STATE OF THE DATA UNION

NASA Office of Space Science and Applications

1992

National Aeronautics and Space Administration
Office of Space Science and Applications
Information Systems Branch

Prepared with the assistance of
Futron Corporation
Under Contract NASW-4493

TABLE OF CONTENTS

FIGURES AND TABLES

PREFACE

This is the first report on the State of the Data Union (SDU) for the NASA Office of Space Science and Applications (OSSA). OSSA responsibilities include the collection, analysis and permanent archival of data critical to space science research. The nature of how this is done by OSSA is evolving to keep pace with changes in space research. Current and planned missions have evolved to be more complex and multidisciplinary, and are generating much more data and lasting longer than earlier missions. New technologies enable global access to data, transfer of huge volumes of data, and increasingly complex analysis. The SDU provides a snapshot of this dynamic environment, identifying trends in capabilities and requirements.

The SDU is envisioned as an annual report. This first report will describe the current space science data environment, and present parameters which capture the pulse of key functions within that environment. It reports on the continuous efforts of OSSA to improve the availability and quality of data provided to the scientific community, highlighting efforts such as the Data Management Initiative. Subsequent reports will focus on evolutionary changes, as well as ongoing status of the parameters included in this report.

The SDU is intended to be an information resource for space research stakeholders in NASA and the science community. In addition, it provides important information to other government agencies and interested parties. It should help ensure the ability of OSSA to better serve the science community, and to meet the increasingly demanding data management needs of future space science and applications missions.

Joe Bredekamp

Chief
Information Systems Branch

EXECUTIVE SUMMARY

Space science data is a critical national asset. From understanding the origins of our solar system and the universe to predicting global environmental change, space science data holds the keys to our past and future. Collecting data and transforming it into knowledge and discovery are fundamental activities of the space science community. In addition, enabling activities in data storage, transmission, and scientific computing are critical to the success of space science endeavors. It is the objective of these supporting activities to ensure the availability and accessibility of data to a broad array of users, and to ensure the preservation of data through a proper flow into permanent archives.

Recent years have seen a transition in the nature of space science research. Operationally, these changes include an increase in the number of multi-agency efforts and international collaborations with correlative science, increased use of data beyond the original experiments, greater diversity in the style of research and modes of operation, and a continuity from project operations to post-mission research. Space science missions are becoming much more data intensive than in the past. Spacecraft are increasing in capability and carrying more instruments. These instruments are increasing in complexity with higher and higher data rates. Mission lifetimes are increasing, with some expected to last 15 years or longer. Annual data volume generated by these missions is rising explosively, from 0.5 terabits in 1989 to an expected volume of more than 2,500 terabytes by the late 1990's, as shown below in Figure 1.

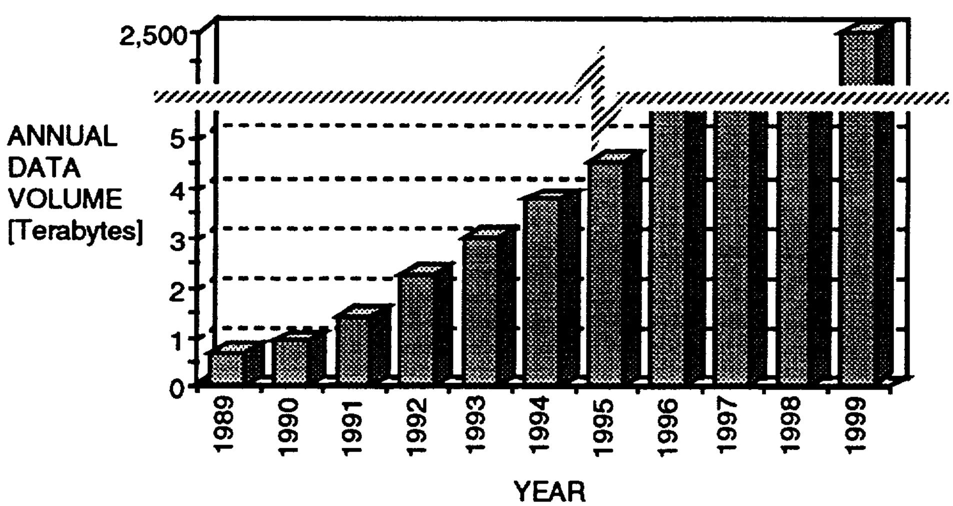


Figure 1 *Projections for annual OSSA mission data volume.*

Users of OSSA science data are growing in sophistication. Their expectations about the quality of data provided and degree of data management and analysis services available are increasing. Technological advances are also changing space science research. Wide area information servers and services enable distributed computing through client/server relationships, and enhance distributed collaboration as networking capabilities and capacities increase. Advanced workstations provide powerful local capability, which can access centralized supercomputing resources as required.

These changes have led to new requirements and demands on the data management system. The data and information systems which support space science research must experience a rise in capability commensurate with ever increasing requirements. Greater integration and coordination of data management efforts is required. This includes better interoperability between data management systems, through standards and guidelines, and integrated policies and architectures. Sharing of infrastructure elements and services will leverage limited resources through economies of scale, to focus limited resources on conducting science. Long range planning for data management must balance needs with available resources to ensure the development of the appropriate technologies in a rapidly evolving environment.

The data management environment of OSSA involves many stakeholders, who in conjunction with the data comprise the OSSA Data Union. Central to the data union are scientists who transform data into scientific knowledge and discovery. Primary users develop experiments and generally have initial rights to data generated by their experiments, whereas secondary users or retrospective investigators perform continuing analysis of science data. Non-science users such as journalists are also considered. OSSA data systems maximize the efficient use of resources through a hybrid mix of distributed and centralized elements. Requirements unique to discipline research are supported by a discipline data system (DDS) for each of the six science divisions. Data system elements are also provided as required for each of a number of projects within each division. Elements which support the needs of all divisions are centralized and provided as OSSA institutional support in the areas of data management and archiving, scientific computing, networking, and applied research & technology. These centralized support areas consolidate requirements and provide economies of scale for common OSSA data management needs. Underlying all of the stakeholders in the data union is the data itself, including science data, mission data and ancillary data necessary for calibration or analysis of science data. Figure 2 illustrates all of these data union elements.

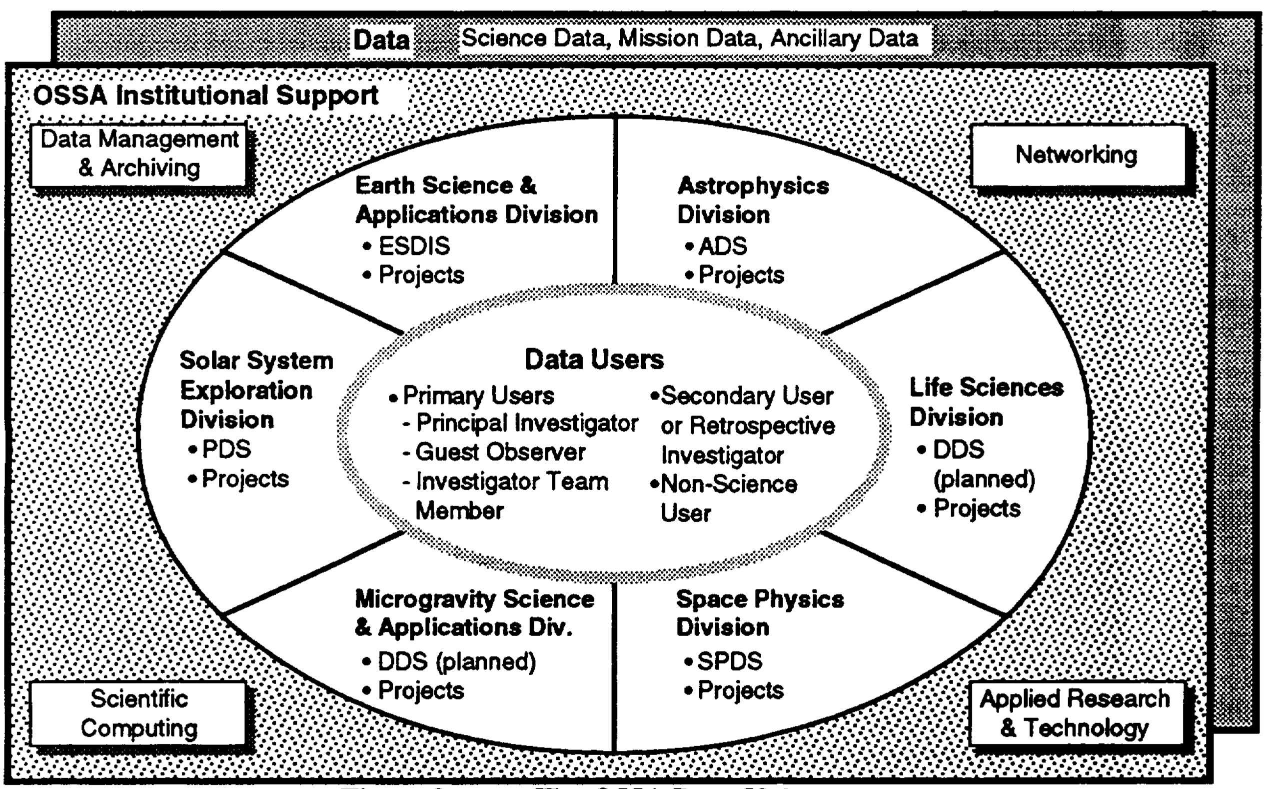


Figure 2 *The OSSA Data Union*

There have been many studies and reports in recent years that have played important roles in addressing these changes, and shaping the data union. From the CODMAC reports in the 1980's to the more recent GAO reports, and ongoing initiatives in Data Management and High Performance Computing, attention is being focused on meeting increasingly demanding user needs and ensuring preservation of critical data within limited budgets. NASA's Information Systems

Strategic Planning Project (ISSP) was a year long effort culminating in a strategy for meeting the evolutionary needs of space science research. Finally, a new OSSA data management policy directive was approved in March of 1992.

The new policy stresses the importance of data management planning, addressing the complete life cycle of data from initial collection and processing through permanent archiving. It stresses the continued trend to decentralize where appropriate through discipline based data management systems. In order to ensure appropriate input and involvement by the research community, the policy recommends establishing review processes throughout the data life cycle. In addition, the new policy emphasizes the need for continuous infrastructure and technology enhancements.

In the drive to decentralize many data management functions, OSSA's approach to data access and storage has evolved towards an environment in which:

(1) Science investigators initially access data from the project data repository.
(2) Data then flows from the project data repository to a discipline data archive for broader access by the scientific community.
(3) Then the data typically goes to the NSSDC for permanent retention and ongoing access.
(4) The NASA Master Directory provides cross cutting information regarding identification and location for all data of OSSA interest.

In this context, the combination of DDS's and NSSDC make up the OSSA archive environment. This distributed architecture enables the DDS's to enhance intra-discipline research, and promote increased researcher participation through improved accessibility to discipline science data. Additionally, initiatives in the areas of data format standards, network interconnectivity, user access services, etc. will enhance inter-disciplinary research and further promote increased researcher participation.

The NASA/OSSA Data Management Initiative (DMI) was initiated in 1991 as an integration and extension of previously related activities. It is a multi-year, multi-million dollar effort that will ensure archiving of appropriate data from past missions, and creation of an infrastructure to enable the orderly archiving of data from future missions. The program will also ensure that data are preserved, inventoried and documented to facilitate broad future access by the science community. In support of these goals, there are three principal activities associated with the DMI, including:

1) Identification and community assessment and prioritization of data sets in need of "restoration"
2) Restoration and/or archiving of appropriate data sets
3) Creation/Improvement of the capabilities and capacities of the Discipline Data Systems (DDS) and of the NSSDC, and of the procedures and tools whereby those entities assure the routine flow of increasing volumes of the right data into the OSSA archive environment (and retrievability of there from that environment).

There are six program Divisions within OSSA, each of which emphasizes and applies a different scientific discipline to successfully accomplish the goals of OSSA. These goals are pursued through an integrated program of ground-based laboratory research and experimentation; suborbital flight of instruments on airplanes, balloons, and sounding rockets; flight of instruments and the conduct of life sciences and microgravity research on the Shuttle/Spacelab system and on Space Station Freedom; and development and flight of automated Earth-orbiting and interplanetary spacecraft. The number of space missions in support of OSSA research has been increasing along with the amount of data being generated. Figure 3 indicates the number of space missions to be operational per year for each discipline division through the year 2000. Only those missions currently operational or under development are included in this projection.

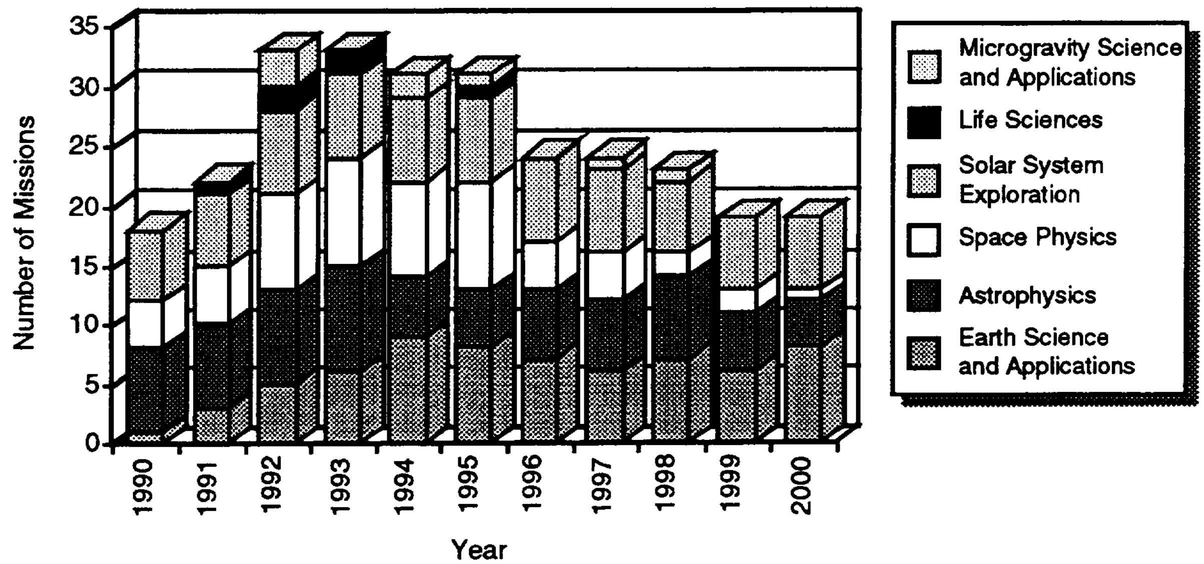


Figure 3 *Projections for annual OSSA operational missions.*

As stated in the roles and responsibilities of the new data policy directive, a long term goal for each of these disciplines is to have an integrated discipline data system (DDS) which will support the data management activities of the division. Some of the divisions have currently operational data systems, while others are beginning to plan their DDS. A summary of divisions and their respective DDS is shown in Table 1 below.

Table 1 *OSSA Discipline Data Systems and status.*

Code	Division	Discipline Data System	Status
SB	Life Sciences Division	TBD	Under Definition
SE	Earth Science & Applications Division	Earth Science Data & Information System (ESDIS)	Operational
SL	Solar System Exploration Division	Planetary Data System (PDS)	Operational
SN	Microgravity Science & Applications Division	TBD	Under Definition
SS	Space Physics Division	Space Physics Data System (SPDS)	Under Development
SZ	Astrophysics Division	Astrophysics Data System (ADS)	Operational

Each discipline is developing a data management strategy guided by the advice and counsel of their respective science communities. These strategies will define their data management environment, and address standard approaches to use of discipline resources as well as institutional resources such as networking, computing and archiving. Data management and archiving issues will be afforded appropriate emphasis and priority from the onset of mission planning. Projects will address these issues and document them in Project Data Management Plans, which will be reviewed by discipline divisions as part of the new start approval process. The status of PDMPs for each of the disciplines is shown below, in Figure 4. The figure distinguishes between projects in Phase C/D and operational projects, although PDMPs are required for all these projects.

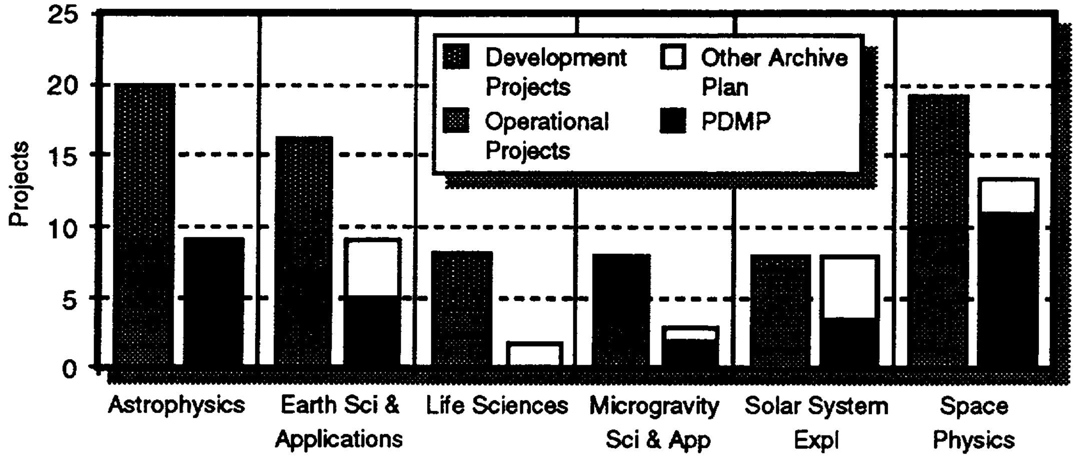


Figure 4 *Current Data Management Planning Status by Discipline*

INTRODUCTION

As OSSA moves into the data-intensive era of the 1990s, timely and responsive data and information systems support increases in importance as a crucial element of overall success in achieving science mission objectives. Data volume alone represents a significant challenge, with the flow of science data into archives expected to increase by several orders of magnitude over the next decade. Furthermore, new trends in the character of space research will drive the evolution of data and information systems. Broad scientific questions to be addressed will be increasingly multidisciplinary in nature, will involve widely dispersed investigator teams, and will require the combination and analysis of data from many different sources. The importance of data products will extend well beyond a particular flight mission, and researchers will increasingly use data sets to address scientific questions and study phenomena not anticipated during initial mission planning.

These changes lead to new requirements on data management systems requiring greater integration and coordination. This includes better interoperability between data management systems through standards and guidelines, and integrated policies and architectures. Sharing of infrastructure elements and services leverages limited resources to focus more on conducting science. Finally, there is the continuing challenge of selecting and applying the appropriate technologies in a rapidly evolving environment. There have been many studies and reports in recent years that have played important roles in addressing these changes, and shaping the data union as shown in Figure 5.

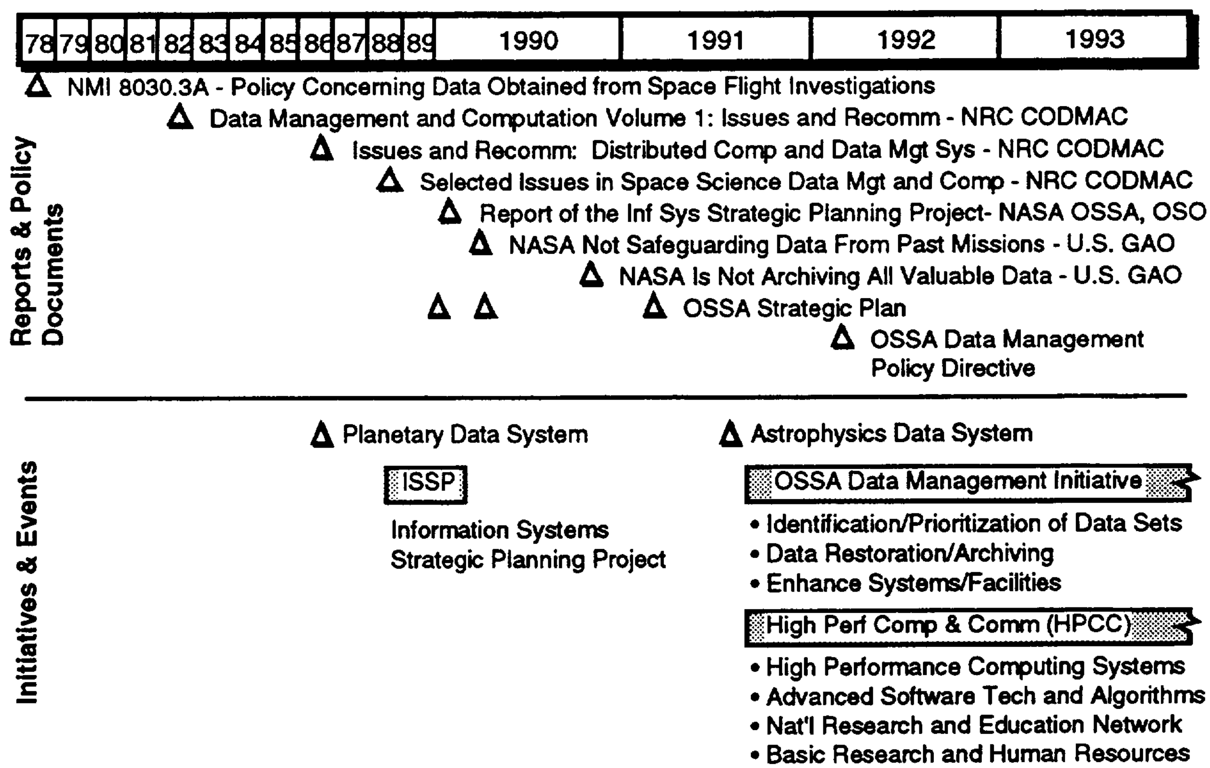


Figure 5 *Events and Initiatives shaping the Data Union*

From the CODMAC reports in the 1980's to the more recent GAO reports, and ongoing initiatives in Data Management and High Performance Computing, attention is being focused on meeting increasingly demanding user needs and ensuring preservation of critical data within limited budgets. NASA's Information Systems Strategic Planning Project (ISSP) was a year long effort culminating in a strategy for meeting the evolutionary needs of space science research. Finally, a new OSSA data management policy directive was approved in March of 1992.

The new policy stresses the importance of data management planning, addressing the complete life cycle of data from initial collection and processing through permanent archiving. It stresses the continued trend to decentralize where appropriate through discipline based data management systems. In order to ensure appropriate input and involvement by the research community, the policy recommends establishing review processes throughout the data life cycle. In addition, the new policy emphasizes the need for continuous infrastructure and technology enhancements.

Space science investigations are complex in nature, involving the interaction of many elements of the data union. Though each mission is unique, there are fundamental elements, functions, and services that are common across space science missions. In general, space science investigations can be divided into three segments of operation: mission operations, science operations, and continuing research. During each segment of the project, fundamental functions and services are performed as part of the integrated investigation. Those functions and services performed during the mission and science operations phases are generally project specific, whereas those performed during continuing research are not. A diagram depicting the OSSA architecture for Mission Operations & Data Analysis (MO&DA) is shown in Figure 6.

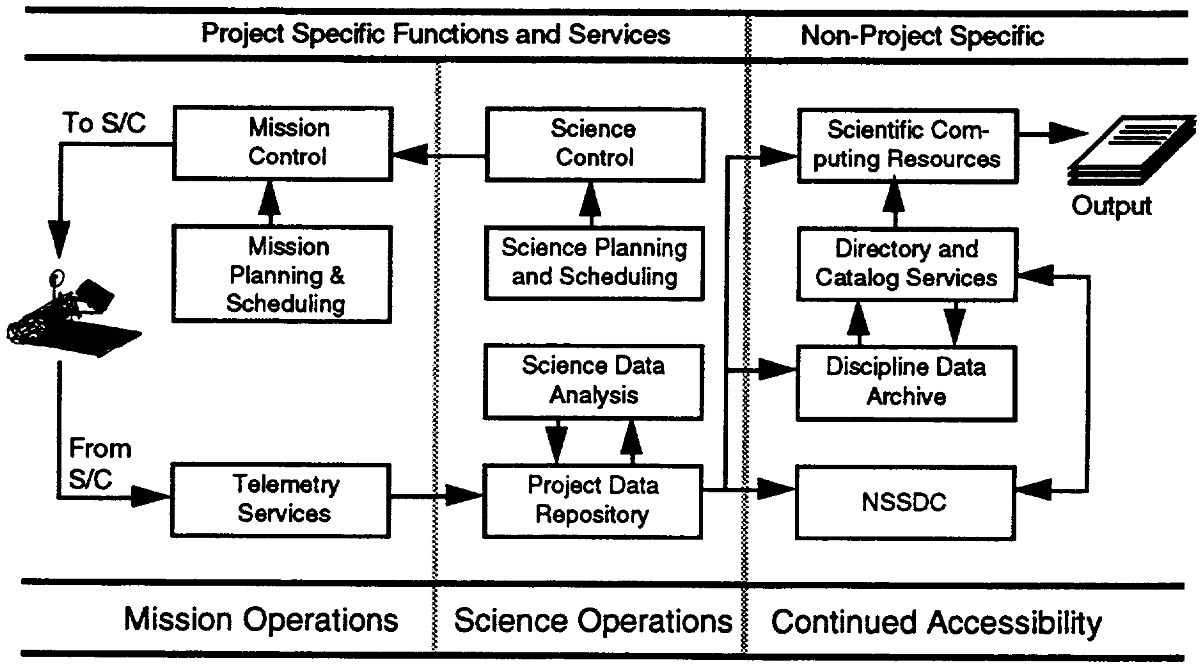


Figure 6 *OSSA MO&DA generalized information flow.*

Mission operations involve the safe and efficient operation of the spacecraft and associated payloads during the active flight portion of the investigation. The principal functions and services associated with mission operations include telemetry services, mission planning and scheduling, and mission control.

Science operations involve the functions and services required to ensure the production of valuable science data or samples during the active flight portion of the investigation. Principal functions and services provided as science operations include science planning and scheduling, science control, project data archive, and science data analysis.

The third segment of the integrated architecture is continuing accessibility. This is the continued derivation and dissemination of useful science knowledge and insight resulting from the data collected during mission and science operations. This segment of operations can go on indefinitely beyond the mission life, performing retrospective analysis and correlative studies. Permanent archive of the original data is critical for this subsequent analysis that may improve results by the use of improved calibration algorithms, or support new investigations that were not envisioned or planned for when the data was collected. The functions and services provided during continuing research include directory and catalog services, scientific computing resources, discipline data archives, and other archives and databases. OSSA's approach to data management has evolved towards a system in which:

(1) Science investigators initially access data from the project data repository.
(2) Data then flows from the project data repository to a discipline data archive for broader access by the scientific community.
(3) Then the data typically goes to the NSSDC for permanent retention and ongoing access.
(4) The NASA Master Directory provides cross-discipline information regarding identification and location for all data of OSSA interest.

In this context, the combination of DDS's and NSSDC make up the OSSA archive strategy. This development within NASA now provides a framework for improved community participation in mission data and information access. This trend facilitates a distributed data management architecture within OSSA where the DDS's enhance intra-discipline research, and promote increased researcher participation through improved accessibility to discipline science data. Additionally, continuing activities in the areas of data format standards, network interconnectivity, user access services, etc. will enhance inter-disciplinary research and further promote increased researcher participation.

As data progresses throughout its life cycle, the involvement of various elements of the data union varies. This progression may be characterized by many phases including development, operations, various stages of research, storage and archive. A representative illustration of the progressing phases of data activity, overlaid against the functions performed on the data and the users of the data, is shown in Figure 7.

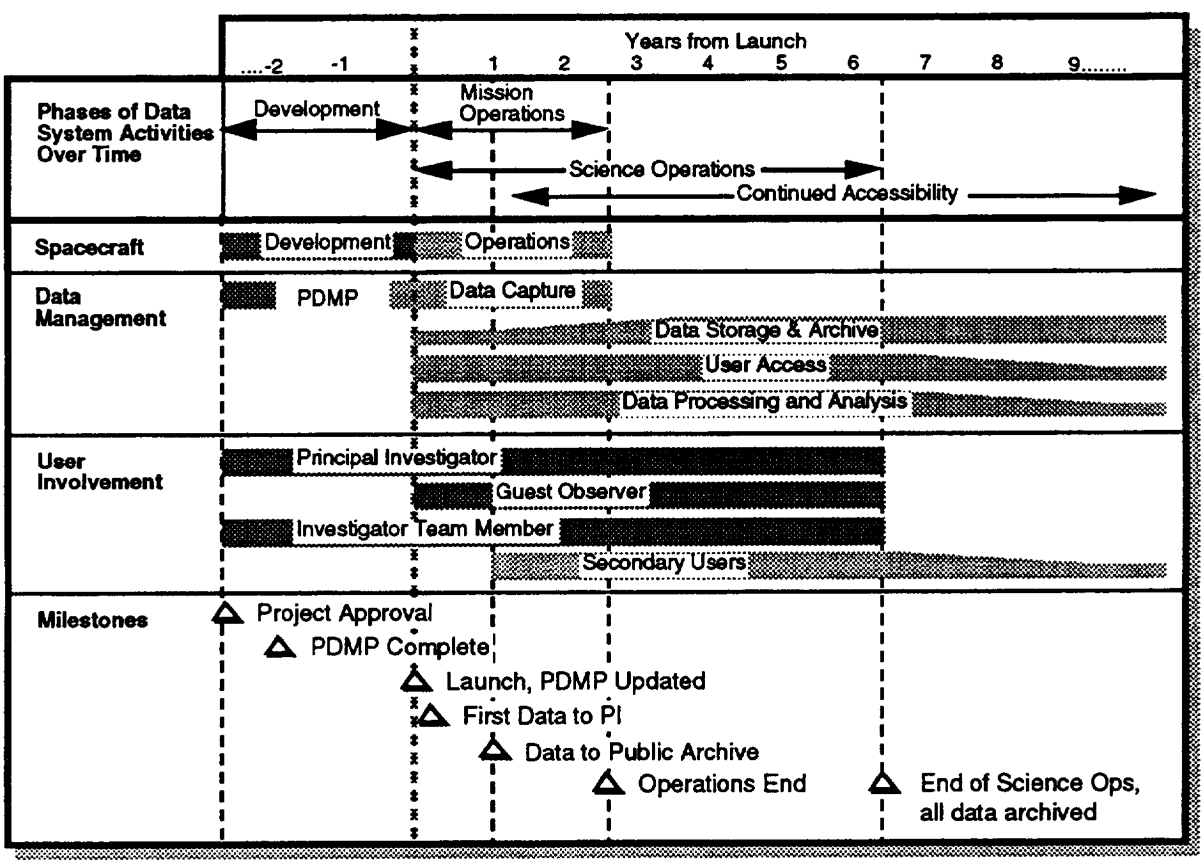


Figure 7 *Data Management over time*

USERS

The purpose of developing science missions and collecting data from observations is so that scientists can use the data to gain insight and make new scientific discoveries. There are a number of users of OSSA scientific data, each with different requirements and needs. Principal categories of usage are:

- Primary User - Science investigators who plan and design the experiments, and have an immediate need for access to the data being generated. They have initial access to the data.
  - Principal Investigator - Often works with co-investigators, responsible for planning, development, and integration of experiments and instruments, data analysis, and the selection and preparation of the analyzed data for archiving. Principal Investigator is usually tied to a particular instrument.
  - Guest Observer - Has access to observation time, to generate specific space science data to conduct independent investigations, although seldom participates in initial mission planning or instrument design.
  - Investigator Team Member -- Member of an investigator team who shares data acquired by the instruments of the team.
- Secondary User or Retrospective Investigator - Members of the general science community, could include discipline peers, or interdisciplinary scientists. Usually conduct their analysis using data that has been archived, as well as data provided or published by the PI. Secondary users also work in collaboration with primary users.
- Non-Science User - Includes general public, Public Affairs/Outreach or curious individuals seeking data for information purposes rather than further scientific investigation.

OSSA DISCIPLINE DIVISIONS

OSSA has the responsibility for using the unique environment of space to conduct scientific study of the universe, to understand how the Earth works as an integrated system, to solve practical problems on Earth, and to provide the scientific foundations for expanding human presence beyond earth. Missions to perform these activities are conducted by the six science discipline Divisions within OSSA. As stated in the roles and responsibilities of the new data policy directive, a long term goal for each of these disciplines is to have an integrated DDS which will support the data management activities of the division. Some of the divisions have operational data systems, while others are beginning to plan their DDS. Each discipline is developing a data management strategy that will define their data management environment, and address standard approaches to use of discipline resources as well as institutional resources such as networking, computing and archiving.

Each of the OSSA discipline divisions will be discussed in the following sections. The composition and status of each DDS will be described, in addition to ongoing missions and those under development.

ASTROPHYSICS DIVISION

The Astrophysics Division uses space missions in Earth orbit and, perhaps someday on the moon, to observe the universe and develop physical models of the phenomena observed. The program has been implemented in close coordination with the astronomical community, especially through the cognizant committees of the National Academy of Sciences. Themes of the Astrophysics program are three-fold: Cosmology, the study of the origins, structure and eventual fate of the universe; Astronomy, research into the origin and evolution of galaxies, stars, planets and life; and Physics, studies to understand the physics of matter under the extreme conditions found in astrophysical objects.

To address these themes, a strategy of observations across the electromagnetic spectrum has been developed including four Great Observatories. The Hubble Space Telescope (HST) and the Compton Gamma Ray Observatory (GRO) are currently operational. The Advanced X-ray Astrophysics Facility (AXAF) and the Space Infrared Telescope Facility (SIRTF) are planned for implementation during the 1990s. Additional astrophysics data is generated by other division-funded free flyer and shuttle-attached instruments, from astrophysics instruments flown on past missions and funded by the division or predecessors, and from ground based research.

In March of 1988, NASA released the report of the Astrophysics Data System Study addressing the data management needs of the astrophysics science community. This report recommended that the data sets and those most knowledgeable about them remain in the same physical location. It suggested that a change needed to come in the form of a link between these multiple locations and their data sets. The study recommended that in the future these locations be linked via high speed communications networks, and that the various data sets should be accessible through a common set of tools. The results of the study were used to develop the Astrophysics Data System.

The Astrophysics mission operations and data analysis (MO&DA) programs are in a period of transition resulting from changes in operations including incorporation of extensive guest observer (GO) programs and extended mission time scales. These changes have prompted a review and revision of division policies for MO&DA which are being compiled in a Flight Program MO&DA Policies and Guidelines document. This document will help investigators improve their activities in the preparation, implementation and management of Astrophysics missions and the collection, reduction, analysis, dissemination, and archiving of mission science data.

Astrophysics Data System

The Astrophysics Data System (ADS) was conceived to provide the scientific community with an efficient and effective means to access NASA's Astrophysics data holdings. User requirements and design concepts for the ADS were developed during a series of workshops in 1987 which resulted in the March 1988 report of the Astrophysics Data System Study. Beginning development in 1989, the ADS became operational in the second quarter of 1991, consisting of the following elements:

a) Operational sites providing authentication, routing, and other project services
b) A set of host nodes (suppliers of data and/or services) along with the databases, data archives, catalogs, and other services these sites provide
c) A distributed set of users.

The project adopted a distributed database system that could be accessed via NSI from the user's home institution. The ADS currently includes eight physically distributed nodes that are interconnected via NSI as depicted in Figure 8. Additional nodes are planned at the Center for EUV Astrophysics in Berkeley, CA providing information in extreme ultraviolet astronomy, and at the GRO Science Center in Greenbelt, MD providing information in gamma ray astronomy. Most currently important data sets are held at sites responsible for their source missions and/or instruments. Many data sets from early missions are held at NSSDC, as well as duplicate data sets from selected current missions.

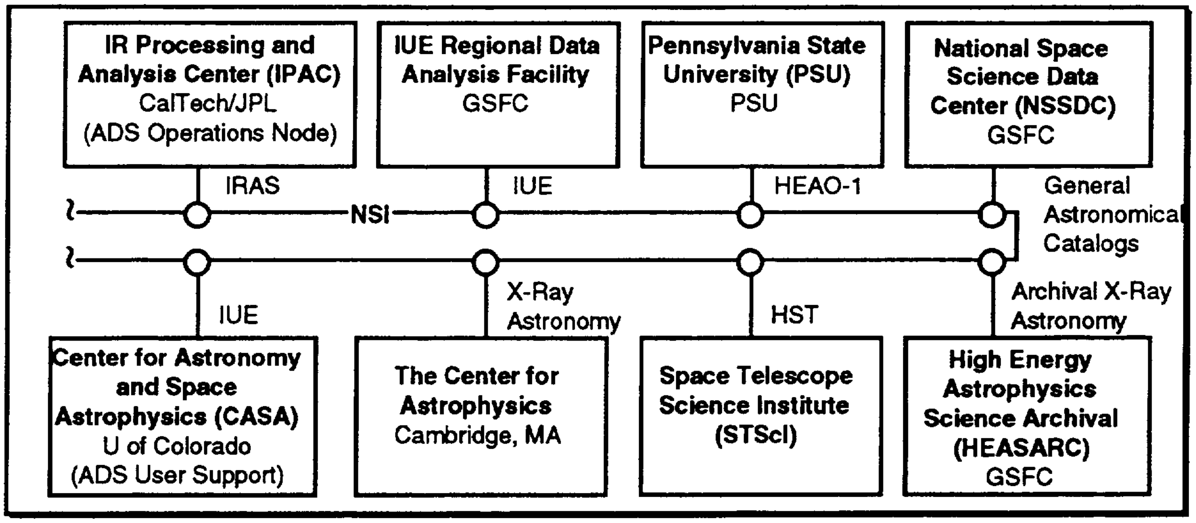


Figure 8 *The ADS is a distributed system.*

The distributed database system allows the scientists to simply access data and software tools at any of the NASA Astrophysics Science Operations and Data Centers, employing a client-server mode of operation. Key features of the ADS include the following:

- Free use to qualified researchers
- Interaction with all NASA Astrophysics data and operations centers
- Single, integrated, uniform user interface
- Multispectral, multidisciplinary data system
- Multiple platform support (e.g., UNIX, PC)
- Internet connectivity
- Access to textual information, astronomical databases, data archives, observation logs and plans.

As a precursor to equivalent facilities for other wavelength bands, the High Energy Astrophysics Science Archival Research Center (HEASARC) was established at GSFC to manage data from x-ray and gamma ray missions. The HEASARC is building a cadre of scientists to help create an archive of the right data, documentation, ancillary data, software, etc., and to be able to support researchers' findings, access, and use of the archived data. The HEASARC uses NSSDC data management expertise and hardware and software systems to the maximum extent feasible so that it can focus on scientific data and scientific user issues.

The Space Telescope Science Institute Facility (STScI) is located at Johns Hopkins University in Baltimore, Maryland. Responsibilities include obtaining, reviewing, and prioritizing observation proposals. In addition, it provides long-range science planning, participates in real time science operations, and performs science data calibration, data analysis, science instrument trend analysis, science data archiving, and data distribution for the selected observers.

The Infrared Processing and Analysis Center (IPAC) facility is jointly operated by the California Institute of Technology (CalTech) and the Jet Propulsion Laboratory (JPL). This facility is the ADS operations node. It is responsible for supporting the astrophysics researcher through archiving, reducing and analyzing the IRAS data sets and integrating their data processing tools. Additional responsibilities include: project management, system operations, system integration oversight, overall systems engineering, participate in software integration, and direct user support.

The Pennsylvania State University (PSU), Dept. of Astronomy and Astrophysics has taken responsibility for providing access to and for maintaining the HEAO-1 data sets. PSU is responsible for the organization and coordination of the ADS User's Group.

The Center for Astronomy and Space Astrophysics (CASA) is located in Boulder, Colorado. CASA is responsible for the following: Integrated System Test and Quality Assurance, User Documentation Integration and Validation, and Training Material Generation and Integration.

The next step in ADS development will be to provide access to data archives as well as databases. The ADS will allow users to transfer data to their own data processing facilities in standards formats such as the flexible image transport system (FITS). Additionally, the ADS plans to offer users information searching and retrieval services. Such services may include a searchable archive of abstracts from astronomy and astrophysical literature, on-line access to documentation on NASA missions, mission operations centers, and data products, and complete descriptions of data processing algorithms used to create data products.

Astrophysics Projects

Table 2 summarizes the Operational and Planned (Phase C/D) projects of the Astrophysics Division. The table provides status of the Program Data Management Plan (PDMP) for each project, and indicates whether other archive plans are in place and the principal planned archive location.

Table 2 *Astrophysics projects and status.*

Project Name	Launch Date	Mission Life Nom/Pot'l	PDMP Status	Archive Location	Comments
Operational					
ASTRO-1	12/2/90	10 days	Signed PDMP	NSSDC	ASTRO Series
COBE	11/18/89	2 yr / 4 yr	Signed PDMP	NSSDC	
EUVE	6/4/92	2 yr / 4 yr	Signed PDMP	SDSF (ADS), NSSDC	300 GB planned
GRO	4/5/91	12 yr	Signed PDMP	NSSDC	300 GB planned
HEAO-2	11/13/78		No PDMP	SAO, NSSDC	
HST	4/24/90	15 yr	Signed PDMP	STScI, NSSDC	
IRAS	1/25/83		Signed PDMP	NSSDC	
IUE	1/26/78	15 yr	No PDMP	NSSDC	
ROSAT	6/1/90	2.5 yr	Signed PDMP	ROSAT Data Archive Facility, NSSDC	375 GB planned
Development *Phase C/D*					
ASTRO-2	9/94	10 days	Signed PDMP	NSSDC	Astro Series 40 GB planned
ASTRO-D	2/6/93	2 yr	Draft PDMP	NSSDC	300 GB planned
AXAF	3/99	15 yr	No PDMP		Prelim SIRD
DXS	1/13/93	6 days	No PDMP		Hitchhiker Program
IEH/SHUTTLE			No PDMP		
Radioastron	1996	3 yr	No PDMP		
Spartan 204/SHUTTLE			No PDMP		
SPECTRUM-X	1995	5 yr	No PDMP		
SWAS	6/95	3 yr	No PDMP	SAO	Element of SMEX, Operations Concept Document (OCD) Available
XMM	1998		No PDMP		
XTE	4/96	2 yr	No PDMP		Explorer Program, SIRD available

SOLAR SYSTEM EXPLORATION DIVISION

The fundamental goals and approaches of the Solar System Exploration Division focus on scientific research in the following areas:

Origin and Evolution: To determine the present nature of the solar system, its planets, moons, and primitive bodies, and to search for other planetary systems in various stages of formation, in order to understand how the solar system and its objects formed, evolved, and (in at least one case) produced environments that could sustain life.

Comparative Planetology: To better understand the planet Earth by determining the general processes that govern all planetary development and by understanding why the "terrestrial" planets of the solar system are so different from each other.

Pathfinders to Space: To establish the scientific and technical data base required for undertaking major human endeavors in space, including the survey of near-Earth resources and the characterization of planetary surfaces.

These goals are consistent with recommendations by the Committee on Planetary and Lunar Exploration of the National Academy of Sciences and the Solar System Exploration Committee of the NASA Advisory Council. Solar system exploration is conducted in three distinct stages:

(1) Reconnaissance, involving flyby missions

(2) Exploration, generally conducted with orbiting spacecraft and atmospheric probes

(3) Intensive study, involving soft landers, sample returns, and human exploration.

Over the past three decades, the reconnaissance phase (initial robotic mission flybys) of solar system exploration was completed, with the exception of the Pluto-Charon system. A more capable robotic exploration phase has been underway for several years for the Moon and Mars with the Surveyor and Viking missions, respectively. Finally, an intensive study phase of the moon was initiated during the Apollo era. In the coming decades, efforts will include missions to both the outer and inner planets, as well as to the small bodies (e.g., asteroids) of the solar system. Also, in preparation for missions with humans, both the Moon and Mars will be studied extensively by robotic spacecraft, either on the surfaces or from low orbits.

In response to the needs for broad access to planetary data by the research community, the Solar System Exploration Division has established the Planetary Data System (PDS).

Planetary Data System

The goal of the PDS is to provide the best planetary data to the most users forever. It provides cost-effective archiving and access to high quality planetary science data sets, expert scientific help to the community in using the data, and a mechanism to develop the technologies needed to support such a scientific information system. In addition, the PDS has developed a generalized science catalog, nomenclature and data standards, data distribution technology, and distributed software tools to assist in the archive process.

The Planetary Data System (PDS) is based upon a widely distributed, electronically connected architecture that includes seven discipline nodes and a central node. The Central Node located at the Jet Propulsion Laboratory (JPL) in Pasadena, California provides overall project management, coordinates and distributes data standards, evaluates emerging technologies and provides an interface to planetary missions. The discipline nodes are located at institutions around the country, each having an expertise in a particular area of planetary science. A schematic of the PDS is shown in Figure 9.

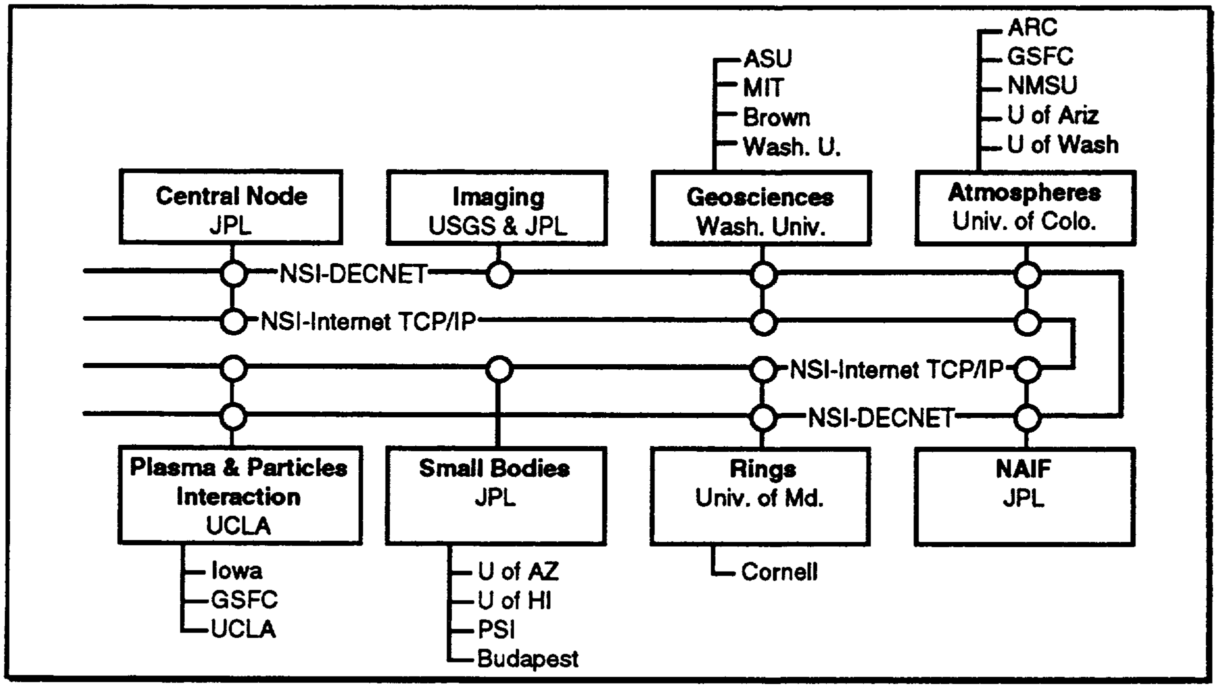


Figure 9 *PDS Architecture.*

The NSSDC supports the PDS by providing a safe and separate deep archive for the long term backup of planetary data on magnetic tapes. PDS nodes distribute products at no charge to users electronically, on tape and on CD-ROM. Priority is given to users funded by the Solar System Exploration Division. The NSSDC currently provides distribution to other requesters at a nominal charge.

PDS works with flight projects during all phases of a project to help it plan, design and implement a quality archive. During the planning phase several archive planning documents are prepared by the mission and reviewed by the PDS. These documents include the PDMP, the Archive Policy and Data Transfer Plan, and the Science Data Management Plan. PDS helps projects design their archive products and coordinate the validation or peer review of data products prior to their submittal to archive. Currently the majority of PDS archive products are published and distributed on CD-ROM.

PDS maintains catalogs at both the Central Node and Discipline Nodes that describe the data sets and data products available for distribution. PDS also provides information about the archive data products to the NSSDC which is used to populate the Master Directory.

The Navigation Ancillary Information Facility (NAIF) at JPL provides the planetary science community with datasets and transportable software tools appropriate for computing, archiving, accessing and distributing the ancillary observation geometry data needed for full interpretation of science instrument observations of solar system bodies. The ancillary data and allied software are distributed to flight project science teams while a mission is in its active stage, and are subsequently archived within the PDS. NAIF also provides ephemeris products and allied software to support other OSSA discipline division activities such as HST, Galileo, etc.

The Geosciences node at Washington University is responsible for working with planetary missions to help ensure that the data of relevance to the geosciences disciplines are properly documented and archived in the PDS. The node also restores and publishes selected data sets from past missions on CD-ROMs for delivery to the planetary science community. Additional services include information and expert assistance on its data holdings. Archivals include derived image data, geophysics data, microwave data, spaceborne thermal data and spectroscopy data.

The Plasma and Particles Interaction node at UCLA is responsible for the acquisition, preservation and distribution of fields and particle data from all planetary missions (excluding Earth observations). Services provided to the research community include a menu-based interactive on-line interface for selection, viewing, and ordering of data sets, limited ability to build customized data sets, and indirect access to science consultants.

The Small Bodies node at JPL, along with its subnodes have expertise on comets, asteroids, and interplanetary dust. There are a number of on-line tools available including access to NAIF software, and epheride calculation of comets or asteroids.

The Planetary Rings node is responsible for restoring, archiving, and publishing data sets describing planetary ring systems. This includes all of the relevant data sets acquired from previous and future spacecraft, as well as from Earth-based observatories. In order to support research with these data sets, the node also provides expert assistance, catalog information to help scientists pinpoint the most relevant data, and a suite of software tools to simplify data manipulation.

The Planetary Atmospheres node is a consortium of research groups led by the University of Colorado, with five subnodes. Researchers associated with the node curate data sets within their sub-disciplines and provide a wealth of research expertise to space scientists and other users of the PDS.

Solar System Exploration Projects

Table 3 summarizes the Operational and Planned (Phase C/D) projects of the Solar Systems Exploration Division. The table provides status of the Program Data Management Plan (PDMP) for each project, and indicates whether other archive plans are in place and the principal planned archive location.

Table 3 *Solar System Exploration projects and PDMP status.*

Project Name	Launch Date	Mission Life Nom/Pot'l	PDMP Status	Archive Location	Comments
Operational					
GALILEO	10/18/89	8 yr	Signed PDMP	PDS	
MAGELLAN	5/4/89	4 yr / 10 yr	Signed PDMP	PDS	3,070 GB planned
MARS OBSERVER	9/25/92	4 yr	Other plan in place	PDS	Science Data Mgmt. Plan; Archive Policy & Data xfer Plan in place
PIONEER 10, 11	3/3/72 4/6/73	20 yr +	Other plan in place	Original data in NSSDC, new versions in PDS	
VOYAGER 1,2	9/5/77 8/20/77	30 yr	Other plan in place	Original data in NSSDC, new versions in PDS	Archive Policy & Data Transfer Plan in Place (updated Nov. 1990)
Development *Phase C/D*					
CASSINI	10/97	13 yr	PDMP in progress	PDS	Draft Archive Policy & Data Transfer Plan

SPACE PHYSICS DIVISION

The Space Physics Division (SPD) investigates the origin, evolution, and interactions of space plasmas in the heliosphere and the cosmos. The objectives of the division, endorsed by the Committee on Solar and Space Physics of the National Academy of Sciences, are to understand the following: solar and heliosphere physics (understanding the Sun, both as a star and as the dominant source of energy, plasma, and energetic particles in the solar system); the physics of the magnetospheres and ionospheres of planets, especially earth; the upper atmosphere and solar-terrestrial coupling (understanding the interactions between the solar wind and other particles and the electromagnetic fields and atmospheres of solar system bodies). Comparative planetary studies of interactions of the solar wind with other bodies allows solar terrestrial interactions to be placed in a broader context.

The Space Physics Division is responsible for assuring the coordinated, effective, and appropriate management of data for space physics missions. Associated with these responsibilities, the Space Physics Division has identified four major information systems challenges for the near future:

1. Identifying space physics data sets from past missions; then, prioritizing and implementing the restoration and archiving of relevant data sets.
2. Developing the archives of data and information from currently operating flight projects and making them accessible to scientists.
3. Managing the highly diversified and sophisticated data and information streams from new flight projects.
4. Capturing supporting and correlative data and information and incorporating mathematical and numerical models and tools essential to the full realization of science return on all flight projects.

The Space Physics Division is proposing a Space Physics Data System (SPDS) as a logical framework within which to address these challenges and discharge its data management responsibilities.

Space Physics Data System

Data from space physics missions will be provided to the science community through the Space Physics Data System (SPDS). An SPDS Steering Committee (SPDS/SC) was established in 1990 to begin definition of the SPDS. Several meetings were held in 1990 and 1991, including input from representatives of data systems already in existence. It was concluded by the SPDS/SC that the SPDS should be a widely distributed system in which most of the resources would be in the hands of the scientists who were actively analyzing, maintaining and distributing the data. The fundamental objectives of the SPDS are as follows:

- Curate those data deemed important to the SPD mission by the space physics community
- Provide information about those data to the potential users
- Provide access to the data and ancillary information
- Enable the efficient and cost-effective analysis of data, particularly in a synergistic or correlative manner
- Act as a focus for information, evaluation and guidelines in the appropriate use of new data management technologies in the space physics community.

Initially SPDS will use existing, project-specific data systems linked to each other and the user community by the NASA telecommunications infrastructure. Participation in the initial SPDS confederation will be driven from the grass roots and will be stimulated by the ability to perform correlative analysis. The SPDS will, in time, develop an architecture with nodes defined by science discipline and should become a common framework and common resource for other data systems subsequently developed by individual institutions, laboratories, universities and missions. As the scientific productivity of SPDS is demonstrated to a broader community with cost-effective access to more datasets, it can be expected that the future institution-level data systems will develop to be more readily compatible with SPDS. SPDS will also have a key role in working with future

SPD flight projects to best assure their adequate and coordinated long-term data management planning.

The SPDS is currently being further defined by the space physics division, with support from the NSSDC. There are a number of factors being addressed in defining the full SPDS, including the following:

- The system in all phases of design, implementation, operation and management must be kept closely coupled to the science community
- The system design and implementation strategy must be able to meet the constraints of
  - The need to evolve and adapt
  - Limited funding throughout the system lifetime
  - Rapid changes in feasible and/or cost-effective technologies
- The overall system design and the implementation strategies must be orderly, well-considered, appropriately complete and both technically and operationally robust.

SPDS community workshops are planned for next year, with the goal of developing an SPDS project plan and initiating systems engineering and design for the data system. Development of SPDS design and implementation strategies will continue through FY94. A preliminary concept for the SPDS configuration is shown in Figure 10, indicating preliminary nodes. An NRA is planned for permanent SPDS nodes, with awards in late FY94 and initial funding in mid FY95.

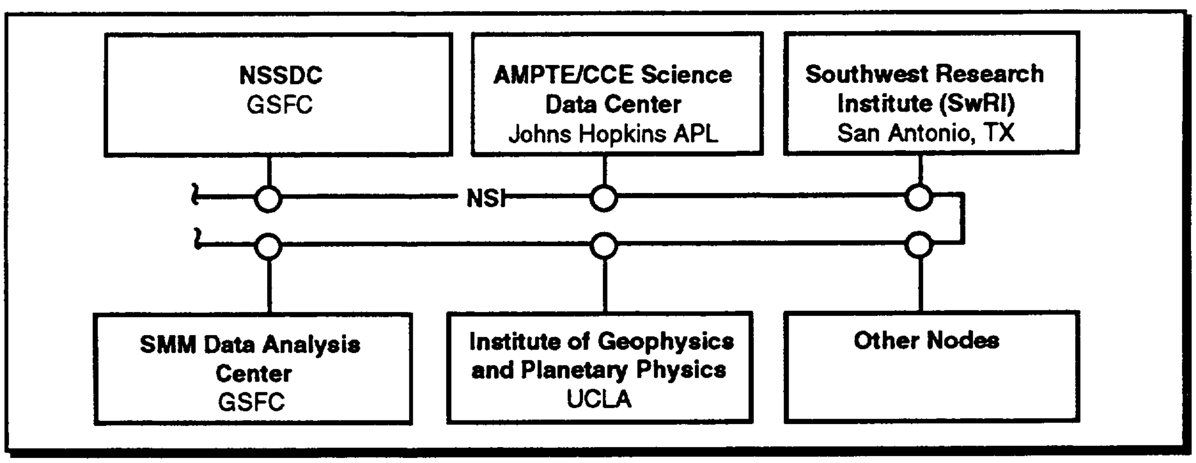


Figure 10 *A preliminary concept for the SPDS is a distributed architecture of multiple nodes.*

The SPDS is envisioned to be a network of data centers united under a common theme. Existing institutional capabilities are to be leveraged in getting SPDS started. Some of these capabilities include the NASA Master Directory, NDADS, NASA/OSSA Office of Standards and Technology (NOST), and the NSSDC.

Space Physics Projects

Table 4 summarizes the Operational and Planned (Phase C/D) projects of the Space Physics Division. The table provides status of the Program Data Management Plan (PDMP) for each project, and indicates whether other archive plans are in place and the principal planned archive location.

Table 4 *Space Physics projects and PDMP status.*

Project Name	Launch Date	Mission Life Nom/Pot'l	PDMP Status	Archive Location	Comments
Operational					
Atmospheric Balloons	Various	10-90 days each	Other plan in place	Wallops, NSBF	SIRD
CRRES	7/25/90	1 yr / 4 yr	Signed PDMP	NSSDC	Element of ISTP, Joint with DoD in conjunction with sounding rockets
Geotail	7/14/92	2 yr / 3 yr	Signed PDMP	ISTP CDHF, NSSDC	Element of ISTP/COSTR 138.7 GB EDR planned 0.5 GB KPD planned
ICE/ISEE 3	8/12/78	10 yr / 25 yr	Other plan in place	NSSDC	
IMP-8	10/25/73	19 yr	Draft PDMP, Other plan in place	NSSDC	
SAMPEX	6/19/92	3 yr	Draft PDMP	UMSOC, NSSDC	Element of SMEX 112 GB planned
Suborbital	Various		No PDMP	Wallops Flight Facility	
TSS	7/16/92	36 hr	No PDMP		
Ulysses	10/6/90	5 yr	Draft PDMP	NSSDC	
Yohkoh	8/30/91	3 yr	Signed PDMP	NSSDC	Formerly named Solar-A, 150 GB planned
Development *Phase C/D*					
ACE	6/97	1 yr / 3 yr			
CLUSTER	12/95	2 yr	Signed PDMP	NSSDC	Element of ISTP/COSTR
EHIC (TIROS-J)		3 yr	No PDMP		Secondary payload on Tiros-J
FAST	9/94	1 yr	Draft PDMP		Element of SMEX
POLAR	5/94	1.5 yr / 3 yr	Signed PDMP	NSSDC	Element of ISTP/GGS 366.8 GB EDR planned 11.1 GB KPD planned
SOHO	7/95	2 yr 5 mo	Signed PDMP	NSSDC	Element of ISTP/COSTR
SPARTAN-2 (ASP)	3/23/93	9 days	No PDMP		
WIND	8/93	2 yr / 3 yr	Signed PDMP	NSSDC	Element of ISTP/GGS 94.9 GB EDR planned 0.4 GB KPD planned
WISP			No PDMP		

EARTH SCIENCE AND APPLICATIONS DIVISION

The overarching goal of the Earth Science and Applications Division, as formulated by the Earth System Sciences Committee of the NASA Advisory Council, is to:

Obtain a scientific understanding of the entire Earth system on a global scale by describing how its component parts and their interactions have evolved, how they function, and how they may be expected to continue to evolve on all timescales.

The challenge associated with these goals is to:

Develop the capability to predict those changes that will occur in the next decade to century, both naturally and in response to human activity.

The study of terrestrial phenomena should not be limited to particular components solely in the atmosphere, in the oceans, on land, or within the biosphere. Instead, such study must be directed at understanding the responsible physical, chemical, and biological processes that operate to unify the Earth environment as a system. These processes must then be cast in the form of algorithms for assimilation into global models. Finally, these models must be tested against comprehensive, long-term global-scale data sets in order to validate their accuracy as descriptive and predictive tools.

The Earth Observing System (EOS) will carry out multidisciplinary Earth science studies employing a variety of remote sensing techniques. Its primary goal is the generation of long term Earth science data sets of measurements and processes in the areas of:

- Agriculture
- Forestry
- Geology
- Hydrology
- Oceanography
- Snow and Ice
- Troposphere and Upper Atmosphere
  - Chemistry
  - Radiation
  - Dynamics

The intent of the EOS program is to obtain data sets that pertain to global studies of the Earth as a system, emphasizing the interactions and couplings of the atmosphere-ocean-land-cryosphere system.

Earth Science Data and Information System

The goal of the Earth Science Data and Information System (ESDIS) is to achieve greater interoperability and resource sharing among NASA Earth science data systems, the Earth science data centers of other agencies, and Earth scientists themselves.

Present activities include the NASA Ocean Data System (NODS), the NASA Climate Data System (NCDS), the Pilot Land Data System (PLDS), the Crustal Dynamics Data and Information System, and the Synthetic Aperture Radar (SAR) Data Catalog System, together with the processing and archiving of data from the Coastal Zone Color Scanner, the Earth Radiation Budget Experiment, and the Total Ozone Mapping Spectrometer. Building upon this extensive experience, ESAD provides information systems prototypes for a wide variety of future Earth science missions, including the Earth Probe series of Explorer-class missions and the Earth Observing System mission. Present ESAD activities, such as the UARS and TOPEX/Poseidon flight project data system developments, form the basis for an Earth science data management infrastructure that will support data systems for future flight projects.

The goal of the NASA Ocean Data System (NODS) is to support the Oceans research community by providing interactive access to data sets, catalogs and documentation from spaceborne ocean viewing sensors. The NODS provides support in the acquisition, archiving, and distribution of physical data of the oceans. The NODS is part of the existing core of the JPL Distributed Active Archive Center (DAAC).

The NASA Climate Data System (NCDS) is an interactive scientific data management system composed of an integrated set of software tools for locating, accessing, manipulating, and displaying data from research missions. The NCDS was a part of the NSSDC, but has recently become a part of the GSFC DAAC. The NSSDC still provides a range of data and computer support to the NCDS.

The Pilot Land Data System (PLDS) is a distributed information management system designed to support NASA's land science community. The PLDS was a part of the NSSDC, but has recently become a part of the GSFC DAAC. The NSSDC still provides a range of data and computer support to the PLDS. The PLDS provides a wide range of services including:

- Management of information about scientific data
- Access to a library of scientific data
- A data ordering capability
- Communications
- CD-ROM publication
- Browse facility
- Science project support
- User assistance

The most important future system is the Earth Observing System Data and Information System (EOSDIS), which is being implemented as a part of the EOS program. EOSDIS will begin by building upon the efforts mentioned above. EOSDIS will provide the underlying infrastructure for the EOS mission by providing planning, scheduling, command, control, and monitoring services for the EOS spacecraft and instruments, and also by supporting the production, archiving, distribution, and analysis of Earth science data. EOSDIS will be a widely distributed network of spacecraft and instrument control centers, data processing facilities, and data archives. The EOS program will generate new data at the rate of about 2 terabytes per day, deriving from 50 Megabits per second of raw data from the spacecraft alone. EOSDIS will manage these data, creating a comprehensive, global, 15-year data set containing over 10 petabytes (10 to the power of 15 bytes) of data.

The development of EOSDIS will be evolutionary, starting with pathfinder data sets drawn from existing data systems. The principal components of the EOSDIS are the Distributed Active Archive Centers (DAACs) identified in Table 5 below, and Affiliated Data Centers (ADCs).

Table 5 *Distributed Active Archive Centers (DAACs) of the EOSDIS.*

Responsible Organization	Location	Data Topics
GSFC	Greenbelt, MD	Upper Atmosphere, Meteorology, Global Biosphere, Geophysics
JPL	Pasadena, CA	Ocean Circulation, Air-Sea Interaction
LaRC	Hampton, VA	Radiation Budget, Aerosols, Tropospheric Chemistry
University of Colorado, National Snow and Ice Data Center (NSIDC)	Boulder, CO	Cryosphere (non-SAR)
USGS / Earth Resources Observation System (EROS) Data Center	Sioux Falls, SD	Land Processes Imagery
University of Alaska / Fairbanks (UAF)	Fairbanks, AK	Sea Ice (SAR), Polar Processes Imagery
MSFC	Huntsville, AL	Hydrologic Cycle
Oak Ridge National Laboratory	Oak Ridge, TN	Biochemical Dynamics

An initial capability, called Version 0 EOSDIS, will be put in place in 1994. Version 0 will continue to provide the operational capabilities at the DAACs, will improve access to existing data, and will have working prototypes of integrating elements providing users with an Earth sciences view. EOSDIS will then gradually grow in its capabilities to permit acquiring, processing, archiving, and distributing the large volumes of data from the EOS instruments as they are deployed. It will also evolve into NASA's Earth Science Data and Information System for the archival and distribution of existing data and data from future non-EOS missions.

Earth Science and Applications Projects

Table 6 summarizes the Operational and Planned (Phase C/D) projects of the Earth Science and Applications Division. The table provides status of the Program Data Management Plan (PDMP) for each project, and indicates whether other archive plans are in place and the principal planned archive location.

Table 6 *Earth Science and Applications projects and PDMP status.*

Project Name	Launch Date	Mission Life Nom/Pot'l	PDMP Status	Archive Location	Comments
Operational					
ATLAS-1	3/24/92	10 days	Draft PDMP	NSSDC	included SSBUV
ERBS	10/5/84	7 yr / 11 yr	Other plan in place	NSSDC	
LAGEOS-II	9/92	10,000 yr	Other plan in place	Crustal Dynamics DIS	
San Marco-D	3/25/88	1 yr	Signed PDMP	NSSDC	
SSBUV-3	8/2/91	9 days	Other plan in place	NSSDC	
Topex/Poseidon	8/10/92	3 yr / 5 yr	Other plan in place	NODS @ JPL	Preliminary SIRD
UARS	9/12/91	3 yr	Draft PDMP	NSSDC	
Development *Phase C/D*					
ADEOS Japanese Spacecraft	2/96	3 yr	No PDMP		Earth Probe; NSCAT & TOMS
ATLAS Series	5/93, 6/94, 95, 97 - 01	10 days each mission	Draft PDMP	NSSDC	Includes SSBUV series over a solar cycle
EOS	6/98 First Launch	15 yr 3-5 yr each	Draft PDMP	EOSDIS	5 series of 3-5 spacecraft each, 10 petabytes planned
ERS-2	2/94	3 yr	No PDMP		
LANDSAT-7	12/97	5 yr	No PDMP	EOSDIS	Continuation of series
RADARSAT	12/94	3 yr	No PDMP	Alaska SAR Facility	
SeaWiFS	9/93	3 yr	No PDMP	EOSDIS	
TOMS	7/94	2 yr	No PDMP		Earth Probe
TRRM	8/97	3 yr	No PDMP	EOSDIS	Earth Probe

LIFE SCIENCES DIVISION

The Life Sciences Division is responsible for planning, directing, implementing, and evaluating that part of the overall NASA program that deals with the understanding of how living systems respond to the space environment; the search for the origin, evolution, and distribution of life in the universe; the development of the scientific and technological foundations for expanding human presence beyond Earth orbit and into the solar system; and the provision of operational medical support to all space missions involving humans.

The goals of the Life Sciences Division include ensuring the health, safety, and productivity of humans in space, and acquiring fundamental scientific knowledge concerning space biological sciences. These goals are supported by the following objectives:

1) To provide for the health and productivity of humans in space,
2) To develop an understanding of the role of gravity on living systems,
3) To expand our understanding of life in the universe, and
4) To promote the application of life sciences research to improve the quality of life on Earth.

The data management elements that will support Life Sciences activities are under study. A Life Sciences Data Archive Workshop was held in December 1990, and planning activities continue. The NSSDC will have the master directory entries from the Life Sciences data system. If cost effective, the NSSDC may archive digital or film data for the Life Sciences data system, but the NSSDC will not archive other materials.

Life Sciences Projects

Table 7 summarizes the Operational and Planned (Phase C/D) projects of the Life Sciences Division. The table provides status of the Program Data Management Plan (PDMP) for each project, and indicates whether other archive plans are in place and the principal planned archive location.

Table 7 *Life Sciences projects and PDMP status.*

Project Name	Launch Date	Mission Life Nom/Pot'l	PDMP Status	Archive Location	Comments
Operational					
COSMOS	11/92	19 days	No PDMP		Part of Russian Biosat Program
IML-1	1/22/92	8 days	Other plan in place		
Spacelab-J	9/1/92	7 days	No PDMP		PIP
SLS-1	6/6/91		Other plan in place	Archive Plan in Place	
Development *Phase C/D*					
IML-2	7/94	13 days	No PDMP		
Spacelab-D2	1/27/93	9 days	No PDMP		PIP
SLS-2	6/22/93	13 days	No PDMP		PIP
SLS-3	2Q 95		No PDMP		

MICROGRAVITY SCIENCE AND APPLICATIONS DIVISION

The mission of the NASA Microgravity Program is to obtain new knowledge and increase understanding of gravity-dependent physical phenomena and those phenomena obscured by the effects of gravity in biological, chemical and physical systems, and, where feasible, to facilitate the application of that knowledge to commercially viable products and processes. The OSSA Microgravity Science and Applications Division (MSAD) uses the unique attributes of the space environment to conduct research basic and applied research in three primary areas:

(1) Fundamental science, which includes the study of the behavior of fluids, transport phenomena, condensed matter physics, and combustion science.
(2) Materials science, which includes electronic and photonic materials, metals, alloys, glasses, and ceramics.
(3) Biotechnology, which focuses on macromolecular crystal growth and cell science.

Participants in these research areas are a diverse community of university, government and industry investigators. Many of the participants are already engaged in extensive ground-based research programs, brought together by the common need to perform experiments in the microgravity space environment in order to advance their individual areas of research.

The program uses the future capabilities of Space Station Freedom, together with free-flying platforms and Extended Duration Orbiter missions resulting in the number of MSAD investigations increasing significantly over the next decade, which will correspondingly multiply the quantity of data and other products generated, archived and subsequently accessed by the science community.

The data management elements that will support Microgravity Science and Applications activities are being addressed by the MSAD archiving study. Phase one of the study was completed in November 1992. It is expected that a distributed system will be developed, with several nodes. Node locations may include MSFC, LeRC, and NIST in Maryland. The NSSDC will have the master directory entries from the MSAD data system. If cost effective, the NSSDC may archive digital or film data for the MSAD data system, but the NSSDC will not archive other materials.

Microgravity Science and Applications Projects

Table 8 summarizes the Operational and Planned (Phase C/D) projects of the Microgravity Science and Applications Division. The table provides status of the Program Data Management Plan (PDMP) for each project, and indicates whether other archive plans are in place and the principal planned archive location.

Table 8 *Microgravity Science and Applications projects and PDMP status.*

Project Name	Launch Date	Mission Life Nom/Pot'l	PDMP Status	Archive Location	Comments
Operational					
IML-1	1/22/92	8 days	Other plan in place		
USML-1	6/92	13 days	Draft PDMP		PIP
USMP-1	10/8/92	9 days	Draft PDMP		PIP
Development *Phase C/D*					
IML-2	7/94	13 days	No PDMP		
USMP-2	1/94		No PDMP		
USMP-3	1995				
USMP-4	1997				
USMP-5	1998				

OSSA INSTITUTIONAL INFRASTRUCTURE

The effective management and utilization of rapidly growing space science data assets calls for new approaches and modifications to the infrastructure for accomplishing quality research. Ultimate success in meeting this challenge will be measured in terms of responsiveness to science user needs for convenient access both to data and to the tools and capabilities to convert data into meaningful information and use the information for improved scientific insight. The OSSA institutional infrastructure has been established to augment all discipline data systems in providing timely and responsive information systems services to the OSSA science community. The Information Systems Branch in OSSA is responsible for the institutional infrastructure, and promotes policies, standards and practices to facilitate interoperability and resource sharing. Activities within the program include the following:

- Data Management & Archiving
- Networking
- Scientific Computing
- Applied Research & Technology

DATA MANAGEMENT & ARCHIVING

The objective of data management and archiving is to contribute discipline-independent tools, facilities, and expertise to the overall OSSA effort whereby NASA mission data are best exploited for scientific gain in the near and long terms. Principal elements of the data management and archiving function include the NSSDC for OSSA-wide data archiving and dissemination, and the NASA Master Directory for finding data in a widely distributed data environment. The archive environment is in the midst of an extended transition from a centralized architecture (NSSDC only) to a distributed architecture (Discipline Data Systems plus the NSSDC). The national network of OSSA is shown in Figure 11.

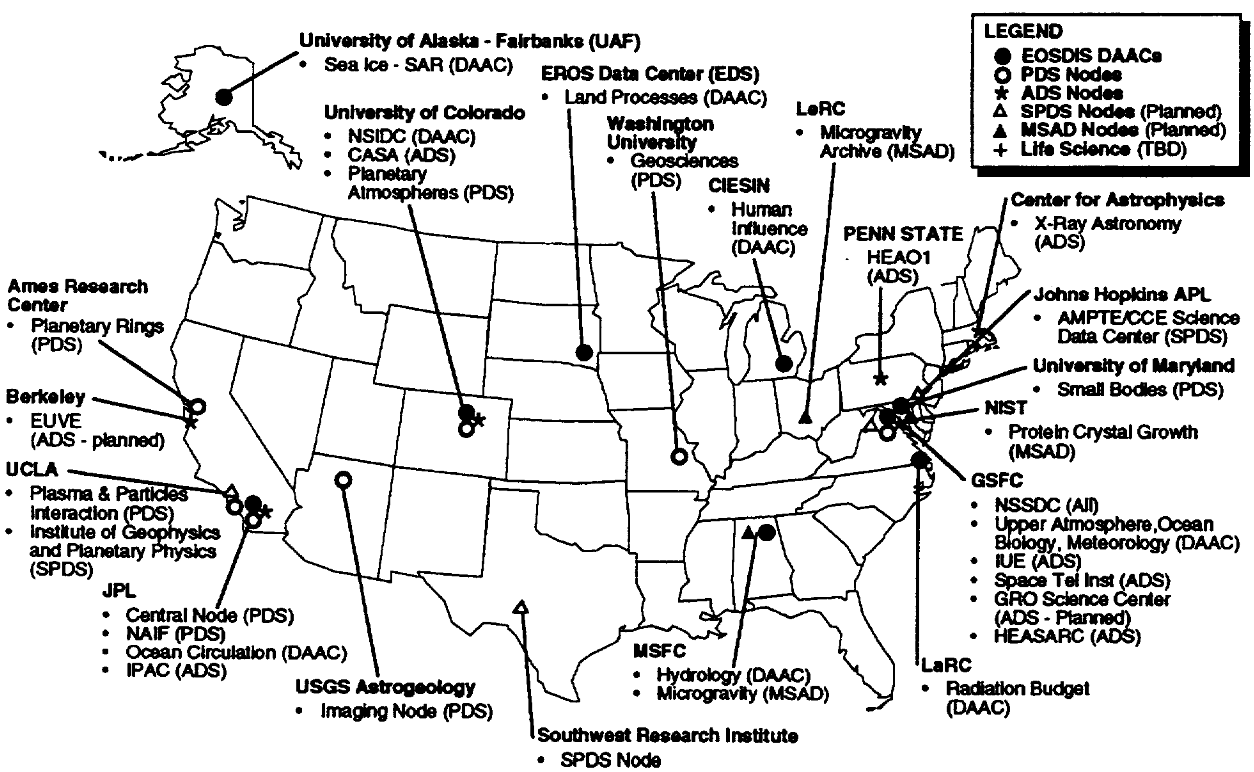


Figure 11 *The OSSA archive environment is distributed across the United States.*

The NASA/OSSA Data Management Initiative (DMI) was initiated in 1991 as an integration and extension of previously related activities. It is a multi-year, multi-million dollar effort that will ensure archiving of appropriate data from past missions, and creation of an infrastructure to enable the orderly archiving of data from future missions. The program will also ensure that data are preserved, inventoried and documented to facilitate broad future access by the science community. In support of these goals, there are three principal activities associated with the DMI, including:

1) Identification and community assessment and prioritization of data sets in need of "restoration"
2) Restoration and/or archiving of appropriate data sets
3) Creation/Improvement of the capabilities and capacities of the Discipline Data Systems (DDS) and of the NSSDC, and of the procedures and tools whereby those entities assure the routine flow of increasing volumes of the right data into the OSSA archive environment (and retrievability of there from that environment).

A complete description of the OSSA DMI can be found in Appendix A.

Effective data management is critical to the success of OSSA's science investigations, and to the near-term and long-term exploitation of the science data. Program Data Management Plans (PDMP's) are a requirement for all OSSA science investigations and should address all the data flow of a given space flight project, from early planning and scheduling through the archiving phase of a project. The essential functions of the PDMP are to:

a. Provide consistent documentation to facilitate planning and implementation of science data management needs.
b. Identify and characterize all project data sets and indicate those which require archiving.
c Specify the time, location, and format for Project data and supporting documentation to flow into the OSSA archive environment.

PDMP requirements relative to the Project Development Cycle are described in the OSSA Program Directive: Policy for the Management of the OSSA Science Data. The first version of the PDMP should be prepared in the same time as the Project Plan, shortly after new start approval for the project. It is envisioned that updates to the document will be made to reflect significant changes in data management planning throughout the period prior to launch, and throughout the mission operations and data analysis phase of the project. The formal definition (non advocate) review or equivalent mechanism conducted before a project receiving new start approval will assess data management plans as well as spacecraft development, instrument plans, etc. The requirements for PDMP development over project phases are shown in Figure 12.

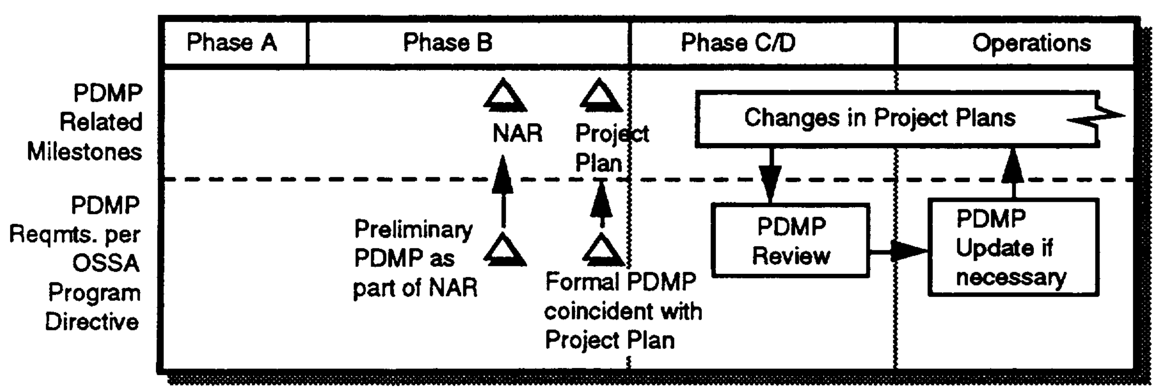


Figure 12 *PDMP Requirements.*

Current PDMP status for relevant OSSA missions, in the appropriate phase of development or operations is shown in Figure 13. From this information, it can be seen that there are currently 84 projects requiring PDMPs per OSSA policy. Of the projects requiring PDMPs, 30 projects have PDMPs while 14 projects have an alternative document that addresses archiving and data management plans. Many of these projects were in operation prior to the requirement for a PDMP.

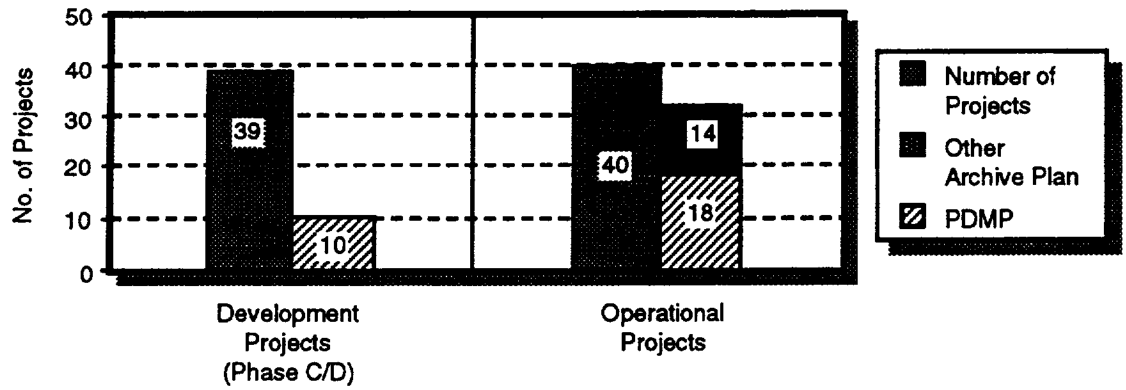


Figure 13 *Data Management Planning Status.*

OSSA is revising the "Guidelines for Developing a PDMP" as part of the overall update of data management policies. These guidelines, originally distributed to flight projects in 1988 to provide uniform guidance for developing plans, have contributed to the progress in generating effective plans. On the basis of experience to date, more specific guidance will be given with key data management parameters to be addressed and tracked as the project develops. These parameters will provide a general overview of project data management requirements. The NIMS database at the NSSDC will be a primary resource for compiling data from all OSSA PDMPs. It incorporates parameters that can be used for planning at the spacecraft, experiment, and dataset levels. The PDMPs should address those parameters identified in the guidelines document so that the PDMP becomes a standard source of information for planning.

National Space Science Data Center (NSSDC)

The NSSDC was established in 1966 at GSFC, as the principal multidisciplinary data archive center for OSSA. The NSSDC supports each of the Discipline Divisions in their data management and archiving, although specific support varies depending upon the needs of the division. Table 9 summarizes the type of support provided by NSSDC for each division.

Table 9 *NSSDC support to each of the Discipline Divisions is tailored to their needs.*

	INFO SVCS. (NASA Master Directory, ETC.)	STANDARDS & TECHNOLOGY	DATA HANDLING		
			Offline Hold/ Dissem.	Online Access	Value-Added
Life Sciences Division	Y	Y	TBD (Offline Dissemination Maybe)		
Earth Sci & App Division	Y (1)	Y	Y (3)	GRID-TOMS	N
Solar Sys Expl Division	Y	Y	Y	N	N
Microgravity Sci & Ap Div	Y	Y	TBD		
Space Physics Division	Y	Y	Y	Y	Y (4)
Astrophysics Division	Y	Y (2)	Y	Y	Y (5)

(1) NSSDC Maintains/Operates "Global Change" Master Directory
(2) NSSDC/NOST Operates a FITS user support office
(3) NSSDC expects to shed most earth science data & responsibility within two years
(4) OMNI, COHO, SSC, CDAW, CD-ROM (partial POP support)
(5) Astronomical data center

In a recent reorganization at GSFC, the NSSDC was transferred to the Space Sciences Directorate as of 31 May 1992. The NSSDC was issued a new charter in the reorganization, as part of the Space Science Data Operations Office. The new charter is as follows:

- Maintain and operate NASA Master Directory and other OSSA-wide information services
- Standards and Technologies - Be pool of expertise for OSSA data environment; recommend OSSA-wide standards
- Data - Support the discipline divisions in discharging their data management, archiving, and dissemination responsibilities
- Discharge responsibilities of World Data Center-A for Rockets and Satellites

The Earth Science & Applications Division currently has the most data holdings in the NSSDC, but these data are in the process of being transferred to the EOS DAACs over the next several years. Other divisions with large amounts of holdings in the NSSDC include the Solar System Exploration, Astrophysics, and Space Physics Divisions. Table 10 summarizes the data holdings for each type of media (by discipline) at the NSSDC, as of 1/1/92.

Table 10 *Data holdings at NSSDC, by storage media for each discipline.*

	Planetary	Earth Science	Astronomy	Space Physics	Other
Magnetic Tapes	9,177	51,868	4,778	17,106	1,167
Optical Disks		210	20	11	
Floppy Disks		18		10	
CD-ROMS	56	1	1		
Microfilm	3,249	792	5,942	20,087	1,194
Microfiche	15,355	2	18,312	10,949	640
Slides	33	812	89	38,474	364
Film (Frames)	305,080	235,284	63,459	3,860	4,456
Film (Feet)	42,814	4,300		2,200	4,400

The NSSDC provides access to a variety of on-line services through the NSSDC On-Line Data and Information Services (NODIS). The number of sessions has more than doubled over the past year. This trend is evident in Figure 14.

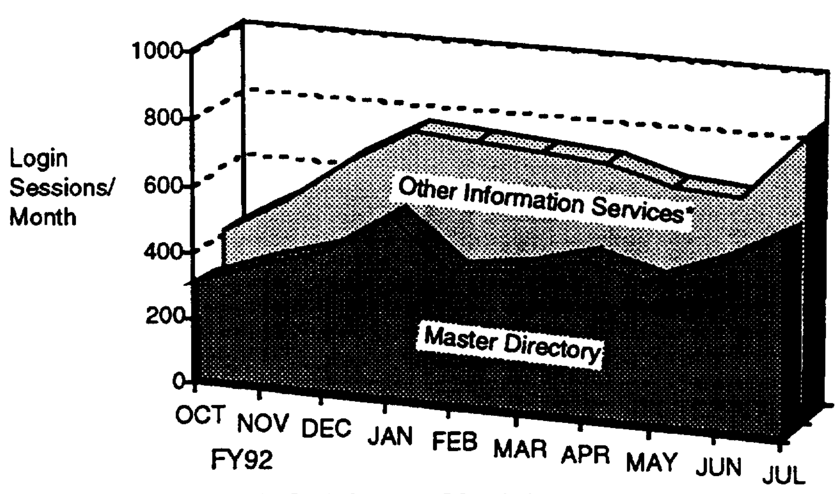


Figure 14 *The use of on-line services at NSSDC was steadily increasing in 1992.*

NODIS is a menu-driven utility that can be reached via dial-up or network. An array of individual services are available through NODIS, including the NASA Master Directory (MD), Personnel Information Management System (PIMS), the American Institute of Aeronautics and Astronautics (AIAA) Canopus newsletter, and the Astronomical Data Center (ADC) on-line information system for astronomical catalogs.

The NASA Master Directory is used more than any other information services through NODIS. It identifies, describes, and points to accessible data worldwide, that may be of interest to OSSA research activities. There are currently 1,600 entries included in the MD database. Included in this are many entries that are aggregated over multiple individual data sets. In addition to data sets, the directory also contains supplementary information about other data information systems and data archives, organized data collecting campaigns and projects, data sources such as spacecraft or Earth-based observing platforms, and data sensors that were used to acquire the data.

More than 90% of currently archived OSSA data has been identified in the MD, which is accessed nearly 10,000 times annually. The NASA Master Directory and the Global Change Master Directory are currently the same directory. The two may diverge in the future, pending development of greater differences in content and requirements. A new client-server architecture has been developed for the MD, as well as a new user interface. An International Directory Network is operational, which can provide directory access to an array of foreign users such as ESRIN, CNES, Canada, Russia, and the Chinese Academy of Science. The distribution of entries in the MD is shown in Figure 15, according to discipline.

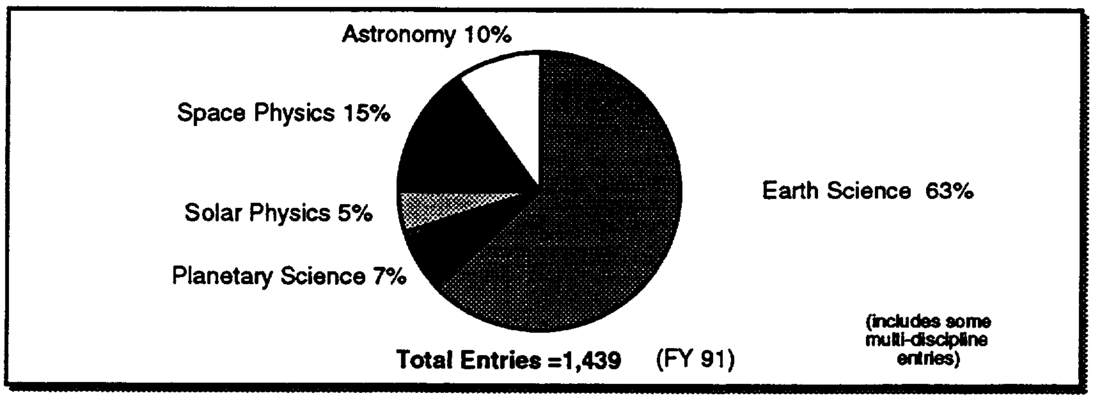


Figure 15 *The NASA Master Directory includes entries for many disciplines.*

In addition to the on-line services provided to the user community by NSSDC for data distribution, there is also hardcopy distribution. As the availability and speed of network access is increasing, more and more requests for services will be provided electronically. The number of data requests made to NSSDC for electronic and hardcopy distribution are shown in Figure 16 below.

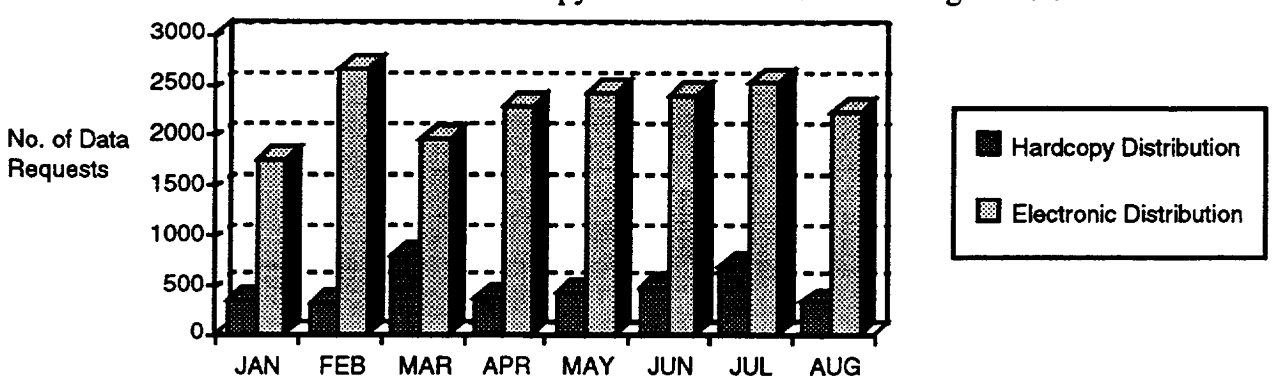


Figure 16 *Electronic distribution of NSSDC data in 1992 was four times as frequent as hardcopy distribution.*

There are a number of various media types being used by the NSSDC for storage of scientific data. The archive advanced technology task is exploring new data storage media and automated tools which will enable OSSA to archive the enormous quantities of expected data within limited budgets. The types of media and the associated future use for each are summarized in Table 11 below.

Table 11 *Storage media use at the NSSDC is changing as new technology is introduced.*

Media Type	Future Status
Round Magnetic 7 Track Tapes	To be eliminated
Round Magnetic 9 Track Tapes	Continued use expected
3480 tape cartridges	Use to increase as medium used internally to NSSDC
8 mm exabyte and 4 mm DAT	Use to increase as transfer medium
12 inch Write once read many (WORM)	SONY use up, OPTIMEM use down
Compact disk read-only memory (CD-ROM)	Increased use expected

The increased use of CD-ROM as a media type for distribution and archiving is shown dramatically in Figure 17. The high data storage density, ease of transport, and expanding use of CD technology have combined to make this dramatic growth in use possible. Technologies such as CD-ROM are making data much more accessible to scientists.

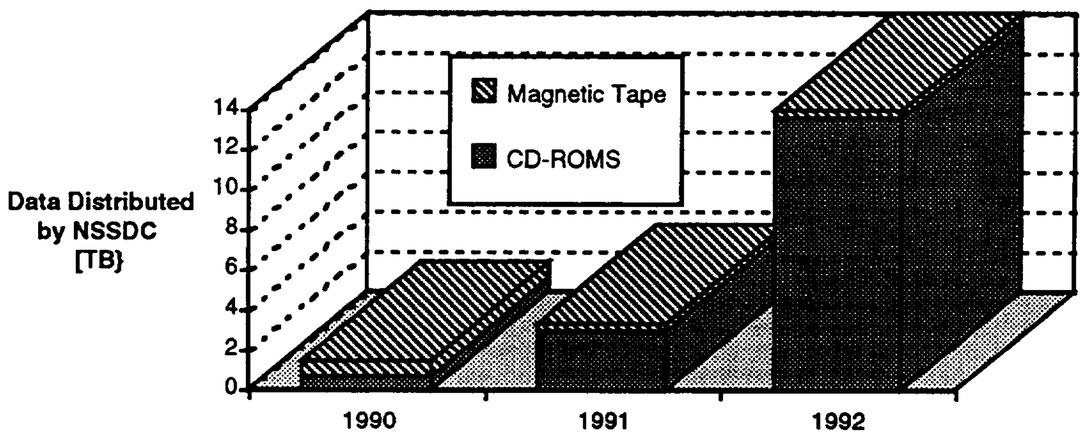


Figure 17 *Data distribution has increased dramatically with the introduction of CD-ROMs.*

In addition to storage media, techniques and tools of accessing those media are being investigated. The NSSDC Data Archive and Distribution Services (NDADS) provides an automated data retrieval request service. Using a robotic "jukebox", NDADS provides near-line access to selected project data. This means that data access approaches the speed of on-line systems, without the need for huge amounts of on-line storage capacity. Data for the robotic jukebox is currently written on up to 182 12-inch WORM disk platters, providing a total capacity of approximately 1.2 Terabytes. Software modifications and additional hardware will enable expansion of up to 8.6 TB storage in the NDADS near-line environment. The introduction of near-line technology through NDADS has dramatically increased the use of network data transfers in the last few years, over those transfers that were provided through the NSSDC computing facility (NCF). This is shown in Figure 18.

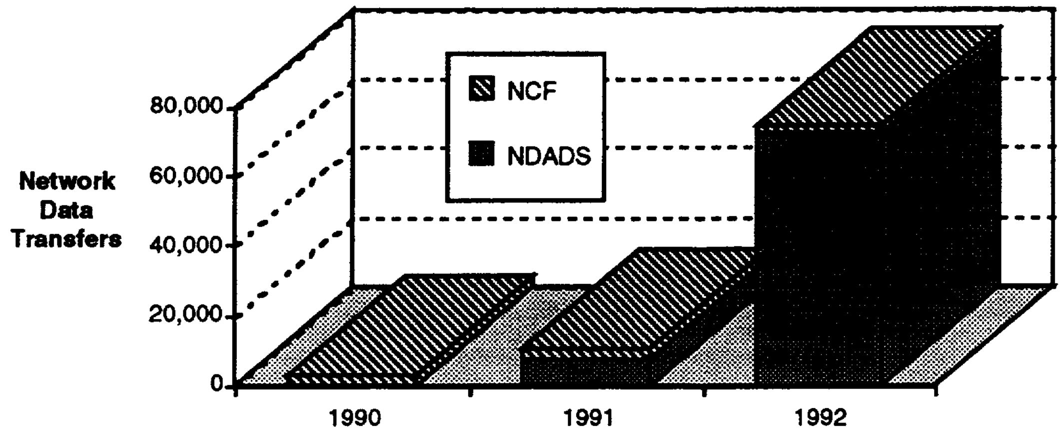


Figure 18 *The NDADS system has dramatically increased network data transfers.*

Standards

There is a NASA/OSSA Office of Standards and Technology (NOST) at GSFC that is responsible for providing a focal point for standards and new technology information. NOST takes an active role in promoting the use of standards, with the goal of improved interoperability among data systems, leading to increased data access by scientists. NOST participates in the development of new standards, assists the OSSA science community in the use of standards, and makes information on standards and new technologies available to the community.

NOST sponsored a workshop in June 1992 to address a variety of formats which are under development. In addition to the NASA/OSSA community, participants included other U.S. agencies such as NOAA and USGS, as well as international representatives from the Canadian Centre for Remote Sensing, and ESA. Ten formats were discussed as shown in Table 12, with proponents or developers giving presentations on each.

Table 12 *There are a number of standard data formats being proposed by the science community and considered by OSSA.*

BUFR	Binary Universal Form for the Representation of Meteorological Data
CDF	Common Data Format
CEOS SS	Committee on Earth Observing Satellites Superstructure
FITS	Flexible Image Transfer System
GRIB	Gridded Binary
HDF	Hierarchical Data Format
NETCDF	Network Common Data Format
PDS Labels	Planetary Data System Labels
SDTS	Spatial Data Transfer Standard
SFDU	Standard Formatted Data Units

The workshop was successful in increasing the NASA/OSSA community's understanding of the similarities and differences among various widely promoted formats. A workshop report and formats comparison document will be available soon to assist users in understanding and selecting appropriate formats.

NETWORKING

The changes in the data union towards a distributed system of interconnected nodes has made networking a critical component of OSSA science activities. Networking capability for the OSSA science community is provided by the NASA Science Internet (NSI). NSI is NASA's worldwide science communications network, as illustrated in Figure 19. The objectives of the NSI networking function are to provide computer networking services, management and operations support, and technical assistance to the OSSA science community. In addition, it provides rapid and reliable access to data resources, computing facilities and collaborators.

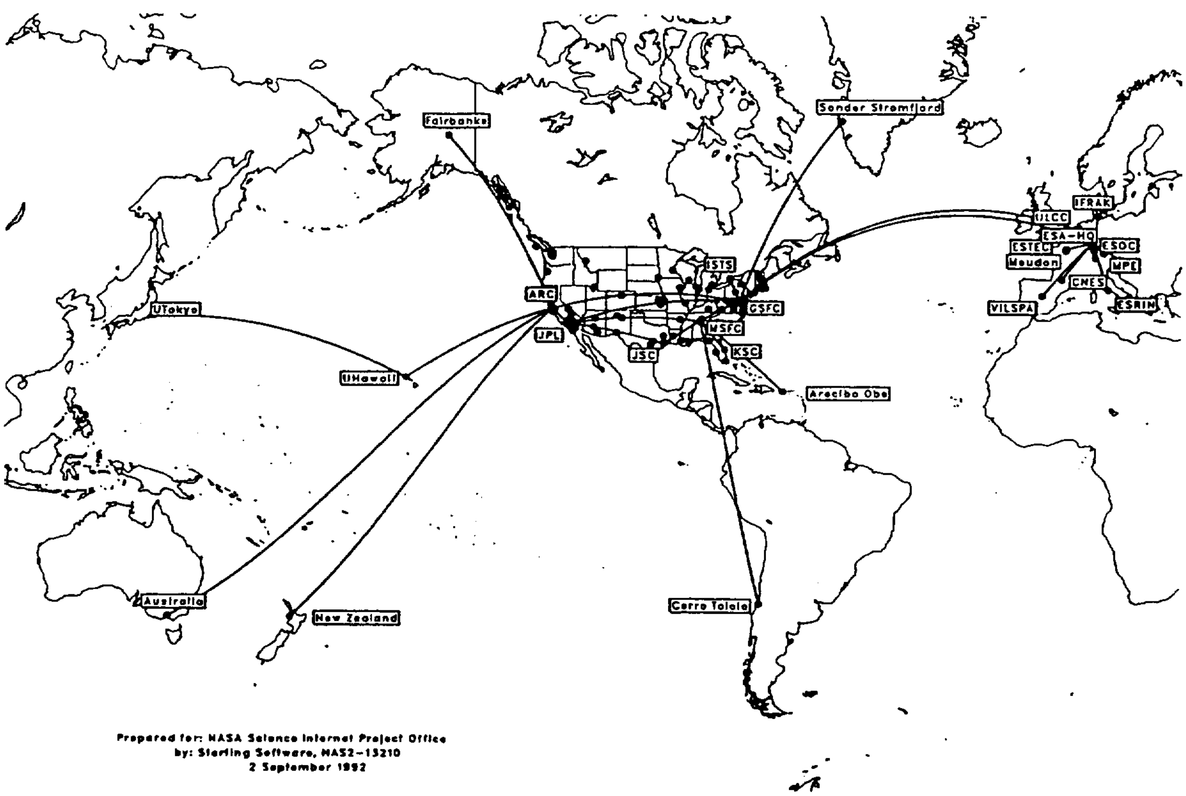


Figure 19 *The NSI is NASA's worldwide science communications network.*

NSI's approach to meet these objectives is to integrate various networks including the Office of Space Operations' Program Support Communications Network (PSCN), NASA Center networks, the NSFnet and Regional university research networks, and the research networks of Europe and the Pacific Rim. The NSI can then directly support NASA space missions and science discipline programs, as well as review and validate requirements through OSSA discipline managers. In addition, such connectivity enables the unified access to widely distributed information sources in support of the science community. NASA's NSI is a partner with the Department of Energy (DOE) ESNET and the National Science Foundation (NSF) NSFnet to establish the dominant federal partners in providing a national infrastructure serving US science research.

NSI has become a critical component of OSSA's information systems infrastructure, and is recognized by users as essential for NASA's continued successful leadership in space science research. NSI currently connects about 4,500 scientists at over 300 research institutions worldwide. Recent growth in the number of nodes and sites on NSI is shown in Figure 20.

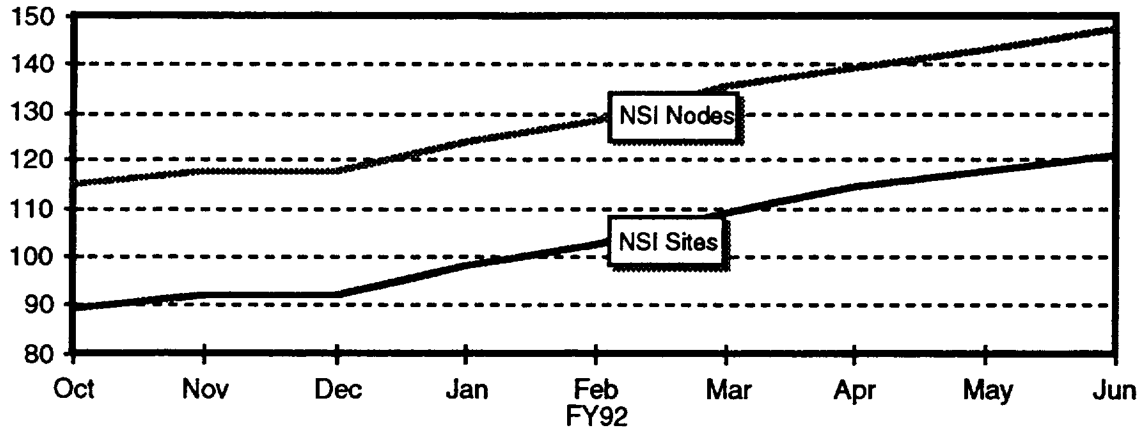


Figure 20 *NSI Nodes and Sites growth.*

Along with an increase in sites and users comes an increase in requirements for NSI services. The number of requirements has approximately doubled each of the past two years, with an increase of all requirements prior to FY90 of 329, to a total of 1,944 by the end of FY92. These trends can be seen in Figure 21. Of the 1,944 total requirements, 845 have been completed or processed, while 1,099 are active. Certain large programs place an unusually high number of requirements on the NSI, as EOSDIS did in FY91. However, there were still a significant number of additional requirements in FY92.

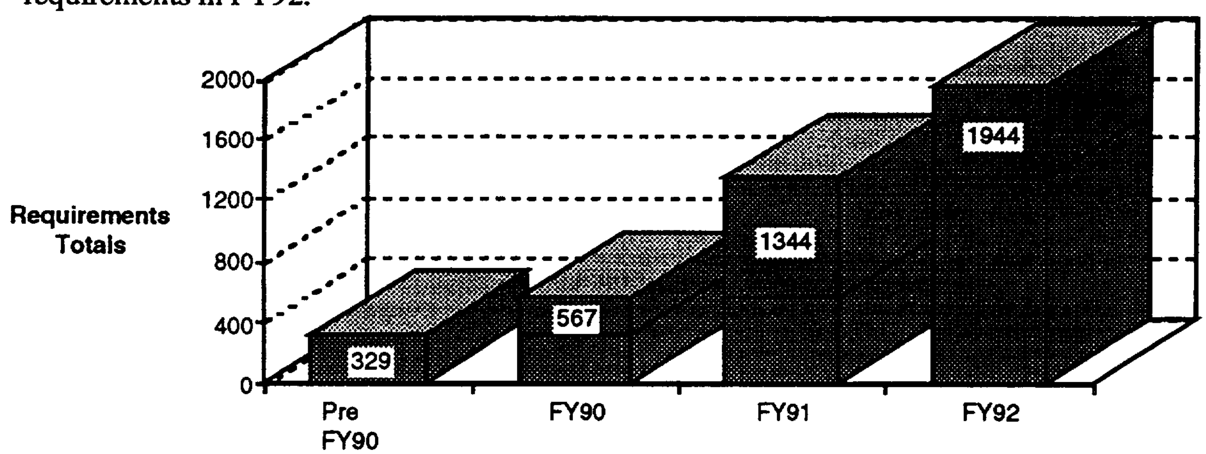


Figure 21 *NSI requirements have been significantly increasing.*

NSI provides quality service to science users, backed by nationally recognized technical excellence and leadership in networking to ensure high quality, reliable end-to-end service. NSI has adopted a continuous improvement methodology for achieving better cost and performance. Reductions in requirements implementation has dropped from 22 to 20 months for Internet and from 26 to 23.5 months for PSCN, in the last six months of FY92. In spite of the increased number of users and nodes being supported by NSI, the number of network problems being reported is decreasing.

High Performance Networking

OSSA is participating in the National Research and Education Network (NREN), an element of the High Performance Computing and Communications (HPCC) initiative. Through the NREN, the NSI community will have the benefits of T3 (45 Mbps) service and performance rather than the existing T1 (1.5 Mbps) capability. These improvements will maximize use of existing network resources, with initial deployment expected in March 1993. This will greatly enhance the data transmission capability of the existing NSI network, at an estimated cost savings to the science community of at least 50% per network access.

SCIENTIFIC COMPUTING

The objectives of OSSA scientific computing efforts are to provide the computational environment necessary to support the requirements of OSSA missions and the associated science community, and to develop the strategy and plans that will assure the continual evolution of that environment in consonance with emerging trends in science methodology and technology. The community of OSSA computer users is continuing to grow, and their demands for computational capability is growing as well. Although budgetary resources are constrained, application of advanced technology and leveraging other efforts have enabled computational resources to keep up with these increased demands. The growth in demands and plans to accommodate those demands for key resources over the next several years are shown in Figure 22.

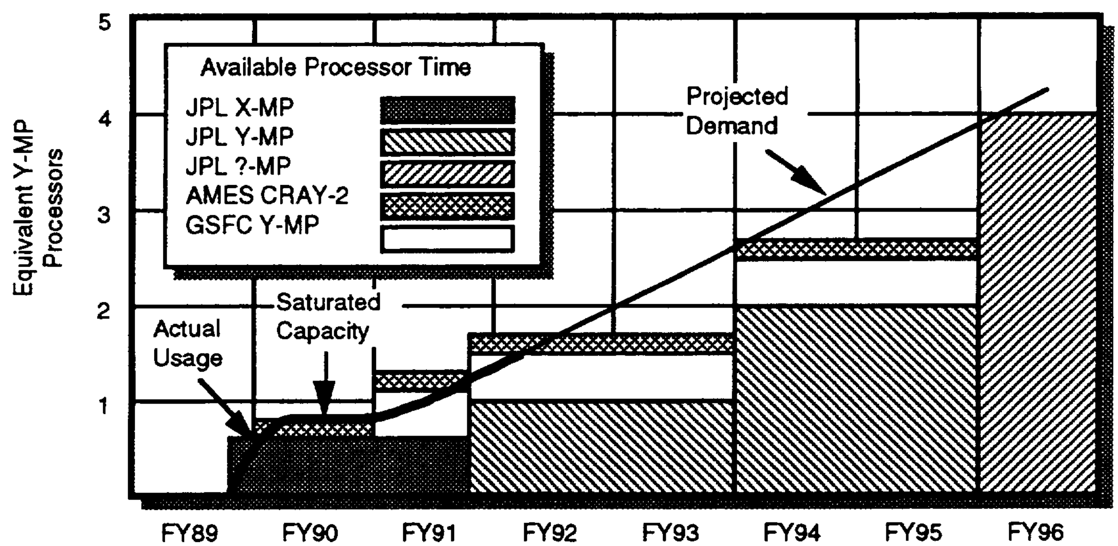


Figure 22 *Scientific computing capability must grow rapidly to meet increasing demands.*

Principal elements of scientific computing are the NASA Center for Computational Science (NCCS) at GSFC, the JPL Supercomputing Project at the Jet Propulsion Laboratory, and the Concurrent Computing Testbed Activity at the Jet Propulsion Laboratory.

The NCCS is a central scientific computing facility providing services to the Goddard science community as well as to the OSSA community in general. Principal elements of the facility are a Cray Y-MP 8/464 supercomputer and a 3 terabyte mass data storage and distribution system. The facility currently serves some 1,500 users annually, with remote usage accounting for some 25% of the cycles used. Plans are underway to enhance the OSSA supercomputing capacity.

The JPL Supercomputing Project includes a Cray Y-MP 2E/116 with limited on-line storage. This facility currently supports 450 members of the JPL research community and has been an enabler of science research and data analysis local to the JPL facility. Supplemental computing capacity is provided to users through the Ames' NAS facility and the NCCS.

The concurrent computing testbed activity is a small research activity examining effective utilization of loosely coupled heterogeneous computing systems in the solution of OSSA related problems. User interfaces for these environments are being examined as well as scientific research distributed over computing engines of different architectures.

Supercomputer utilization of the JPL and GSFC computing resources for FY92 is shown below in Figure 23.

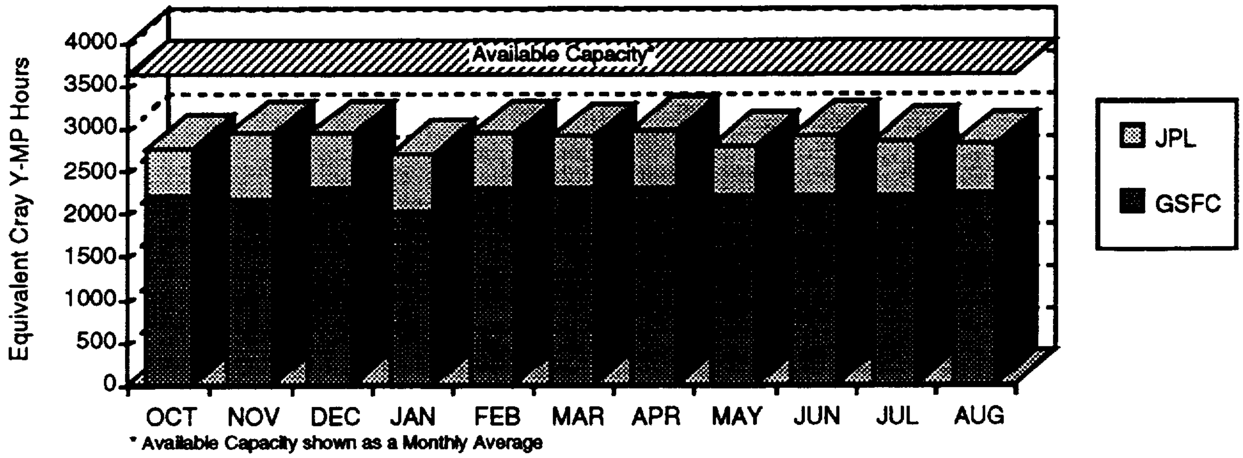


Figure 23 *Supercomputer utilization for FY92.*

High Performance Computing

OSSA is participating in the High Performance Computing and Communications (HPCC) Program. The goal of this program is to accelerate the development of application of high performance computing technologies to meet NASA science and engineering requirements. In earth and space science, the specific benefits are multidisciplinary modeling and monitoring of the earth and its global changes and assessments of their impact on the future environment. This will be realized through development of algorithms and architecture testbeds capable of fully utilizing massively-parallel concepts that are scalable to sustained teraFLOPS performance. In addition, the creation of a generalized software environment will be needed for massively parallel computing applications. The impact of these technologies will then be demonstrated and applied to NASA research on Earth and space sciences physical phenomena.

APPLIED RESEARCH & TECHNOLOGY

The objective of the applied research & technology area is to apply advanced information systems technology as appropriate to improve support to OSSA science programs in meeting the needs of science users. The trend in space science information systems has been higher data rates, greater data volumes, increased reliance on supercomputers, hierarchical management, development of tools to support coordinated science operations and analysis, development of master directories, increased need for onboard operations, and increased concern for security. Future needs in information systems will include interoperability with international systems, device independent system architectures, standard interfaces, development of tools to take advantage of standards, and testbeds.

Elements of the Applied Research and Technology effort addressing these trends include applied research to enhance science data management, analysis and visualization. JPL is leading efforts in visualization and animation of science data from the solar system. Development work in visualization is also underway at the Center of Excellence in Space Data and Information Sciences (CESDIS) at GSFC. NASA Research Announcements (NRA's) are being used in open solicitation to broaden participation in this area to universities and industry. An important role for the Information Systems Branch in this area is to provide an infrastructure for testing, evaluating, inserting, and maintaining new technologies that are being developed. In addition, efforts are being made to coordinate and leverage these research and development activities with NASA's Office of Aeronautics and Space Technology, the National Science Foundation, industry and others.

OSSA will accelerate the development and application of high performance computing technologies through the NASA High Performance Computing and Communications Program (HPCC) Earth and Space Science (ESS) applications project. The approach to HPPC in OSSA includes use of a NASA Research Announcement (NRA) to select Grand Challenge Applications and principal investigator teams that require teraFLOPS computing capability for addressing NASA science problems. Successive generations of scalable computing systems will be developed as testbeds for the Grand Challenge Applications. The investigators and testbeds will be interconnected through high speed network links, and a software development environment and computational techniques will be provided to support the investigators. In collaboration with the investigator teams, OSSA will conduct evaluations of the testbeds across applications and architectures, leading to a selection of the next generation of scalable teraFLOPS testbed.

Major programs and activities in the applied research & technology area include the following.

JPL Visualization Activities

- Visualization tools are being utilized effectively by several projects, including Magellan Venus flyovers, and UARS ozone results
- Successful demonstration of Explorer image rendering software, integrated with Hypercube testbed and Solar System Visualization tools (Explorer was renamed Surveyor)
- Several visualization tasks under development
  - Parallel Methods in Image Analysis
  - The Linked Windows Interactive Data System
  - Three Dimensional Interactive "Explorer"
  - Visiting the Planets

Applied Information Systems Research Program (AISRP)

- 22 Investigations have been selected and are underway via the initial solicitation under this program.

- Linkwinds was demonstrated, marking the first time a Multi-User System Environment (MUSE) capability has been shown cross continent
- Copies of Linkwinds have been installed at several test sites including UCSD, Univ. of Colorado, Oregon State University, and UARS investigators.
- Continuing research on software called DataHub for visualization and analysis using large distributed databases.

Center of Excellence in Space Data & Information Systems (CESDIS)

CESDIS brings together computer scientists from university, industrial, and government laboratories to conduct computer science research having application to Earth and space science, focus attention on accessing, processing, and analyzing data from space observing systems, and collaborate with NASA space and Earth scientists. The CESDIS accomplishes these goals by:

- Funding research projects
- Supporting research personnel for projects funded by NASA through other programs
- Providing a computer science research environment
- Conducting workshops and conferences
- Administering fellowships
- Developing areas for collaborative efforts
- Producing technical reports.

Some CESDIS projects are:

- Communication protocol research
- Highly concurrent synchronization mechanisms
- Multi-Resolution Common Data Format (MR-CDF)
- Adaptive storage management
- Computer Assisted Analysis of Auroral Images Obtained from high altitude Polar satellites
- Parallel compression of space and earth data

Appendix A

DATA MANAGEMENT INITIATIVE (DMI)

The NASA/OSSA Data Management Initiative (DMI) is a multi-year, multi-million dollar effort that will ensure archiving of appropriate data from past missions, and creation of an infrastructure to enable the orderly archiving of data from future missions. The program will also ensure that data are preserved, inventoried and documented to facilitate broad future access by the science community.

The DMI program, coordinated by OSSA's Information Systems Branch (ISB) was proposed as a $4M new start for FY92. The program was funded and began in FY92 at the $1.8M level despite the absence of "new" dollars in the budget. Future funding for this activity out of the ISB budget has been projected as $1.4M, $1.1M and $750k for FY93-95 respectively. The Science divisions are investing similar amounts in associated efforts to enhance the data environment.

The funding in FY92 allowed three types of activities associated with the DMI to begin, including:

1) Identification and community assessment and prioritization of data sets in need of "restoration"
2) Restoration and/or archiving of appropriate data sets
3) Creation/Improvement of the capabilities and capacities of the Discipline Data Systems (DDS) and of the NSSDC, and of the procedures and tools whereby those entities assure the routine flow of increasing volumes of the right data into the OSSA archive environment (and retrievability of there from that environment).

Initial efforts in the early years of the DMI will focus on identification and restoration/archiving of appropriate data sets. Funding has been provided in FY92 at the sites of four space physics groups and six astrophysics groups for restoration and reformatting. In addition, the restoration program at JPL is continuing. Improvements to data management systems and facilities will be the focus of the DMI once important data sets have been safely archived. However, definition and development of new DDS's in Life Science, Microgravity, and Space Physics has begun under the DMI in 1992.

Identification and Assessment

A comprehensive information base is being developed at NSSDC which will identify all extant OSSA-mission data sets. The information base is already populated with information concerning NSSDC-held data and many other OSSA data sets which were identified in the OSSA Data Census coordinated by NSSDC in 1990. A summary of the data in this information for each data set is as follows:

- Data Set Identification
- Suitability for archiving
- Community recommendations concerning data set archiving
- Archiving status
- Other related information

OSSA Discipline Divisions are each orchestrating efforts to identify potentially archive-desirable data sets not yet identified in the NSSDC information base. Each Division has identified one person to be responsible for coordination of the Division's DMI activities, including assessing and prioritizing data sets for restoration and archiving.

Restoration

Once data sets have been assessed and prioritized, the process of restoration and archiving can begin. The restoration process will address three principal activities:

- Ensuring the usability of the data once archived.

- Reformatting of data (to standard formats, adding labels, etc.)
- Migrating data bits to new media

To ensure usability, documentation and/or software may be required if and when persons having initial expertise with the data are no longer available to support correct use. No OSSA-wide standards currently exist for data format standards, and the Divisions will determine what reformatting is required. Migration of data bits from old media to new media will occur at central cites such as the NSSDC or DDS facilities, or at distributed PI or other sites if appropriate capabilities exist. Sources of data for the migration will include old data volumes already held at central sites and data sent there by PI's, as well as data held by PI's. A goal of the data restoration activity is to migrate data sets which are now held in the distributed OSSA community by PI's, to the public OSSA archive environment of the appropriate DDS, the NSSDC, or both.

The Science Digital Data Preservation task is a part of the DMI at JPL. The focus of Phase I has been on creating an inventory and assessing tapes that are at least five to fifteen years in age and degenerating. This has been done on 135,000 tapes that JPL has retrieved from the Federal Records Center. These 7-track and 9-track tapes include data from Viking (40,000 tapes), Voyager, Mariner, and other missions. Of these tapes, 50,000 have been identified as critical to save for future science use, under the guidance of SDEB, PDS and PSDSG. Other tapes are either duplicates, of lower priority, or can no longer be read. The Phase I Final Report will be published in October 1992. Phase II work will begin the process of converting 9-track tapes to CD write once media at the rate of about 3,000 tapes per year, with expected funding levels. Conversion of the 7-track tapes has been postponed.

Appendix B
EDUCATIONAL OUTREACH PROGRAMS

There are a number of educational outreach programs sponsored by OSSA to facilitate access to scientific data and increase minority participation in scientific research.

Minority University-Space Interdisciplinary Network (MU-SPIN)

The networking capability provided by the NSI enables increased University participation in OSSA science activities. The Minority University-Space Interdisciplinary Network (MU-SPIN) program is a networking and education initiative for Historically Black Colleges and Universities (HBCU), Minority Universities (MU), and other universities with large minority student enrollment. The program's main goal is to interconnect the computing facilities of HBCU's and MU's with the NSI, and to promote awareness and usage of wide area networking technology in support of collaborative interdisciplinary scientific research among faculty, students, and NASA scientists. The program consists of four major components:

1) Wide area networking
2) Faculty/student development
3) The residence program
4) User Working Groups

NASA/University Joint Venture (JOVE)

JOVE was initiated as a pilot program with six universities in 1989, to develop aerospace research capabilities and to promote science and engineering education. JOVE concentrates on institutions of higher education that have had little or no involvement in the Nation's aerospace program. JOVE is a capability building program, aimed at curriculum development, the enhancement of student research potential, and outreach programs to students and the broader community served by participating universities. Under this program, NASA makes space science data and NASA resources available to the university researchers in exchange for the university providing faculty and student support on a matching funds basis to carry out the research. Participation in JOVE has grown significantly since it began in 1989, as shown below.

Year	1989	1990	1991	1992
Participating Universities	6	15	30	55

Graduate Student Researchers Programs (GSRP)

In 1980, NASA initiated the Graduate Student Researchers Program (GSRP) to cultivate additional research ties to the academic community and to support promising students pursuing advanced degrees in science and engineering. Since then, approximately 1,200 students have completed the program's requirements. In 1987, the program was expanded to include the Underrepresented Minority Focus (UMF) Component. This program was designed to increase minority participation in graduate study and research and, ultimately, in space science and aerospace technology careers. Approximately 230 minority students have completed the program's requirements while making significant contributions to the nation's aerospace efforts. Continuing to expand fellowship opportunities, NASA announced in 1990 the Graduate Student Fellowships in Global Change Research (GSGCR). Designed to support the rapid growth in the study of Earth as a system, approximately 150 fellowships have been awarded since its inception. And, in 1992, NASA announced opportunities in the multiagency High Performance Computing and Communications (HPCC) Program designed to accelerate the development and applications of massively parallel processing. Approximately five new fellowships will be awarded yearly.

Appendix C

OSSA POLICY ON SCIENCE DATA MANAGEMENT

OFFICE OF SPACE SCIENCE AND APPLICATIONS PROGRAM DIRECTIVE

Responsible Offices:

Life Sciences Division
Earth Sciences and Applications Division
Solar Systems Exploration Division
Flight Systems Division
Microgravity Science and Applications Division
Space Physics Division
Astrophysics Division

Subject: Policy for the Management of the Office of Space Science and Applications' Science Data

1. Purpose

The purpose of this Program Directive is to establish NASA's policy for, and delineate responsibilities and authorities relative to, the continuing management of the Office of Space Science and Applications' science data. It replaces NASA Management Instruction (NMI) 8030.3A, "Policy Concerning Data Obtained from Space Science Flight Investigations", dated May 2, 1978, in satisfying this function which was first established in a NASA Policy Directive, dated January 7, 1967 (NPD 8030.3). Among other important modifications, NMI 8030.3A established the requirement for all space flight projects to develop Project Data Management Plans (PDMPs). As defined in NMI 8030.3A, the PDMP was essentially conceived as a data archiving plan. The increasing complexity of NASA science investigations and the volume of data that they generate (among other factors) emphasizes the need for increased emphasis and priority for data management planning early in the project's life. Additionally, these data management planning activities must address the total flow of research data not just archiving. This Program Directive expands the scope of the PDMP to include planning for data management throughout the project planning and implementation phases.

2. Scope

This program directive is applicable to the management of all science data resulting from Office of Space Science and Applications sponsored research missions and programs.

3. Policy

a. Science data generators and users shall serve as a primary source of requirements, as well as final judge of the quality and value of scientific data. Advice and guidance shall be obtained from the science and applications research community in the planning and implementation of NASA's data management systems.

b. NASA shall establish and maintain archives to preserve and make accessible all valuable NASA science data and information. This system of data archives shall include easily accessible information about NASA's data holdings, guidance, and aids for locating and obtaining the data. A review process, including scientific community representation, shall be established to determine what data should be archived and to assure conformance with completeness and quality standards.

c. National and international standards for media, formats, and communication of data sets shall be used to the greatest extent possible. NASA shall participate in the development and implementation of standards. NASA unique standards shall be used only if adequate national or international standards are lacking. The intent of this policy is to standardize the interfaces between the users and NASA's data and information systems, not to standardize the systems themselves.

d. Project, discipline, and OSSA-wide data management activities shall be reviewed periodically to assess status and progress relative to Agency and OSSA goals, objectives and standards, and the needs of the science community. Once archived, data sets and supporting information shall be periodically reviewed to assess their value for continued retention by NASA. This process shall also prevent the loss of important data sets.

e. All data being captured by NASA science projects and space flight missions shall be addressed in a Project Data Management Plan (PDMP) to assure the availability of data and supporting information on a timely basis for use by the science community. The formal definition (non advocate) review or equivalent mechanism conducted prior to a project receiving new start approval will assess data management plans as well as spacecraft development, instrument plans, operations plans, etc. After new start and budget approval a formal Project Data Management Plan will be prepared and approved coincident with the Project Plan signed by the Associate Administrator and Field Center Director. Project Data Management Plans must be updated as significant changes occur that impact the project's plans for data management, and PDMPs should be reviewed periodically to determine if updates are required. For programs in which selected investigators have initial periods of exclusive data use, data should be made openly available as soon as that period expires. In such cases, the duration of all exclusive use periods shall be explicitly defined.

f. NASA shall periodically conduct a review of its data repositories and archives to determine the state of data and to assure conformance with applicable government standards for data storage.

g. Recognizing the pivotal importance of technology in meeting its future needs for data and information systems, OSSA shall establish an active process to maintain an awareness of emerging applicable technologies, infuse them into its systems, and stimulate new technology development where warranted.

4. Responsibilities and Authorities

a. Associate Administrator for Space Science and Applications

The Associate Administrator for Space Science and Applications is responsible for maintaining and ensuring the implementation of NASA's data management policy, including issuing implementing instructions and guidelines.

b. Assistant Associate Administrator for Science and Applications

The Assistant Associate Administrator for Science and Applications shall serve as chairperson of the Information Systems Management Board, which is chartered to coordinate OSSA's data management activities, and to identify issues, set priorities, and provide recommendations to the Associate Administrator for Space Science and Applications on these activities.

c. OSSA Discipline Division Directors

The Directors of OSSA's Science Discipline Divisions are responsible for the overall administration of their Division's data management activities in accordance with this Program Directive and the decisions of the Information Systems Management Board. Data acquired from both flight projects and non-satellite programs will be addressed as part of this responsibility. The primary objective of this activity is to assure the continuing value of OSSA's science data by providing data management procedures, systems and services that are responsive to the needs of the project, discipline, NASA, and broad research communities.

d. OSSA Flight Systems Division Director

The primary responsibilities of the Information Systems Program within OSSA's Flight Systems Division are to formulate and coordinate OSSA wide data management policy and to provide the supporting infrastructure across the discipline efforts. This includes providing a broad range of data management capabilities which transcend discipline-specific data management activities and serving as OSSA's point of contact for data management activities.

Appendix D

STATUS OF DATA MANAGEMENT ISSUES RAISED BY GAO

There have been two recent reports issued by the United States General Accounting Office (GAO) which have addressed issues in NASA's management of space science data:

- Space Operations: NASA Is Not Properly Safeguarding Valuable Data From Past Missions, GAO/IMTEC-90-1, March 1990
- Space Operations: NASA Is Not Archiving All Potentially Valuable Data, GAO/IMTEC-91-3, November 1990

In general, NASA found that the reports provided a useful assessment of some of the key issues in data management, and agreed with many of the shortcomings identified in the reports. There was a concern about balance however, particularly in the first report, regarding recognition of on-going NASA initiatives already addressing the identified shortcomings.

A summary of the status of ongoing NASA data management activities relative to GAO recommendations from both reports was provided to GAO on 26 June 1991. Significant progress since that time has continued, within budgetary constraints, on all issues and recommendations in these reports. The following material is not a comprehensive description of the OSSA data management and archiving program, but reports on the current status of those elements which are particularly relevant to recommendations made by GAO in its two reports.

Those elements of the OSSA data management and archiving program which are relevant to the eight recommendations from the report published March 1990 will be discussed first.

1. *Conduct a through inventory of all NASA's space science data stored at NASA centers and contractors, universities, research institutions, and other federal agencies.*

   NASA has continued to conduct several data inventory activities resulting in a high level inventory of digital data sets. These activities have included the following:

   - OSSA Data Census Phase 1 and Phase ll
   - NSSDC Database
   - PSASS Survey
   - PDS Catalog
   - Science Digital Data Preservation Task

   The OSSA Data Census was an extensive two-phase census effort initiated in August 1989 with our principal investigator community. The objective was to identify data sets held from previous missions that should be included in archives and made available to the general research community. This recent census was an update to a previous census completed in March 1981. A report on Phase One was issued in March 1990 and Phase Two was issued October 1991. The census results confirm NASA's understanding that most extant data from inactive missions, which ought to be archived, are already archived.

   The information system at NSSDC has been upgraded to accommodate census information and to permit ongoing tracking of OSSA archiving status. This comprehensive information base will identify all extant OSSA-mission data sets. The information base is already populated with information concerning NSSDC-held data and many other OSSA data sets which were identified in the OSSA Data Census. A summary of the data in this information for each data set is as follows:

- Data Set Identification
- Suitability for archiving
- Community recommendations concerning data set archiving
- Archiving status
- Other related information

All new projects are tagged and updated for inclusion in the Master Directory which is now fully operational. The MD is providing on-line information to the research community about existing data sets, including archive locations, etc. It is populated with descriptions of at least 90 % of NASA mission data.

As Phase I of the Science Digital Data Preservation Task, JPL has inventoried and evaluated all of its 135,000 institutionally managed tapes.

The inventory of data is continuing through the DMI. Distributed data sets not currently in the OSSA archive environment are being collected and archived as appropriate.

2. *Assess, in cooperation with the scientific community, the inventoried data for its scientific value and integrity of its storage media.*

OSSA has now performed an institution-wide survey of its digital data holdings, which has provided the information necessary for an assessment of storage media and archive desirability.

In conjunction with the science community, OSSA is continuing to identify and assess the value of data products for inclusion in archives. As part of the Data Management Initiative, there is an active, ongoing effort with steering groups from each of the science disciplines to oversee and evaluate the quality and value of data sets within their discipline community. OSSA Discipline Divisions are each orchestrating efforts to identify potentially archive-desirable data sets not yet identified in the NSSDC information base. For each data set, this information base contains information as to the suitability for archiving, recommendations from the science community and archiving status. Each Division has identified one person to be responsible for coordination of the Division's DMI activities, including assessing and prioritizing data sets for restoration and archiving.

The Data Set evaluation performed by the PDS of the Solar System Exploration Division is a good example. This evaluation included :

- Sorting data sets by PDS Discipline Node, assigning distributed data sets relating to a specific node to that node for review.
- Each node was requested to furnish the following information:
  - High level review for correctness of node assignment, duplicates, or missing data sets
  - Detailed analysis by data set of disposition (archive, redundant), and priority for preservation/restoration

"Irreplaceable data" will be identified and priorities established for preservation and/or restoration.

3. *Copy valuable data from deteriorating tapes to archival quality magnetic tapes or other storage media suitable for long-term retention of digital data, and release unneeded tapes for reuse or disposal.*

There are a number of ongoing efforts to convert data stored on deteriorating tapes to other media such as Compact Disk Read Only Memory (CD-ROM) and optical disks. Principal data restoration programs are underway at NSSDC and at JPL, and will be widely pursued throughout the NASA data environment as part of the Data Management Initiative. As outlined in the Data Management Initiative, once data sets have been assessed and prioritized, the process of restoration and archiving can begin. The restoration process underway addresses three principal activities:

- Ensuring the usability of the data once archived.
- Reformatting of data (to standard formats, adding labels, etc.)
- Migrating data bits to new media

To ensure usability, documentation and/or software may be required if and when persons having initial expertise with the data are no longer available to support correct use. No OSSA-wide standards currently exist for data format standards, and the Division's will determine what reformatting is required. Migration of data bits from old media to new media will occur at central sites such as the NSSDC or DDS facilities, or at distributed PI or other sites if appropriate capabilities exist. Sources of data for the migration will include old data volumes already held at central sites and data sent there by PI's, as well as data held by PI's. A goal of the data restoration activity is to migrate data sets which are now held in the distributed OSSA community by PI's, to the public OSSA archive environment of the appropriate DDS, the NSSDC, or both.

Selected examples of data restoration for popular data sets include the Voyager encounter data on a series of CD-ROMS, and data from Dynamics Explorer, Coastal Zone Color Scanner and Total Ozone Mapping instruments on the Nimbus spacecraft which have been converted to optical disks. Not only do these new media ensure the permanent archiving of these data sets, but they also enable much more rapid and much broader access. In addition to these older data sets, the Magellan data has been stored on CD-ROM and is being widely distributed. We will continue and expand these efforts, guided by the needs and direction of the scientific community.

At NSSDC, approximately 20,000 tapes have had their data extracted and written on to new media (mostly 6250-bpi tape and 3480 tape cartridges). The recovery rate of data from these old tapes, which are 10 to 25 years old, is over 98%.

At JPL, a Science Digital Data Preservation task has been ongoing which addresses Data Preservation and Data Restoration.

The Science Digital Data Preservation task was a part of this portion of the DMI at JPL. The focus of Phase I has been on creating an inventory and assessing tapes that are at least five to fifteen years in age and degenerating. This has been done on 135,000 tapes that JPL has retrieved from the Federal Records Center. These 7-track and 9-track tapes include data from Viking (40,000 tapes), Voyager, Mariner, and other missions. Of these tapes, 50,000 have been identified as critical to save for future science use, under the guidance of SDEB, PDS and PSDSG. Other tapes are either duplicates, of lower priority, or can no longer be read. Principal efforts under Phase I were:

- Establishment of the Science Data Evaluation Board
- Publication of "Data Disposition Policy and Procedures", with approvals from ADLs
- Completed inventory and assessment of 135,000 tapes

- Identified and prioritized appropriate tapes for archive with integral participation of the science community
- Initiated OSSA-wide planetary data survey

The Phase I Final Report was published in October 1992. Phase II work will focus on the conversion of identified tapes to archival quality storage media, in order of priority. The process will convert 9-track tapes to CD write once media at the rate of about 3,000 tapes per year, with expected funding levels. Conversion of the 7-track tapes has been postponed.

4. *Archive valuable scientific data in facilities that meet National Archives and Records Administration (NARA) regulations.*

NASA has continued to make significant and affordable upgrades at several facilities, bringing them into better compliance with NARA regulations. More expensive options, including the leasing of NARA compliant storage facilities, are under study as part of the Data Management Initiative.

The following table summarizes compliance levels for selected facilities. This summary is an update to a table in the March 1990 report.

	NSSDC		TSSF		MIPL		IPAC	
NARA Regulations	11/90	9/92	11/90	9/92	11/90	9/92	11/90	9/92
Temperature Control	non	partial	non	partial	partial	partial	partial	
Humidity Control	non	partial	non	partial	non	partial	partial	
Test/certify media	non				non	partial	non	
Off-site backup	partial	partial			partial	partial		
Security	non		partial	partial	partial	partial	non	partial
Samples	non				non	non	non	non
NIST & Industry Guidelines								
Tape handling	partial	partial	non	partial	partial	partial	non	partial
Fire protection	partial	partial	partial	partial	partial	partial	partial	
Water protection	non	partial	non	partial	partial	partial	non	non
Hardware maintenance								
Tape maintenance	non	partial			non	non	non	
Tape transportation	non	non	non	non	non	partial	non	partial

Levels of compliance: non · partial · full

Data Archive compliance with industry and government standards has improved significantly.

Specific examples of what has been done to improve compliance at the NSSDC and TSSF at GSFC are included as follows.

Goddard Space Flight Center (GSFC) - Tape Staging and Storage Facility (TSSF):

- An electronic card/key system was installed at the entrance to the TSSF
- Two temperature/humidity recorders installed

- One water-level sensor installed which is connected to the central security system
- Box Edge Protectors installed on all "boxed" tapes stored on pallets
- Two cameras, one video switcher, one monitor, and one VCR have been installed to observe and record activity at the loading dock
- Fourteen motion detectors were added
- Smoke and fire detectors increased from fifteen to sixty
- Door contact sensors were increased from ten to twenty-eight
- Strobe lights were added to fire alarm horns for the hearing impaired
- Building security was augmented with a guard on second shift.

The above equipment is tested monthly and calibrated, when required, every six months.

GSFC - National Space Science Data Center (NSSDC):

- Combustibles removed from tape storage areas
- Hydro-thermograph in tape archive area is calibrated monthly
- Plastic sheets are protecting tapes in the event of water leakage in the ceiling of archive area
- 25 % of archive area is cleaned, on a rotating basis, every two weeks
- A cardkey activated door was installed on the tape library at the NSSDC in May 1992
- Most NSSDC computers have been moved to an area controlled via key-card access in Building 28
- Backup copies are routinely generated for all incoming tapes and for all volumes created in the Data Restoration program
- Electronic/interactive system designed and implemented for NSSDC staff members to log in/out tapes from archive area
- The NSSDC is now buying more expensive tapes which are pre-certified by the supplier rather than using tapes out of NSSDC store stock to ensure data integrity
- Data sampling is regularly being conducted on a large portion of tapes as part of the data restoration process
- Tape maintenance is being performed as part of the data restoration process
- Temperature and humidity gauges are calibrated every six months
- Tapes are no longer stored in the hallway. All tapes are now under lock and humidity control. G-13 was set up as an FRC storage area complete with a separate humidity-chart recorder
- The number of tapes resident in the computer room has been reduced
- A "dual technology" approach for archiving data has been implemented, using the round and square tape method
- An effort has begun to store one copy of the data off-site
- An additional fire alarm bell was installed in the library
- The tape cleaner/certifier was recalibrated
- Transportation personnel were "educated" regarding the proper care and handling of magnetic tapes between GSFC and the FRC
- IDA is used to determine the age, location and usage of the data.

Advances in storage technology can ease burdens for data storage. An example is that higher storage density among storage media, leads to lower requirements for storage volume. A real world impact of this is that the Science Digital Data Preservation Task has reduced the need for large, costly tape storage facilities by reducing the volume of tapes through conversion to higher density media and disposing of duplicate data. Of 135,000 tapes which were archived at the Federal Records Center, only 50,000 really need to be archived. These 50,000 will be restored onto only 3,000 CD-WO disks and 17,000 analog tapes.

5. *Develop and Implement agency-wide tape management and maintenance standards which include all NARA regulations and NIST guidelines.*

NASA has completed a handbook on Records Management, titled "NASA Records Management Guide" which includes NASA-wide standards for tapes and other archive media. The guide on Records Management was completed in July, 1992 and distributed to NASA, GSA, and NARA. It contains an extensive section on electronic records and contains specific agency tape management and maintenance procedures. In addition, NASA is working with agencies and industry to develop management and maintenance standards for other types of media that are becoming part of NASA's data environment.

NARA published, in the Federal Register, its revised Electronic Records Management procedures as 36 Code of Federal Regulations (CFR) Part 1234 on May 8, 1990 and NASA announced these regulations to its field installations on May 18, 1990. In addition, the General Services Administration (GSA) published revised guidelines on tape management on January 30, 1991 in Federal Information Resources Management Regulation (FIRMR) Bulletin B-1 entitled, "Electronic Records Management." These publications from NARA and GSA were much more current than those from the National Institute of Standards and Technology (NIST) and were used by NASA until NASA's handbook on Records Management was completed.

6. *Ensure that the offices and officials responsible for managing science data are identified and their responsibilities clearly defined.*

The revised NASA/OSSA Science Data Management Directive which was issued in March 1992, explicitly addresses the roles and responsibilities of the offices and officials responsible for managing science data. This effort was coordinated by the Information Systems Branch with the members of the Associate Administrator's Office and the data management committees of each OSSA Discipline Division.

Excerpted from that document:

4. Responsibilities and Authorities

a. Associate Administrator for Space Science and Applications
The Associate Administrator for Space Science and Applications is responsible for maintaining and ensuring the implementation of NASA's data management policy, including issuing implementing instructions and guidelines.

b. Assistant Associate Administrator for Science and Applications
The Assistant Associate Administrator for Science and Applications shall serve as chairperson of the Information Systems Management Board, which is chartered to coordinate OSSA's data management activities, and to identify issues, set priorities, and provide recommendations to the Associate Administrator for Space Science and Applications on these activities.

c. OSSA Discipline Division Directors
The Directors of OSSA's Science Discipline Divisions are responsible for the overall administration of their Division's data management activities in accordance with this Program Directive and the decisions of the Information Systems Management Board. Data acquired from both flight projects and non-satellite programs will be addressed as part of this responsibility. The primary objective of this activity is to assure the continuing value of OSSA's science data by providing data management procedures, systems and services that are responsive to the needs of the project, discipline, NASA, and broad research communities.

d. OSSA Flight Systems Division Director
The primary responsibilities of the Information Systems Program within OSSA's Flight Systems Division are to formulate and coordinate OSSA wide data management policy and to provide the supporting infrastructure across the discipline efforts. This includes providing a broad range of data management capabilities which transcend discipline-specific data management activities and serving as OSSA's point of contact for data management activities.

7. *Ensure that NASA officials responsible for overseeing NASA IRM periodically review NASA's data management and archiving to ensure compliance with NARA regulations.*

In FY 1990, IRM reviews included some elements which addressed data management and archiving. In FY 1991, NASA formally incorporated tape archiving requirements as part of the NASA IRM Review program, making physical inspection of tape storage facilities part of the regularly scheduled IRM reviews at its field installations. In addition, NARA, as part of its visits to the field installations to validate record holdings, conducts inspections of tape storage facilities. To date, NASA Headquarters has conducted inspections at the Kennedy Space Center and at the Jet Propulsion Laboratory in September 1990 and December 1990, respectively. Kennedy Space Center was found to have an excellent tape storage facility and the Jet Propulsion Laboratory was found to have made substantial progress in correcting its tape storage deficiencies. The Marshall Space Flight Center was inspected by NASA Headquarters in September 1991 and four additional field installations were inspected in FY 92.

8. *Ensure that NASA's data management and archiving are allocated adequate resources to properly store and maintain NASA's space science data holdings.*

The NASA/OSSA Data Management Initiative (DMI) is a multi-year, multi-million dollar effort that will ensure archiving of appropriate data from past missions, and creation of an infrastructure to enable the orderly archiving of data from future missions. The program will also ensure that data are preserved, inventoried and documented to facilitate broad future access by the science community.

The DMI program, coordinated by OSSA's Information Systems Branch (ISB) was proposed as a $4M new start for FY92. The program was funded and began in FY92 at the $1.8M level despite the absence of "new" dollars in the budget. Future funding for this activity out of the ISB budget has been projected as $1.4M, $1.1M and $750k for FY93-95 respectively. The Science divisions are investing similar amounts in associated efforts to enhance the data environment.

The funding in FY92 allowed three types of activities associated with the DMI to begin, including:

1) Identification and community assessment and prioritization of data sets in need of "restoration"
2) Restoration and/or archiving of appropriate data sets
3) Creation/Improvement of the capabilities and capacities of the Discipline Data Systems (DDS) and of the NSSDC, and of the procedures and tools whereby those entities assure the routine flow of increasing volumes of the right data into the OSSA archive environment (and retrievability of there from that environment).

Initial efforts in the early years of the DMI will focus on identification and restoration/archiving of appropriate data sets. Funding has been provided in FY92 at the sites of four space physics groups and six astrophysics groups for restoration and reformatting. In addition, the restoration program at JPL is continuing. Improvements to data management systems and facilities will be the focus of the DMI once important data sets have been safely archived. However, definition and development of new DDS's in Life Science, Microgravity, and Space Physics has begun under the DMI in 1992.

Those elements of the OSSA data management and archiving program relevant to the five recommendations from the November 1990 report will be discussed in the following paragraphs.

1. *Require NSSDC to identify and, if warranted and cost effective, obtain all outstanding archival data from past missions not yet delivered to its archives.*

   NASA has completed its two-phase data census to identify all data products from previous missions that should be archived. The census was initiated through the NSSDC and conducted in conjunction with working groups from each of the science disciplines. An element of the working groups contribution was to develop criteria and procedures for reviewing and establishing priorities to move selected data sets either to the NSSDC or to the appropriate discipline data center. During the first phase, data held by scientists and facilities associated with JPL and GSFC were surveyed. Phase one of the census was summarized in a report dated March 12, 1990. This report identified 294 data sets from 72 spacecraft. Phase two of the census surveyed more than 200 former Principal Investigators in addition to those involved in Phase one, in an effort to identify data suitable for archiving. While the survey response was less than 100%, it was determined that the extant data from over 80 % of the inactive investigations for which there were responses were fully archived at NSSDC.

   Data sets suitable for archiving were identified from the following missions:

Mission	Data	Source
IMP 7&8	750 library tapes	CalTech
HEAO-3	600 tapes	CalTech
OGO-5	100 tapes	Univ. of Chicago
Spacelab-2	100 tapes	Univ. of Chicago
IMP 6&8	2,555 digital experimenter tapes, 11,000 analog tapes, film data	Univ. of Iowa
Viking Lander 1&2	Meteorology data	Univ. of Washington
OGO-4	50 books of strip charts	NRL
Mariner 4	Wind data report	MIT
Solrad-HI	30 tapes	MIT
Viking Orbiter 1&2	300 tapes	Stanford

   Additional activities to archive outstanding data are ongoing as part of the Data Management Initiative.

2. *Revise data management policy to (1) recognize the need to archive selected original data of potential long-term scientific value, and (2) specify archiving requirements for data produced by life science, microgravity, aircraft, balloon, and sounding rocket missions, and data from NASA instruments flown on Shuttle missions and foreign spacecraft.*

   A revised policy on science data management for the Office of Space Science and Applications (OSSA) was issued in March 1991. The need to archived selected original data of potential long-term scientific value is recognized in paragraph 3(b) of this policy:

"NASA shall establish and maintain archives to preserve and make accessible all valuable NASA science data and information. This system of data archives shall include easily accessible information about NASA's data holdings, guidance, and aids for locating and obtaining the data. A review process, including scientific community representation, shall be established to determine what data should be archived and to assure conformance with completeness and quality standards."

The scope and effect have been updated to deal with the full range of OSSA science programs, including life sciences, microgravity, and suborbital programs. This is reflected in Section 2 of the Policy Directive on Scope:

"This program directive is applicable to the management of all science data resulting from Office of Space Science and Applications sponsored research missions and programs."

3. *Ensure that all missions develop and submit approved PDMPs.*

The new OSSA Program Directive: "Policy for the Management of NASA Science Data" specifies that:

> *"All data being captured by NASA science projects and space flight missions shall be addressed in a Project Data Management Plan to assure the availability of data and supporting information on a timely basis for use by the science community...."*

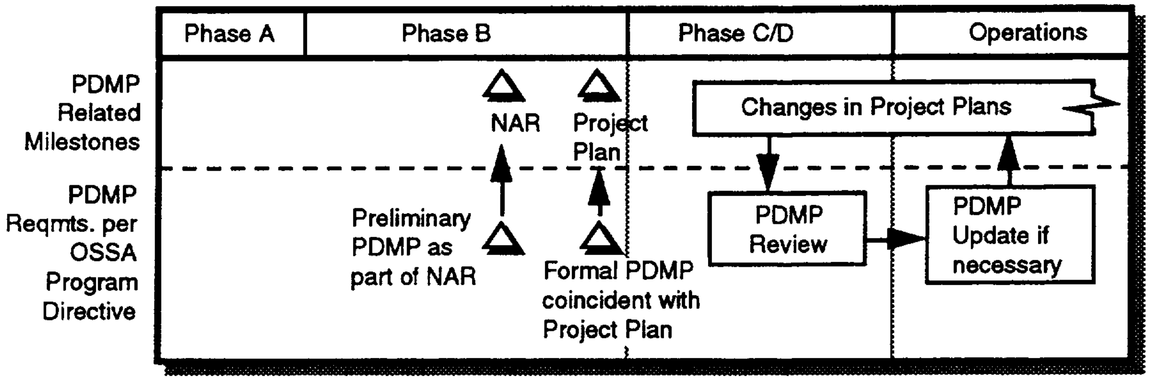
This directive emphasizes the importance of data archiving, but also addresses the broader problem of data management, with a strong emphasis on early planning. It calls for completion of PDMPs earlier in the project life cycle. The first version of the PDMP will be prepared in the same time frame as the Project Plan, shortly after new start approval for the project. It is envisioned that updates to the document will be made to reflect significant changes in data management planning throughout the period prior to launch, and throughout the mission operations and data analysis phase of the project. The formal definition (non advocate) review or equivalent mechanism conducted prior to a project receiving new start approval will assess data management plans as well as spacecraft development, instrument plans, etc.

PDMPs should be consistent with the framework established within this policy as well as with the data management plans of the relevant disciplines. Science discipline divisions are responsible for reviewing these plans as part of the new start approval, examine them on an on-going basis as projects develop, and, in the context of integrating project data plans, build toward a total research capability and for that discipline. Disciplines will be guided by the advice and counsel of their respective science communities to ensure the adequacy of plans for the flow of science data into discipline-oriented archives that serve the entire community. In addition, an OSSA-level review will be conducted annually to provide an integrated assessment of plans.

The essential functions of the PDMP are to:

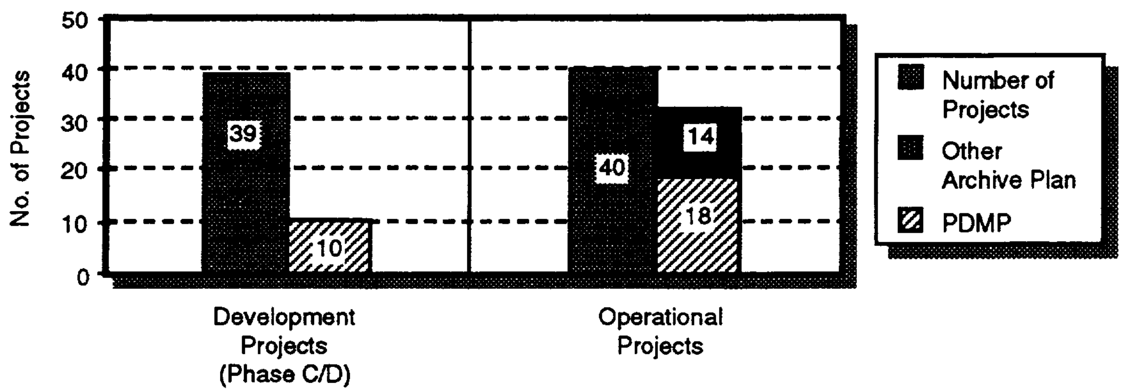
a. Provide consistent documentation to facilitate planning and implementation of science data management needs.
b. Identify and characterize all project data sets and indicate those which require archiving.
c Specify the time, location, and format for Project data and supporting documentation to flow into the OSSA archive environment.

The requirements for PDMP development over project phases are shown in the figure below.



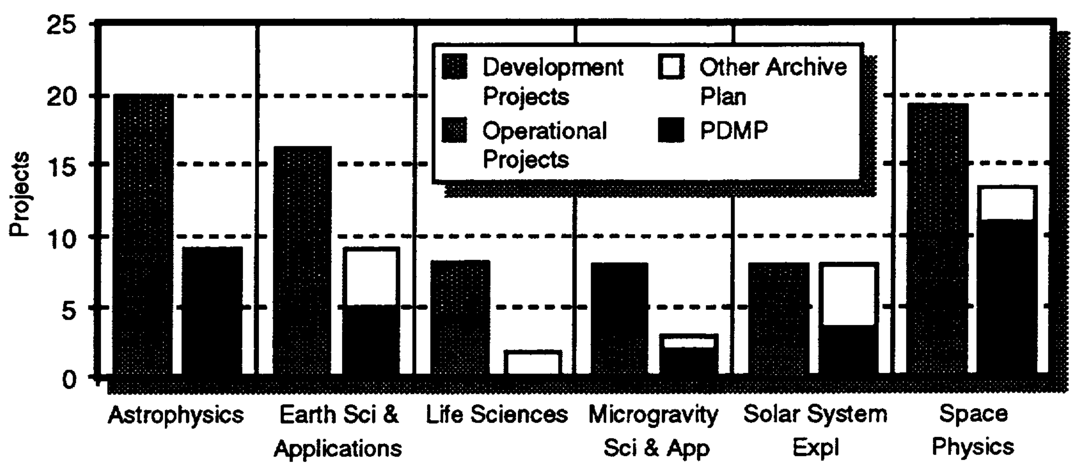
PDMP Requirements.

Current PDMP status for relevant OSSA missions, in the appropriate phase of development or operations is shown in Figure 12. From this information, it can be seen that there are currently 79 projects requiring PDMPs per OSSA policy. Of the projects requiring PDMPs, 28 projects have PDMPs while 14 projects have an alternative document that addresses archiving and data management plans. Many of these projects were in operation prior to the requirement for a PDMP.



Data Management Planning Status.

The status of PDMPs for each of the disciplines is shown below. The figure distinguishes between projects in Phase C/D and operational projects, although PDMPs are required for all these projects.



Current Data Management Planning Status by Discipline

OSSA is revising the "Guidelines for Developing a PDMP" as part of the overall update of data management policies. These guidelines, originally distributed to flight projects in 1988 to provide uniform guidance for developing plans, have contributed to the progress in generating effective plans. Based on experience to data, more specific guidance will be given in terms of key data management parameters to be addressed and tracked as the project develops. These parameters will provide a general overview of project data management requirements. The NIMS database at the NSSDC will be a primary resource for compiling data from all OSSA PDMPs. It incorporates parameters which can be used for planning at the spacecraft, experiment, and dataset levels. The PDMPs should address those parameters identified in the guidelines document so that the PDMP becomes a standard source of information for planning.

4. *Establish and enforce an internal controls system to ensure that original data are not destroyed until NSSDC has received all appropriate archival data.*

Procedures for assuring that lower level, "original" data products are retained until higher level data products are generated and archived have been evaluated in the context of the overall revision of data management policies. This procedure will be one of the key checklist items within the PDMP for flight projects (Note: In the architecture envisioned for future archives, data will flow to discipline archive centers as well as to the NSSDC.)

In addition, NASA is considering as agency policy to capture all "original" data (i.e., time-ordered experiment data with space network downlink artifacts removed) on a routine, production basis. This would serve as an institutional backup and provide a source outside individual project data systems to recover and reprocess data.

NASA, in conjunction with NARA, has updated NASA's Records Disposition Handbook, 1441JA, last published on December 1, 1970. This handbook contains all of NASA's record holdings, the length of their retention, and their disposition and destruction authorities. NARA archivists have conducted week long evaluations at each of NASA's field installations to validate record holdings, and schedules for retention, disposition and destruction.

NARA completed its review of NASA in June 1992. NASA expects to issue its printed Records Disposition Handbook in early 1993 after NARA approval (due by the end of 1992) . After approval by NARA it will be promulgated to all its field installations for implementation. Records management has been made an internal controls assessable unit, and enforcement will be through NASA's internal controls program.

5. *Determine what additional actions could be taken to (1) involve scientists more in the development and operation of mission data management systems, and (2) more strongly encourage missions to include participation of outside scientists on mission-level data management committees.*

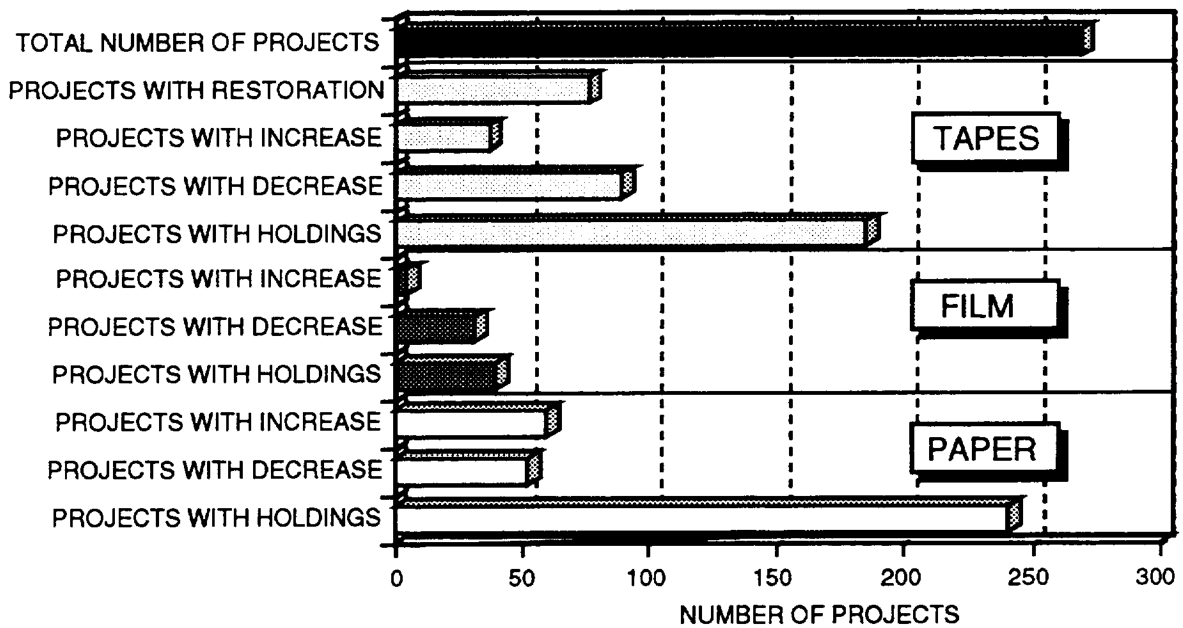
NASA recognizes science user involvement as a vital element for success and has implemented ways to strengthen that involvement. NASA is committed to maintaining an infrastructure for the preservation and distribution of data beyond the original mission science teams, and has involved the general science community in determining what data sets and products should be preserved and made accessible through open archives.

NASA currently has in place an advisory committee structure to drive specific data mission planning at all levels of program activity. There are broad advisory groups such as Space Science and Applications Advisory Committee (SSAAC), disciplinary advisory groups such as the Life Sciences Subcommittee (LSS) and mission-specific advisory groups. These advisory groups provide advice and guidance in all aspects of data management for all programs and projects within each discipline. Representation on these groups is reviewed to ensure the appropriate balance, and that the general science community interests are adequately represented.

Adequacy of the current groups, and the possibility of creating new groups or expanding the charters of current groups, to satisfy the requirements for community participation in the Data Management Initiative (assessing science value of census-identified data sets, etc.) is being discussed between the Information Systems Branch and OSSA Discipline Divisions.

NSSDC Holdings Summary

There are currently 269 science projects with holdings at the NSSDC. The status of the holdings for these projects was given in the table in Appendix IV of the November 1990 GAO report "Space Operations: NASA Is Not Archiving All Potentially Valuable Data." The figure below summarizes the changes in the holdings of these projects by three major media categories of tapes, film or paper. For each media, the number of projects with holdings in that media is provided, as well as the number of projects which have experienced an increase or decrease in holdings at the NSSDC since November 1990. For tapes, the number of projects which have undergone restoration of data from old tapes is shown as well.



Appendix E

GLOSSARY

Active Data Base	Subsets of data or complete data bases that are being actively used by the science community in ongoing research. Generally under the control of, and housed with the science community.
Ancillary Data	Non-science data needed to generate Level 1 data sets. Consists of instrument gains, offsets; pointing information for scan platforms, etc.
Correlative data	Other science data needed to interpret spaceborne data sets. May include ground-based data observations such as soil type or ocean buoy measurements of wind drift.
Data	Information of use to scientific investigations.
Data Acquisition	Process whereby basic data is received by a system.
Data Analysis	Process by which higher-level data products are derived from basic data acquired by instruments. Data analysis functions include modeling, manipulation, data interpretation, and data presentation.
Data Archive	Long-lived collections of science, operational and related ancillary data, maintained as a national resource at a data center, supported with adequate cataloging, protection, and distribution functions. It provides long-term access to data by the general space science community.
Data Base	The actual data, either part of an archive, repository, or active data base that is needed to do scientific research.
Data Catalog	Descriptions of data base in sufficient detail to retrieve subsets of data. Searchable by data fields or attributes, down to some level of granularity. Used to look or browse through a data base.
Data Directory	Top-level index containing information about location, ownership, contents of data. Used as first step in determining what types of data exist for given time, period, location, etc.
Data Handling	The process of data acquisition including onboard encoding and compression of data generated by flight sensors, data preprocessing on the ground to remove the artifacts of data transmission and conversion of raw data to Level 0 data, and management of this process to assure completeness and accuracy of
Data Retrieval	Process whereby data is transferred from a data storage center to a science user.
Data Retrieval System	Use of processing algorithms and software in order to access the archived/stored data.
Data Repository	Short-term data base that serves as a way station or clearinghouse for data - such as a mission data base to support operations and compilation of initial results. Temporary buffers for new data, usually existing only as long as the mission producing the data.
Data Set	The accumulation of data products, supplemental data, software, and documentation that will completely document and support the use of those data products. A data set can be part of a data set collection, can reside on a single physical volume or across multiple volumes.
Data Storage	Process whereby basic data or processed data is transferred to a stable medium prior to actual usage.
Data Storage Center	Archiving center where data is available for access by the science community.
Data Transfer	Process whereby data flows between systems/elements.
Decommutation	Process whereby the downlink data stream is split into data streams that contain data from only one or from select payloads or systems.
Element	Physical part of a system, which performs a function or functions.
Functions	Characteristic action of an element or group of elements of any system.

Level 0 Data	Reconstructed unprocessed instrument data at full resolution. Edited Data corrected for telemetry errors and split or decommutated into a data set for a given instrument. Sometimes called Experimental Data Record. Data are also tagged with time and location of acquisition.
Level 1A Data	Reconstructed unprocessed instrument data at full resolution, time-referenced, and annotated with ancillary information including radiometric and geometric calibration coefficients and geo-referencing parameters (i.e., platform ephemeris) computed and appended but not applied to the Level 0 data. Calibrated Data - Level 0 data that are still in units produced by instrument, but that have been corrected so that values are expressed in or are proportional to some physical unit such as radiance. No resampling, so Level 0 data can be reconstructed.
Level 1B Data	Level 1A data that have been processed to sensor units (i.e., radar backscatter cross section, brightness temperature, etc.). Not all instruments will have a Level 1B equivalent. Resampled Data - have been resampled in the time or space domains in such a way that the original edited data cannot be reconstructed. Could be calibrated in addition to be resampled (can also meet Level 1A definition).
Level 2 Data	Derived environmental variables (e.g., ocean wave height, soil moisture, ice concentration) at the same resolution and location as the Level 1 source data.
Level 3 Data	Variables mapped on uniform space-time grid scales, usually with some completeness and consistency (e.g., missing points interpolated, complete regions mosaiced together from multiple orbits)
Level 4 Data	Model output or results from analyses of lower level data (i.e., variables that are not measured by the instruments, but instead are derived from these measurements)
Metric	A quantitative parameter used to assess the performance of systems or functions/elements
Productivity	A metric relating output to the resources required for generating that output.
Raw Data	Telemetry data with data embedded
Secondary User	A researcher not involved with instrumentation design, development, or data acquisition. A secondary user would normally go to a data archive to obtain the required data set.
System	An integrated set of elements which performs the necessary functions to accomplish the desired operation.
Telemetry Services	Those activities required to convert the spacecraft downlink into data that is useful to the experimenter or investigator.
User Description	Description of why the data were acquired, any peculiarities with the data sets, and enough documentation to allow secondary user to extract information from the data.

Appendix F
ACRONYMS

ACE	Advanced Composition Explorer
ADC	Astronomical Data Center Affiliated Data Center
ADEOS	Advanced Earth Observation System
ADS	Astrophysics Data System
ARC	NASA Ames Research Center
ASP	Attached Shuttle Payload
ASTRO	Astronomical Laboratory
ATLAS	Atmospheric Laboratory for Applications and Science
AXAF	Advanced X-ray Astrophysics Facility
BBXRT	Broad Band X-ray Telescope
BUFR	Binary Universal Form for the Representation of Meteorological Data
CARB	Center for Advanced Research in Biotechnology
CASA	Center for Astronomy and Space Astrophysics
CDDIS	Crustal Dynamics Data Information System
CDF	Common Data Format
CDHF	Central Data Handling Facility
CD-ROM	Compact Disk Read-Only Memory
CEOS SS	Committee on Earth Observing Satellites Superstructure
CIESIN	Consortium for International Earth Sciences Information Network
COBE	Cosmic Background Explorer
COD-MAC	Committee on Data Management and Computation
COSTR	Collaborative Solar-Terrestrial Research
CRISTA	Cryogenic Infrared Spectrometer Telescope for Atmosphere
CRRES	Combined Release and Radiation Effects Satellite
DAAC	Distributed Active Archive Center
DDS	Discipline Data System
DIS	Data and Information System
DMI	Data Management Initiative
DXS	Diffuse X-ray Spectrometer
EDR	Experimental Data Record
EHIC	Energetic Heavy Ion Composition
EOC	Earth Observation Contrl
EOS	Earth Observing System
EOSDIS	EOS Data and Information System
ERBE	Earth Radiation Budget Experiment
EROS	Earth Resources Observation System
ESAD	Earth Science and Applications Division
ESDIS	Earth Science Data and Information System
EUVE	Extreme Ultraviolet Explorer
FAST	Fast Auroral Snapshot Explorer
FITS	Flexible Image Transfer System
FUSE	Far UV Spectroscopy Explorer
FY	Fiscal Year
GAO	Government Accounting Office
GGS	Global Geospace Science
GO	Guest Observer
GP-B	Gravity Probe B
GRIB	Gridded Binary
GRO	Gamma Ray Observatory
GSFC	NASA Goddard Space Flight Center
HDF	Hierarchical Data Format
HEAO	High Energy Astrophysics Observatory
HEAS-ARC	High Energy Astrophysics Science Archival Research Center
HNC	Heavy Nuclei Collector
HPCC	High Performance Computing and Communications
HST	Hubble Space Telescope
ICD	Interface Control Document
ICE	International Cometary Explorer
ICF	Instrument Control Facility
IEH	International Extreme-UV Far-UV (Hitchhiker)
IML	International Microgravity Laboratory
IMP-8	Interplanetary Monitoring Platform-8
IPAC	Infrared Processing and Analysis Center
IRAS	Infrared Astronomical Satellite
ISB	Information Systems Branch
ISO/OSI	International Standards Org./Open Systems Interconnection
ISSP	Information Systems Strategic Planning Project
ISTP	International Solar-Terrestrial Physics
IUE	International Ultraviolet Explorer
JPL	Jet Propulsion Laboratory
JSC	NASA Johnson Space Center
KPD	Key Parameter Data

LAGEOS	Laser Geodynamics Satellite
LaRC	NASA Langley Research Center
LeRC	NASA Lewis Research Center
MD	Master Directory
MIT	Massachusetts Institute of Technology
MO&DA	Mission Operations & Data Analysis
MSAD	Microgravity Science & Applications Div.
MSFC	NASA Marshall Space Flight Center
NAIF	Navigation Ancillary Information Facility
NASA	National Aeronautics and Space Administration
NCCS	NASA Center for Computational Sciences
NCDS	NASA Climate Data System
NDADS	NSSDC Data Archive and Distribution Services
NETCDF	Network Common Data Format
NEW-PIMS	Neutral Environment With Plasma Interaction Monitoring System
NIST	National Institute of Standards and Technology
NMI	NASA Management Instruction
NMSU	New Mexico State University
NODS	NASA Ocean Data System
NOST	NASA/OSSA Office of Standards and Technology
NRA	NASA Research Announcement
NSBF	National Scientific Balloon Facility
NSCAT	NASA Scatterometer
NSF	National Science Foundation
NSI	NASA Science Internet
NSP	NASA Support Plan
NSSDC	NASA Space Science Data Center
OAST	Office of Aeronautics and Space Technology
OCD	Operations Concept Document
ORFEUS-SPAS	Orbiting and Retrievable Far and Extreme UV Spectrometer
OSSA	Office of Space Science and Applications
PDMP	Project Data Management Plan
PDS	Planetary Data System
PDS Labels	Planetary Data System Labels
PI	Principal Investigator
PIP	Payload Integration Plan
PLDS	Pilot Land Data System
PSASS	Planetary Science Analysis Support System
PSCN	Program Support Communications Network
PSU	Pennsylvania State University
Radarsat	Radar Satellite
ROSAT	Roentgen Satellite
SAMPEX	Solar, Anomalous, and Magnetospheric Particle Explorer
SAO	Smithsonian Astrophysical Observatory
SAR	Synthetic Aperture Radar
S/C	Spacecraft
SDTS	Spatial Data Transfer Standard
SDU	State of the Data Union
SeaWiFS	Sea-Viewing Wide Field Sensor
SFDU	Standard Formatted Data Units
SIRD	Support Instrumentation Requirements Document
SIRTF	Space Infrared Telescope Facility
SLS	Space Life Sciences Laboratory
SMEX	Small Class Explorers
SOHO	Solar and Heliospheric Observatory
SPAN	Space Physics Analysis Network
SPD	Space Physics Division
SPDS	Space Physics Data System
SPDS/SC	SPDS/Steering Committee
SPTN	Shuttle Pointed Autonomous Research Tool For Astronomy
SRL	Space Radar Laboratory
SSBUV	Shuttle Solar Backscatter UV Instrument
STScI	Space Telescope Science Institute
SWAS	Submillimeter Wave Astronomy Satellite
SwRI	Southwest Research Institute
TDRSS	Tracking and Data Relay Satellite System
TOMS	Total Ozone Mapping Spectrometer
TOPEX	Ocean Topography Experiment
TRMM	Tropical Rainfall Measurement Mission
TSS	Tethered Satellite System
TSSF	Tape Staging and Storage Facility
UAF	University of Alaska / Fairbanks
UARS	Upper Atmosphere Research Satellite
UCLA	University of California - Los Angeles
UHRXS	Ultra High Resolution Extreme Ultra Violet Spectroheliograph
UMSOC	University of Maryland Space Operations Center
USGS	United States Geological Survey
USML	United States Microgravity Laboratory
USMP	United States Microgravity Payload
WISP	Waves in Space Plasma
WORM	Write Once, Read Many
XTE	X-Ray Timing Explorer

Appendix G

REFERENCES

1) "Data Management and Computation, Volume 1: Issues and Recommendations", CODMAC, 1982

2) "Issues and Recommendations Associated with Distributed Computation and Data Management Systems for the Space Sciences", CODMAC, 1986

3) "Space Research Data Management in the National Aeronautics and Space Administration", George H. Ludwig, NASA Technical Memorandum 89403, November 1986

4) "Earth System Science: A Closer View", Report of the Earth System Sciences Committee, NASA Advisory Council, NASA, January 1988

5) "Astrophysics Data System Study", March 1988

6) "Report of the Information Systems Strategic Planning Project", NASA OSSA/OSO, January 1990

7) "NASA Is Not Properly Safeguarding Valuable Data From Past Missions", GAO/IMTEC-90-1, March 1990

8) "NASA Is Not Archiving All Potentially Valuable Data", GAO/IMTEC-91-3, November 1990

9) "Space Physics Missions Handbook", NASA OSSA Space Physics Division, February 1991

10) "1990 Annual Statistics and Highlights Report", NSSDC, Doc. No. NSSDC/ WDC-A-R&S 91-04, March 1991

11) "Office of Space Science and Applications Strategic Plan 1991", April 1991

13) Joseph H. King, "Report on Phase Two of 1990 OSSA Data Census", NSSDC, Doc. No. NSSDCA/ WDC-A-R&S 91-23, October 1991

12) "OSSA Information Systems Program Overview Presentation", April 1992

14) "Space and Earth Sciences Information Systems: Issues and Opportunities for the 90's", January 1992

15) "1991 Annual Statistics and Highlights Report", NSSDC, Doc. No. NSSDC/ WDC-A-R&S 92-06, April 1992

16) "OSSA Information Systems Branch Strategies For The 1990's", April 1992

17) "Flight Project Data Book, 1992 Edition", NASA OSSA, 1992

18) "Mission Requirements and Data Systems Support Forecast", NASA GSFC, Doc. No. 501-803, June/July 1992