Ensuring the Climate Record from the NPP and NPOESS Meteorological Satellites

DETAILS

71 pages | 8.5 x 11 | null
ISBN null | DOI 10.17226/12263

AUTHORS

Committee on Earth Studies, Commission on Physical Sciences, Mathematics and Applications, National Research Council

Ensuring the Climate Record from the NPP and NPOESS Meteorological Satellites

Committee on Earth Studies
Space Studies Board
Division on Engineering
and Physical Sciences

National Research Council

APPENDIXES

National Academy Press, 2000

Ensuring the Climate Record from the NPP and NPOESS Meteorological Satellites

NOTICE: The project that is the subject of this report was approved by the Governing Board of the National Research Council, whose members are drawn from the councils of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine. The members of the committee responsible for the report were chosen for their special competences and with regard for appropriate balance.

Support for this project was provided by National Aeronautics and Space Administration contract NASW-96013, and National Oceanic and Atmospheric Administration contracts 50-DGNE-5-00210 and 50 DKNA-6-90040. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the sponsor.

International Standard Book Number 0-309-06994-7

Copies of this report are available free of charge from:

Space Studies Board
National Research Council
2101 Constitution Avenue, NW
Washington, DC 20418

Printed in the United States of America

engineering programs aimed at meeting national needs, encourages education and research, and recognizes the superior achievements of engineers. Dr. William A. Wulf is president of the National Academy of Engineering.

The **Institute of Medicine** was established in 1970 by the National Academy of Sciences to secure the services of eminent members of appropriate professions in the examination of policy matters pertaining to the health of the public. The Institute acts under the responsibility given to the National Academy of Sciences by its congressional charter to be an adviser to the federal government and, upon its own initiative, to identify issues of medical care, research, and education. Dr. Kenneth I. Shine is president of the Institute of Medicine.

The **National Research Council** was organized by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and advising the federal government. Functioning in accordance with general policies determined by the Academy, the Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in providing services to the government, the public, and the scientific and engineering communities. The Council is administered jointly by both Academies and the Institute of Medicine. Dr. Bruce M. Alberts and Dr. William A. Wulf are chairman and vice chairman, respectively, of the National Research Council.

Ensuring the Climate Record from the NPP and NPOESS Meteorological Satellites

Membership and Acknowledgment of Reviewers

COMMITTEE ON EARTH STUDIES

MARK R. ABBOTT, Oregon State University, *Chair*
JOHN R. CHRISTY, University of Alabama, Huntsville
CATHERINE GAUTIER, University of California, Santa Barbara
CHRISTOPHER O. JUSTICE, University of Virginia
RALPH F. MILLIFF, National Center for Atmospheric Research
SCOTT PACE, RAND
DALLAS L. PECK, U.S. Geological Survey (retired)
MICHAEL J. PRATHER, University of California, Irvine
R. KEITH RANEY, Johns Hopkins University Applied Physics Laboratory
DAVID T. SANDWELL, Scripps Institution of Oceanography
LAWRENCE C. SCHOLZ, West Orange, New Jersey
CARL F. SCHUELER, Raytheon Santa Barbara Remote Sensing
GRAEME L. STEPHENS, Colorado State University
FAWWAZ T. ULABY, University of Michigan
SUSAN L. USTIN, University of California, Davis
FRANK J. WENTZ, Remote Sensing Systems
EDWARD F. ZALEWSKI, University of Arizona

ARTHUR A. CHARO, Senior Program Officer
THERESA M. FISHER, Senior Project Assistant

SPACE STUDIES BOARD

JOHN H. McELROY, University of Texas at Arlington (retired), *Chair*
MARK R. ABBOTT, Oregon State University
FRAN BAGENAL, University of Colorado
JAMES P. BAGIAN, Veterans Health Administration's National Center for Patient Safety
JAMES L. BURCH, Southwest Research Institute
ROBERT E. CLELAND, University of Washington
STEVEN H. FLAJSER, Loral Space and Communications Ltd.

COMMISSION ON PHYSICAL SCIENCES, MATHEMATICS, AND APPLICATIONS

ACKNOWLEDGMENT OF REVIEWERS

This report has been reviewed by individuals chosen for their diverse perspectives and technical expertise, in accordance with procedures approved by the National Research Council's (NRC's) Report Review Committee. The purpose of this independent review is to provide candid and critical comments that will assist authors and the NRC in making the published report as sound as possible and to ensure that the report meets institutional standards for objectivity, evidence, and responsiveness to the study charge. The contents of the review comments and draft manuscript remain confidential to protect the integrity of the deliberative process. The committee wishes to thank the following individuals for their participation in the review of this report:

Jeff Dozier, University of California at Santa Barbara,
David M. Glover, Woods Hole Oceanographic Institution,
Moshe Pniel, Jet Propulsion Laboratory,
Eugene M. Rasmusson, University of Maryland, and
John Townshend, University of Maryland.

Although the individuals listed above have provided many constructive comments and suggestions, responsibility for the final content of this report rests solely with the authoring committee and the NRC.

Ensuring the Climate Record from the NPP and NPOESS Meteorological Satellites

Foreword

Satellites planned for the National Polar-orbiting Operational Environmental Satellite System (NPOESS) will serve civilian and defense needs for timely environmental data. These data notably support weather forecasting and "nowcasting," but, if appropriately calibrated and archived, they could support the examination of global change as well. The Department of Defense and the National Oceanic and Atmospheric Administration are the principal agencies responsible for NPOESS, but the National Aeronautics and Space Administration provides the technology for new generations of sensors and is the developer and operator of research satellites addressing global environmental change. The processing and archiving of data from NPOESS and its precursor mission, the NPOESS Preparatory Project (NPP), and the merger of these data with research satellite data for long-term preservation and access are the focus of this report.

Assessing and—ultimately—forecasting global environmental change is a difficult scientific problem with important consequences for public policy. The resolution of both the scientific and public policy questions requires long-term data sets that are carefully defined and well calibrated. The only efficient and cost-effective means to obtain a major part of the satellite data segment of the long-term record will be to employ the data available from NPP and NPOESS. Ensuring the suitability of these data for analyses of global change will, however, require early consideration by both instrument developers and data archivists.

The key recommendations of this brief study concern the creation of a long-term archive that will enable preservation of the climate record from NPP and NPOESS. Long-term archives must not, however, be dusty repositories of data; they should instead be active centers for data study and utilization, with intensive involvement by the research community. The stored data need to be robust and capable of improvement as scientific understanding progresses and computer technology evolves. The data must also be readily accessible to the research community and affordable for them. The challenges associated with the creation of such an active data archive can be met only through collaboration between the operational and research communities. The nature of the required collaboration is discussed in the report that follows.

John H. McElroy, *Chair*
Space Studies Board

Ensuring the Climate Record from the NPP and NPOESS Meteorological Satellites

Executive Summary

INTRODUCTION

Researchers studying the issues surrounding global climate change have a particular need for the kind of repetitive, long-term, high-quality measurements that can be provided from the vantage point of space. Operational weather satellites provide perhaps the only means for securing these measurements. The next generation of operational sensing systems is currently being designed, and the National Polar-orbiting Operational Environmental Satellite System (NPOESS), scheduled for launch beginning in 2009, is an important component of this operational monitoring system. NPOESS is being developed with the goal of meeting the converged operational data needs of the National Oceanic and Atmospheric Administration (NOAA) and the Department of Defense (DOD), as well as some of the data needs of the National Aeronautics and Space Administration (NASA) Earth observation programs.

In a joint mission to facilitate the transition of appropriate "research" satellite measurements into the operational domain, NASA and the NPOESS Integrated Program Office (IPO) are developing the NPOESS Preparatory Project (NPP). NASA and NOAA are supporting the NPP as part of a program of risk reduction demonstration and validation for NPOESS sensors, algorithms, and processing. The NPP satellite, scheduled for launch in 2005, will include critical sensors that are planned for flights on NPOESS. In addition, the NPP mission is expected to provide an early test of space and ground segments for NPOESS.

The NPOESS IPO has begun working with the members of the climate research community to define operational climate measurement needs. The IPO has also begun to assess the implications of these needs for NPOESS instrument design. However, it is equally important to ensure that the data systems will meet climate research needs. At the request of NOAA and NASA, the Space Studies Board's Committee on Earth Studies conducted a short-duration study of issues related to ensuring the climate record from the planned NPP and NPOESS satellites (see Appendix A for a statement of task). This report presents the committee's recommendations; it draws heavily on background material presented at the 2-day workshop that the committee hosted on February 7-8, 2000, and on discussions during and after the workshop.[1] It also draws on investigations by the committee for the two-part report *Issues in the Integration of Research and Operational Satellite Systems for Climate Research*.[2]

Climate Data Records

In briefings to the committee, NASA and NOAA officials acknowledged that there is no operational ground system infrastructure for U.S. climate data and services. The climate research community therefore requires satellite data from NPP and NPOESS that can be used to generate climate data records (CDRs), data whose quality is known quantitatively and for which temporal and spatial biases are minimized (or at least quantified). CDR production will require considerable scientific insight, including the blending of multiple data sources, error analyses, and access to raw data sets. Moreover, information on sensor design, operation, and calibration will also be necessary to develop a consistent CDR across multiple sensors.

NASA intends for environmental data records (EDRs)[3]—the priority data products that will be produced from NPOESS data—to be utilized to the maximum extent possible to meet CDR requirements. However, NASA also expects to cap the resources available for CDR processing to ensure that the EDR production requirements are met. Although NPP- and NPOESS-derived EDRs may have considerable scientific value, CDRs are far more than a time series of EDRs. While the lines may be indistinct, there remain fundamental differences between products that are generated to meet short-term needs (EDRs) and those for which consistency of processing over years to decades is an essential requirement. Given the experience of climate researchers, it is unlikely that the standard EDRs will meet the quality requirements for CDRs, particularly in the area of data refinement and reprocessing as algorithms mature. Moreover, production and refinement of CDRs through reprocessing may be difficult (or unaffordable) in the present plans.

Long-Term Archiving and the National Climatic Data Center

NOAA is the federal agency with responsibility for archiving environmental satellite data, and its National Climatic Data Center (NCDC) is a potential repository of data to support climate research in the coming NPP/NPOESS era. Currently, NCDC has a total digital archive of approximately 700 terabytes (TB). Data from the scheduled launch in 2005 of the NPP satellite will add another 90 TB annually; if managed by NCDC, data from NPOESS in the 2009 time frame would add yet another 228 TB annually.

There are currently no funds to archive NPP or NPOESS data, and although NASA and NOAA have a memorandum of understanding (MOU) regarding the eventual archiving of Earth Observing System (EOS) data by NOAA, it is on a best-effort basis. Although plans are being discussed, there is no implementation strategy within NASA or NOAA to archive even the raw data records (RDRs, analogous to Level 1 data) from these missions. There are also no plans to store sensor design information and calibration and ancillary data necessary to develop CDRs. Ominously, from a climate research perspective, the prospects for developing CDR and long-term archive (LTA) plans for NPP and NPOESS are ever more doubtful, given that plans for the EOS Terra and Aqua (formerly known as AM-1 and PM-1) data sets are not yet firmly in place.

Guiding Principles

NASA and NOAA have experienced both success and failure in recent attempts to develop data systems. The committee believes that much can be learned from these experiences but notes that the fundamental objective of establishing a set of data systems and services to meet the needs of climate research will require more than MOUs and larger magnetic tape silos. New services must be supported that are not available in the present mix of NASA research mission data systems and NOAA long-term archives. New scientific and policy demands are being placed on these systems, and new management and technical approaches must be established. Increasing funding is a necessary condition; however, the committee does not believe that funding alone is sufficient. *While encouraged by NASA and NOAA's recent attention to preserving the*

climate record of NPP, NPOESS, and EOS, the committee believes that an enormous investment in Earth observations is at serious risk.
Based on an examination of prior studies, as well as discussions at its February 7-8, 2000, workshop, the committee identified a set of principles it believes can help ensure the preservation of the climate record from NPP and NPOESS:

Accessible and policy-relevant environmental information must be a well-maintained part of our national scientific infrastructure.

The federal government should (1) provide long-term data stewardship, (2) certify open, flexible standards, and (3) ensure open access to data. The government does not necessarily need to control the implementation of every task and service for a climate data system. Rather, it should undertake those activities and services that cannot be done in a competitive academic or commercial environment.

Because the analysis of long-term data sets must be supported in an environment of changing technical capability and user requirements, any data system should focus on simplicity and endurance.

Adaptability and flexibility are essential for any information system if it is to survive in a world of rapidly changing technical capabilities and science requirements.

Experience with actual data and actual users can be acquired by starting to build small, end-to-end systems early in the process. EOS data are available now for prototyping new data systems and services for NPP and NPOESS.

Multiple sources of data and services are needed to support development of climate data records (CDRs). The quality of the CDRs will improve as more research groups work with the various input data sets, and the overall system will be more robust if it does not rely on a monolithic implementation. Fostering open competition for services promotes innovation and new ideas.

Science involvement is essential at all stages of development and implementation. Having climate data record developers and users assisting in the specification, design, building, and testing of the system will help ensure its usefulness to the research community.

RECOMMENDATIONS

The committee's study of issues related to the preservation of the climate record from future NASA and NOAA satellites was necessarily brief and drew heavily on previous work and the 2-day workshop. In addition, the committee drew on the lessons learned to date with NASA's EOS Data and Information System. Underlying the committee's recommendations is its belief in the critical need and unique potential for data from NPP and NPOESS to satisfy the demands of the climate research community. In particular, the committee believes that prudent planning and modest investments early in the program will allow the NPOESS system to continue essential climate research-quality data records and develop new records based on the rich blend of planned instruments.

Climate research will require a variety of services, ranging from careful long-term stewardship of the basic data sets to intensive data analysis and algorithm refinement. The committee believes these complex scientific and information system activities are best broken into two more manageable pieces-the long-term archive (LTA) and the active archive-instead of formed into a comprehensive, single-system solution. Climate research requires an integration of the more stable, long-term functionality of the LTA and the flexibility of the active archives to pursue and develop new capabilities.

The first four recommendations of the committee are presented in order of priority. The last four, which are not prioritized, focus on programmatic and management structures to meet these essential requirements for a climate data system.

Recommendation 1. NOAA should begin now to develop and implement the capability to preserve in perpetuity the basic satellite measurements (radiances and brightness temperatures).

The development of long-term, consistent time series based on CDRs requires access to the lowest level of data available. In general, this means the raw data records (RDRs), or Level 1A data. The low-level data can be used to develop refined CDRs as scientific and technical understanding of Earth processes and sensor performance improves over time. The committee recommends that NOAA do the following:

Archive both current and future data sets, including those from both research and operational satellite missions, in an LTA.

Archive information on sensor development, calibration, operation product validation, and appropriate metadata along with the basic radiances.

Migrate data sets to new, computer-compatible media on a regular basis, such that data sets are refreshed every 2 to 3 years consistent with the pace of technology evolution.

Organize data in the LTA based on user access patterns to optimize data retrieval.

Recommendation 2. NOAA should guarantee climate researchers affordable access to all RDRs in the long-term archive, with an emphasis on large-volume data access.

Development of CDRs requires access to enormous data volumes, but it is likely that only a small number of researchers will need such extensive access to the raw data. Thus, a well-designed set of basic services would meet this basic function without being too costly. The committee recommends that NOAA act on the following:

Award the LTA functions on a competitive basis to both government and private organizations to promote innovation.

Start the development of the LTA immediately with a simple set of end-to-end capabilities to gain experience and modify the plans and implementation accordingly. (End-to-end is defined as being from sensor aperture to the desktop of the climate information user.)

Recommendation 3. NASA, in cooperation with NOAA, should support the development and evaluation of CDRs, as well as their refinement through data reprocessing.

Because the CDR process is driven by science understanding, there will be a continuing need for the involvement of researchers. The NOAA/NASA Pathfinder shows that the agencies can generate critical data sets for transitioning research products into operational data products. Over the next decades, the committee expects that a few experimental CDRs may become effectively "operational" products and will be produced by NOAA. The committee recommends that NASA, in cooperation with NOAA, take action as follows:

Periodically select science investigations and provide adequate support to develop and evaluate new CDRs.

Preserve sensor calibration and operating information, as well as metadata and ancillary data fields, in a manner that allows reprocessing the CDRs.

Evaluate on a regular basis the organization of data sets in the LTA in light of actual data usage patterns to improve reprocessing and access efficiency.

Recommendation 4. NOAA and NASA should define and develop a basic set of user services and tools to meet specific functions for the science community, with NOAA assuming increasing responsibility for this activity as data migrates to the long-term archive.

NASA's Distributed Active Archives Centers, as well as components of NASA's Earth Science Information Partners, are gaining experience with responding to data requests and setting up user services. Although the focus is on the order entry process (catalog, data location, browse, etc.), more attention needs to be given to quality assurance and the order fulfillment process (metadata, subsetting, electronic data delivery, etc.). Emphasis should be given to reducing cost through automation. It is essential that the large-volume data sets from the archive be affordable for the science user community. The committee recommends that NOAA and NASA do the following:

Select teams on a competitive basis that will identify and provide specific user services and tools (see Appendix D). As part of an ongoing process of system evaluation and improvement, these teams would assist in identifying and providing essential user services. Based on a rigorous analysis of a user model for climate research, they would make recommendations on characteristics such as data subsetting and browse capabilities.

Support and maintain a balance between internal and external expertise at the government data centers.

Examine the feasibility of providing open electronic access to a rolling archive of RDRs and EDRs through the NESDIS Central that is planned for NPOESS and NPP.

Recommendation 5a. NASA, in cooperation with the Integrated Program Office, should develop the NPOESS Preparatory Project as an integral component of a climate data system.

NPP represents a unique opportunity to test both scientific and programmatic interfaces related to an integrated data systems strategy. It will bridge the gap between the NASA research missions and the NPOESS operational missions. There is potential to begin the development of long-term, high-quality CDRs and an associated data system for climate research, but it is an opportunity that could be missed. The committee recommends that NASA, in cooperation with the NPOESS IPO, proceed as follows:

Develop and implement a prototyping activity to link the NPP Science Data Segment with the NOAA LTA. This activity should start with NASA EOS data sets, including Atmospheric Infrared Sounder (AIRS)/Advanced Microwave Sounding Unit (AMSU)/MODIS in anticipation of the NPP data sets, including CrIS/ATMS/VIIRS.

Put aside reserve funds from data system development to support evolutionary development activities as the program matures to ensure that the system is not locked in with no resources for subsequent enhancement.

Develop prototype user services for NPP climate data records.

Recommendation 5b. Select, on a competitive basis, and then support an NPP science team as soon as possible.

The team should consist of sensor experts, algorithm developers, and science data users. Because the functions will require different levels of involvement in the sensor development and operation process, they will require different levels of support. The team would advise on the NPP data system needs, including scientific data processing, archiving, and distribution requirements.

Recommendation 6. NOAA, in cooperation with NASA, should invest in early, limited capability prototypes for both long-term archiving and the NPP data system.

Data systems that do not develop, test, and evaluate on a frequent, regular basis are nearly always late and over budget. System development costs generally increase as the cube of the number of years in development. A climate data system will build on existing components and existing capabilities, but new functions and new interfaces must be developed and implemented to meet the requirements for climate research. The committee recommends that NOAA, in cooperation with NASA, take action as follows:

Competitively select and support a science data team to assist a NOAA long-term archiving program on the following:

—Archive requirements for long-term data sets, including RDRs, metadata, and ancillary data fields;
—Archiving of CDRs, algorithms, and processing environments;
—Data structure and organization to facilitate access and reprocessing;
—Flexible, open standards to facilitate data access and refinement;
—Data reprocessing priorities; and
—Minimal user services and tools.

Require the NPOESS total system performance requirements (TSPR) contractor to work with the science data team to facilitate CDR production and archiving from both NPP and NPOESS.

Develop flexible standards and formats that allow new services to be developed in the future.

Begin to develop a small number of CDRs using the LTA services.

Recommendation 7. NASA and NOAA should develop and support activities that will enable a blend of distributed and centralized data and information services for climate research.

NASA and NOAA should consider a hybrid mode of operation rather than building a rigid, centralized system or relying on structure to emerge from an uncoordinated set of data systems. The government should ensure and manage the activities it does best, while fostering innovation and flexibility in those parts of the overall system that do not need to be closely managed. The committee recommends that NASA and NOAA proceed as follows:

Implement the NPP data system as a federation of linked activities, such as that proposed in the NewDISS framework.

Where appropriate, build on existing and planned capabilities, including EOSDIS, the Earth Science Information Partners, NASA's Distributed Active Archive Centers, and NOAA's Data Centers, and develop new capabilities as user experience is gained.

[1]See Appendix B for the workshop agenda and a list of participants.

[2]National Research Council (NRC), Space Studies Board. 2000. *Issues in the Integration of Research and Operational Satellite Systems for Climate Research: I. Science and Design*, in press; National Research Council (NRC), Space Studies Board. 2000. *Issues in the Integration of Research and Operational Satellite Systems for Climate Research: II. Implementation*, in press.

[3]As defined by the NPOESS IPO, EDRs are data records that contain the environmental parameters or imagery required to be generated as user products as well as any ancillary data required to identify or interpret these parameters or images. EDRs are generally produced by applying an appropriate set of algorithms to raw data records.

Ensuring the Climate Record from the NPP and NPOESS Meteorological Satellites

1 Introduction

Researchers studying the issues surrounding global climate change have a particular need for the kind of repetitive, long-term, high-quality measurements that can be provided from the vantage point of space. Operational weather satellites provide perhaps the only means for securing these measurements. The next generation of operational sensing systems is currently being designed, and the National Polar-orbiting Operational Environmental Satellite System (NPOESS), scheduled for launch beginning in 2009, will be an important component of this operational monitoring system. NPOESS is being developed with the goal of meeting the converged operational data needs of NOAA and DOD, as well as some of the data needs of NASA's Earth observation programs.

In a joint mission to transition appropriate research satellite measurements into the operational domain, NASA and the NPOESS Integrated Program Office are developing the NPOESS Preparatory Project (NPP). NOAA is supporting the NPP as part of its risk reduction demonstration and validation for NPOESS sensors, algorithms, and processing. Included in the NPP is the planned launch in 2005 of critical sensors that are planned for flights on NPOESS. The intent is to develop an operational prototype for the provision of satellite-based climate data as well to conduct an early test of space and ground segments for NPOESS.

At this time, attention is being given to defining operational climate measurement needs and assessing their implications for instrument design. However, it is equally important to ensure that the data systems will meet climate research needs. At the request of NOAA and NASA, the Space Studies Board's Committee on Earth Studies conducted a short-duration study of issues related to ensuring that the planned NPP and NPOESS satellites will produce a suitable climate record. This report presents the committee's recommendations; it draws heavily on the background material presented at a 2-day workshop that the committee hosted on February 7-8, 2000,[1] and on the discussions that came out of that workshop. In addition, it draws on the committee's investigations in connection with a recent two-part report: *Issues in the Integration of Research and Operational Satellite Systems for Climate Research*.[2] Of particular relevance to the present report are discussions in the aforementioned reports of how to execute long-term science monitoring in an operational framework and the particular needs of climate researchers for long-term, well-calibrated, reliable data sets.

Investments of several billion dollars have been made in the past 10 years for both NASA and NOAA satellite data systems. The committee believes it is imperative that the nation build on these investments as well as learn from past successes and failures. Any data and information system that intends to serve climate research needs should, if possible, be designed and implemented in the context of existing and planned systems. Moreover, a climate research information strategy will need to address two challenges that have not yet been addressed by the present constellation of satellite data systems. First, the volumes of data that must be archived and made accessible for decades to centuries are far larger than present long-term archives such as the archive at the NOAA National Climatic Data Center (NCDC). Second, much of climate research requires long-term, consistent (or at least well-characterized) time series.[3]

NOAA is the federal agency with the responsibility for archiving environmental satellite data, and its NCDC is a potential repository for data to support climate research in the coming NPOESS era. NCDC presently archives about 2 terabytes (TB) per year of Polar-orbiting Operational Environmental Satellite (POES) data. However, as a repository for NPP and NPOESS data, it would need to archive approximately 2 TB per *day* of data. There are presently no funds for NOAA to archive NPP or NPOESS data. In addition, there are no firm plans to archive sensor design information and the calibration and ancillary data necessary to develop what the committee terms "climate data records" (CDRs).

Climate research will require a variety of services, ranging from stable, long-term stewardship of the basic data sets to intensive data analysis and algorithm refinement. As discussed in subsequent chapters, the committee believes it would be best to break these complex scientific and information system activities into two more manageable pieces—the long-term archive (LTA) and the active archive—instead of attempting to build a comprehensive, single-system solution.

The next chapter of this report, Chapter 2, provides details of NASA and NOAA preliminary data system plans for NPP and NPOESS, including their plans for long-term and active archives. However, as noted in the text, these plans are not necessarily focused on the requirements of CDRs. Chapter 3 discusses the requirements for climate data services in the context of the existing plans for production, archiving, and distribution of CDRs. The committee's findings and recommendations are presented in Chapter 4. The acronyms and abbreviations used in the report are listed in Appendix C.

[1]See Appendix B for the workshop agenda and a list of participants.

[2]National Research Council, Committee on Earth Studies. 2000. *Issues in the Integration of Research and Operational Satellite Systems for Climate Research: I. Science and Design*, in press; *Issues in the Integration of Research and Operational Satellite Systems for Climate Research: II. Implementation*, in press.

[3]National Research Council, Board on Sustainable Development. 1999. *Global Environmental Change: Research Pathways for the Next Decade*. Washington, D.C.: National Academy Press.

Ensuring the Climate Record from the NPP and NPOESS Meteorological Satellites

2 Data Systems Plans

INTRODUCTION

This chapter focuses on the current data processing plans of the various organizations involved in the National Polar-orbiting Operational Environmental Satellite System (NPOESS) program, including the NPOESS Integrated Program Office (IPO) as well as NASA and NOAA's National Environmental Satellite, Data, and Information Service (NESDIS). The present satellite data systems, such as EOSDIS and the system planned for NPOESS, are designed primarily to meet needs for rapid access to small data sets, such as those for short-term forecasting. This chapter will concentrate on the present and planned operational capabilities, evaluated in the context of climate research requirements. The background material in this chapter draws heavily on presentations to the committee at its February 7-8, 2000, workshop.

NPOESS AND NPP PLANS

The NPOESS satellites are being designed with a wide variety of sensors that will provide data that will be used to generate 61 required environmental data record (EDR) products.[1] The NPOESS sensors will be supported by a sophisticated ground data processing architecture, which the NPOESS IPO calls the Interface Data Processing Segment (IDPS). The IDPS will process data from NPOESS that will be generated at some 10 times the combined data rates of the current polar weather satellites. It will also process a plethora of data from auxiliary sources (e.g., climatology records and other satellite data, such as those from the Geostationary Operational Environmental Satellite) to produce the EDR products.

Interface Data Processing Segment

The NPOESS IPO is working with the NPOESS Preparatory Project (NPP) at NASA under two competing contracts for the Program Definition and Risk Reduction phases.[2] The IPO expects to develop an Interface Data Processing Segment (IDPS) under a Total System Performance Responsibility (TSRP) contract starting in April 2002. Information on the IDPS was presented to the committee at its February workshop; it is summarized here to elucidate the context and limitations of the IPO's capabilities to service the science data user during the NPOESS era.[3]

Goals and Objectives

The IDPS would incorporate early delivery and risk-reduction relative to the nominal 2008 C1 NPOESS satellite launch by providing service for the prototype NPP spacecraft containing the NASA-supplied Advanced Technology Microwave Sounder (ATMS), the IPO-supplied Cross-Track Infrared Sounder (CrIS), and the Visible IR Imaging Radiometer Suite (VIIRS), as well as an instrument of opportunity yet to be selected. As such, the IPO is working with NASA to develop an NPP prototype IDPS capability, known for now as the Science Data Segment (SDS). The NPP prototype IDPS (alias SDS) is intended to provide early user evaluation of NPOESS data products and to allow algorithm and sensor verification and opportunities for sensor calibration and validation. Further, more than half the 61 NPOESS EDRs would be supplied by the prototype NPP capability. This would allow algorithm modifications and improvements prior to the first NPOESS launch (C1); more specifically, it would provide prototype demonstration and use of the NPOESS imager and sounder EDRs. It is important to recognize that the NPOESS program is focused entirely on the generation of EDRs; therefore, insofar as the EDRs do not meet the requirements of climate science, much of the SDS will not meet them either.

Context and Interfaces

The NPOESS IPO is currently conducting a phase I ground system demonstration and prototype effort with two competing contractors. Phase I efforts are providing early definition of the IDPS external interfaces, an end-to-end processing demonstration, and overall ground processing architectures and trade-offs. How these efforts progress over the next 2 years will determine if the IPO can realize its desire for an efficient, flexible, and modular system that will allow future upgrades at reasonable cost. Phase II, engineering and manufacturing development (EMD), is scheduled to start in April 2002, at which time details of the IDPS design will be available.

Key Components and Functional Flow

The key components of the IDPS include the following:

- Data ingest of raw data records (RDRs);
- Data processing of RDRs to sensor data records (SDRs) and EDRs;
- Data and EDR quality and assurance;
- Data management, data storage, and user access, retrieval, and distribution; and
- Remote field terminals.

In accord with the NPOESS mission, the IDPS is being designed to emphasize operational user support. The program's mandate to support the operational user and the constraints imposed by the budget have resulted in plans to limit the IDPS to ingestion and processing of data on a near-real-time basis and an ability to store data for no more than 3 to 5 days, typically, and then only because some EDRs are expected to require the application over that limited time period of data acquired earlier. This limitation clearly conflicts with the needs of research data users.

Open Items

Although the IPO is well under way in its planning, the following key aspects of the IPO Interface Data Processing Segment have not been well defined:

- Comprehensive postlaunch calibration and validation of the sensors, the EDR algorithms, and the data products and

- Detailed NPP IDPS system definition.

Unique Aspects of the NPOESS Preparatory Project Data Processing Plans

NPP has been under definition for about 2 years, managed by NASA Goddard Space Flight Center. It is intended to provide data continuity between the EOS mission and the first NPOESS spacecraft launch, which will take place sometime between 2005 and approximately 2008, and to provide further risk reduction and prototype operations for critical aspects of the NPOESS mission.[4] A prototype data system is being developed by the IPO with NASA and NOAA participation. The system is intended to demonstrate many aspects of the NPOESS IDPS discussed above, as well as extra climate data processing capabilities through the planned Science Data Segment. NOAA's NESDIS hopes to also provide archiving for the anticipated data needs of the research community that are not planned as part of the NPOESS IDPS. (These plans are summarized below.) Funding for these efforts is expected to be provided by NASA through the NPOESS Preparatory Project, but only for about 3 years beyond the NPP satellite launch.

NPOESS Preparatory Project Mission Overview

NPP is a joint effort of NASA, NOAA, and the NPOESS IPO. [5] Funded and managed largely by NASA, it has two key objectives:

1. Data continuity for EOS Terra and Aqua through the NPOESS C1 launch and

2. Risk reduction for the NPOESS C1 mission.
If the only purpose of NPP were EOS data continuity, then NASA might focus on spacecraft development and plan to use EOSDIS to process the NPP data, letting IPO develop the IDPS to process NPOESS data starting in 2008. However, NPOESS risk reduction suggests that additional effort should be devoted to preparation for NPOESS operational data processing. Accordingly, a prototype IDPS is to be available before the NPP launch to exercise critical features of the operational IDPS, as well as to provide data continuity for the research community, which will be transitioning to dependence on NPOESS a few years later.
NASA is managing and funding much of the NPP effort; it is also working with the research user community because a portion of the NPP mission's data processing effort is intended to focus on provision of NPP-unique climate data processing. The NPP mission is also expected to include a NESDIS-supplied Archive and Distribution Segment (ADS) as per a pending MOU between NASA and NOAA's NESDIS.

NPP Mission Priorities

NPP will serve as a bridging mission from research-oriented EOS measurements to the NPOESS operational measurements, at the same time as it supports the NPOESS research component. Additionally, NPP will provide an opportunity to validate NPOESS EDRs as well as some yet-to-be-defined CDRs (see Box 2.1) and to improve operational algorithms for future use by NPOESS. It intends to rely heavily on operational agency units, such as the IPO and NOAA NESDIS, for data services. NASA, on the other hand, plans to fund CDR-specific algorithm development through competitive Announcement of Opportunity (AO) contracts to the science community.

A key aspect of the NPP mission is recognition of the difference between an operational and a research mission[6] and the limitations that will be imposed on the operational NPOESS mission. For example, NPOESS is expected to provide EDRs within 20 minutes of acquiring data from the spacecraft, so it will be forced in many cases to compromise the quality of the EDRs that might have been produced if tardy, but higher quality, auxiliary data were included in a more relaxed EDR production time line. While the operational community may generally find immediate data products more valuable than higher-quality deferred data products, the research community almost invariably is willing to suffer a significant product delay to ensure maximum quality.

As already noted, long-term archiving is not currently planned for the operational NPOESS IDPS. The committee also notes that operational algorithms tend to be empirical and subject to regional tuning for better short-term results, while research algorithms are more physically based. For this reason, the committee believes there is also a need to provide a separate RDR to CDR processing capability for research applications. However, this capability is not currently part of the NPOESS IDPS concept.

NASA has appointed an interim science panel (ISP) to help it define and develop a parallel NPP Science Data Segment to be "attached" to the developing NPOESS IDPS, with initial funding provided by NASA. It will then select an NPP science team (to replace the ISP) through a standard Announcement of Opportunity (AO) once the NPOESS sensor downselect process has been concluded in late 2000. At its February 2000 workshop, the committee was told that the science team would have about 20 members and would provide guidance to NASA on sensor and mission design, science algorithm development and CDR definition, and NPP satellite postlaunch algorithm validation and CDR production. The science team would also provide guidance to the IPO for improved algorithms to benefit operations as well as research applications. Appendix D describes the committee's views on science teams and their anticipated role.

A NASA goal with respect to the science team effort is to delimit the AO selection based on the differentiation of EDRs and CDRs. Indeed, a primary criterion for selection will be the ability of the AO respondents to demonstrate that the proposed CDRs would contribute to climate research in a way that the NPOESS EDRs could not, while also showing that the proposed CDRs could be obtained through algorithmic evaluation of the NPOESS RDR, SDR, and EDR data stream. The committee supports the formation of science teams and has developed a set of working principles for implementing them. Its views are summarized in Appendix D.

BOX 2.1
What Is a Climate Data Record?

Climate research and monitoring often require the detection of very small changes against a naturally noisy background. For example, sea surface temperatures can change by several kelvin between daytime and nighttime or from year to year, whereas the climate signal of interest may change only 0.1 K over a decade. Moreover, changes in sensor performance or data processing algorithms often introduce changes greater than the climate signal. In addition to noise, spatial and temporal biases in the measurements confound climate researchers. A climate data record (CDR) is a time series that tries to account for these sources of error and noise, producing a stable, high-quality data record with quantified error characteristics. A CDR is suitable for studying interannual to decadal variability.

A CDR requires considerable refinement of the raw data, generally the blending of multiple data streams. These streams may come from multiple copies of the same sensor, or they may be ancillary data fields that are used to "correct" the primary data stream. Thorough analysis of sensor performance and improved processing algorithms are also required, as are quantitative estimates of spatial and temporal errors.

Starting in late 1978, nine polar-orbiting satellites carried identical copies of the Microwave Sounding Unit (MSU) to measure atmospheric temperatures. The last MSU now occupies the afternoon orbit slot (NOAA-14), while the morning slot is monitored by the Advanced Microwave Sounding Unit (AMSU) on NOAA-15. Constructing CDRs from MSU instruments has revealed that even though the instruments are essentially identical, differences among them are as large as the climate signal being sought. Once in space, each was found to have a unique response to variations in direct solar heating. Others experienced shifts in responses to onboard calibration targets. And one was found, after launch, to have been improperly calibrated in the laboratory. A final complication was due to the fact that the frequencies monitored with the new AMSU were slightly different from those monitored with the legacy MSUs.

Scientists who were interested in stable, long-term temperature records from the MSU were required to commit considerable resources to discover problems and test adjustments. Individuals with important knowledge about the instrument were difficult to track down. Laboratory calibration information was needed and was challenging to find and decipher. When it came to verification, it was fortunate that two instruments were simultaneously operational, allowing an intercomparison of temperatures to assess the impact of the corrected data (i.e., the Earth became the common calibration target). Finally, the independent ground-based weather balloon network provided a level of confidence in the corrections.

The final NPP priority, consistent with NASA's lead role in IPO technology insertion, is to engage NASA in oversight of the prelaunch characterization and calibration of NPOESS sensors, as well as vicarious postlaunch calibration and validation field campaigns. NASA intends to build on its calibration support activity for the EOS Moderate Resolution Imaging Spectroradiometer (MODIS) instrument and use the AO-funded science team to develop research-quality Level 1 algorithms.

Science Data Segment

According to presenters at the committee's February 2000 workshop, the SDS—at least during the NPP mission, while it is funded by NASA—will perform the following tasks:

Produce and archive consistent, research-quality Level 1 products;

Reprocess Level 1 data as needed after validation and feedback from Level 2+ CDR developers;

Provide optional routine processing for CDR investigators funded through the AO process;

Reprocess Level 2+ products based on Level 1 revisions, validation, and lessons learned;

Allow for the acquisition and retention of ancillary and auxiliary data;

Provide supplemental processing to aggregate research-quality EDRs into weekly/monthly, climate-grid-scale CDRs; and

Distribute data only to AO investigators and a long-term archive (LTA).

Further information about the NPP Science Data Segment is provided in Box 2.2.

NESDIS DATA PROCESSING PLANS FOR NPOESS

NOAA's NESDIS and its National Climate Data Center (NCDC) at Asheville, North Carolina, are working with the IPO and NASA through the IPO's team of IDPS advisory representatives to determine how the long-term archiving of CDRs and other data might be accommodated during the NPP and NPOESS missions.[7]

NESDIS/NASA Plans for a National Climate Data Archive and Services

The basis for a national climate data archive is well established and dates to at least the National Climate Program Act of 1978 (see Appendix E).[8] This act, which continues in force, tasks the Department of Commerce with providing, among other things, "systems for the management and active dissemination of climatological data and information." A statement issued in 1991 by the White House Office of Science and Technology Policy[9] required that each global change parameter have at least one explicitly designated archive, and a Presidential Decision Directive issued in 1996[10] also required the U.S. government to produce and archive long-term environmental data sets.

BOX 2.2
Science Data Segment

According to NASA officials, the NPP Science Data Segment (SDS) design is intended-to the maximum extent possible within the constraints of the NPOESS budget-to at least partially transition from a NASA-funded operation to normal NPOESS operations that will utilize the Interface Data Processing Segment (IDPS). To enhance this possibility, NASA intends environmental data records (EDRs) generated by the NPOESS IDPS to be directly utilized to meet climate data record (CDR) requirements. NASA expects to cap the resources available for CDR processing to ensure that the EDR production requirements are met. However, given the experience of climate researchers (see Box 2.1), the committee believes it is unlikely that the standard EDRs will meet the quality standards necessary for CDRs, particularly in the area of data refinement and reprocessing as algorithms mature. Such reprocessing may require significant computational and storage capabilities, exceeding the requirements for EDR production. Therefore, it is probable that much of the CDR production will take place outside the context of NPP.

NASA intends that the SDS only provide the necessary additional processing capability not already found in the planned NPOESS IDPS. The SDS is being planned as a relatively low-cost appendage to the IDPS to provide CDR-specific processing, in addition to long-term archiving. The committee was also briefed on plans for the development of a permanent Climate Data Handling System (CDHS) that would be appended to the NPOESS IDPS. However, details of how the CDHS would be funded and maintained over the life of the NPOESS mission are as yet unknown.

There are currently three notional SDS architectures. One is a centralized architecture, characterized by a single CDHS located at a NASA facility (Goddard Space Flight Center, for example). At the other extreme is an option for fully distributed processing, characterized by having all the processing done by the principal investigators (PIs) selected through the NPP CDR investigator AO process. In this case, all the hardware and software for the CDHS function would be located at the PIs' home institutions, and communications would be handled via networks and other data distribution methods. The third architecture—a hybrid option—is characterized by a central CDHS that would handle various "common" PI requirements, as well as capabilities at the home institutions of the PIs that would handle PI-unique requirements.

NASA presenters at the February 2000 workshop also noted that NPP would need to operate an interim ground station data acquisition capability from 2005 to 2008. This capability would incorporate direct broadcast for regional applications and would also facilitate the commercialization of NPOESS data. The ground station would perform front-end data processing to Level 0 (frame synchronization, error detection and correction, and packet reassembly), provide Level 1b+ toolkits, and provide product archive and distribution to create an infrastructure similar to that anticipated during the NPOESS era. The committee notes that the present POES system supports an enormous installed base of high-resolution picture transmission (HRPT) ground stations worldwide. The next generation of Earth remote sensing satellites, including EOS and NPOESS, will utilize higher downlink frequencies, such as X band, which are beyond the capabilities of the existing HRPT stations. Such stations have played on important role in near-real-time data transmission.

Plans for the Earth Observing System (EOS) resulted in several NASA and NOAA agreements related to data archiving and dissemination. These include a 1989 NASA/NOAA memorandum of understanding applicable to ocean and atmospheric data. This MOU stated that NOAA would use its "best efforts to . . . assume responsibility . . . for active long-term archiving and appropriate science support activities . . . for the EOS program."[11] In 1997, the NASA associate administrator for Earth science and the NOAA NESDIS assistant administrator agreed in an exchange of letters that the two agencies would begin to develop plans for a ground system infrastructure for U.S. climate data and services.[12] Both agencies recognized NOAA as the U.S. agency charged with climate data stewardship.

In 1997 and 1998, NASA and NOAA participated in a long-term archive cost modeling exercise.[13] They also participated in a workshop on science and global change LTA principles and objectives,[14] which they supported through the U.S. Global Change Research Program (USGCRP) Program Office, and in an LTA prototype development effort started in 1998 (and still in progress) to test the sociotechnical feasibility of a NASA-NOAA LTA. Planning for LTA implementation is currently under way. In September 1999, NASA and NOAA entered into negotiations on a broader MOU describing a "partnership in generating long-term climate records from Earth observation satellite data."[15] NOAA and NASA representatives at the February 2000 workshop said that they plan to complete an LTA implementation plan by September 2000.

Plans for Archive and Distribution Segment at the National Climatic Data Center

The National Climatic Data Center (Table 2.1) is recognized by both NASA and NOAA, and by the NPP mission planners and the IPO, as the most likely repository of CDRs and other records to support climate research in the NPOESS era. However, no funding for this activity has been allocated. Nevertheless, the incremental costs of the LTA proposed for the climate research aspects of NPP and NPOESS operations are expected to be lowest if the NCDC is given enough funding to supply the capability.

NCDC currently provides substantial online data access from an active (online) archive, as well as from backup, on-site and off-site, offline archives. Technological obstacles to the kind of archiving envisioned for the NPOESS era are driven primarily by projected growth by a factor of more than 20 in the amount of digital data to be collected and archived annually over the next 15 years. At the committee's February 2000 workshop, NOAA and NASA officials presented the Archive and Distribution Segment (ADS) envisioned for the NPP Climate Data Handling System and the eventual supplemental SDS and ADS to be appended to the NPOESS operational IDPS. ADS would draw specifically on NCDC's considerable experience in establishing and managing active and offline archives for business, research, and government policy analysis over several decades.[16]

TABLE 2.1 NOAA Data Centers

Data Center	Host Institution or Location	Specialty
National Climatic Data Center	Asheville, N.C.	Climate of United States Archive of weather data
National Geophysical Data Center	Boulder, Colo.	DMSP archive[a] Glaciology World Data Center-A for Marine Geology and Geophysics Paleoclimatology Solar-terrestrial physics Solid Earth geophysics
National Ocean Data Center	Silver Spring, Md.	Coastal oceanography Ocean climate Biological oceanography

National Snow and Ice Data Center	University of Colorado	Snow and ice, cryosphere World Data Center-A for Glaciology

[a]Defense Meteorological Satellite Program.

NASA DATA PLANS

EOS Data and Information System

NASA's Earth Observing System Data and Information System (EOSDIS) manages data from NASA's Earth science research satellites and field measurement programs, providing limited data archiving, distribution, and information management services. EOSDIS is the layer that integrates the seven Distributed Active Archive Centers (DAACs) (Table 2.2) to provide seamless access to the users. Through the EOSDIS Core System (ECS) contractor, EOSDIS provides the necessary hardware and software to the DAACs to capture, process, and distribute data from the EOS satellites. Each DAAC is responsible for archival and management of data in a given scientific discipline. In their reports on site visits to the seven DAACs, NRC review panels conclude that most DAACs are serving the user community quite well and that closing any one of them would reduce NASA's ability to meet its scientific objectives.[17]

Since the inception of EOSDIS, the concept behind it was revised several times to reflect then-current thinking on the appropriate balance between centralized data management centers and small, heterogeneous, but flexible, federated data centers. Moreover, the EOSDIS model continues to change in response to the need to accommodate data from the small PI-led Earth System Science Pathfinder (ESSP) missions[18] and the experience with the Earth Science Information Partners (ESIPs) (see Box 2.3).

NewDISS

NASA's Earth Science Enterprise (ESE) is in the midst of planning its data and information systems strategy for the next 6 to 10 years. This strategy is being designed to suit the new approach to Earth observations, which will include exploratory missions as well as systematic observations. To execute this plan, the ESE intends to rely on a constellation of comparatively small missions organized around critical scientific questions. This approach should be contrasted with the more centralized approach of the first series of comparatively large EOS satellites and its associated data and information system, EOSDIS.

Replacing EOSDIS in the new strategy is the New Data and Information System and Services (NewDISS). NASA views NewDISS as an essential element of a plan to bring order to its Earth observations program, which will now consist of a collection of PI-driven missions, NASA facility missions, and long-term operational missions. Absent such a plan, it would be extremely difficult to address Earth system science questions that demand access to a variety of data sets.

NewDISS intends to preserve a high level of diversity in data system participants and technical approaches rather than attempt to develop a unified system. This will require a delicate and evolving balance between preserving flexibility and maintaining a degree of predictability to support the Earth science community. The NewDISS concept would develop a set of interfaces (or rules of engagement) that would facilitate the insertion of new services and encourage new developers while maintaining critical services. Using the Internet as a model, the NewDISS concept would work to define these interfaces. Moreover, under NewDISS, the proper role of the federal government would be defined, especially as the guarantor of open access and standards and the steward of long-term data archives.

TABLE 2.2 NASA Distributed Active Archive Centers

Distributed Active Archive Center	Host Institution	Scientific Specialty/ Terra Instruments
Alaska SAR Facility	University of Alaska	Sea ice, polar processes/none
EROS Data Center	U.S. Geological Survey	Land processes/ASTER, MODIS
Goddard Space Flight Center	NASA	Upper atmosphere, atmospheric dynamics, global biosphere, hydrologic processes/ TRMM, MODIS
Langley Research Center	NASA	Radiation budget, aerosols, tropospheric chemistry/ CERES, TRMM, MISR, MOPITT
National Snow and Ice Data Center	University of Colorado	Snow and ice, cryosphere/ MODIS
Oak Ridge National Laboratory	Department of Energy	Biochemical fluxes and processes/ none
Physical Oceanography	Jet Propulsion Laboratory NASA-Caltech	Oceanic circulation, air-sea interactions/ none
Socioeconomic Data Archive Center	CIESIN Columbia University	Socioeconomic data and applications/ none

ASTER = Advanced Spaceborne Thermal Emission and Reflection Radiometer
CERES = Clouds and the Earth's Radiant Energy System
CIESIN = Center for International Earth Science Information Network
MOPITT = Measurement of pollution in the troposphere
TRMM = Tropical Rainfall Measuring Mission

BOX 2.3
The Federation of Earth Science Information Partners

The Federation of Earth Science Information Partners was created by NASA's Earth Science Enterprise (ESE), in part to test assertions by scientists that the organization and even the management of data and information systems could be best accomplished by a federation of scientific users rather than a centralized activity directed by a federal agency.[1] It consists of various autonomous elements that interact to deliver a complete suite of user services.

Some Earth Science Information Partners (ESIPs) focus on basic data production and processing under rigorous standards of quality control, others focus on the development of new data products and services, while still others are developing services for the general public and the commercial sector. Experience to date with the Federation is encouraging, and NASA is examining this model as it develops NewDISS, a successor to EOSDIS that will be designed to better handle the PI mode of data processing expected from new, smaller, missions.

1John Townshend, University of Maryland, "The Federation of Earth Science Information Partners," Presentation to the EOS Investigator Working Group, April 2000, Tucson, Ariz. Briefing charts available online at http://www.esipfed.org/.

At the committee's February 2000 workshop, NASA officials stated their intention to have NewDISS build upon the experience with the ESE Federation, which was created in response to the report of the Committee on Global Change Research (NRC, 1999). They also intend to examine the lessons learned from EOSDIS, particularly the need to foster competition in data system design and implementation.

NASA Long-Term Archive

Long-term archival is presently the subject of planning by the NASA Distributed Active Archive Centers. The overall goal is to develop a strategy to archive NASA data within a joint NOAA/NASA framework. The current NASA plan for EOS data is to provide medium-term access through the existing DAACs (Table 2.2) while the data are being used by the funded NASA science teams (3 to 10 years). The longer-term archiving and distribution of EOS data is still uncertain. Three LTA options are being considered:

1. Archive in place at the DAACs;
2. Transfer NASA data to NOAA archives[19] (e.g., NCDC; see Table 2.1); and
3. A hybrid of the first two options.

NASA has established study teams to explore various implementation approaches, including the data sets that should be considered, system phasing, archive services that might be provided, formats and interoperability, and joint management. NASA intends to establish a prototype in the next 12 months to test various concepts.

SUMMARY

Planning for NPP and NPOESS ground systems is just beginning. At its February workshop, the committee heard from NASA, NOAA, and the IPO that they plan to work together to meet the operational needs of their users. The committee was also made aware of planning activities intended to support the needs of research users to the extent possible in an operational program that is already working under a constrained out-year budget.

Management of the projected volumes of data from EOS, NPP, and NPOESS will present significant challenges. Although there are substantial technical challenges, the committee believes the management and organizational challenges could be even larger-that is, having NASA and NOAA translate their *plans* into clear and explicit agreements on agency roles, responsibilities, and resource obligations. Absent such agreements, there is a danger that agencies will simply pass requirements from one agency to another, a clearly inadequate solution to the many challenges posed by the torrent of data anticipated from NPP and NPOESS.

REFERENCE

National Research Council (NRC), Board on Sustainable Development, Committee on Global Change Research. 1999. *Global Environmental Change, Research Pathways for the Next Decade*. Washington, D.C.: National Academy Press.

[1]The NPOESS IPO defines EDRs as data records that contain the environmental parameters or imagery required to be generated as user products as well as any ancillary data required to identify or interpret these parameters or images. EDRs are generally produced by applying an appropriate set of algorithms to raw data records (RDRs). Information about the NPOESS EDRs can be found in the NPOESS Integrated Operational Requirements Document, which is available on the World Wide Web at http://npoesslib.ipo.noaa.gov/ElectLib.htm.

[2]The contractors are TRW and Lockheed Martin Missiles and Space Company.

[3]Reginald Lawrence, NPOESS IPO, presentation to the committee on February 7, 2000.

[4]Like POES, the NPOESS system must be designed as a launch-on-demand system. Consequently, a satellite must be operational after a launch failure or an on-orbit failure demands a replacement satellite. Based on this strategy, the first NPOESS satellite, referred to as C1, must be available by the fourth quarter of fiscal year (FY) 2008 to support any potential launch or on-orbit failure of POES-N' and DMSP F-20, which are currently scheduled for launch in January 2008 and July 2008, respectively. The first planned NPOESS launch is scheduled for the first quarter of FY 2009.

[5]Information on NPP in this section and the next draws on presentations to the committee at the February 7-8, 2000, workshop by Daniel DeVito, Robert Murphy, and Joy Henegar, all of NASA.

[6]National Research Council (NRC), Space Studies Board. 2000. *Issues in the Integration of Research and Operational Satellite Systems for Climate Research: I. Science and Design*, in press.

[7]This section draws on presentations by Martha Maiden of NASA and Frank Crowe and Tom Karl of NCDC at the February 7-8, 2000, workshop.

[8]The National Climate Program Act of 1978 (15 USC 2901) is available online at http://www4.law.cornell.edu/uscode/15/ch56.html#PC56. Section 108 of Public Law 101-606, "Global Change Research Act of 1990," November 16, 1990, also refers to the 1978 Act. See NRC (1999), Appendix A

. [9]Office of Science and Technology Policy, "Data Management for Global Change Research," Policy Statement, July 2, 1991. Available online at http://www.gcrio.org/USGCRP/DataPolicy.html.

[10]National Science and Technology Council, "National Space Policy," PDD NSTC-8, September 19, 1996. Available online at http://sun00781.dn.net/spp/military/docops/national/nstc-8.htm.

[11]The MOU is reproduced as Appendix B of U.S. Global Change Research Program, *Global Change Science Requirements for Long-Term Archiving: Report of the Workshop*, October 28-30, 1998, Boulder, Colo., March 1999.

[12]Tom Karl of NCDC and Martha Maiden of NASA, "Overview of NESDIS/NASA Plans for National Climate Data and Services Including Long Term Archive," Presentation at the committee's February 7-8, 2000, workshop.

[13]Ibid.

[14]U.S. Global Change Research Program, *Global Science Requirements for Long-Term Archiving: Report of the Workshop*, October 28-30, 1998. Boulder, Colo., March 1999. This workshop is notable for having developed key LTA guiding principles and objectives, including the following:

• Facilitate the best possible science and highest-quality assessments to support business and policy decisions.
• Document Earth system variability on global, regional, and local scales.
• Ensure archive holdings are actively promoted and readily available.

[15]Tom Karl and Martha Maiden, Presentation (see footnote 12) at the committee's Feburary 7-8, 2000, workshop.

[16]NOAA representatives at the committee's February 2000 workshop indicated that the ADS needs to provide the following specific capabilities:

• Data ingest, validation and archive: (1) data records, metadata, ancillary data, calibration coefficients; science products, reprocessed products, EDRs from Centrals, and concurrent ingestion of IDPS and SDS products and (2) data format, content, and metadata validation, confirmation to IDPS and SDS, and data provider coordination.
• Archive management: Fulfill orders, back up and restore archive, and periodically validate archive integrity.
• User interface and customer service: Electronic interactive query, search, and order services; verify payment; maintain user registration and accounts; and provide status and account information.
• User product generation: Generate and ship products in a timely manner.
• User product tracking and report generation: Maintain order status, generate statistical reports (such as data volume received, retransmitted, and archived and orders received, fulfilled, and pending), and report on turnaround time.

[17]National Research Council (NRC), Board on Earth Sciences and Resources. 1999. *Review of NASA'S Distributed Active Archive Centers*. Washington, D.C.: National Academy Press.

[18]The ESSP program is characterized by relatively low to moderate cost, small to medium-size missions that are capable of being built, tested, and launched in a short time. A description can be found online at http://essp.gsfc.nasa.gov/.

[19]The NOAA data centers have a wider array of responsibilities, ranging from delivery of operational weather data to the National Weather Service, to the analysis and archival of weather and climate data, physical oceanography data collected by ships and satellites, coastal observations, solar-terrestrial observations, glaciology, and even marine geology and geophysics. Unlike NASA, which has a primary focus on scientific research, the NOAA data programs are primarily related to operations and user services.

Ensuring the Climate Record from the NPP and NPOESS Meteorological Satellites

3
Essential Services for Climate Data

This chapter presents the committee's perspective on the requirements for essential data system services for climate research and on how well current plans meet these requirements. In writing this chapter, the committee drew on discussions and presentations at its workshop on February 7-8, 2000; it also referred to related work completed by the NRC and others.[1,2,3] While the earlier work focused on broad questions of what should be archived in a climate data archive, the focus of this study is on near-term steps that can be taken to ensure that the climate data record from NPOESS and NPP is preserved and that basic services are provided. Appendix E provides historical information relevant to this chapter.

GUIDING PRINCIPLES

The committee believes adherence to the following principles can help ensure the preservation of the climate record from NPP and NPOESS:

Accessible and policy-relevant environmental information must be a well-maintained part of our national scientific infrastructure. High-quality data and information on climate change are the foundation on which policy decisions will be made. Collecting, managing, archiving, and distributing environmental data must be given sufficient priority to ensure that the requisite data foundation exists.

The federal government should (1) provide long-term data stewardship, (2) certify open, flexible standards, and (3) ensure open access to data. The government does not necessarily need to control the implementation of every task and service for a climate data system. Rather it should undertake those activities and services that cannot be done in a competitive academic or commercial environment.

Because the analysis of long-term data sets must be supported in an environment of changing technical capability and user requirements, any data system should focus on simplicity and endurance. Complex systems often become point designs; they meet current requirements but cannot incorporate any changes in tools or requirements.

Adaptability and flexibility are essential for any information system if it is to survive in a world of rapidly changing technical capabilities and science requirements. The system should not just react to change but instead should continually track technology and system performance so that it can respond proactively.

Experience with actual data and actual users can be acquired by starting to build small end-to-end systems early in the process. EOS data are available now for prototyping new data systems and services for NPP and NPOESS. A waterfall design process, whereby the system is not available for testing and evaluation until it is complete, is nearly certain to fail in a changing world.

Multiple sources of data and services are needed to support development of climate data records (CDRs). There is no single approach for deriving a particular CDR, such as atmospheric temperature or ozone density. The quality of the CDRs will improve as more research groups work with the various input data sets, and the overall system will be more robust if it does not rely on a monolithic implementation. Fostering open competition for services promotes innovation and new ideas. Access to the raw data will be needed to enable CDR development;

Science involvement is essential at all stages of development and implementation. Scientific advice and involvement will be necessary for CDR design and production as well as for data system implementation. Having climate data record developers and users assisting in the specification, design, building, and testing of the system will help ensure its usefulness to the research community.

COMPONENTS OF A CLIMATE DATA SYSTEM

Developing a system appropriate for climate data records poses unique challenges related to the complexity and timescale of the data. Data sets (and information about the data sets) must be preserved for decades, and data must be archived in a manner that facilitates reanalysis. In addition to locating and delivering data, a climate data system must accommodate the new CDRs that will be developed as climate research evolves.

Although these services appear to be identical to those provided by any data system, the scale of the problem, as well as the frequency of change, is much different from that of the typical satellite mission data system. For example, many NASA missions pass their data to the National Space Science Data Center at the Goddard Space Flight Center with the expectation that use of the data for long-term studies will be infrequent. The NOAA meteorological satellite data are stored in the National Climatic Data Center (NCDC) with the expectation that individual users will request only subsets of the data. On the other hand, as described below, active archives are designed for changes in algorithms, services, and user requirements; in general, they are better suited to the needs of casual users of small volumes of data. The following sections discuss an architecture that meets these varying needs, an architecture that relies on a blend of active archives and long-term archives to provide an overall set of climate data services.

Long-Term Archive Services for Climate-Related Data

The committee views a long-term archive (LTA) as more than a static repository for data—it sees it as allowing a broad set of users to examine, retrieve, or copy the data. While users of an archive might simply want to retrieve data, it is more likely that they would require further data processing. Issues related to the refinement or reprocessing of the data are key drivers in determining an appropriate architecture and implementation strategy for an LTA. Certain processing may be done best within the LTA, while other processing could be done in the specialized facilities of the users. The committee's views on the necessary functions of an LTA, which are based on its prior work[4] and information gained at the February 7-8, 2000, workshop, may be summarized as follows:

Ensure the long-term survival of the data by (1) ingesting, storing, migrating media, and cataloging all the data and (2) including all ancillary data and calibration and characterization of the instruments. History has shown that review and retrospective reprocessing of data are important elements in understanding climate and the effects of natural and anthropogenic changes on climate trends.

Provide additional information, algorithms, and such higher level products as can aid in the future reprocessing of the data.

Distribute data to the large-scale users and primary processors by (1) streaming data for continual and scheduled transfers, (2) allowing the primary user and/or processor to hold the data for distribution, and (3) allowing competitive bidding for these services by universities and other groups.

Data-use analysis has shown that a small number of users account for the bulk of data transferred.[5] As data transfer bandwith is a key design parameter, this effect must be accounted for in developing the LTA.

What Should Be Stored?

The distinctions among science, weather, and climate data are fuzzy at best. Much of the data can be used for several purposes and the distinction is more in the requirements for minimum latency and regularity of processing. The value of meteorological data used for weather prediction is ephemeral and the data must be processed quickly. However, the data may retain their value for other uses. While it may be valid to distinguish between operational and research processing for a facility, with the advent of the NPOESS system and the planned use of the operational system to support more research, that distinction is also less clear. Thus, EOS, NEXRAD,[6] NPP, and NPOESS data may all reside in an LTA for "climate."

The needs of climate researchers are embodied in the CDR concept. Although the NPP- and NPOESS-derived EDRs may have considerable scientific value, CDRs are far more than a time series of EDRs. Even the EOS data products will require continuous assessment and refinement as knowledge of both the algorithms and the sensor improves over time. Although the lines may be indistinct, there remain fundamental differences between products that are produced to meet short-term needs and those where consistency of processing over years to decades is essential.

Management of the LTA

Participants at the committee's February 2000 workshop considered several approaches to managing the LTA. A particular challenge was establishing a balance between centralized oversight and local processing and dissemination. A loosely connected system with specialized elements appears to be the optimal approach: for example, a central LTA supported by individual, active data archive centers to work with specific data streams and data sets. The active data archives could be modeled on NASA's Distributed Active Archive Centers (DAACs) if problems identified in a recent NRC review[7] are addressed.

Although NOAA and NASA are seen as the responsible agencies, the committee believes there should be a competitive approach to implementing the individual elements. The committee believes NOAA should implement the central LTA, which is responsible for the long-term preservation of the data. However, other elements could be implemented by a mix of academic, nonprofit, and commercial organizations. Implementation must be incremental to foster evolutionary growth and continual innovation. Such an approach should also result in lower costs.

The National Climatic Data Center and the Long-Term Archive

The NOAA National Climatic Data Center is a potential repository of data to support climate research in the coming NPOESS era. NCDC is currently assigned the mission of managing the nation's resource of global climatological in situ and remotely sensed data and information to "promote global environmental stewardships; to describe, monitor and assess the climate; and to support efforts to predict changes in the Earth's environment."[8] NCDC receives, processes, controls the quality of, archives, and distributes weather- and climate-related data from a variety of sources, including satellites, ground radar, and in situ measurement systems. NCDC currently houses approximately 700 TB of data; by 2010, the total archive at NCDC may exceed 5000 TB (see Box 3.1).

Storage and protection of the data are only two of the functions of an LTA. Access to the data and the ability to reprocess the data as scientific understanding improves are also required. The reprocessing requirement is particularly challenging; the committee does not believe that the NOAA National Climatic Data Center has sufficient resources to fully address this challenge now. Further, it is evident that NCDC or future LTA sites will face even more demanding requirements as NPP and NPOESS data become available.

Increased amounts of long-term storage and media migration are already taken into account in NCDC plans to accommodate increased volumes of data. However, the committee finds that there needs to be additional work on the treatment of several categories of data, including the following:

Recent data in active use that might be best held in a separate active archive;

Aging data that are used in periodic reassessments and other studies; and

Older data that have not been accessed in some stated interval of time.

Previous workshops and studies provide some guidance on these issues; however, the committee recommends consideration of a standing LTA science team to ensure a user voice in the LTA design.

BOX 3.1
Near-Term Projections of Future Needs at NCDC

The technical and financial demands associated with managing the climate-related data anticipated from NPP and NPOESS may be seen by examining operations at the current repository of such data, the NOAA National Climatic Data Center.

NCDC has a current annual budget of approximately $20 million, which supports a civil service staff of 175 plus 60 contract personnel. The present total digital archive at NCDC is approximately 700 terabytes (TB), and officials expect approximately 80 TB of data to be added in calendar year 2000. (About three-fourths of this amount is expected to come from ground NEXRAD radars.) NCDC officials told the committee that current funding is inadequate to properly manage this quantity of data. Large increases (about 14 TB annually) in data will come from the Initial Joint Polar System (IJPS) when the European satellite MetOp is launched in 2003. NCDC also expects more data (an increase of 50 percent) from NEXRAD as capabilities are developed to ingest these data via telecommunications from each site.

NCDC is currently working with NASA on the long-term archival of EOS data; officials expect by 2003 to be ingesting an amount of EOS data equivalent to the amount of NEXRAD data that is currently being ingested. Further, it is anticipated that this load will increase by approximately an order of magnitude by the end of the decade. The scheduled launch in 2005 of the NPOESS Preparatory Project satellite will add another 90 TB of data annually. NPOESS operations, which might begin in 2009, would add another 228 TB of data annually.

It is anticipated that by 2005 NCDC will be tasked to ingest and make available more than 1 TB per day, or about 500 TB of data annually, which will increase to about 1000 TB annually by 2010. By that time, the total archive at NCDC is expected to be approximately 5000 TB.

NCDC estimates that it will require between $10 million and $15 million of additional funding each year to be able to handle the expected increase in data and information.[1] Further, officials note that these estimates assume minimum levels of user services. In particular, there would be no browsing, reprocessing, or subsetting capability for users to interact with the data. Such higher levels of service would require significant additional expenditures, in the range of $3 million to $5 million annually, according to NCDC officials.

[1]According to officials at NCDC, these additional funds will pay for more personnel and a larger information technology infrastructure, including increased communications capacity, online and near-online storage capability, drives, cabinets, tapes, and software.

SOURCE: Tom Karl, director of NCDC, Private communication, July 2000.

Active Archive Services for Climate Data

Making data available to the casual, small-volume user requires a different set of services than those provided by an LTA. These services include rapid access to subsets of the data, sophisticated search tools to locate coincident data, and a focus on providing online search and order fulfillment. These services support the production of customized data products and allow climate researchers to test new algorithms. Thus, long-term archive services are only one component of a climate data system, and climate researchers require "active archives" that are comparable to those of the research satellite missions. Climate research will require an effective integration of these active and long-term archive services, much as it requires an integration of research and operational satellite missions.

Active archive services have sometimes driven system design. In EOSDIS, the requirements to meet the data needs of the casual and small-volume user grew and began to drive design. The small-volume user can drive costs because the many small data extractions and associated packaging and shipping are personnel-intensive and can require expensive hardware and software. This type of service accounts for the largest number of transactions and involves random access to the data rather than sequential transfers. As the system grows, such usage could become a large factor in the overall usage.

The committee believes the seven DAACs in NASA's current system have been reasonably successful in serving their user communities. The DAACs have developed a working system for delivering reasonably large volumes of data to their user communities, largely by responding to evolving science requirements. Based on the experience of the DAACs, as well as the lessons of EOSDIS (see Box 3.2), the committee finds as follows:[9]

Where new active archives need to be developed, a rapid development approach should be employed to produce a working system that delivers data to the user community in about a year.

Active archives can play a vital role in both the development and operation of LTA facilities.

User representation should be ensured by making sure that each active archive has a user working group that is made up of representatives selected from the archive's users. The user working group should be able to propose new services and to endorse or veto changes that the archive staff may suggest.

The active archives and long-term archive facilities should participate in a federation that would work to resolve issues of common concern, such as ensuring that the active archives can back up and restore data in the LTA and that various user communities can obtain help in translating each other's terminology and data structures, and to encourage tool sharing among the community members.

Both active archives and long-term archive facilities need to be flexible enough to accommodate changes in the technology they use on timescales between 2 and 5 years.

CONCLUDING OBSERVATION

Finally, the committee notes that the President's budget for fiscal year 2001 includes some $4 million for activities at NCDC related to development of the LTA. Even if this work is fully funded by Congress, it should be recognized that substantially greater investments will be required to develop the LTA and to address the issues raised in this chapter and elsewhere in this report.

BOX 3.2
Lessons from EOSDIS

A number of participants at the committee's February 2000 workshop expressed the view that the EOS Data and Information System (EOSDIS) had "failed" as a result of having been overspecified in design and overcentralized in execution. The system is thought to have been overly specified because it could not take full advantage of very rapid advances in computer and data storage technology, including the software associated with databases. The concern with overcentralization refers to the degree of central control in a system that was highly distributed. Workshop participants argued that requirements for processing and disseminating a particular data stream are best performed by the people who generate and use the data.

NASA implemented EOSDIS as single contract, and this also contributed to the ongoing difficulties with the system. Building monolithic systems in an environment where technology is changing rapidly and user requirements are poorly known or changing is an impossible task. For example, the EOSDIS contractor consistently underestimated advances in computing capability and frequently invested in technology that then became obsolete. If NASA had funded contractors for specific functions and relied on a systems integrator to link the components, the system might have been more resilient.

The EOSDIS experience of cost and schedule overruns may also be informative regarding LTA development cost. Cost and development time grow rapidly with the size of the programs; therefore, it is important to attempt to minimize the size of program elements to be developed. In addition, because users of the resultant systems may not understand or express their requirements fully, the functions and requirements tend to grow during development. One of the presenters at the CES workshop argued that the problem of understanding the user and avoiding requirements creep is best addressed at the level of the individual user. Further, it was stated that attempting to address these issues from a large central organization has rarely worked in the past.

[1]National Research Council (NRC), Board on Sustainable Development, Committee on Global Change Research. 1999. *Global Environmental Change, Research Pathways for the Next Decade*. Washington, D.C.: National Academy Press.

[2]U.S. Global Change Research Program (USGCRP), *Global Change Science Requirements for Long-Term Archiving: Report of the Workshop*, October 28-30, 1998. Boulder, Colo., March 1999.

[3]H. Jacobowitz, ed. 1997. *Climate Measurement Requirements for the National Polar-orbiting Environmental Satellite System (NPOESS)*, Workshop Report, NOAA. College Park, Md., February 27-29, 1997.

[4]National Research Council (NRC), Space Studies Board. 2000. *Issues in the Integration of Research and Operational Satellite Systems for Climate Research: I. Science and Design*, in press; National Research Council (NRC), Space Studies Board. 2000 *Issues in the Integration of Research and Operational Satellite Systems for Climate Research: II. Implementation*, in press.

[5]B.R. Barkstrom, "Understanding Data Users and Predicting Their Behavior," Presented at the committee's February 7-8, 2000, workshop but dated July 16, 1996.

[6]The Next Generation Weather Radar system (NEXRAD) comprises approximately 160 Weather Surveillance Radar-1988 Doppler (WSR-88D) sites throughout the United States and selected overseas locations. Information about NEXRAD is available online via links within the NCDC site at http://www.ncdc.noaa.gov/ol/radar/radarresources.html#WHATIS.

[7]National Research Council (NRC), Board on Earth Sciences and Resources. 1999. *Review of NASA'S Distributed Active Archive Centers*. Washington, D.C.: National Academy Press.

[8]From the home page of NCDC, available online at http://www.ncdc.noaa.gov/.

[9]The committee notes the similarity of these recommendations to recommendations made nearly two decades ago by the National Research Council's Committee on Data Management and

Computation (see National Research Council, Committee on Data Management and Computation (CODMAC). 1982. *Data Management and Computation, Volume I: Issues and Recommendations*. Washington, D.C.: National Academy Press).

Ensuring the Climate Record from the NPP and NPOESS Meteorological Satellites

4 Conclusions and Recommendations

CONCLUSIONS

While encouraged by NASA's and NOAA's recent attention to preserving the climate record of NPP and NPOESS (and also EOS), the committee believes that an enormous investment in Earth observations is at serious risk. In briefings to the committee, NASA and NOAA officials acknowledged that there is no operational ground system infrastructure for U.S. climate data and services. (NOAA has requested funding for a prototype activity, the National Environmental Data Archive System—NEDAAS.)

The climate research community requires that satellite data from NPP and NPOESS be able to generate climate data records, data whose quality is known quantitatively and whose temporal and spatial biases are minimized (or at least quantified). CDR production will require considerable scientific insight, including the blending of multiple data sources, error analyses, and access to raw data sets. Moreover, information on sensor design, operation, and calibration will also be necessary to develop a consistent CDR across multiple sensors. Production and refinement of CDRs by reprocessing data may be difficult (or unaffordable) under present plans.

Archives often attempt to maximize the number of users to justify continuing financial support, but this may not be necessary for supporting the development of CDRs. Moreover, efforts to facilitate broad use may divert attention from CDR issues. For example, archives may try to provide a broad range of data services, ranging from data visualization to data analysis. The EOS Data and Information System (EOSDIS) was driven in its early development by the concept of one-stop shopping, whereby users could locate and retrieve all of the data they needed from a single site, although the data were actually stored in many distributed archives. Thorough analysis and design of the data system are essential, especially if a key objective is the provision of long-term data sets to guide public policy. Design of archive services must be firmly based on actual user requirements in order to control costs.

A look at the recent history of NASA and NOAA data systems reveals both successes and failures. Much can be learned from previous experience, but the fundamental objective of establishing a set of data systems and services to meet the needs of climate research will require more than MOUs and larger magnetic tape silos. New services must be supported that are not available in the present mix of NASA research mission data systems and NOAA long-term archives. New scientific and policy demands are being placed on these systems, and new management and technical approaches must be established. Although increasing funding is a necessary condition, it alone is not sufficient.

RECOMMENDATIONS

The committee's study of issues related to preservation of the climate record from future NASA and NOAA satellites was necessarily brief and drew heavily on previous work and on the 2-day workshop. Underlying the committee's recommendations are strongly held views about the critical need and unique potential for data from NPP and NPOESS to contribute to the work of the climate research community. In particular, the committee believes that prudent planning and modest investments early in the program will allow the NPOESS system to generate and continue essential climate-quality data records and to develop new records based on the rich blend of planned instruments. In putting forward its recommendations, the committee also drew on the lessons learned from the data system for the EOS Data and Information System (EOSDIS). Climate research will require a variety of services, ranging from careful long-term stewardship of the basic data sets to intensive data analysis and algorithm refinement. This complex set of scientific and information system activities should be broken into manageable pieces rather than used to form a comprehensive, single-system solution. Two critical functions are the long-term archive (LTA) and active archive functions. Although there are strong similarities, there are also fundamental differences in the nature of the interactions between the data system and its users and the degree of flexibility and stability. Climate research requires an integration of the more stable, long-term functionality of the LTA and the flexibility to pursue and develop new capabilities in the active archives.

The first four recommendations are presented in order of priority. They can be summarized as follows:

1. Preserve the basic data sets.
2. Provide access to these data sets.
3. Develop CDRs based on these data sets.
4. Provide basic user services and tools to the science research community.

The next four recommendations, which are not listed in order of priority, focus on programmatic and management structures to meet the essential requirements for a climate data system. They can be summarized as follows:

5a. and 5b. Include the NPOESS Preparatory Project (NPP) as part an emerging climate system.
6. Invest in early prototypes of LTA and active archive functions.
7. Develop LTA and active archives as a set of complementary services.

Recommendation 1. NOAA should begin now to develop and implement the capability to preserve in perpetuity the basic satellite measurements (radiances and brightness temperatures).

The development of long-term, consistent time series based on CDRs requires access to the lowest level of data available. In general, this means the raw data records (RDRs), or Level 1A data. The low-level data can be used to develop refined CDRs as scientific and technical understanding of Earth processes and sensor performance improves over time. The committee recommends that NOAA do the following:

Archive both current and future data sets, including those from both research and operational satellite missions, in an LTA.

Archive information on sensor development, calibration, operation product validation, and appropriate metadata along with the basic radiances.

Migrate data sets to new, computer-compatible media on a regular basis, such that data sets are refreshed every 2 to 3 years consistent with the pace of technology evolution.

Organize data in the LTA based on user access patterns to optimize data retrieval.

Recommendation 2. NOAA should guarantee climate researchers affordable access to all RDRs in the long-term archive, with an emphasis on large-volume data access.

Development of CDRs requires access to enormous data volumes, but it is likely that only a small number of researchers will need such extensive access to the raw data. Thus, a well-designed set of basic services would meet this basic function without being too costly. The committee recommends that NOAA act on the following:

Award the LTA functions on a competitive basis to both government and private organizations to promote innovation.

Start the development of the LTA immediately with a simple set of end-to-end capabilities to gain experience and modify the plans and implementation accordingly. (End-to-end is defined as being from sensor aperture to the desktop of the climate information user.)

Recommendation 3. NASA, in cooperation with NOAA, should support the development and evaluation of CDRs, as well as their refinement through data reprocessing.

Because the CDR process is driven by science understanding, there will be a continuing need for the involvement of researchers. The NOAA/NASA Pathfinder (see Box 4.1) shows that the agencies can generate critical data sets for transitioning research products into operational data products. Over the next decades, the committee expects that a few experimental CDRs may become effectively "operational" products and will be produced by NOAA. The committee recommends that NASA, in cooperation with NOAA, take action as follows:

Periodically select science investigations and provide adequate support to develop and evaluate new CDRs.

Preserve sensor calibration and operating information, as well as metadata and ancillary data fields, in a manner that allows reprocessing the CDRs.

Evaluate on a regular basis the organization of data sets in the LTA in light of actual data usage patterns to improve reprocessing and access efficiency.

Recommendation 4. NOAA and NASA should define and develop a basic set of user services and tools to meet specific functions for the science community, with NOAA assuming increasing responsibility for this activity as data migrates to the long-term archive.

NASA's Distributed Active Archives Centers, as well as components of NASA's Earth Science Information Partners, are gaining experience with responding to data requests and setting up user services. Although the focus is on the order entry process (catalog, data location, browse, etc.), more attention needs to be given to quality assurance and the order fulfillment process (metadata, subsetting, electronic data delivery, etc.). Emphasis should be given to reducing cost through automation. It is essential that the large-volume data sets from the archive be affordable for the science user community. The committee recommends that NOAA and NASA do the following:

Select teams on a competitive basis that will identify and provide specific user services and tools (see Appendix D). As part of an ongoing process of system evaluation and improvement, these teams would assist in identifying and providing essential user services. Based on a rigorous analysis of a user model for climate research, they would make recommendations on characteristics such as data subsetting and browse capabilities.

BOX 4.1
NOAA/NASA Pathfinder Experience

In the early 1990s, NASA and NOAA developed a joint initiative to provide critical data sets for the global change community. This initiative was known as the NOAA/NASA Pathfinder and was supported largely by the EOS Program Office at NASA. Pathfinder projects involved the global change research community in identifying their most pressing data needs.

Pathfinder projects aimed at providing improved access to and processing of large time-series data sets from different satellite systems (for example, Advanced Very High Resolution Radiometer (AVHRR), TIROS Operational Vertical Sounder (TOVS), Geostationary Operational Environmental Satellite (GOES), Special Sensor Microwave Imager (SSM/I), and Landsat). Priority data sets were identified and a science working group was formed for each product to advise on product specifications, consensus algorithms, and data processing requirements. Emphasis was given to reprocessing decadal data records with raw data amenable to calibration, allowing calibration between multiple instruments. Joint Announcements of Opportunity from NASA and NOAA provided open competition for the generation of new Pathfinder products. A Pathfinder benchmark period (April 1987 to November 1988) was established for all the projects to facilitate complementary analyses and intercomparison studies.

The Pathfinder projects substantially improved the availability of time-series data for the science community and demonstrated how NASA and NOAA could collaborate on joint data initiatives to serve the global change research community. For example, the AVHRR GAC (Global Area Cover) Land Pathfinder undertook the first major reprocessing of the entire AVHRR land data record and the associated transfer of the data record to a new media. Future joint initiatives between NASA and NOAA in the context of the NPP and NPOESS data system development would benefit from the lessons learned from the various Pathfinder data activities.

Support and maintain a balance between internal and external expertise at the government data centers.

Examine the feasibility of providing open electronic access to a rolling archive of RDRs and EDRs through the NESDIS Central that is planned for NPOESS and NPP.

Recommendation 5a. NASA, in cooperation with the Integrated Program Office, should develop the NPOESS Preparatory Project as an integral component of a climate data system.

NPP represents a unique opportunity to test both scientific and programmatic interfaces related to an integrated data systems strategy. It will bridge the gap between the NASA research missions and the NPOESS operational missions. There is potential to begin the development of long-term, high-quality CDRs and an associated data system for climate research, but it is an opportunity that could be missed. The committee recommends that NASA, in cooperation with the NPOESS IPO, proceed as follows:

Develop and implement a prototyping activity to link the NPP Science Data Segment with the NOAA LTA. This activity should start with NASA EOS data sets, including Atmospheric Infrared Sounder (AIRS)/Advanced Microwave Sounding Unit (AMSU)/MODIS in anticipation of the NPP data sets, including CriS/ATMS/VIIRS.

Put aside reserve funds from data system development to support evolutionary development activities as the program matures to ensure that the system is not locked in with no resources for subsequent enhancement.

Develop prototype user services for NPP climate data records.

Recommendation 5b. Select, on a competitive basis, and then support an NPP science team as soon as possible.

The team should consist of sensor experts, algorithm developers, and science data users. Because the functions will require different levels of involvement in the sensor development and operation process, they will require different levels of support. The team would advise on the NPP data system needs, including scientific data processing, archiving, and distribution requirements.

Recommendation 6. NOAA, in cooperation with NASA, should invest in early, limited capability prototypes for both long-term archiving and the NPP data system.

Data systems that do not develop, test, and evaluate on a frequent, regular basis are nearly always late and over budget. System development costs generally increase as the cube of the number of years in development. A climate data system will build on existing components and existing capabilities, but new functions and new interfaces must be developed and implemented to meet the requirements for climate research. The committee recommends that NOAA, in cooperation with NASA, take action as follows:

Competitively select and support a science data team to assist a NOAA long-term archiving program on the following:
—Archive requirements for long-term data sets, including RDRs, metadata, and ancillary data fields;
—Archival of CDRs, algorithms, and processing environments;
—Data structure and organization to facilitate access and reprocessing;
—Flexible, open standards to facilitate data access and refinement;
—Data reprocessing priorities; and
—Minimal user services and tools.

Require the NPOESS total system performance requirements (TSPR) contractor to work with the science data team to facilitate CDR production and archiving from both NPP and NPOESS.

Develop flexible standards and formats that allow new services to be developed in the future.

Begin to develop a small number of CDRs using the LTA services.

Recommendation 7. NASA and NOAA should develop and support activities that will enable a blend of distributed and centralized data and information services for climate research.

NASA and NOAA should consider a hybrid mode of operation rather than building a rigid, centralized system or relying on structure to emerge from an uncoordinated set of data systems. The government should ensure and manage the activities it does best, while fostering innovation and flexibility in those parts of the overall system that do not need to be closely managed. The committee recommends that NASA and NOAA proceed as follows:

Implement the NPP data system as a federation of linked activities, such as that proposed in the NewDISS framework.

Where appropriate, build on existing and planned capabilities, including EOSDIS, the Earth Science Information Partners, NASA's Distributed Active Archive Centers, and NOAA's Data Centers, and develop new capabilities as user experience is gained.

Ensuring the Climate Record from the NPP and NPOESS Meteorological Satellites

A
Statement of Task

Background Researchers studying the issues surrounding global climate change have a particular need for the kind of repetitive, long-term, high-quality measurements that can be provided from the vantage point of space. Operational weather satellites provide perhaps the only means for securing these measurements. The next generation of operational sensing systems are currently being designed, and the National Polar-orbiting Operational Environmental Satellite System (NPOESS), scheduled for launch beginning in 2009, provides an important component of this operational monitoring system. NPOESS is being developed with the goal of meeting converged operational data needs of NOAA and DOD, as well as some of the data needs in NASA Earth observation programs.

As part of its planning to transition appropriate research satellite measurements into the operational domain, NASA, working with the NPOESS Integrated Program Office, is developing the NPOESS Preparatory Program (NPP). NOAA is supporting the NPP program as part of its risk reduction demonstration and validation for NPOESS sensors, algorithms, and processing. NPP will provide a launch in 2005 of critical sensors (VIIRS, ATMS, and CrIS) that are planned for flights on NPOESS. The intent is to develop an operational prototype for the provision of satellite-based climate data as well to provide an early test of space and ground segments for NPOESS.
At this time, attention is being given to defining operational climate measurement needs and assessing their implications for instrument design. However, it is equally important to ensure that the data systems will meet climate research needs. Effective data management systems are a significant challenge to the federal agencies.

Plan The Committee on Earth Studies (CES) will provide a preliminary assessment of data processing, management, and archiving issues that should be considered in the near term as plans are developed to maximize the utility to climate researchers of data anticipated from several planned satellites. These include the satellites in the NPOESS Preparatory Program (NPP), the National Polar-Orbiting Environmental Satellite System (NPOESS), and, if approved, the Windsat program.

The committee's report will focus on the broad requirements for the data system for these satellites. As the report will be a preliminary assessment, it will also include recommendations on how to establish a process by which additional input on data and data management needs can be obtained. Where possible, the committee will note where incremental investments can provide significant improvements towards meeting the needs of climate research.
The committee's report will address issues that include:
—How to ensure that data sets of known quality will be readily available to the climate research community (along with associated meta-data on instrument calibration and performance);
—How to ensure that original multi-year radiance data are both affordable and easily retrievable from long-term archives.

In this context, issues such as consensus algorithms, data streams, quality assurance, media evolution, on-line and automated data retrieval, data set reprocessing and cataloguing will be addressed. Related issues for consideration include the importance of establishing user models for the climate community, which will, for example, explore climate data user access patterns as an input to system design.

Schedule In consultation with other relevant NRC units, the Committee on Earth Studies will lead a 2-day workshop in Washington on February 7 and 8, 2000, that will provide a forum for preliminary discussion of the data system needs for long-term satellite-based climate observations. The workshop will use current plans for data management for NPOESS and NPP and lessons learned from previous experience with operational and research satellite data management systems as a point of departure. Participants at the workshop will include agency officials, climate-change researchers, and scientists familiar with the workings of existing data archives. The committee will then meet in closed session on February 9 and 10, 2000, to begin drafting its report.

Ensuring the Climate Record from the NPP and NPOESS Meteorological Satellites

B
Workshop Agenda and Participants

Workshop on Climate Data Processing
and Archive Strategy for NPP and NPOESS
February 7-8, 2000, Washington, D.C.

AGENDA

Monday, February 7, 2000

8:30 a.m.
I. Introduction: Goals of Workshop, Ground Rules—*Mark Abbott*

8:45 a.m.
II. Welcome, Opening Remarks from Senior Agency Officials
NOAA—*Tom Karl*
NASA—*Mike Luther*
IPO—*Capt. Craig Nelson*

9:00 a.m.
III. NPOESS Data System Plans and Requirements—*Reggie Lawrence*

Focus on operational requirements as well as the kind of requirements/goals IPO has to ensure so that new instrument data streams will able to achieve climate research quality

9:30 a.m.
IV. NESDIS/NASA Plans—*Tom Karl/Martha Maiden*

Ongoing planning for future Climate Data and Services archive program housed at NOAA NESDIS: Long-term archiving, reprocessing, costs
NASA: Post-EOSDIS, linkage with NOAA, support for analysis/reprocessing

10:00 a.m.
Break

10:30 a.m.
V. NPP Present Plans-Pathfinder for NPOESS?—*Dan DeVito*

Toward a joint NASA/NPOESS/NOAA plan

Mission System Overview—*Dan DeVito*
Science Cal/Val—*Bob Murphy*
Interface Data Processor Segment in the NPP Era for Risk Reduction—*Reggie Lawrence*
Science Data Segment for NPP—*Joy Henegar*

In Situ Terminal and Processing for NPP—*Pat Coronado*
Archive and Distribution Segment for NPP—*Mike Crowe*

12:00 noon
Lunch

1:00 p.m.
VI. Theoretical and Empirical Approaches to Data Center User Access Modeling—*Bruce Barkstrom*, LARC/NASA

1:30 p.m.
VII. CES Data System Chapter, Summary of Issues—*Frank Wentz/David Sandwell*

2:00 p.m.
VIII. Charge to Working Groups

What investments can be made in the short-term to build a climate research data system? Some specific issues could include the following:

EOS data will soon be flowing, and can use those data will be used as a prototype for developing the technologies and actual architectural elements of the long-term archive climate data and products. Are there priorities or suggestions on what would be the most useful place to start? In what interactions would the community be willing to participate?

For example, scenarios such as a subset of MODIS products end-to-end flowing from NASA GSFC to NOAA NCDC across a big pipe, testing data compression forms, and data access patterns. Workshop discussions could focus on a particular DAAC/NOAA data center interaction across a range of products, instruments, data centers.

What are the priorities for which data sets we should focus on for reprocessing? What are the critical science plans that would determine how often we should reprocess? How do we begin to get a long time series across precursor —POES/GOES-EOS-NPP-NPOESS or get a better handle on the problems?

5:00 p.m.
Recess

Tuesday, February 8, 2000

8:30 a.m.
Plenary Session

9:00 a.m.
Breakout Groups

12:00 noon
Lunch

1:00 p.m.
Breakout Groups

3:00 p.m.
Plenary Session

5:00 p.m.
Adjourn

Agenda Revised 2/4/2000

PARTICIPANTS

Committee on Earth Studies:

Mark R. Abbott, Oregon State University, Chair
John R. Christy, University of Alabama, Huntsville
Catherine Gautier, University of California, Santa Barbara (unable to attend)
Christopher O. Justice, University of Virginia
Ralph F. Milliff, National Center for Atmospheric Research
Scott Pace, RAND
Dallas L. Peck, U.S. Geological Survey (retired)
Michael J. Prather, University of California, Irvine
R. Keith Raney, Johns Hopkins University Applied Physics Laboratory (unable to attend)
David T. Sandwell, Scripps Institution of Oceanography
Lawrence C. Scholz, West Orange, New Jersey
Carl F. Schueler, Raytheon Santa Barbara Remote Sensing
Graeme L. Stephens, Colorado State University (unable to attend)
Fawwaz T. Ulaby, University of Michigan (unable to attend)
Susan L. Ustin, University of California, Davis (unable to attend)
Frank J. Wentz, Remote Sensing Systems
Edward F. Zalewski, University of Arizona

Guest Participants:

Bruce Barkstrom, NASA LARC
Tony Busalacchi, NASA GSFC
Peter Cornillon, University of Rhode Island
Mike Crowe, NOAA NCDC
Dan DeVito, NASA GSFC
Jim Dodge, NASA HQ.
Vicki Elsbernd, HQ NASA
Dave Emmitt, Simpson Weather Associates, Inc.
Bob Evans, University of Miami
Wayne Faas, NOAA NCDC
Capt. Dennis Feerick, NOAA IPO
Sylvia Graff, NOAA NESDIS
Sara Graves, University of Alabama-Huntsville
Joy Henegar, NASA GSFC
Tom Karl, NOAA NCDC
Kathie Kelly, University of Washington
Steve Kempler, NASA GSFC
Herb Kroehl, NOAA NGDC
Reggie Lawrence, NOAA IPO
Mike Luther, NASA HQ
Martha Maiden, NASA HQ
Stephen Mango, NOAA IPO

Gail McConaughy, NASA GSFC
Richard McGinnis, NASA LARC
Frank Muller-Karger, University of South Florida
Bob Murphy, NASA GSFC
Capt. Craig Nelson, NOAA IPO
Joyce Penner, University of Michigan
Rama Ramapriyan, NASA GSFC
Steve Running, University of Montana
Linda Statler, NOAA NCDC
Ray Taylor, NASA GSFC
David Vercelli, NOAA NESDIS
Greg Williams, NASA HQ

National Research Council Staff:

Joseph K. Alexander, Space Studies Board
Ina B. Alterman, Space Studies Board
Arthur A.Charo, Space Studies Board
Elbert (Joe) Friday, Jr. Board on Atmospheric Sciences and Climate

Ensuring the Climate Record from the NPP and NPOESS Meteorological Satellites

C
Acronyms and Abbreviations

ADS	Archive and Distribution Segment
AIRS	Atmospheric Infrared Sounder
AMSU	Advanced Microwave Sounding Unit
AO	Announcement of Opportunity
ASTER	Advanced Spaceborne Thermal Emission and Reflection Radiometer
ATMS	Advanced Technology Microwave Sounder
AWS	Air Weather Service
CDHS	Climate Data Handling System
CDR	climate data record
CERES	Clouds and the Earth's Radiant Energy System
CIESIN	Center for International Earth Science Information Network
CrIS	Cross-Track Infrared Sounder
DAAC	Distributed Active Archives Center
DOD	Department of Defense
ECS	EOSDIS Core System
EDR	environmental data record
EMD	engineering and manufacturing development
EOS	Earth Observing System
EOSDIS	Earth Observing System Data and Information System
ESE	Earth Science Enterprise
ESIP/ESSIP	Earth Science Information Partner/Earth System Science Information Partnerships
ESSP	Earth System Science Pathfinder
FAA	Federal Aviation Administration
GOES	Geostationary Operational Environmental Satellite
HRPT	high-resolution picture transmission
IDPS	Interface Data Processing Segment
IPO	Integrated Program Office
ISP	Interim Science Panel
LTA	long-term archive
MISR	Multi-angle Imaging Spectroradiometer
MODIS	Moderate Resolution Imaging Spectroradiometer
MOPITT	measurement of pollution in the troposphere
MOU	memorandum of understanding
NASA	National Aeronautics and Space Administration
NCAR	National Center for Atmospheric Research
NCDC	National Climatic Data Center
NEDAAS	National Environmental Data Active Archive System
NESDIS	National Environmental Satellite, Data, and Information Service
NewDISS	New Data and Information System and Services
NEXRAD	Next Generation Weather Radar System
NOAA	National Oceanic and Atmospheric Administration

NPOESS	National Polar-orbiting Operational Environmental Satellite System
NPP	NPOESS Preparatory Project
NWS	National Weather Service
PI	principal investigator
POES	Polar-orbiting Operational Environmental Satellite
RDR	raw data record
SDR	sensor data record
SDS	Science Data Segment
SWT	science working team
TB	terabyte
TRMM	Tropical Rainfall Measuring Mission
TSPR	Total System Performance Responsibility
USGCRP	U.S. Global Change Research Program
VIIRS	Visible IR Imaging Radiometer Suite

Ensuring the Climate Record from the NPP and NPOESS Meteorological Satellites

D
The Role of Science Teams

The advantages of developing close working relationships among members of the science community and those directly involved with a particular instrument and its data system are well known. *The committee believes these close working relationship can be provided to a large degree by the creation of science teams for the instruments, their associated climate data records (CDRs), and the long-term archive (LTA).* Relative to the cost of the instruments and the data system, the committee sees this as being a small but essential investment. The committee advises as follows:

Science teams should be recruited competitively through Announcements of Opportunity and the proposal peer review process with 3 to 5 years of support envisioned.

Renewals should be dependent on the team members' contribution and on the need to maintain continuity. Consideration should be given to scientific publications and presentations at national and international science meetings.

There should be frequent meetings to facilitate contacts with engineering and agency representatives and to guide dataset preparation, development, and distribution.

Science teams should provide outreach to the broader science community, thus helping to strengthen the user community for the instrument.

IMPLEMENTATION PRINCIPLES

The science team construct has been applied successfully by NASA to satellite missions, often on a sensor-by-sensor basis. The committee's recommendations for developing climate processing and archival strategies for NPP and NPOESS include competitions to select the science working teams (SWTs), which will become integral parts of the program to also develop CDRs and support the LTA (see Figure D.1). The SWTs should follow some principles of implementation that are based on past experience. First, an SWT should be selected through an open and competitive process (e.g., an Announcement of Opportunity, peer-reviewed proposals, or selection by an advisory panel). SWT memberships should be planned to last 2 to 5 years in order to strike a balance between refreshing the intellectual pool regularly and attracting potential principal investigators (PIs) by offering a multiyear commitment that will allow in-depth investigations. An SWT should be limited to 15 or 20 PIs such that its meetings will be interactive and permit efficient interfaces with agency and engineering personnel. SWT meetings should be held frequently (as often as twice a year) to maximize science input that will guide system refinements and data product development and distribution. Subsets of the team may need to meet more often. Renewal awards for SWT PIs should weight heavily the PI contribution to the team, his or her record on publications germane to the sensor system or data archive in question, and PI participation in national and international science meetings where data products and science applications are reported.

The committee proposes that there should be two science teams: one associated with the instrument and the other associated with the long-term archive. A third joint working group established from members of these two teams would focus on the development, production, and management of CDRs (see Figure D.1).

BENEFITS AND ORGANIZATION OF SCIENCE TEAMS

The committee believes that competitively chosen science working teams for the NPOESS Preparatory Project would have a number of benefits because they would do the following:

Facilitate early science efforts (e.g. on prototype systems and/or synthetic datasets) that could contribute directly to engineering and systems analyses.

Optimize algorithms through competition (e.g. retrieval algorithms, extrapolations, etc.)

Provide a conduit to the user community.

Provide timely notice to the research community, which would rapidly expand the user base.

Exploit the science perspective for system refinements (i.e. for follow-on missions), validation, and error detection.

As stated in the recommendations contained in Chapter 4 and reiterated in the Executive Summary, the development of an end-to-end climate data system for NPP, including the implementation of the NPP SWT, should begin now, with EOS and other relevant existing sensor data serving as prototypes for NPP systems. In this way, the NPP SWT could influence engineering and systems analyses before the NPP launch. The NPP SWT should be encouraged to develop synthetic data sets from heritage sensor systems. Such data sets could demonstrate science applications and build the case in the science community for NPP and NPOESS as viable sources of CDR. Just prior to and just after launch, competition within an SWT would serve to optimize retrieval algorithm development and efficiency. Similarly, early science results presented by SWT members would advertise NPP and NPOESS CDR applications and rapidly expand the science user base. The SWT perspective would be critical for quality assessment and validation efforts, including in situ measurement comparisons when feasible. SWT experience would also lead to system refinements for follow-on missions (e.g., NPP to NPOESS).

Instrument Team

The committee views the NPP instrument team as requiring a mixture of expertise:

Instrument/sensor experts;

Retrieval algorithm and calibration experts;

Science applications experts (climate, with a global-scale emphasis);

Data systems experts (help develop CDRs with application experts and calibration and validation information); and

Validation expert with in situ experiment capability.

The particular blend of expertise is expected to evolve along with the program. At first, SWT expertise will likely be weighted toward sensor systems. Retrieval algorithm competitions will become important immediately before and after launch, and applications and data systems experts will be required from launch through (repeating) validation exercises. The same blend of expertise will be needed for the development of CDRs, with leadership roles probably being assumed by applications and data systems experts.

Climate Data Archive Team

The committee sees several key advantages of having a science working team for the climate data archive (CDA). These include gaining a science perspective on the minimum levels of service required to implement a useful and affordable data system and the establishment of science-based priorities for various CDRs within the NPP and NPOESS datasets. In addition, a CDA team would be an efficient means of exposure to new archive hardware and software technologies, including storage media, dataset migration, and retrieval strategies. Like the instrument team, a CDA SWT would serve as a conduit to the broader user community. It would disseminate expertise on key CDRs and serve to relieve agency responsibility for documentation by publishing in the peer-reviewed scientific literature.

The committee anticipates that the CDA SWT would require a blend of expertise:

Archive hardware experts;

Archive software experts (database systems, statistical methods for large datasets);

Climate data record experts who would develop CDRs with the help of the above-mentioned instrument team; and

Applications experts (climate data analyses, general circulation model validation).

Data system hardware and software experts should interact with experts on CDR applications. The former should collaborate with NPP SWT experts in the definition of specific CDRs for NPP and NPOESS. In addition, the CDA team might include applications experts in the fields of general circulation model validations.

CDR Joint Science Group

The CDR joint science group would consist of climate data record experts from both teams and from other groups outside the teams that are involved in the development, generation, and archival of CDRs. The task of this group would be to coordinate CDR-related activities within the two teams, to plan for the data archive needs, and to provide a forum for the development of guidelines and best practices for CDR generation and management.

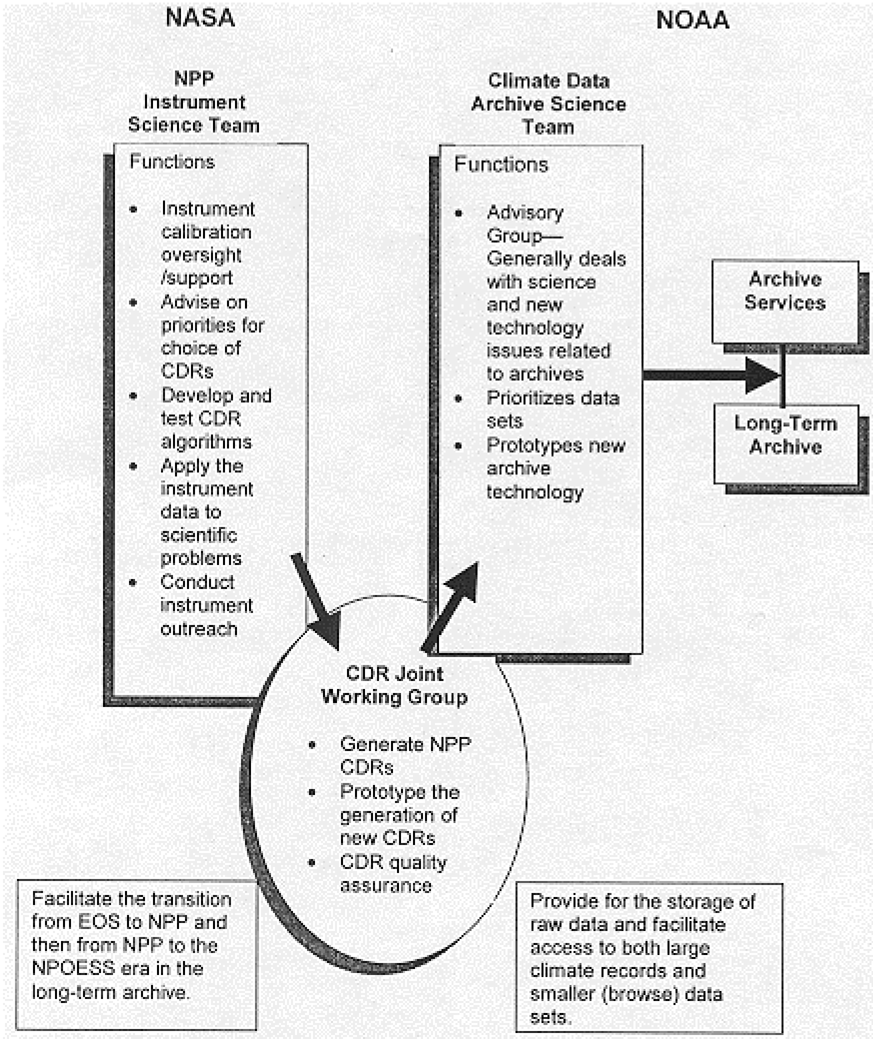

Figure D.1 Roles and responsibilities of the instrument, CDR generation, and data archive science teams.

Ensuring the Climate Record from the NPP and NPOESS Meteorological Satellites

E
The Basis for a National Climate Data Archive

This appendix, which provides background on and examines several issues in the historical quest for a national climate data system, is intended to be a brief overview and relies heavily on two key documents (NRC, 1999; USGCRP, 1999).

NATIONAL CLIMATE DATA ACT OF 1978

The basis for a national climate data archive is well established and dates to at least the National Climate Program Act of 1978.[1] This act continues in force, and the Department of Commerce is tasked to provide, among other things, "systems for the management and active dissemination of climatological data and information." NOAA is currently executing its role as guardian and distributor of the data; however, as documented in this report, increased support will be required to prevent the loss of future data.

AGREEMENTS BETWEEN NASA AND NOAA

Historically, NASA has had the responsibility for space science data. It processes, disseminates, and sometimes retains the data from its programs. In the EOS era, an attempt was made to create a data system (EOSDIS) to maintain the large volume of data expected from the EOS spacecraft. The problems of EOSDIS have been reviewed by several committees and are not discussed here. A 1989 MOU[2] between NASA and NOAA called for NASA to transfer all "oceans and atmospheric" data to NOAA and for NOAA to provide information on its GOES, POES, and related European satellite data for inclusion in the EOSDIS Information Management System. With the gradual demise of EOSDIS, this latter requirement may have become moot, but the intent was clear that NOAA is the administrator of the data.
In 1999, NASA and NOAA drafted a second MOU that provides for a partnership, "Generating Long-Term Climate Records from Earth Observing Satellite Data." This is the basis for ongoing cooperation between the NCDC and various NASA science projects. It forms the basis of the agreements on NPP and NPOESS (see Chapter 2).

The establishment of the NewDISS as a replacement for EOSDIS is expected to lead to a distributed system with a number of elements that vary in their approach to data production and storage. This "federation" of centers avoids several problems of EOSDIS and other centralized software projects. Each development is more manageable and is completed in a shorter time. This approach allows evolutionary incorporation of new technology into the system and prevents early obsolescence. These specialized active archive centers can be built quickly and eliminated as the programs they represent are completed. However, it puts more pressure on a LTA to absorb and maintain the data sets. Detailed plans for NewDISS and its relationship to a future climate LTA were not available to the committee at the time this report was written.

NATIONAL SPACE POLICY

The President, through his Science Advisor, generates space policy statements that set the policy and overall goals of the U.S. Space Program. The policy statement in 1996 (NSTC PDD-8) again indicated that NOAA has responsibility for the data on the civil side. The following quote allows for a data archival, though it does not require one:

> The Department of Commerce (DoC), through the National Oceanic and Atmospheric Administration (NOAA), has the lead responsibility for managing Federal space-based civil operational Earth observations necessary to meet civil requirements. In this role, the DoC, in coordination with other appropriate agencies, will: (a) acquire data, conduct research and analyses, and make required predictions about the Earth's environment; (b) consolidate operational U.S. Government civil requirements for data products, and define and operate Earth observation systems in support of operational monitoring needs; and (c) in accordance with current policy and Public Law 102-555 provide for the regulation and licensing of the operation of private sector remote sensing systems.

DATA MANAGEMENT FOR THE GLOBAL CHANGE RESEARCH PROGRAM

The U.S. Global Change Research Program (USGCRP) is often identified as a vehicle for climate and related research. The program commands the attention of a large segment of the scientific community interested in climate and Earth observation from space. However, the USGRCP is not a single entity or agency of the government but a collection of interested parties. As such, it is not an effective body for the management of a climate program but would be an effective voice in providing guidelines, requirements, and direction for the research.

An extensive review of the USGCRP was carried out recently by the NRC Committee on Global Change Research (NRC, 1999). The NRC review was concerned largely with the future direction of global change research; however, issues related to data system problems were also addressed. In particular, Chapter 9, "Processing and Distributing Earth Observations and Information," supports the long-term functional requirements of the EOSDIS; it also outlines several objectives for the data system.

Several of the findings and recommendations of the 1999 NRC report specifically address data systems.[3] The committee concurs with the NRC guidance; further, it believes this guidance is consistent with that presented by participants at the February 2000 workshop. Specifically, workshop participants advised that in designing a climate data LTA for NPP and NPOESS the following should be done:

A general set of functional requirements should be provided;

Interfaces should be specified at a high level to ensure interoperability and to foster use of the data;

Responsibility for development should be given to interested (users) groups that understand the data; and

Development of the archive should be accomplished via the combination of smaller segments.
The committee notes that these steps are consistent with a strategy that gives NOAA overall responsibility for the data archive, with ancillary agreements between NASA and NOAA as needed.

REFERENCES

National Research Council (NRC), Board on Sustainable Development. 1999. *Global Environmental Challenge: Research Pathways for the Next Decade*. Washington, D.C.: National Academy Press.

U.S. Global Change Research Program (USGCRP). *Global Change Science Requirements for Long-Term Archiving, Report of the Workshop*, October 28-30, 1998, Boulder, Colo., March 1999.

[1]Section 108 of Public Law 101-606, "Global Change Research Act of 1990," November 16, 1990 refers to the National Climate Program Act of 1978 (15 USC 2901 et seq.) See NRC 1999, Appendix A. See also http://www4.law.cornell.edu/uscode/15/ch56.html#PC56.

[2]A copy of this MOU is included in USGCRP, 1999.

[3]Finding 5, Recommendation 5, and discussion on p. 531 of Chapter 11, "Findings and Recommendations," in NRC (1999).