Titina C. Eason, Evaluator  
Roberta H. Gaston, Evaluator

---

**Atlanta Regional  
Office**

Lee A. Edwards, Acting Assistant Director

---

**Los Angeles Regional  
Office**

Allan Roberts, Assistant Director  
Larry J. Bridges, Site Senior  
Benjamin H. Mannen, Evaluator

---

# Related GAO Products

---

Space Data: NASA's Future Data Volumes Create Formidable Challenges  
(GAO/IMTEC-91-24, Apr. 8, 1991).

Environmental Data: Major Effort Is Needed to Improve NOAA's Data Management and Archiving (GAO/IMTEC-91-11, Nov. 20, 1990).

Space Operations: NASA Is Not Archiving All Potentially Valuable Data  
(GAO/IMTEC-91-3, Nov. 2, 1990).

Space Data: Information on Data Storage Technologies (GAO/IMTEC-90-88FS, Sept. 12, 1990).

Space Operations: NASA Is Not Properly Safeguarding Valuable Data From Past Missions (GAO/IMTEC-90-1, Mar. 2, 1990).

---

### **Ordering Information**

**The first copy of each GAO report and testimony is free. Additional copies are \$2 each. Orders should be sent to the following address, accompanied by a check or money order made out to the Superintendent of Documents, when necessary. Orders for 100 or more copies to be mailed to a single address are discounted 25 percent.**

**Orders by mail:**

**U.S. General Accounting Office  
P.O. Box 6015  
Gaithersburg, MD 20884-6015**

**or visit:**

**Room 1000  
700 4th St. NW (corner of 4th and G Sts. NW)  
U.S. General Accounting Office  
Washington, DC**

**Orders may also be placed by calling (202) 512-6000  
or by using fax number (301) 258-4066.**

**United States  
General Accounting Office  
Washington, D.C. 20548**

