United States General Accounting Office

GAO

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

September 1991

EARTH OBSERVING SYSTEM

Information on NASA's Selection of Data Centers

RELEASED

GAO/IMTEC-91-67

GAO

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

Information Management and
Technology Division

B-245033

September 18, 1991

The Honorable George E. Brown, Jr.
Chairman, Committee on Science,
Space, and Technology
House of Representatives

Dear Mr. Chairman:

On June 20, 1991, you requested that we provide information about the National Aeronautics and Space Administration's (NASA) selection of seven Distributed Active Archive Centers (DAACs) for the Earth Observing System Data and Information System (EOSDIS). EOSDIS is a large, distributed information system project expected to cost between $3.6 and $4 billion through fiscal year 2000. The selection of DAAC sites is important because it will influence how broadly and closely EOSDIS will be linked with the scientific community. This report discusses (1) the process NASA used to select the seven DAAC sites, (2) other data centers with significant earth science holdings, and (3) the roles and functions of data centers within EOSDIS. Details of our objectives, scope, and methodology are provided in appendix I.

Results in Brief

According to EOS program officials, the process NASA used to select the seven DAAC sites was largely subjective. They said that no systematic, objective analyses were made in which candidate sites were weighed against the agency's stated criteria. A number of data centers with significant earth science data holdings were not selected as DAACs, including one other NASA center as well as facilities operated by the National Oceanographic and Atmospheric Administration (NOAA), the National Science Foundation, the Environmental Protection Agency, and the Departments of Energy and Agriculture. NASA plans to negotiate agreements with some of these other agencies to have their facilities serve as affiliated data centers, providing EOSDIS with specially processed data unavailable elsewhere.

As integral parts of EOSDIS, the DAACs will be responsible for processing, archiving, and distributing data from the EOS observatory instruments. Each DAAC will handle data related to certain specific scientific disciplines. NASA has allowed for future changes to the number or location of the DAACS, but currently has no plans to make any such changes.

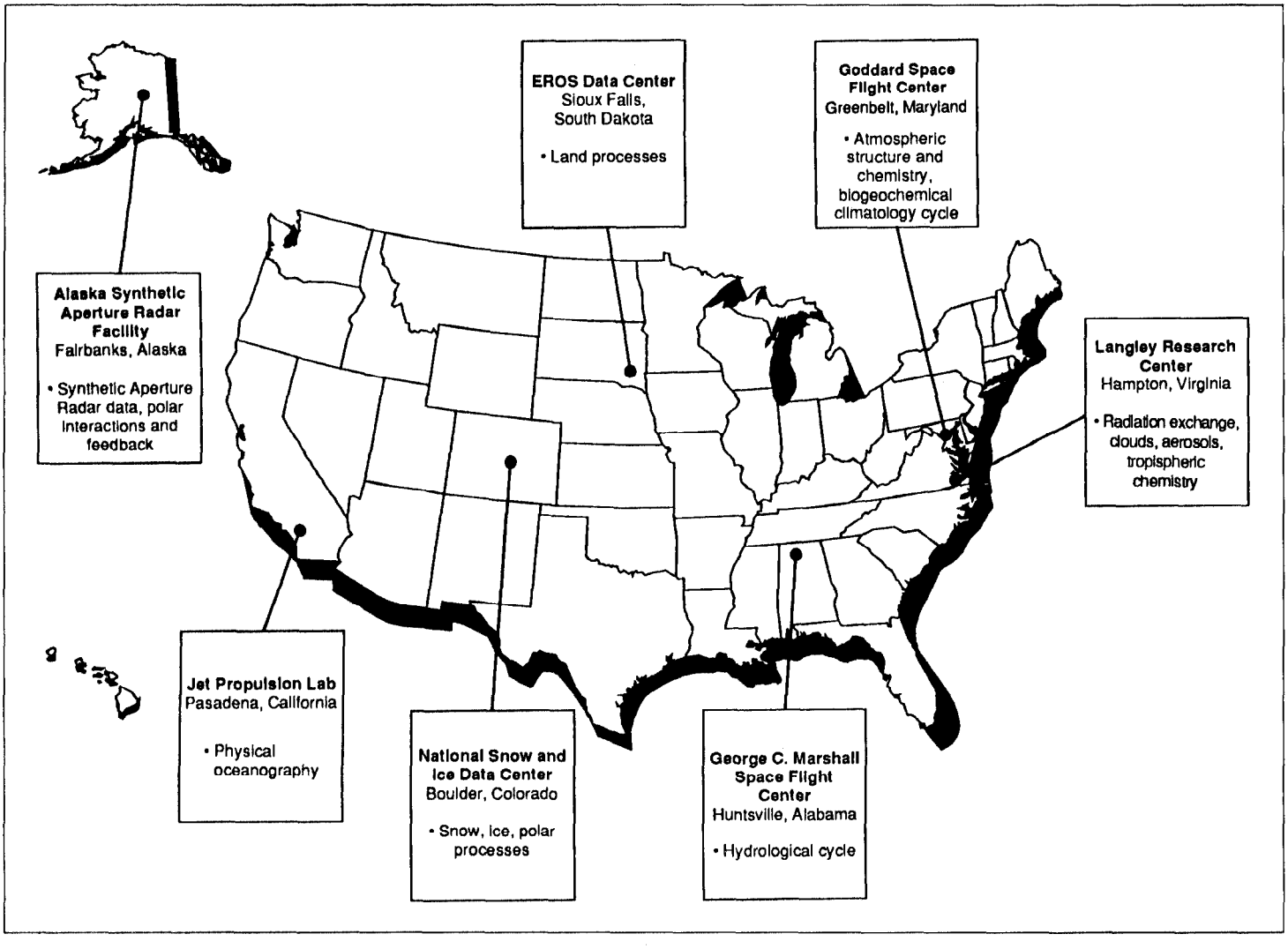
Background

The goal of EOS is to create a unified scientific observing system that will permit interdisciplinary and multidisciplinary studies of the earth's atmosphere, biosphere, oceans, land surfaces, polar regions, and internal processes over a 15-year period. EOS, which is estimated to cost between $16 and $18 billion through fiscal year 2000, is the centerpiece of NASA's Mission to Planet Earth Program, which aims to obtain a scientific understanding of the entire earth as an integrated system and to determine the processes that contribute to environmental balance or change.

Central to EOS is a series of space-based observatories containing a variety of instruments that will collect data about the earth on a continual basis beginning in 1999. Those data will be relayed to EOSDIS which, in turn, will process it into a form usable by the scientific research community. In addition to providing data from the EOS observatories themselves, EOSDIS is intended to give scientists access to a variety of other earth science data, from other remote sensing satellite missions as well as from ground-, ocean-, and air-based observations. NASA has estimated that EOSDIS will serve nearly 100,000 users from federal, state, and local government agencies; academia; foreign countries; and industry.

NASA decided to spread the data-handling functions of EOSDIS among seven DAAC sites located around the country. NASA plans to contract for EOSDIS development and will provide the system (hardware and software) to each DAAC. NASA's decision is in keeping with recommendations of the scientific community to diversify EOSDIS and to build on existing data systems and expertise. The planned DAACs are shown, with their designated scientific disciplines, in figure 1.

Figure 1: DAAC Sites and Assigned Disciplines



DAACs will play a critical role in the design and operation of EOSDIS, a system described by the National Research Council as a pioneering effort, far exceeding any existing civilian information system.[1] DAAC activities will be funded from the EOS program budget. As an element of EOSDIS, each DAAC will be responsible for processing, archiving, and distributing data from the EOS observatory instruments related to its designated scientific discipline. Thus some DAACs will process more data than

[1]The U.S. Global Change Program: An Assessment of FY 1991 Plans, National Academy of Sciences' National Research Council, National Academy Press, 1990.

others. Goddard Space Flight Center, for example, will process many times more data than Marshall Space Flight Center because more instruments are planned within Goddard's designated scientific disciplines, and those instruments will produce data at higher rates. In addition to DAACS, NASA intends to negotiate with certain other data centers to obtain data needed by EOSDIS, such as NOAA's National Climatic Data Center and National Oceanographic Data Center. These sites, to be called Affiliated Data Centers (ADCS), are expected to be funded by their operating agencies rather than NASA, with one exception.[2] The roles and functions of DAACS and ADCS are discussed in greater detail in appendix II.

DAAC Selection Criteria and Process

NASA has articulated three specific criteria for the selection of DAAC sites. According to a NASA program document, a potential DAAC site should possess (1) a critical mass of in-house scientific expertise in the use of remotely sensed data, (2) an existing management and technical infrastructure, and (3) a long-term institutional commitment to the processing and archiving of EOS and related non-EOS data.[3] According to program officials, the first criterion is consistent with recommendations by the National Academy of Sciences to locate space science data management at sites where the research community is active. The second and third criteria reflect program managers' concerns that the chosen sites have both the capability and the desire to commit to a long-term relationship with the EOS program.

According to EOS program officials, the number of DAAC sites, their locations, and their assigned scientific specialties for the EOS program were all decided upon informally, without systematic or objective analyses of potential sites against the stated criteria. Instead, the program officials relied on their own experience and knowledge of the capabilities at potential sites in determining where DAACS should be located. NASA's reasons for choosing each of the DAACS are detailed in appendix III.

A number of other earth science data centers exist, some of which may become ADCS. Most are funded by other government agencies, such as NOAA, the National Science Foundation, the Environmental Protection Agency, and the Departments of Energy and Agriculture. NOAA, in particular, has extensive holdings related to climatology and oceanography and also has expertise in developing computer procedures to process

[2]The exception is the Consortium for International Earth Science Information Networks (CIESIN), which recieves funding from NASA but is independently managed.

[3]Early EOSDIS Program Plan, draft version, August 1990.

weather data. NASA recognizes that EOSDIS will need ready access to data held at non-EOSDIS locations, such as NOAA's data centers, and plans to negotiate access to these data by establishing these centers as ADCs. However, because NASA has not yet completed a thorough study of the non-EOS data that EOSDIS will need to fully achieve its goals, it does not yet know the full extent to which it will need access to data outside of EOSDIS.

Future Changes in DAAC Selections

NASA officials said that they currently have no plans to increase the number of DAAC sites or change any of the selected locations. However, changes could be made in the future. In its request for proposals for the main EOSDIS hardware and software contract, NASA listed the seven DAAC sites as a "minimal" configuration, and stated that the number could be increased over the life of the contract. Further, the request states that "the Contractor's design shall accommodate changes in the number and location of DAACS."

Our work was performed in accordance with generally accepted government auditing standards. We discussed the report's contents with NASA officials; however, as requested by the Committee, we did not obtain written comments from NASA on a draft of the report.

As arranged with your office, unless you publicly announce the contents of this report earlier, we plan no further distribution of it until 30 days from the date of this letter. We will then give copies to other appropriate congressional committees; the Administrator, NASA; and will make copies available to other interested parties upon request. Please contact me at (202) 275-4649 if you have any questions concerning this report. Major contributors are listed in appendix IV.

Sincerely yours,

Samuel W. Bowlin
Director, Defense and Security
Information Systems

Contents

Abbreviations

ADC	Affiliated Data Center
CIESIN	Consortium for International Earth Science Information Networks
DAAC	Distributed Active Archive Center
EOS	Earth Observing System
EOSDIS	EOS Data and Information System
EROS	Earth Resources Observations System
GAO	General Accounting Office
GOES	Geostationary Operational Environmental Satellite
IMTEC	Information Management and Technology Division
NASA	National Aeronautics and Space Administration
NOAA	National Oceanographic and Atmospheric Administration

Objectives, Scope, and Methodology

In June 1991, the House Committee on Science, Space, and Technology requested that we provide information about NASA's selection of the seven EOSDIS DAACs. To do this we interviewed the EOS program managers involved in making the selections and reviewed related program documentation. We interviewed the managers of the DAACs at Goddard Space Flight Center, Earth Resources Observations System Data Center, and Marshall Space Flight Center about the capabilities they have in place and their future plans. We also interviewed officials at NOAA's National Oceanographic Data Center and National Environmental Satellite Data and Information Service. We obtained copies of draft self-assessments of capabilities at each of the DAACs and estimates of data-processing volumes expected during the EOS era. We did not independently assess NASA's specific selections for DAACS, nor did we evaluate sites that were not chosen as DAACs. We also did not assess the appropriateness of choosing a seven-DAAC configuration.

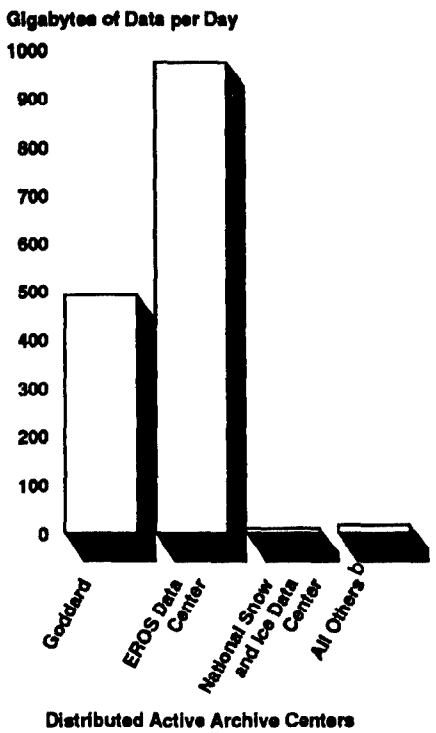
We performed our work from June through August 1991, at various sites including NASA headquarters, Washington, D.C., and the Goddard Space Flight Center, Greenbelt, Maryland.

Roles and Functions of EOSDIS Data Centers

DAAC Roles and Functions

The core responsibility of each DAAC will be to process, store, and distribute the stream of data that it will be receiving from designated instruments on the EOS observatories. Some DAACs will process more data than others. Representative data volumes and processing loads projected for the first EOS platform (EOS-A) and the free-flying Synthetic Aperture Radar instrument are shown in figures II.1 and II.2.

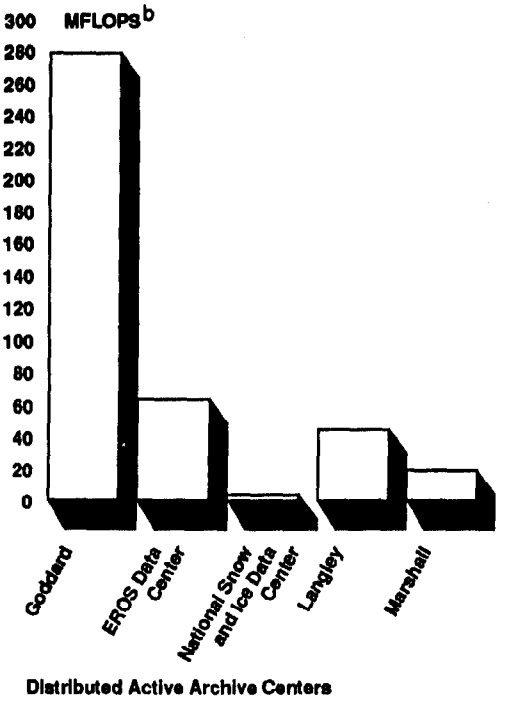
Figure II.1: Volume of EOS Data to Be Stored Daily at Each DAAC[a]



[a]Includes data from EOS Platform A and EOS Synthetic Aperture Radar Instrument.

[b]Includes Jet Propulsion Laboratory, Alaska, Langley, and Marshall.

Figure II.2: Processing Load for EOS-A Data by DAAC[a]



[a]Does not include Jet Propulsion Laboratory and Alaska. The daily processing load at Jet Propulsion Laboratory is only 0.9 MFLOPS. No data are available for Alaska, where only data from the Synthetic Aperture Radar instrument, a separate satellite from the EOS observatories, will be processed.

[b]Million floating point operations per second. The processing loads shown here are the long-term average rates at which the DAAC computer systems must process data to keep up with the input flow.

NASA plans to work with the DAACs to build prototypes of some of the functions needed for EOSDIS and to develop an early capability to process, archive, and distribute data in the way that will be needed once the data stream from the EOS observatories begins. Beginning in 1996, NASA's main EOSDIS contractor will provide hardware and software to each of the DAAC sites to support the fully operational EOSDIS. This hardware and software will duplicate a system that will be originally installed and tested by the contractor at Goddard Space Flight Center.

The systems to be developed at each DAAC include a product generation system, a data archive and distribution system, and an information management system. The product generation system will be responsible for processing the raw instrument data it receives into standard products within 4 days of the space-based observation. This system also will be able to reprocess data when necessary without interrupting the

normal processing of new data. The data archive and distribution system will archive and distribute to the user community the large quantities of data that are produced by the product generation system, as well as an assortment of other supporting data, both from within the EOS program and from a variety of other sources. The information management system will be the user interface for the DAAC, offering users access to all data throughout EOSDIS, including data resident at a particular DAAC, plus a variety of other user services, such as help in locating and ordering data.

An internal network will link the seven DAACs together into one EOSDIS with a single user interface. This means that a scientist signing on to the system at any of the individual DAAC sites would be presented with the same complete access to all EOSDIS data sets. The user would not have to know at which DAAC any particular data set was physically located to access it.

The DAAC systems will reside at facilities provided by the individual sites. NASA is in the process of requesting each of the DAAC sites to provide specific plans for accommodating their DAAC functions. Over the next year or two the EOS project hopes to sign memoranda of understanding with each of the DAACs to solidify their relationships with the project.

ADC Roles and Functions

A variety of other earth science data centers, not funded by NASA, perform data management functions analogous to those of a DAAC. Some of these centers will have an important role in the success of EOSDIS, either because they have data that will be needed in conjunction with EOS data or because they provide special non-EOSDIS services required by the EOS program. For example, CIESIN, a group of Michigan academic institutions, will provide decisionmakers with EOS data that may be important for policy decisions. In these cases, NASA intends to negotiate a working agreement with the data center to establish a degree of connectivity and interoperability with EOSDIS in order to meet the specific data access requirements of EOS. Such centers would be referred to as Affiliated Data Centers (ADCs). They will not be as closely linked to the development of EOSDIS as the DAACs will be, nor will they, with the exception of CIESIN, be funded by NASA. Several sites have already been designated as ADCs, and more could be chosen in the future. Descriptions of current ADCs appear in appendix III.

According to EOS program officials, NASA intends to offer ADCs copies of the information management system software used at DAACs, which is

the software that provides data access and interoperability among the seven separate DAAC sites. NASA would also provide detailed specifications for establishing an interface with EOSDIS. It will be up to the ADCs to determine whether their systems need to be modified to connect with EOSDIS and whether such modifications are worthwhile.

Profiles of EOSDIS Data Centers

Individual DAACs and ADCs are discussed briefly below. DAAC discipline assignments, current data systems, and data holdings are summarized in table III.1.

Table III.1: DAAC Discipline Assignments, Current Data Archive Systems, and Data

DAAC	Discipline Responsibilities	Current Data Archive Systems	Types of Current Data	Current Data Volumes (in gigabytes)
Goddard	Atmospheric structure and chemistry, biogeochemical cycles, and climatology	NASA Climate Data System, Pilot Land Data System, Coastal Zone Color Scanner	Atmosphere, biosphere, land, ocean	2,848
EROS Data Center	Land processes	Global Land Information System	Land	77,000
Jet Propulsion Laboratory	Physical oceanography	NASA Ocean Data System	Sea surface, sea ice, phytoplankton, water	0.06
National Snow and Ice Data Center	Snow, ice, polar processes	Cryosphere Data Management System	Snow, ice	66
Alaska	Synthetic Aperture Radar data, polar interactions and feedback	Archive and Operations System	None	None
Langley	Radiation exchange, clouds, aerosols, tropospheric chemistry	None	Radiation exchange, cloud, aerosol, trace gas, tropospheric chemistry	331
Marshall	Hydrologic cycle	WetNet	Cloud liquid water, lightning, marine wind speed, temperature, water vapor	84

Distributed Active Archive Centers (DAACs)

Goddard Space Flight Center

NASA originally planned to centralize EOSDIS at Goddard. Contractor studies conducted during an early phase of the EOS program showed that this would be the most efficient and least expensive option. NASA eventually decided to distribute EOSDIS functions after being prompted to do so by a panel of scientists chartered to advise NASA about EOSDIS. The panel argued that it was more important to link EOSDIS development to the earth science and data center expertise that was distributed around the country. Nevertheless, Goddard will remain a very large DAAC. Goddard

supports a variety of earth science research programs and maintains several earth science data systems, including the NASA Climate Data System, Pilot Land Data System, and Coastal Zone Color Scanner system. These systems currently hold approximately 1,200 gigabytes[1] of data. An additional 1,640 gigabytes of earth science-related data are stored at Goddard's National Space Science Data Center. Goddard is designated the Global Change DAAC and will be responsible for atmospheric structure and chemistry, biogeochemical cycles, and climatology.

Earth Resources Observations System (EROS) Data Center

The EROS Data Center, operated by the U.S. Geological Survey, was selected so that NASA could establish a partnership with the Geological Survey in global change research. The EROS Data Center is the repository of Landsat satellite data and other land-oriented holdings, such as topographic and cartographic data. EROS Data Center has considerable data-handling experience, including recently implementing the Global Land Information System, a data distribution system that provides functions similar to those planned for EOSDIS. The EROS Data Center will be responsible for processing, distributing, and archiving all EOS data dealing with land processes.

Jet Propulsion Laboratory

The Jet Propulsion Laboratory was chosen as a DAAC largely because it has developed and operates the NASA Ocean Data System, which serves the oceanographic community. This system maintains data from the Seasat satellite mission and is slated to receive data from two upcoming oceanographic satellite missions prior to EOS. Currently the system stores about 60 megabytes of oceanographic data. As a DAAC, the Jet Propulsion Laboratory will support oceanography, geophysical science, and interdisciplinary studies by acquiring, processing, and distributing geophysical data derived from satellite and other sources.

National Snow and Ice Data Center

The National Snow and Ice Data Center, located at the University of Colorado and funded by NOAA, was selected because it is already the snow and ice data center for NOAA and has a data system, funded by NASA, to handle various snow and ice data types. Ground-based ice data are collected and archived at the center, which supports in-house ice research. About 67 gigabytes of data are currently at the center. As a DAAC, the center will be responsible for snow, ice, and polar processes.

[1]One gigabyte of data is approximately one billion bytes. A byte is equivalent to 1 character in a text file, such as the letter "a."

Alaska Synthetic Aperture Radar Facility

The Alaska facility, built by NASA, was selected because it is the receiving and processing station for Synthetic Aperture Radar data from several pre-EOS missions. As a DAAC, it will allow polar scientists to address issues related to global change at high latitudes.

Langley Research Center

Although Langley does not currently archive or distribute earth science data to the scientific community, EOS program officials selected it as a DAAC because of its experience as a processing center for data from earlier NASA missions that studied aerosols and radiation to and from the earth. For internal use by project scientists from those missions, Langley currently maintains approximately 331 gigabytes of data. As the radiative processes and tropospheric chemistry DAAC, Langley will be responsible for radiation exchange, clouds, aerosols, and tropospheric chemistry data.

Marshall Space Flight Center

The selection of Marshall was based on its expertise with microwave and lightning remote sensing data as well as its involvement in WetNet, an experimental closed system of 40 users that distributes hydrologic data by subscription. Approximately 84 gigabytes of data are currently archived at Marshall. As a DAAC, Marshall will continue to be responsible for hydrologic cycle data.

Affiliated Data Centers (ADCs)

The following sites have already been chosen as ADCs. Additional sites could be added in the future.

National Oceanographic and Atmospheric Administration (NOAA)

NOAA produces many atmospheric and oceanographic data sets and maintains a number of data bases of key importance to the EOS program. Some NOAA data sets and standard data products will be required by EOSDIS on a routine basis to support production or validation of EOS standard products. Further, NOAA will need special, real-time access to some EOS data for its operational activities.

Various NOAA organizations produce or hold data of known or potential utility to the EOS program. The Office of Satellite Data Processing and Distribution, for example, will be the source of data from NOAA's operational environmental satellites that will be required on a routine basis

for EOSDIS. NOAA's National Climatic Data Center holds the largest collection of climatic data in the world and collects, processes, archives, and distributes climate data from worldwide sources, including meteorological data ranging from Thomas Jefferson's weather observations to data acquired by active spacecraft. The National Climatic Data Center's historical data holdings may play an important role in global climate research.

NOAA's National Oceanographic Data Center collects oceanographic data from more than 860,000 stations around the world, as well as from U.S. and foreign spacecraft. It also maintains one of the world's largest oceanographic data bases. NOAA also operates the National Geophysical Data Center, which maintains an extensive archive of geophysical and related data, and the National Meteorological Center, which produces weather and climate forecasts that are of critical importance to global change research.

NASA is planning to negotiate an agreement with NOAA regarding its participation as an ADC in EOSDIS. NASA envisions establishing a single link with NOAA as a whole rather than separate links with NOAA's various data centers.

Consortium for International Earth Science Information Networks (CIESIN)

The primary role of CIESIN is to make the data from EOSDIS accessible to policymakers. It will provide to nonresearch users, such as government decisionmakers, environmentalists, and farmers, information on global change issues and an ability to effectively access EOSDIS. CIESIN will not archive any low-level scientific data but, rather, will maintain the higher level products that it generates, as well as other high-level data, such as demographic or socioeconomic information useful for global change research related to human activities.

University of Wisconsin

The University of Wisconsin's Space Science and Engineering Center archives and distributes data from NOAA's Geostationary Operational Environmental Satellites (GOES). It also develops and tests new algorithms for deriving information from the raw satellite data. EOSDIS will provide a link to this archive in order for EOSDIS users to be able to search for and request off-line delivery of GOES data.

Other ADCs

Plans call for other sites being designated as ADCs as the need for their data within EOSDIS becomes more apparent. NASA is still in the process of assessing what existing earth science data will be most valuable for the EOS program.

Major Contributors to This Report

Information Management and Technology Division, Washington, D.C.

Ronald W. Beers, Assistant Director
John A. de Ferrari, Evaluator-in-Charge
Elizabeth L. Johnston, Computer Scientist
Michael P. Fruitman, Supervisory Reports Analyst

Ordering Information

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

U.S. General Accounting Office
P.O. Box 6015
Gaithersburg, MD 20877

Orders may also be placed by calling (202) 275-6241.

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

Official Business
Penalty for Private Use $300

First-Class Mail
Postage & Fees Paid
GAO
Permit No. G100