United States General Accounting Office

GAO

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

bruary 1992

EARTH OBSERVING SYSTEM

NASA's EOSDIS Development Approach Is Risky

145956

GAO/IMTEC-92-24

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

Information Management and
Technology Division

B-247178

February 25, 1992

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

Dear Mr. Chairman:

This report responds to your request that we review the National Aeronautics and Space Administration's (NASA) development of the Earth Observing System Data and Information System (EOSDIS). Because the Committee intends to undertake extensive oversight of EOSDIS, you asked that we examine several key aspects of NASA's system development approach, including the extent to which EOSDIS (1) will use prototyping to reduce system development costs and risks, and (2) may require technologies that are beyond the current state of the art. The report identifies weaknesses in NASA's development strategy, which could prevent future global change researchers from gaining access to and making optimal use of valuable as well as costly data.

As agreed 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 appropriate congressional committees; the Administrator, NASA; and other interested parties. Copies will also be made available to others upon request.

This work was performed under the direction of Samuel W. Bowlin, Director, Defense and Security Information Systems, who can be reached at (202) 336-6240. Other major contributors are listed in the appendix.

Sincerely yours,

Ralph V. Carlone

Ralph V. Carlone
Assistant Comptroller General

Executive Summary

Purpose

NASA has initiated a program, called the Earth Observing System (EOS), to collect an integrated set of data to study the earth's atmosphere, biosphere, oceans, and land surfaces as a complete system over a 15-year period. The data from EOS will be processed, archived, and distributed by the EOS Data and Information System (EOSDIS). NASA estimates that EOSDIS development will cost around $3 billion through fiscal year 2000.

Given the unprecedented scope of the EOS effort and the critical importance of EOSDIS, the House Committee on Science, Space, and Technology asked GAO to examine several key aspects of NASA's development approach, including the extent to which EOSDIS (1) will use prototyping to reduce system development costs and risks, and (2) may require technologies that are beyond the current state of the art.

Background

Central to the EOS program 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 1998. The data from these spacecraft will be sent to seven distributed data centers that, as the core of EOSDIS, will process, archive, and distribute that data to the scientific community.

NASA bills EOSDIS as a comprehensive system that will bring together data from many sources to serve the needs of scientists performing integrated, interdisciplinary studies of the earth. In addition to data from the EOS observatories, EOSDIS is intended to be flexible enough to include previously archived data; new measurements from other non-EOS spacecraft; various ground-, ocean-, and air-based measurements; and data processing software developed by the scientific user community. As such, EOSDIS will be the one system responsible for archiving and distributing all NASA earth science data. The goal of EOSDIS is to make this vast wealth of data easily accessible and affordable to a broad range of researchers and to aid these researchers in producing useful global change data products and models. The intended scope of EOSDIS far exceeds that of any previous civilian data management system.

Results in Brief

NASA's development strategy does not adequately identify and mitigate the significant technological risks inherent in a project of the size, scope, and technical complexity of EOSDIS. NASA is not devoting enough attention to prototyping key aspects of the EOSDIS concept, nor is it adequately planning for the development and incorporation of critical advanced technologies.

NASA's near-term EOSDIS prototype projects do not fully address critical areas where technical feasibility is in question nor are they substantial enough to allow users to assess key EOSDIS functions. Furthermore, certain key advanced technologies, including new data base search techniques and data storage and retrieval methods, are not specifically addressed in NASA's official development strategy, even though they have been identified as critical to the long-term success of EOSDIS. Without specific plans and resources in place to assure that emerging user needs are satisfied and key technical challenges are aggressively pursued and resolved, NASA is running an unnecessarily high risk that EOSDIS may not meet future global change research needs and may need costly modifications to be useful.

Principal Findings

Early EOSDIS Prototype Projects Do Not Address Major Development Risks

NASA has initiated a number of prototyping projects within the early pre-contract phase of EOSDIS development, known as Version 0. However, these projects do not address critical areas where technical feasibility is in question, nor are they substantial enough to allow users to assess key EOSDIS functions. Without more complete and rigorous early prototyping, NASA will be making its EOSDIS procurement decision with scant feedback on the technical feasibility or user acceptance of major design concepts. As a result, the full-scale development of the system will risk not meeting scientific needs and expectations.

Of the $83 million that NASA plans to spend on pre-contract activities in fiscal year 1992, only $19 million is for Version 0. Similar annual amounts are projected for the duration of Version 0. This includes work on three major prototype projects being directed by the project office. These projects are aimed at increasing the interoperability of existing earth science data systems through development of a common user interface, networking, and data format standards. They do not stand as true EOSDIS prototypes because they focus on current systems and do not fully address key technical risks inherent in designing the full-scale system. For example, rather than researching new ways to provide unprecedented levels of access to vast amounts of earth science data, these projects largely enhance current levels of service through incremental improvements in connectivity and interoperability. Furthermore, in building these prototypes, NASA is not taking full advantage of relevant systems experience and knowledge in other agencies and programs. The

National Center for Atmospheric Research (NCAR), for example, has extensive experience in managing large amounts of on-line earth science data on its Mass Store System and could provide expertise in designing Version 0 prototypes. Because of these shortcomings, the results of the Version 0 prototypes may be of little value to the long-range system development effort.

New Technology Will Be Needed to Fulfill EOSDIS Goals

According to specialists in data base research and data management, without several significant technological advances EOSDIS cannot reach its goal of providing ready access to the vast amount of EOS and other earth science data that global change researchers will need. Current data base search techniques, which were generally designed to meet the needs of transaction-based business applications, are inadequate to support the work of interdisciplinary earth scientists. New software technologies must be developed to allow searching and sorting electronic images by the natural features they depict. Efficient methods of displaying the results of large-scale computations and of locating and accessing specific data from within the massive amounts being collected also must be developed.

Advanced data base search techniques, new ways of displaying complex interrelationships among data, and efficient access to data stored in massive archives are all examples of technologies that are critical to the success of EOSDIS. Although NASA has allowed for future incorporation of advanced technology by endorsing an evolutionary approach to EOSDIS development, the program, as currently structured, gives no specific direction in these areas. Furthermore, EOS program officials have not supported proposals by NASA's research organization to begin doing research in these technologies in the near term.

EOSDIS has been initiated and funded on the premise that it will offer substantial improvements over existing earth science data systems, which are limited in their ability to support global change research. Yet without specifically addressing key advanced technology areas in the context of EOSDIS, NASA risks developing a system that is little better than these current systems.

Recommendations

Given the significant technological risks inherent in the EOSDIS concept, GAO recommends that the NASA Administrator not award the planned EOSDIS Core System contract until specific plans have been developed and resources identified for (1) prototyping the full range of critical system elements and (2) guiding and accelerating research into key advanced technologies that will be essential for the system's ultimate success. In addition, GAO recommends that NASA work to maximize cooperation with other federal agencies and earth science programs having data system experience in global change-related areas.

Agency Comments

As requested, GAO did not provide a draft of this report to NASA for its review and comments. However, GAO discussed the report's contents with NASA officials and included their comments as appropriate. NASA officials agreed that Version 0 prototypes do not fully address critical development risks but believe that prototypes developed by the contractor will eventually address these risks. However, NASA could not provide convincing evidence to support its belief. Neither the terms of the contract nor NASA's management approach to date ensure that the contractor will address critical development risks. The officials also disagreed with the statement that they are not taking advantage of other agencies' systems resources and experience. However, GAO's discussions with officials from these agencies suggest that fuller cooperation is possible. With regard to planning for advanced technologies, the NASA officials agreed with our findings and recommendation.

Contents

Abbreviations

DAAC	Distributed Active Archive Center
DADS	Data Archive and Distribution System
EOS	Earth Observing System
EOSDIS	EOS Data and Information System
GAO	General Accounting Office
IMS	Information Management System
IMTEC	Information Management and Technology Division
NASA	National Aeronautics and Space Administration
NCAR	National Center for Atmospheric Research
NOAA	National Oceanographic and Atmospheric Administration
PGS	Product Generation System
RFP	request for proposals

Introduction

Background

NASA is a participant, along with eight other U.S. government agencies, in the U.S. Global Change Research Program, which aims to reduce key scientific uncertainties and improve scientific predictions regarding global change. The program is overseen by the Committee on Earth and Environmental Sciences, which establishes research priorities and coordinates contributions from each of the participating federal agencies.

Within the context of the Global Change Research Program, NASA's contribution is Mission to Planet Earth, a collection of earth-observing projects that are based on the systemwide approach NASA took in designing spacecraft missions to observe other planets in the solar system. The basic concept behind Mission to Planet Earth is to study the processes, cycles, and interactions among the components of the earth's geosphere and biosphere as a totally integrated system, instead of as a set of discrete, but interrelated, components. The centerpiece of Mission to Planet Earth is the Earth Observing System (EOS), which includes a series of satellites, a scientific research program, and a data and information system (EOSDIS). Mission to Planet Earth also includes several other satellite missions, such as the Upper Atmosphere Research Satellite, launched in 1991.

The Earth Observing System (EOS)

EOS, which is estimated to cost $11 billion through fiscal year 2000, is a large and ambitious program that marks a new, integrated approach to remote sensing of the earth. The specific goal of EOS is to create a unified scientific observing system that will permit interdisciplinary[1] studies of the earth's atmosphere, biosphere, oceans, land surfaces, and polar regions over a 15-year period. Figure 1.1 shows visual examples of some of the kinds of data that will need to be correlated and studied as an integrated system.

[1]Studies that draw on data and expertise from more than one earth science discipline. Earth science disciplines include hydrology, meteorology, and geophysics, among others.

Figure 1.1: Earth Images Showing a Variety of Interrelated Global Change Data

Source: NASA

The major component of EOS is its series of space-based observatories, each containing a variety of instruments that will collect data about the earth on a continual basis beginning in 1998. NASA originally intended to launch large observatory platforms with 9 to 12 separate scientific instruments on each. After an independent review in 1991, NASA decided to spread its scientific instruments among more medium to smaller platforms. Regardless of the number or size of the EOS platforms, NASA's strategy is to collect simultaneous or near-simultaneous measurements from a variety of different instruments.

Earth Observing System Data and Information System (EOSDIS)

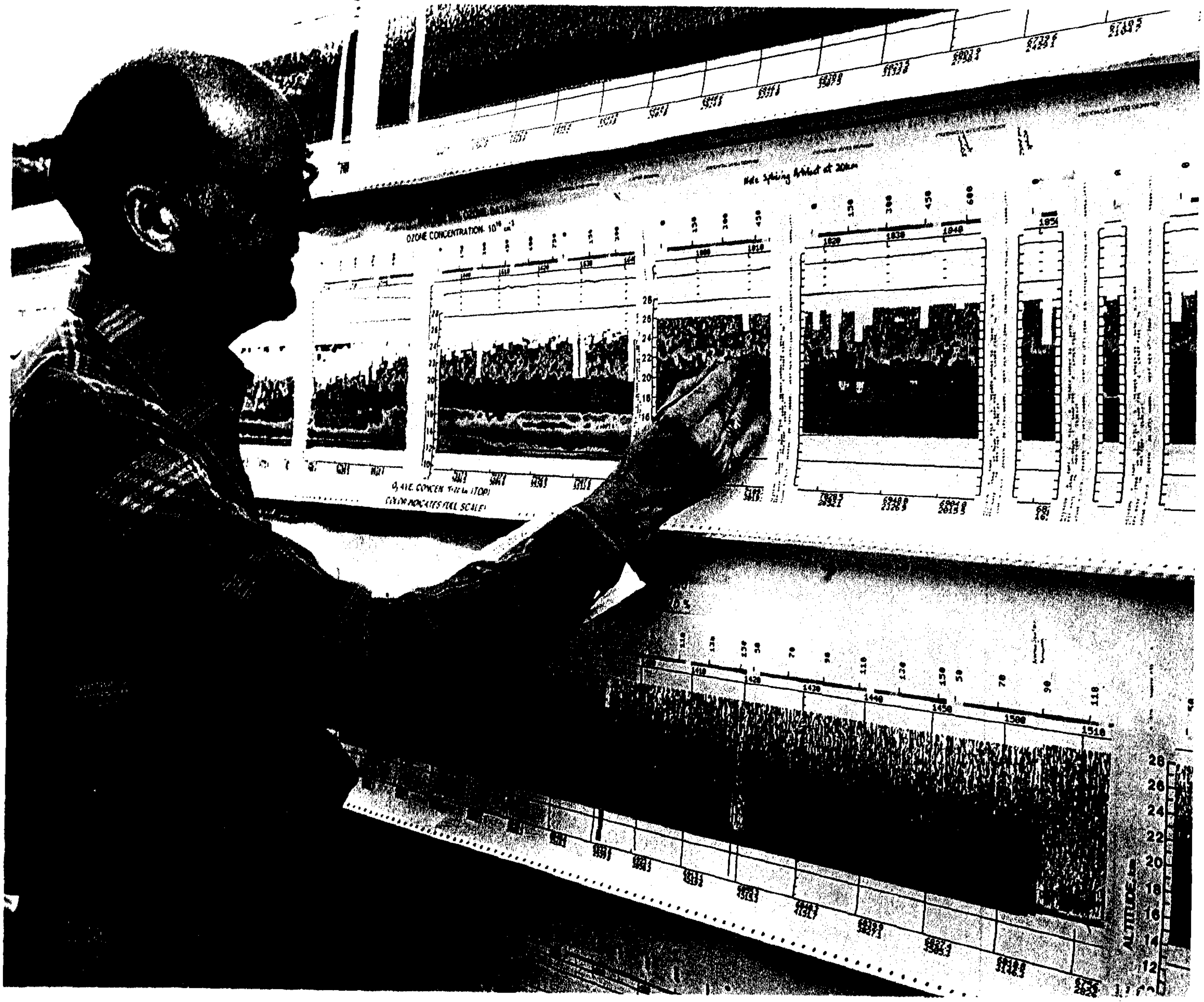
In support of EOS, NASA is developing EOSDIS, a large, distributed information system project that NASA estimates will cost about $3 billion through fiscal year 2000. EOSDIS computing and networking facilities will be responsible for the processing, archiving, and distribution of EOS data in support of (1) data interpretation and modelling and (2) command and control of the EOS observatories.

Beyond its mission-specific task of supporting the data processing needs of the EOS observatories, EOSDIS is intended to play a larger role in improving the way scientists access NASA's earth science data. NASA bills EOSDIS as a comprehensive system that will bring together data from many sources to serve the needs of scientists performing broad-ranging, interdisciplinary studies of the earth. In addition to the data collected by the EOS instruments themselves, EOSDIS is intended to include previously archived data, new measurements from other non-EOS spacecraft, ground-based measurements and other in situ data,[2] and contributions from the scientific user community. As such, EOSDIS will be the one system responsible for archiving and distributing all NASA earth science data and will offer substantial improvements over existing earth science data systems, which are limited in their ability to support global change research.

The goal of EOSDIS is to make its vast wealth of data easily accessible and affordable to a broad range of researchers and to aid these researchers in producing useful global change data products and models. NASA has estimated that there are approximately 10,000 potential EOSDIS users within federal, state, and local government agencies; academia; foreign countries; and industry. Figure 1.2 shows one such user, a NASA atmospheric scientist from the Langley Research Center in Langley, Virginia.

[2] In situ data refers to data obtained in place by various ground-, ocean-, and air-based observations, as opposed to space-based remote sensing data.

Figure 1.2: NASA Scientist Reviewing Laser Ozone Data



Source: NASA

Outside review panels have commended NASA for making an early commitment to the data management and analysis aspects of the EOS program through EOSDIS. In the past, critics have charged that insufficient planning and funding of the ground-based data systems and data analysis activities of some NASA satellite missions have resulted in a poor scientific payoff. In contrast, NASA's goal is to have a full-scale EOSDIS up and running in 1997, 1 year before the launch of the first EOS platform. By piloting the system early using existing earth science data, NASA hopes to be fully prepared for the data generated by the EOS platforms.

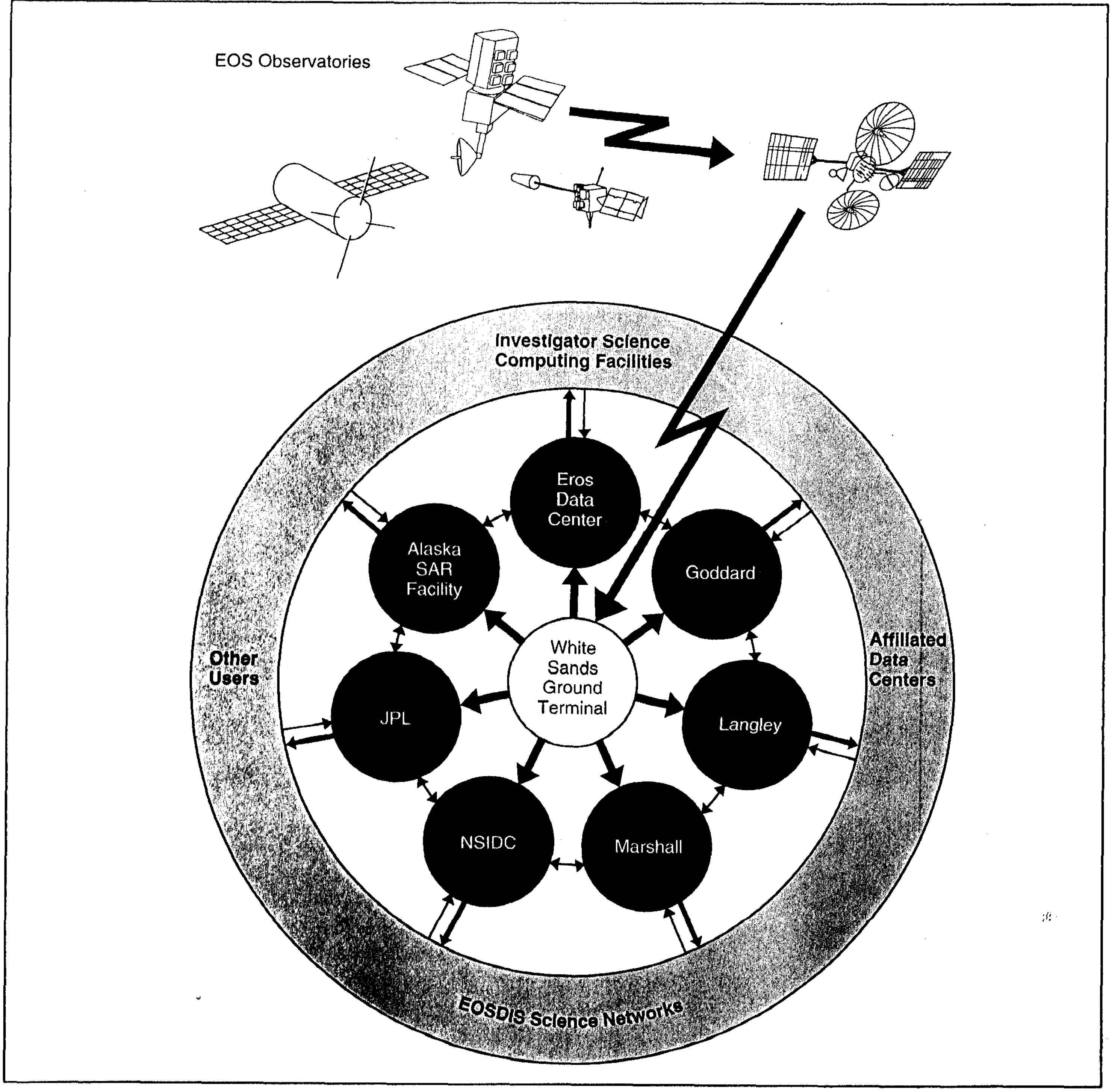
Furthermore, as described above, the ultimate value of EOSDIS extends well beyond being a mere repository for data from the EOS observatory platforms. Scientific advisory committees, such as the NASA Advisory Council, have argued for the need for a major NASA information system to support earth sciences and global change research. Recognizing this larger role for EOSDIS, NASA is budgeting nearly a quarter of all EOS funds for development of the system.

EOSDIS Information Flow

After instruments on-board the EOS platforms make observations, their data will be transmitted to a ground station at White Sands, New Mexico. A computer system at White Sands will do some initial processing to remove errors and artifacts of the space-to-ground transmission process. Then the data will be sent electronically to one of seven Distributed Active Archive Centers (DAAC), depending on the instrument that produced it. The DAAC will be responsible for processing the raw data from the instruments into earth science data usable by the research community at large. NASA's decision to spread the data-handling functions of EOSDIS among seven DAAC sites located around the country rather than centralizing these functions is in keeping with recommendations of the scientific community to diversify EOSDIS and to build on existing data systems and expertise. NASA's process for choosing the seven DAAC sites is discussed in another GAO report.[3] The flow of EOS data from satellite instruments to ground data centers is depicted in Figure 1.3.

[3]Earth Observing System: Information on NASA's Selection of Data Centers (GAO/IMTEC-91-67, Sept. 18, 1991).

Figure 1.3: EOSDIS Information Flow



The data handling functions will be carried out by three inter-connected computer-based systems at each DAAC—a product generation system (PGS), a data archive and distribution system (DADS), and an information management system (IMS). The PGS will be responsible for processing the raw instrument data it receives into pre-defined standard products, including sets of earth science data directly usable by scientists. This system also will be able to reprocess data when, for example, improved algorithms are developed, without interrupting the normal processing of new data. The DADS 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 IMS 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 will be presented with complete access to all EOSDIS data sets. The user will not have to know at which DAAC any particular data set is physically located to access it. NASA has promised to have processed data available to scientists through EOSDIS within 4 days of the initial observations.

There are also several other important elements in the EOSDIS information flow. The scientific community will be responsible for developing the software procedures for converting raw EOS instrument data into usable earth science data. Scientists will develop this software at their own facilities and then turn them over to the DAACs to use in processing data. The scientists will also be responsible for checking the quality of the output from their procedures and for maintaining and enhancing the software.

EOSDIS will need access to other non-EOS data to supplement and validate the data it gets from the EOS platforms. In particular, certain National Oceanographic and Atmospheric Administration (NOAA) data sets and standard data products will be required by EOSDIS on a routine basis to support production or validation of EOSDIS data products. NASA plans to negotiate with NOAA to create a link between NOAA's data centers and EOSDIS to fill this need.

NASA's EOSDIS Development Strategy

NASA's development of EOSDIS has two major components. First is a relatively small-scale in-house program to pilot certain aspects of the future EOSDIS on existing data systems that are in place at the seven DAAC sites. This program is referred to as EOSDIS Version 0 and is budgeted at $19 million in fiscal year 1992. The second component is the large-scale procurement of several versions of an operational EOSDIS through a single, comprehensive contract. NASA has estimated that overall development and operations costs for EOSDIS will be about $3 billion through fiscal year 2000.

Version 0 is a set of early EOSDIS capabilities that are planned to be in place by late 1994, in advance of the main contractor development for EOSDIS. Version 0 is intended to provide some preliminary integration of existing earth science data systems in order to demonstrate the feasibility of interconnecting and interoperating distributed, heterogenous systems and to prototype DAAC data processing concepts not available within current data systems. Specific aspects of Version 0 are discussed in more detail in chapter 2.

Version 1 is scheduled to become operational at the seven DAACs in 1996. It will represent the first major version of the system developed under the EOSDIS Core System contract. Version 2, to be operational in 1997, will be the first full-scale operational EOSDIS. The Core System contract includes all the data processing, archiving, and distribution functions to be carried out at the seven DAACs, as well as an EOS operations center for command and control of the EOS observatories. A request for proposals (RFP) for the EOSDIS Core System was issued in July 1991. NASA expects to choose a contractor in May 1992 and award a contract in late 1992.

Objectives, Scope, and Methodology

On November 7, 1990, the House Committee on Science, Space, and Technology asked us to report on several aspects of NASA's EOSDIS development program. Our first report on this subject was issued on September 18, 1991.[4] Our current objectives were to examine the extent to which EOSDIS (1) will use prototyping to reduce system development costs and risks, and (2) may require technologies that are beyond the current state of the art.

[4]Earth Observing System: Information on NASA's Selection of Data Centers (GAO/IMTEC-91-67, Sept. 18, 1991).

To meet our objectives, we:

- reviewed NASA's EOSDIS program documentation discussing its goals and approach to EOSDIS development;
- reviewed Version 0 prototyping plans and surveyed managers of all seven DAACs concerning their prototyping activities;
- reviewed the RFP for the EOSDIS Core System;
- interviewed EOSDIS program officials at NASA Headquarters, Washington, D.C.; and EOSDIS project officials at NASA's Goddard Space Flight Center, Greenbelt, Maryland;
- interviewed scientific data base systems experts at the National Science Foundation, Washington, D.C.; NASA's Ames Research Center, Moffett Field, California; Goddard; and the University Corporation for Atmospheric Research (Unidata Project), Boulder, Colorado;
- interviewed NOAA representatives involved in EOSDIS-related activities or other global change programs at the National Oceanographic Data Center, Washington, D.C.; the National Geophysical Data Center, Boulder, Colorado; and the Satellite Data Services Division, Suitland, Maryland;
- interviewed data management officials at NCAR and the National Snow and Ice Data Center, in Boulder, Colorado; and reviewed previous reports and documents related to scientific data management prepared by NASA and various scientific groups and committees.

Our audit work was performed in accordance with generally accepted government auditing standards, between January 1991 and January 1992. We discussed the contents of this report with NASA officials and incorporated their comments where appropriate.

Early EOSDIS Prototype Projects Do Not Address Major Development Risks

Prototyping is an engineering technique whereby partial, experimental versions of computer-based systems are rapidly and inexpensively built to validate requirements and test the feasibility of key functions prior to full-scale development. NASA has endorsed prototyping as an essential element of its evolutionary approach to EOSDIS development. However, NASA's near-term Version 0 prototype projects are not substantial enough to support user testing of the design of key system functions, nor do they address critical areas where technical feasibility is in question. Rather than addressing the full range of important high-risk data management areas, the Version 0 prototypes are largely enhancements to existing systems, and many are unlikely to be incorporated into the full-scale contractor-developed system. Furthermore, NASA is not taking full advantage of existing repositories of experience to advance its prototyping efforts. The lack of substantial progress with early prototypes creates uncertainties about the technical feasibility of the system.

Prototyping Lessens System Development Risks

Software prototyping is a way of verifying assumptions made about an automated system before full-scale development begins. Specific objectives of prototyping include: 1) testing and revising requirements so that users and implementers can come to agreement on what the final system should do, 2) testing a system's ability to provide specific functions required in the final version, and 3) modelling system aspects that test current technological limits.

Prototypes are built expressly to test only a subset of the functions of the final system, allowing an aspect or combination of aspects to be carefully tested while other functions are ignored. Therefore, it is less costly and time-consuming to build a prototype to model new or experimental functions in isolation before incorporating those functions in the operational system. And because less effort is required for its development, a prototype can be modified or discarded if it cannot be transferred directly to the operational system. The greatest benefit of prototyping is that it lessens the risk that the final system will not meet users' needs and expectations.

A Major Goal of Version 0 Is to Prototype Key Concepts Early

NASA's evolutionary approach to EOSDIS was developed as a response to criticism from the scientific community that NASA was trying to pre-define all of the requirements for EOSDIS. Version 0 was conceived as the first step in this approach. After a panel of science advisors recommended that EOSDIS designers acquire experience with current earth science systems and build prototypes as part of EOSDIS development, NASA decided to adopt a "build a little, test a little" approach by initiating Version 0, a collection of prototype projects using existing systems and data as a starting point, before the contractor-led development began.

A major goal of Version 0, which began in 1990, is to establish working prototypes of concepts required for the development of the full-scale system, for assessment by representative scientists. Using selected data from existing earth science systems, Version 0 is intended to be operational by mid-1994. Another goal of Version 0 is to facilitate the transfer of knowledge and experience gained from its activities to the contractor-led system development effort.

The National Research Council, in its assessment of EOSDIS plans for fiscal year 1991, recommended that NASA develop a coordinated plan for data system research and prototype development. While noting that certain aspects of EOSDIS were not yet ready for design, the Council suggested that prototyping projects, directed by full-time professionals, should be underway in many areas. Specifically, the Council recommended that prototypes be developed in the following areas:

- data display and user interface,
- browsing capability,
- data formats and media,
- accessibility of data and information,
- cataloging,
- search and query capabilities,
- model and data interaction,
- data structures,
- data reduction algorithms, and
- networking.

The Council stressed that prototyping should address the challenge of the immense size of EOS data sets, which will dwarf any previous experience. It also remarked that prototyping would be needed to learn how scientists will work with EOSDIS so that suitable data management schemes can be selected.

Version 0 Prototyping Efforts Are Inadequate

Although Version 0 prototyping projects have been underway since 1990, these activities are too modest in scope and resources to support validation of the design of key system functions before full-scale development of EOSDIS begins. Of the $83 million that NASA plans to spend on pre-contract activities in fiscal year 1992, only $19 million is for Version 0. These very limited resources cover work on six programwide activities being directed by the EOS Project Office, as well as a variety of smaller projects initiated and conducted independently at each of the seven DAACs. In view of NASA's plan to spend several billion dollars on the overall development of EOSDIS, its investment in early prototyping seems disproportionately small.

One project, an effort to evaluate standard data storage formats, exemplifies problems that result from NASA's inadequate commitment of resources to Version 0 prototyping. To its credit, the EOS project has recognized the importance of establishing standards early in the development of EOSDIS. However, the task of establishing data format standards is large; the earth science community is made up of many smaller disciplines that are used to working independently, using formats tailored to their own needs. Experts have warned that establishing one or two standard formats in such an environment will be extremely difficult. Furthermore, vast amounts of existing earth science data that will need to be incorporated into EOSDIS cannot be easily translated into standard formats because translation is complex, expensive, and often leads to loss or misrepresentation of some of the data originally recorded in the unique format.

Within this daunting context, the Version 0 data format standards project is very modest, with the equivalent of one quarter of a person at each DAAC being assigned to investigate one or more potential standard formats each. The project plans on proposing several standard data formats after some experience is gained in translating data from one format to another. The larger task of developing and prototyping some standard method for accessing the large amount of data stored in non-standard formats is a much greater task and remains unaddressed by the standards project. Further, a variety of other standards will be needed by EOSDIS in addition to data storage formats, including standard formats for metadata, browse data, and quick-look data,[1] and these are also not addressed by the standards project.

[1]Metadata is information that describes the nature of the data sets in the system. Browse data are a summarized or simplified version of the data that can be accessed quickly by scientists wishing to determine if the data are useful to their particular purposes. Quick-look data are also a summary version of the data, but they are intended to allow investigators to assess the quality of the data soon after being obtained.

Version 0 Prototypes Do Not Address Critical Problems

Rather than addressing the full range of critical EOSDIS technical challenges, NASA's Version 0 activities primarily support and enhance existing earth science data systems at the seven DAACs. As such, these projects do not serve as effective prototypes of EOSDIS technology and do not mitigate the technical risks inherent in designing a very large information system that must provide unprecedented levels of access to vast amounts of earth science data.

Two major Version 0 projects—providing network connectivity and developing system interoperability through the Information Management System (IMS)—are case examples. The network connectivity prototype is simply an effort to analyze overall network connectivity requirements and provide dedicated communication lines among the seven DAACs at a rate of 56 kilobits per second. Although this effort can enhance the current operations at the seven DAACS, NASA communications experts have questioned the value of this project as a prototype for the full-scale system for several reasons. First, state of the art technology is not being tested since the data rate of 56 kilobits per second is significantly slower than the 1536 kilobit per second rate currently available on the National Science Foundation's network, which already interconnects all of the DAACs. Second, prototyping of network functions to handle the requirements of the full-scale EOSDIS will be difficult since the current systems require neither the high data volumes nor the fast transfer rates anticipated for the full-scale system. It is unlikely that any of the technology associated with this project will be transferred to the full-scale system.

The stated objective of the second case example, the IMS project, is to facilitate interoperability of existing data systems at the DAACs. Interoperability essentially means that the EOSDIS data base will appear to users as a single, integrated entity stored locally, despite the fact that the data are actually stored in heterogenous systems and geographically dispersed facilities. To achieve meaningful interoperability among existing DAAC systems would require substantial changes to the data storage and access methods at each DAAC, as well as developing new distributed operating system software to manage interactions among these systems.

However, the current IMS project is much more limited in scope and is unlikely to produce results useful in proving the technical feasibility of the full-scale system. Because NASA does not want to disrupt services currently being provided by DAAC systems, the IMS prototype is restricted to the development of a patchwork interface among the various DAAC systems by translating a user's questions about data into various DAAC-unique formats—so that the DAAC systems can interpret them—and then translating the responses back into the user interface format. According to IMS developers, the various parts of the prototype IMS interface are somewhat inconsistent, and the interface may appear complicated and confusing to users. While some gains may be made in improving access to the data currently stored at the DAACs, the IMS prototype will likely be of little help in proving the technical feasibility of the full-scale EOSDIS.

The IMS, as developed for Version 0, does not provide the kind of broad, sophisticated access to data that the full-scale system will need to provide. In fact, the IMS prototype gives users only general information about the limited data holdings currently at the DAACS; it does not allow users to directly access the data or to place orders for the data itself, capabilities already provided individually by existing DAAC systems. Whatever its value in connecting current DAAC systems, the IMS is too limited to serve as a true information management prototype for the full-scale EOSDIS.

According to NASA's development strategy, the EOSDIS Core System contractor will be responsible for initiating and conducting prototyping efforts after the contract is awarded and full-scale development begins. Prototyping is intended to be an ongoing aspect of the contractor's work. However, we believe that devolving responsibility for prototyping to the Core System contractor may make it difficult for NASA to ensure that the full range of critical technological risk areas are addressed in a timely fashion. The contractor's near-term requirement to develop an operational system to support the data processing needs of the EOS satellites could detract from a thorough prototyping program to support the long-term needs of global change researchers. The Core System RFP, while specifying the general need for prototyping, does not indicate the range of key system functions that should be addressed.

NASA Is Not Taking Full Advantage of Available Systems Resources and Experience

In designing and building its Version 0 prototypes, NASA has not taken full advantage of experience available at earth science data and research centers other than the designated DAACs. Several centers outside of the EOS program have unique experience or have done significant research in areas that will be critical for the success of the full-scale EOSDIS. Previous expert panels, including an internal NASA committee as well as the National Research Council, have noted the value of this experience base and urged NASA to make use of it.

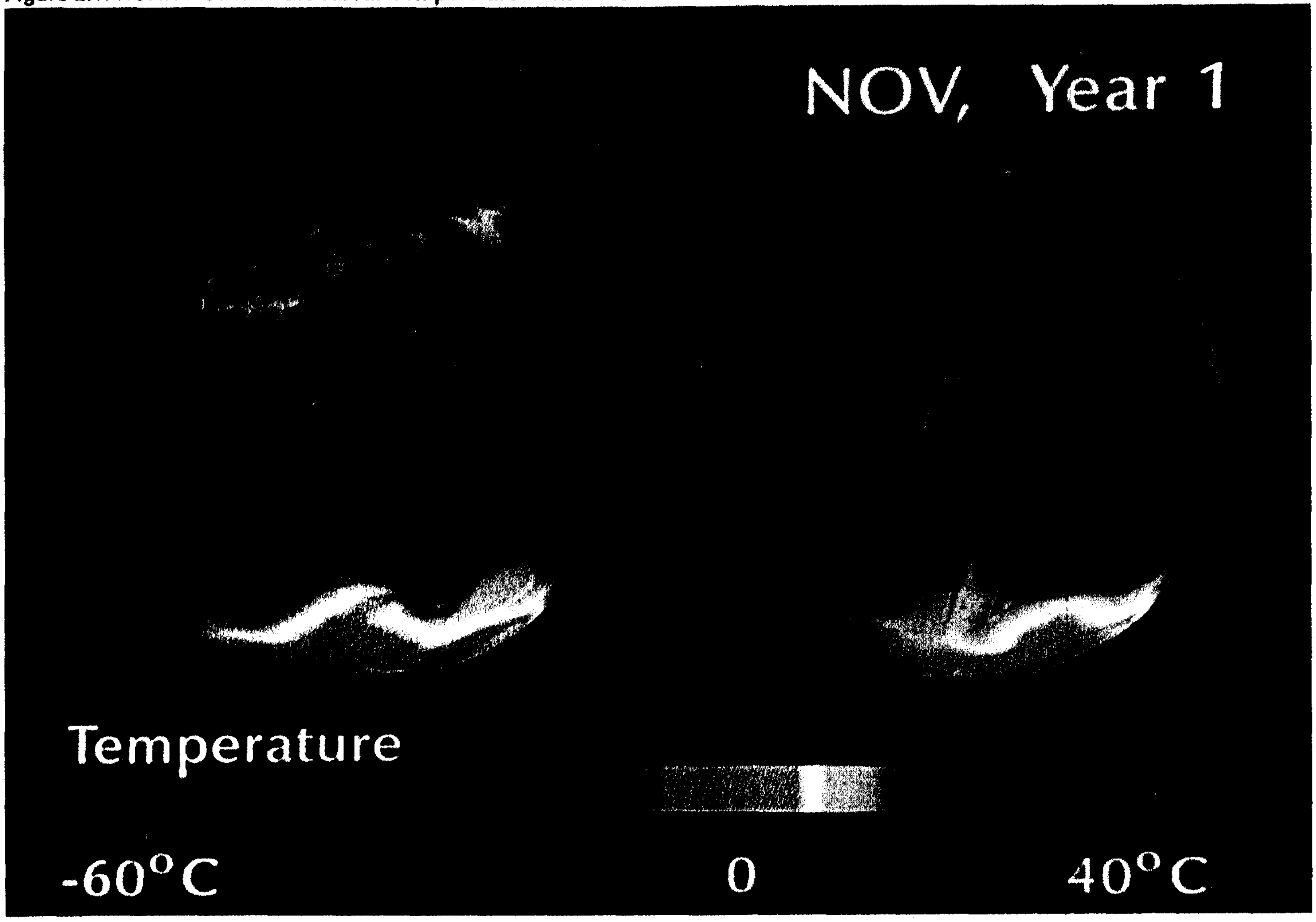
For example, the National Oceanographic and Atmospheric Administration (NOAA) has extensive experience with data management and has prototyped applications of the type that are likely to be required by EOSDIS. NOAA has three National Data Centers[2] and many smaller data repositories. These centers accumulate data at a rate of approximately 6,000 gigabytes per year.[3] Altogether, according to a senior NOAA official, NOAA holds roughly two-thirds of all global change data currently available to researchers. Yet, these centers are not participating in Version 0 activities. Moreover, one of these centers has developed a prototype data base system that provides users with flexible search capabilities such as those needed to do global research. NOAA is also currently sponsoring an interoperability demonstration project, similar to the IMS prototype, at the Goddard Space Flight Center in Greenbelt, Maryland.

The National Center for Atmospheric Research (NCAR) is a National Science Foundation funded center whose mission is to be a national focal point for weather and climate-related studies. Almost all of the priority research areas at NCAR make extensive use of supercomputers and large data sets to model environmental phenomena that will also be the focus of the EOS program, such as global warming and the depletion of the ozone layer. Figure 2.1 shows an example of the results of this modelling process. NCAR operates a large data storage system, called the Mass Store System, holding approximately 20,000 gigabytes of data, that has been cited by the National Research Council for its effectiveness as an earth science data management system. Yet, NASA has not drawn upon NCAR's experience in climate modeling and data management for its Version 0 prototyping activities.

[2]The National Climatic Data Center, the National Geophysical Data Center, and the National Oceanographic Data Center.

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

Figure 2.1: NCAR Model of Global Air Temperature Variations



Source: NCAR

Unidata, a program managed by the University Corporation for Atmospheric Research and sponsored by the National Science Foundation, helps university departments acquire and use atmospheric data. At Unidata, researchers have developed software to facilitate accessing, organizing, storing, analyzing, and displaying scientific data. These are areas of critical concern for EOSDIS that are not being sufficiently addressed in Version 0 prototypes. The EOS Data Panel recommended in 1986 that NASA collaborate closely with Unidata, and, more recently, a

Unidata representative attended a Version 0 workshop on data formats and offered to work with the EOSDIS project on data access standards. However, NASA has not yet acted on Unidata's offer.

New Technology Will Be Needed to Fulfill EOSDIS Goals

According to specialists in data base research and data management both within NASA and in other government organizations, EOSDIS cannot reach its ultimate goal of providing global change researchers ready access to vast amounts of EOS and other earth science data without several kinds of significant technological advances. Outside review panels charged with examining EOSDIS have reached similar conclusions. Although advances have been made in data management techniques for earth science systems, the current state of the art is still inadequate to support sophisticated global change studies. Advanced data base search techniques, automated ways of characterizing scientific data, and efficient access to data stored in massive archives are all examples of technologies that have been identified as critical to the success of EOSDIS. However, the EOSDIS program, as currently structured, does not provide specific direction in these areas. Proposals to do research in these areas in the near term within NASA have not been supported by the Office of Space Science and Applications and currently are not being pursued. As a result, NASA runs the risk that EOSDIS will not achieve its goal of offering scientists the capabilities they need to do important new kinds of global change research.

EOSDIS Goals Are Ambitious

The first order of business for EOSDIS will be to process and distribute the data it receives from the EOS platforms. However, the ultimate goals of EOSDIS are much larger. NASA envisions EOSDIS as a comprehensive system that will bring together data from many sources to serve the needs of scientists performing broad, interdisciplinary studies of the earth. In addition to the data collected by the EOS instruments themselves, EOSDIS will draw together previously archived data, new measurements from other non-EOS spacecraft, ground-based measurements and other in situ data, and contributions from the scientific user community. As such, EOSDIS will be NASA's one integrated system for archiving and distributing all earth science data. Furthermore, NASA acknowledges that EOSDIS will not succeed unless it makes its vast wealth of data easily accessible and affordable to a broad range of researchers and significantly aids those researchers in producing useful global change data products and models.

While these goals are easily stated, they are nevertheless enormously ambitious. To support global change research, EOSDIS users will need to ask broad-based questions of the system and will need ways of readily assessing the wide variety of data sets within EOSDIS to identify data of possible interest. EOSDIS will then need to make that data readily available—through images that can be quickly viewed at the scientist's terminal, for example—so that the utility of the data to the scientist's particular question can be quickly assessed. Current data management technology and practices cannot support these kinds of activities on the scale contemplated for EOSDIS.

Outside Review Panels Have Expressed Caution

Two major review panels charged with examining EOS have expressed concern about NASA's ambitions for EOSDIS, given the limitations of the current technology. The EOS Engineering Review Committee, chartered by the Office of Management and Budget and the National Space Council to examine all aspects of the EOS program, noted that EOSDIS is "the largest data and information system contemplated by this nation" and, while commending NASA for the work it had put into EOSDIS, expressed doubts about "the fundamental underpinning and the sheer scale of the EOSDIS concept."[1] The Committee recommended that NASA carry out a review by technical experts to establish the best way to proceed.

Similarly, the National Research Council, in assessing fiscal year 1991 plans for the U.S. Global Change Research Program, commented that "the intended scope of EOSDIS far exceeds that of any civilian data management system,"[2] and noted that significant advances in data management would be required before such large amounts of scientific data could be made readily available and useful for modeling global change.

Current Systems Are Inadequate for Global Change Research

Many traditional concepts and present practices of data and information management are inadequate for global change studies. The current totality of earth science data has been accumulated in a piecemeal fashion and is fragmented, often poorly documented, and cannot be readily accessed, according to the NASA Advisory Council. While improvements have been made recently in the utility and accessibility of certain earth science data sets, a large gap still exists between current capabilities and the goals of EOSDIS.

Traditionally, earth science data management has consisted of archiving magnetic tapes and maintaining a catalog describing tape holdings by identification number, source, time and location of data generation, and physical location in the archive. A researcher interested in a given set of data needed to know who managed the data, what archive it was part of, specific identifying details about the data, what formats were available,

[1] Report of the Earth Observing System (EOS) Engineering Review Committee, September 1991, p. 25.

[2] The U.S. Global Change Research Program: An Assessment of FY1991 Plans, National Academy of Sciences' National Research Council, National Academy Press, 1990, p. 72.

and the cost, if any, of reproduction. Further, for the data to be useful, the researcher would have to find some way to ascertain the general quality and information content of the data set, either by browsing sample images from the data on a computer terminal—if such browse data were available—or through lengthy discussion about the data with an expert at the archive. When all that was done, the researcher would be ready to order the data for his or her own use.

NASA and other agencies recognize the limitations of this data search and retrieval method, and some significant advances have been made over this method for certain specific collections of data. The Coastal Zone Color Scanner System at Goddard Space Flight Center, for example, allows scientists to search an entire collection of ocean color images on-line and to view less-than-full-resolution versions of images that meet the search criteria. The system also assists the scientist in preparing an electronic order for the original digital data. The order is then filled and sent to the researcher off-line.

The NASA Climate Data System is another example. This system offers climate researchers an interactive, on-line interface allowing them to locate, access, manipulate, and display a variety of climate-related data. Also available is a comprehensive, on-line catalog of data descriptions and an inventory of temporal and volume information to help researchers find and learn about data of potential interest. Data requested by researchers can, in many cases, be brought on-line quickly. The system then provides additional tools for displaying and manipulating the data.

Despite valuable improvements such as these, current systems are still very limited in their data management capabilities. While ready access can be provided to data within a relatively small data base, such as the Coastal Zone Color Scanner System, a very large scale system, such as EOSDIS, could not be accommodated with the same technology. Most NASA Climate Data System data, for example, are stored off-line, with only samples being stored on-line for browsing. Of the approximately 500 gigabytes of data in the system, only about 2 gigabytes are stored on-line. EOSDIS is expected to contain many times more data than the entire NASA Climate Data System, compounding the data management problem.

Current Information Systems Technology Will Not Meet the Advanced Needs of EOSDIS

Many of the limitations of current earth science data systems are due to the limitations of the information systems technology on which they rely. Data management experts closely involved with scientific systems have identified a number of technologies that need further development before a system such as EOSDIS can reach its goals. In reports sponsored by NASA, the National Science Foundation, and the National Academy of Sciences, these experts have cited scientific data base management software, high-volume data storage and access systems, new techniques for data characterization, and data visualization, as well as other technologies, as critical for the development of advanced scientific information systems.

New Data Base Management Systems Are Needed to Handle Scientific Data

Most current data base management systems, including ones that maintain earth science data, are based on the relational model, which has been successful in many commercial applications. However, the relational model was designed largely to meet the needs of transaction-based business applications and, as such, is inadequate to support the significantly different data processing and display requirements of interdisciplinary earth scientists.

Business applications generally involve data bases full of numerous, relatively short records that are frequently updated. The records often can be sorted in many different ways; they do not have to follow any inherent sequence. Relational data base systems arrange these records in tables that allow great flexibility in sorting and provide quick access to make updates to specific records. Thus it is fairly easy, for example, to use relational data base software to access a payroll record in an accounting system and make changes to specific fields, such as a birth date or a salary figure.

Scientific data bases have significantly different qualities from business data bases. For example, sequence order is often important to sets of scientific measurements, whereas the relational model considers the sequencing of the data to be arbitrary. Also, the process of updating and correcting scientific data sets is very different from updating records in a business application in that it tends to be iterative, leading to gradual refinements and improvements, rather than correcting specific errors. Most important, however, is that most earth science data are in the form of images, vast arrays of data with spatial and temporal attributes that are not easily broken down into constituent parts (as opposed to a personnel record, for example, that has discrete fields such as name, address, birth date, etc.). The process of searching scientific data—especially images—for meaningful patterns is quite unlike a relational data base search because the patterns may be hidden within the data sequences and images and therefore not easily sorted and processed.

Current earth science data systems that provide researchers with data search options force their data to fit the relational model. Images are tagged with data fields that a relational data base system can search, such as latitude, longitude, date, and time of observation. However, this means that researchers can only access the data by phrasing their search requests in terms of the tagged information. They cannot directly search the data base for specific kinds of geological formations, for example, or for a certain level of cloud cover, or some other spatial pattern directly related to the scientific content of the data. When the large volumes of disparate data that will be needed to do global change studies are considered, the limitations of the relational model become critical.

New Data Access and Retrieval Methods Are Needed to Cope With Massive Amounts of Data

Very large quantities of data—as much as 2,000 gigabytes per day—are expected to be generated by the EOS observatories. Over the planned 15-year life of the EOS mission, roughly 11 million gigabytes could be collected, more than 1,000 times the amount of text currently stored by the Library of Congress. According to some data management specialists, it will be virtually impossible for scientists to manage so much data unless new automated techniques are developed to ease the work load. The sheer task of storing such a volume of data poses formidable challenges, as we have discussed in a previous report.[3]

Some experimentation has been done at NCAR into ways of making large data sets accessible to researchers. Data management officials at NCAR report that a constant effort is required to monitor incoming data requests and adjust storage locations for different data sets so that the most frequently accessed data is most readily at hand. The NCAR officials identified new software for data file management as being a high development priority, without which a system like EOSDIS, which will hold many times as much data as NCAR's system, will fail.

New Methods Are Needed for Locating Data of Interest

Beyond the logistical problem of physically locating and retrieving specific data from a massive archive is the equally challenging problem of identifying the kinds of useful scientific information that may be locked up

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

within this vast amount of data. Manual review of all the data, including preparing detailed and comprehensive descriptions of its contents, will be infeasible. Instead, new ways of automatically describing the content of the data (data characterization), will be needed.

Data characterization within the earth sciences is an inherently complex process because different researchers may be looking for completely different kinds of information within a given data set. An atmospheric researcher may be interested in studying the cloud formations that are visible in an image, while a geologist may view the clouds as simply obstructions of the ground image hidden beneath. One description of the data that is useful for all researchers is thus hard to create. Nevertheless, computer scientists are inventing ways of addressing the issue.

For example, a researcher at NASA's Ames Research Center has developed a prototype of a system to automatically characterize the scientific content of images taken from space. Using a Landsat image, the experimental system is able to recognize natural patterns in the image solely by performing certain kinds of statistical analysis. Something analogous to this will be required to automatically characterize EOS data. Otherwise, researchers will have no indication of where data of interest is located within the massive volume of images archived within EOSDIS.

New Visualization Techniques Are Needed to Analyze Large Amounts of Data

Once relevant data are identified and retrieved, additional new automated tools will be needed to display and analyze them. Known as data visualization, this area of research addresses ways to allow scientists to combine and correlate large volumes of data and display the results in a meaningful way. Correlating different data sets means being able to view them all from exactly the same perspective, something that cannot be done with most current earth science data. Furthermore, a wide variety of representation techniques will be needed to allow for diverse methods of analyzing the data. A recent study of the research, operations, and information systems needs of the NASA science community listed better data visualization tools as a broadly perceived need, particularly for combining and displaying different types of data. Figure 3.1 shows one way of displaying several separate but interrelated data sets on a single computer screen.

Figure 3.1: A Computer Display Correlating Four Data Sets From the Antarctic

Source: International Business Machines, Inc.

The EOSDIS Program Does Not Address Specific Advanced Technologies That Will Be Needed

Advanced technologies such as automated data characterization, improved data base search and retrieval techniques, and advanced data visualization have all been specifically identified by data management experts as crucial to the success of EOSDIS in supporting global change research. Some research projects have been undertaken in these areas within other NASA organizations, but the EOS program has given them little support. Neither the data characterization prototype at Ames nor the mass data storage experience at NCAR are being drawn on by NASA for use in developing EOSDIS. The major Version 0 prototype efforts do not include projects in these areas.

NASA's Office of Space Science and Applications, which manages EOS and other earth and space science programs, has identified high-volume data management and data visualization as high-priority areas for advanced research. In response to this need, the Office of Aeronautics and Space Technology developed a research program tailored to address these areas. In June 1991, a review panel of technology experts outside of NASA endorsed the proposed new program. However, the Office of Space Science and Applications has not expressed strong support for pursuing the program in the near term within NASA, and as yet it has not been funded.

By not making research into advanced technologies a high priority, NASA is running the risk that EOSDIS, as procured through the EOSDIS Core System contract, will not achieve its goal of offering scientists the capabilities they need to do important new kinds of global change research.

Conclusions and Recommendations

Throughout the 1980s, scientific review panels criticized NASA's past handling of data from its space missions and recommended greater attention to the data management area. In planning for Mission to Planet Earth and the EOS program in particular, NASA has responded to these concerns by making an early and substantial dollar commitment to data processing, distribution, and archiving through EOSDIS. Such a commitment is important because global change researchers will need ready access to vast amounts of EOS and other earth science data.

With EOSDIS, NASA must address both the near-term need to provide data processing support for the EOS satellites and experiments, as well as the long-term requirement to support the global change research community at large by providing broad access to NASA's earth science data holdings. It is vital that NASA not allow the near-term operational requirements to prevent it from building a system that can ultimately provide a "next generation" of capabilities beyond what current earth science data systems provide. NASA's approach to EOSDIS development—combining a small-scale in-house Version 0 prototyping effort with a large systems development contract—does not make adequate provisions for the prototyping of key system features and the development of key advanced information management technologies that will be required to support future global change research.

The EOSDIS Version 0 effort, while conceptually a vitally important development phase, appears to be only a very small element of the EOSDIS project. The prototype projects that have been initiated within Version 0 do not fully address critical system functions, such as data access and retrieval from mass storage systems or true distributed data management. Furthermore, NASA could collaborate more closely with other federal agencies with significant experience in global change-related data management, such as NOAA and the National Science Foundation, on the early development of EOSDIS. Better direction, better cooperation, and a more substantial commitment of resources are needed to make EOSDIS prototyping effective.

Many technological challenges lie ahead before EOSDIS can achieve its ultimate goals of providing broad support for global change research. Some are well-defined as research objectives; however, EOS program managers are not pursuing research projects in these areas. The success of EOSDIS depends upon how quickly and seriously these challenges are addressed. The planned EOSDIS Core System contract as currently structured is not sufficient for ensuring that this happens because it is focused on the delivery of a system that must support the near-term data processing requirements of the EOS satellites rather than the long-term needs of the global change research community at large. We recognize that NASA must meet substantial near-term data processing requirements in support of the EOS satellites. However, our bottom-line concern is that the system developed may do little more than that, leaving global change researchers unable to take full advantage of the data collected in the system. To minimize this risk, NASA needs to add measures to its EOSDIS development strategy that will provide better assurance that the long-term goals of the program will ultimately be met.

Recommendations

Given the significant technological risks inherent in the EOSDIS concept, GAO recommends that the NASA Administrator not award the planned EOSDIS Core System contract until specific plans have been developed and resources identified for (1) prototyping the full range of critical system elements and (2) guiding and accelerating research into key advanced technologies that will be essential for the system's ultimate success. In addition, GAO recommends that NASA work to maximize cooperation with other federal agencies and earth science programs having data system experience in global change-related areas.

Agency Comments

As requested, GAO did not provide a draft of this report to NASA for its review and comments. However, GAO discussed the report's contents with NASA officials and included their comments as appropriate. NASA officials agreed that Version 0 prototypes do not fully address critical development risks but believe that prototypes developed by the contractor will eventually address these risks. However, NASA could not provide convincing evidence to support its belief. Neither the terms of the contract nor NASA's management approach to date ensure that the contractor will address critical development risks. The officials also disagreed with the statement that they are not taking advantage of other agencies' systems resources and experience. However, GAO's discussions with officials from these agencies suggest that fuller cooperation is possible. With regard to planning for advanced technologies, the NASA officials agreed with our findings and recommendation.

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
Lynne L. Goldfarb, Publishing Adviser

Ordering Information

The first copy of each GAO report and testimony is free. Additional copies are $2 each. Orders should be sent to the following address, accompanied by a check or money order made out to the Superintendents 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