ACCESSING EARTH SCIENCE DATA
FROM THE EOS DATA AND INFORMATION SYSTEM

Kenneth R. McDonald
Sherri Calvo

NASA Goddard Space Flight Center
Code 902.1
Greenbelt, MD 20771

INTRODUCTION

NASA's Earth Observing System (EOS) is designed to support the interagency Global Change Research Program through its Scientific Research Program (EOSSRP), Space Measurement System (EOSSMS), and Data and Information System (EOSDIS). The EOSDIS is responsible for the mission and instrument planning, scheduling, and commanding associated with EOS data acquisitions, the routine production, archive, and distribution of EOS data products, and the access to correlative data that may be archived by external data systems and organizations. The concepts related to the functions, architecture, and services of the EOSDIS that have emerged and are evolving are a direct consequence of the characteristics of both the scientific investigations associated with global change research and the community that is conducting those investigations.

A number of factors distinguish global change research from other scientific programs and endeavors. It is interdisciplinary, including studies in all of the Earth sciences and investigations of the interrelationships of different Earth processes. It is not limited to environmental science but also encompasses the analyses of the socioeconomic impacts of global change and the environment's response to human activities. Global change research requires massive sets of geophysical observations from numerous sources over the longest time periods available. The data sources include remotely sensed and in situ observations and predictions from numerical models and analyses. Because of the data demands of global change research, the program must integrate the existing collections of observations that are held by a variety of agencies and organizations with the future acquisitions of the EOSSMS and other programs. The numerous Federal agencies and research institutions involved in the Global Change Research Program are further evidence of the scope of the effort.

The purpose of this paper is to present an overview of the EOS Data and Information System, concentrating on the users' interactions with the system and highlighting those features that are driven by the unique requirements of the Global Change Research Program and the supported science community. However, a basic premise of EOSDIS is that the system must evolve to meet changes in user needs and to incorporate advances in data system technology. Therefore, the development process which is being used to accommodate these changes and some of the potential areas of change will also be addressed.

EOSDIS Program Requirements

Archive Contents

The EOSDIS will hold most of the Earth science data and data products from NASA activities and other data required for the production and effective use of these data. It will hold all of the data and data products from the EOS Space Measurement System and precursor missions.

The majority of NASA's heritage Earth science data will be migrated to EOSDIS and where it is not, EOSDIS will provide pointers and an access path to it. The current projection for the daily EOSDIS data rate by the end of this century is on the order of a terabyte per day, with a total archive of about two petabytes. EOSDIS will also contain metadata and browse products of its holdings, a software library of data production and analysis algorithms and tools, and a documentation library.

<u>User Community</u>

The EOSDIS user community is as extensive and diverse as the disciplines from which it is drawn. The number of users involved directly in Earth science research are estimated to be as many as 10,000 and when the education, applications, and government users are included the number grows to 100,000 or more. The characteristics and therefore the requirements of the users will also vary greatly. They will access the system from workstations, personal computers, and terminals. Some will use the system several times a week while others will only log on occasionally. Their experience level with on-line data systems and with the data itself will range from novice to expert.

<u>Services</u>

The EOSDIS will provide an end-to-end set of data production, management, and distribution services for the supported science community. It will supply the systems to perform the command and control of the EOS satellites and instruments. It will provide or augment the facilities that the science community will use to develop the algorithms that process the EOS data. The systems that will perform the routine generation of EOS products and the creation of special products are part of EOSDIS. The system will perform the archive and distribution functions for the EOS products, including those auxiliary data that are used in their generation. The information management services of the EOSDIS will provide the users with the capability to locate and order any of the data in the archives and will provide access to Earth science data held at external data systems through its interoperability with the interagency Global Change Data and Information System. Finally, the EOSDIS will provide the connectivity between its different elements and the overall management of its resources.

EOSDIS Conceptual Architecture

The process of defining the requirements of the EOSDIS has included the performance of a Phase A Conceptual Design Study and two Phase B Preliminary Design and Resource Estimates. The results of these analyses formed the basis of the conceptual architecture of the EOSDIS (Figure 1) which was included in the Phase C/D Functional and Performance Specification for the EOSDIS Core System (ECS). The architecture divides the EOSDIS into three segments and shows the interfaces between the segments and the external elements, shown as the Science Computing Facilities (SCFs) at the users' home facilities, the external data systems, and the users. The Instrument Support Terminals (ISTs) and the Field Support Terminals (FSTs) represent specialized interfaces that will be developed to support instrument operations and field campaigns, respectively.

The Flight Operations Segment (FOS) manages and controls the EOS platforms and instruments through the EOS Operations Center (EOC) and one or more Instrument Control Centers (ICCs). The ICCs will be used to schedule and command the more complex observatory instruments. The EOC will perform these functions for the survey instruments and perform the overall coordination of the platform and instrument operations. The Communications and System Management Segment consists of the EOSDIS Science Network (ESN) and the System Management Center (SMC). The ESN is responsible for the internal communications between the EOSDIS elements and the SMC monitors the overall resource usage of the system.

Figure 1 - EOSDIS Conceptual Architecture

Shaded areas are responsibility of EOSDIS Core System
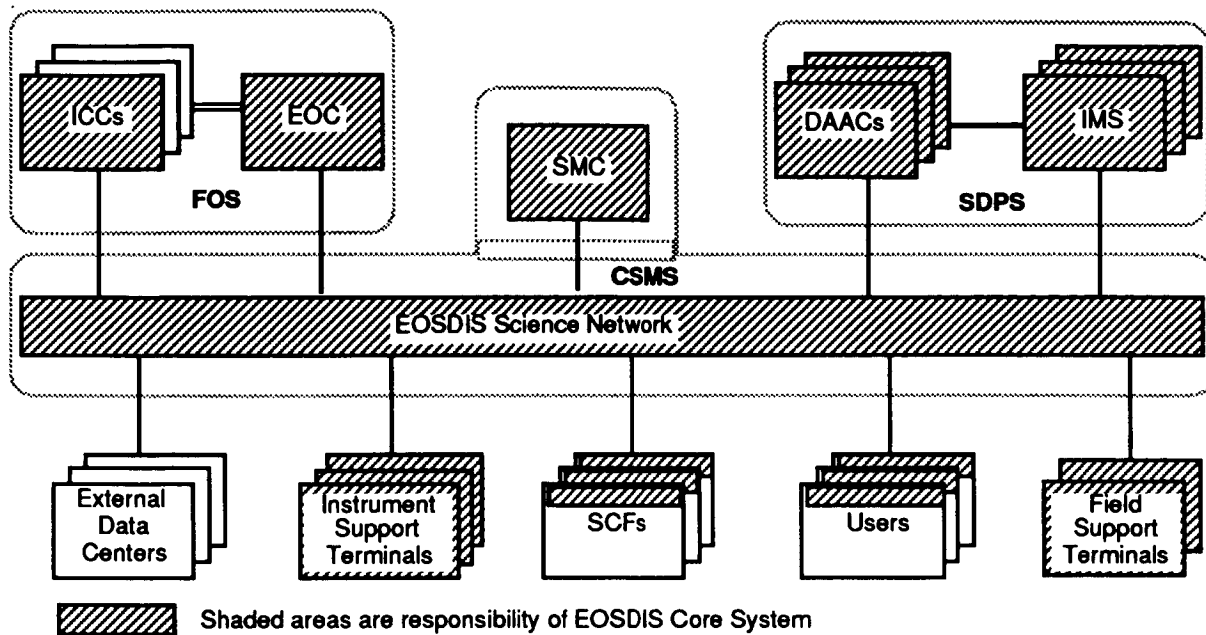
The third segment of the core system is the Science Data Processing Segment (SDPS) which is composed of three functional elements. The Product Generation System (PGS) provides the systems and software to generate the higher level data products from the EOS observations. The Data Archive and Distribution System (DADS) stores the EOSDIS data products and auxiliary data and fills data requests. The third element of the SDPS is the Information Management System (IMS) which is the users' interface to all data and services of the EOSDIS. The IMS manages and provides the users with access to all of the information required to search, select, and order any of the EOSDIS data products and to construct and submit requests for data acquisitions and standard data processing options.

The SDPS is shown as a distributed system with the PGS and DADS elements coupled to form a Distributed Active Archive Center (DAAC) with a corresponding IMS element at each DAAC (the IMS element is somewhat arbitrarily shown outside of the DAAC to depict its system-wide interface to the EOSDIS in addition to its local information management function). This distribution was motivated by several of the Program requirements. First, to centralize a data collection of the size and scope of EOSDIS would be to create a huge, monolithic archive of unprecedented dimensions. The fear that such a center would be unresponsive to the evolving requirements of the user community and the specific needs of the individual user, was quickly pointed out by the EOSDIS science advisory panel. Secondly, building a centralized EOSDIS would diverge from the existing distribution of Earth science data where the archives are held at centers of discipline expertise. Instead, the distributed EOSDIS elements will augment these centers as shown in Table 1. Another advantage that distribution offers is the ability to easily expand the system by the addition of new active archive centers. This option has already been exercised by the addition of the DAAC at Oak Ridge National Laboratory. Finally, the scope of global change research dictates that much of the data required by the science investigators will be held outside of EOSDIS by other agencies and institutions. The technical challenges of providing access to distributed data will have to be addressed.

Table 1 - EOSDIS Distributed Active Archive Centers

Center	Heritage Systems	Areas of Interest
Goddard Space Flight Center	NASA Climate Data System, Pilot Land Data System, Coastal Zone Color Scanner Data System	Upper atmosphere, atmospheric dynamics, global biosphere, and geophysics;
Langley Research Center	ERBE processing	Radiation budget, aerosols, and tropospheric chemistry
EROS Data Center	Global Land Information System, Landsat processing	Land processes imagery
University of Alaska - Fairbanks	Alaska SAR Facility System	Sea ice, polar processes imagery (SAR)
University of Colorado	Cryospheric Data Management System	Cryosphere (non-SAR)
Jet Propulsion Laboratory	NASA Ocean Data System	Ocean circulation and air sea ice interaction
Marshall Space Flight Center	WetNet	Hydrologic cycle
Oak Ridge National Laboratory	Based on CDIAC , ARM , & Surface Water Survey Data Center data management procedures	Biogeochemical dynamics

Access Characteristics

The EOSDIS Program goals and the science user community input have been the basis for the definition of the access requirements and the EOSDIS element designs. The two elements that are directly involved in providing this access are the IMS and the DADS. The DADS data distribution must deliver the data with a response that supports the scientific research process and in the desired format. All of the IMS functions are directed at providing the science users with a functionally complete and robust interface to the data and services of the system.

The IMS design process has identified those functions that the users will require to gain access to the information, data, and services of the EOSDIS. The first function is a user interface that supports a dialog between the user and the system and conveniently supports the interchange of information. The interface to all functions should follow a consistent style and must provide an appropriate level of assistance to guide the user through the session.

The information search functions of the IMS must be capable of answering a wide variety of user queries. At the highest level, the users will want to identify data sets that can be used in their particular area of research. The users will enter information that describes their research such as discipline, parameter, and area of interest into the system, and a directory function will identify those data sets that meet the criteria. The directory will provide overview descriptions of the data sets which may be within the EOSDIS or held by external systems. Users will typically need more specific information to actually order data from the system. An inventory function will identify instances of a data set (granules) that specifically meet the users' data needs. To adequately narrow the search, this function may require users to enter additional criteria such as the specific time and location of interest and data quality requirements. In addition to identifying data sets and data granules, the IMS will perform additional functions that assist in the selection and use of the data. Visual aids will include the display of browse data, which are subset or subsampled versions of the data that have been

produced to allow the user to preview a data granule, and coverage maps showing the geographical location and areal extent of one or more data sets or data granules. Another information access function, referred to as the guide, will provide the capability to access a variety of text and other supplementary information to assist in the selection and use of the data.

The IMS functions also include the generation and delivery of data and service requests to the other elements of the system. Through the user interface of the IMS, the user will be able to construct three types of data requests. For those instruments that have a variable duty cycle, the user can construct a request for a future acquisition. The user interface will also present distribution and delivery options which together with the results of an information search allow the user to construct a request for archived data products. The third type of request is for data processing and will allow the user to select from a set of standard options. In addition to data requests, the IMS will provide the interconnectivity with cooperating external Earth science archives where agreements and standard procedures have been established. Finally, throughout the IMS, an automated "Help" service will be available to provide as much on-line assistance as possible, tailored to the users' needs, and complementary to the user services functions.

The concepts and conceptual designs of the EOSDIS access functions reflect the nature and characteristics of the Global Change Research Program and the Earth science research community. Since the program is interdisciplinary and the community ranges from students in Earth science to the experienced investigator, the system must provide sufficient background information in addition to the minimal search and order functions. Because the users will have different levels of experience with this system and with data systems in general, the help facility must be tailorable to the particular user. The users' local environment will also vary, with some relying on terminals to access the EOSDIS and others using powerful workstations allowing enhanced visualization and other capabilities.

The different characteristics of the scientific investigations will also place widely divergent demands on the system. Some will be global in scale with long time ranges of interest, while others will be site or phenomenon studies with much smaller data requirements. The nature of the study and the preferences of the user will determine the mode in which the user interacts with the system and will be reflected in the degree of coupling between the EOSDIS and the user's local analysis systems. At one extreme, the two are decoupled and the investigator uses the EOSDIS to select and deliver large bulk orders of data that will be managed and processed locally. This corresponds to a "personal library" model and typically relaxes the response time required of the EOSDIS. At the other extreme, the two systems will be completely integrated and the local analysis system will rely on the EOSDIS to manage and supply its data resources. The integrated model may require that smaller data volumes be delivered with each request but also implies significantly more frequent and faster deliveries. This impacts the data storage methods at the DADS and the required network performance.

Development Approach

The EOSDIS Program requirements and goals have had a significant impact on the approach that has been taken to define, design, implement, and operate the system. An overview of the approach is shown in Figure 2. In recognition that the definition and development cycle of such a large, complex system would be a lengthy process and to take advantage of the experience base that exists from the development and operation of the heritage systems at the identified DAAC sites, the concept of EOSDIS Version 0 emerged. The general goal of Version 0 is to provide "lessons learned" for EOSDIS through the integration and augmentation of the capabilities at the DAACs and the development of selected prototypes.

Figure 2 - EOSDIS Development Cycles
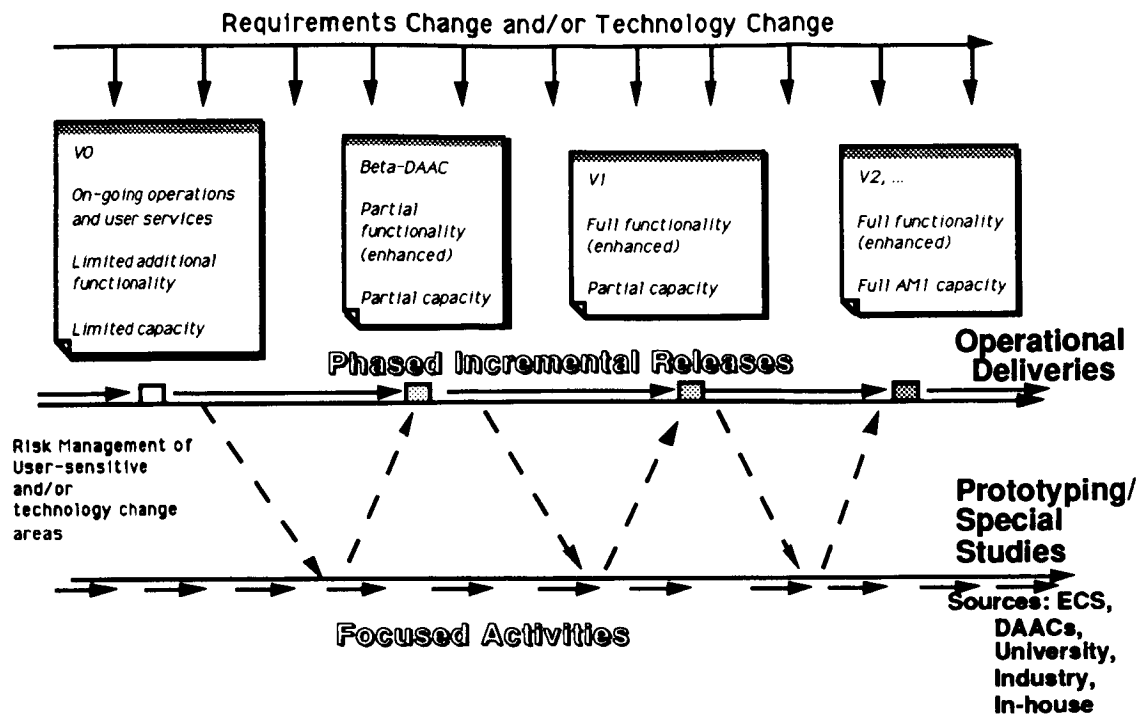
credit G. McConaughy, ESDIS Project

The Version 0 effort consists of DAAC development activities and a number of system-level tasks. Each DAAC is responsible for the design and implementation of the product generation, data archive and distribution, and local information management functions that are required to support their existing data and the data from precursor missions. The development is governed by a set of requirements and an architecture and operations concept that have been developed by the team of EOSDIS Project and DAAC system engineers. In addition to satisfying the local data and information system requirements, this effort provides the mechanism to jointly address the program goals and objectives and a forum to exchange information on technology and approaches.

This system engineering team has also defined the objectives and scope of the information management, networks, and data format system-level tasks. The Version 0 IMS system-level task is working with the DAAC teams to implement an operational prototype that provides a cross-DAAC data search and order capability. This prototype will provide a preliminary version of each of the information access functions and the data access function which allows users to place requests for archived data products. The Version 0 formats task has evaluated many of the data formats that are currently being used by the science community and is working with the DAACs to reach consensus on the standard formats that will be used in the Version 0 timeframe. The networks task is analyzing the bandwidth requirements to perform the Version 0 on-line access and distribution functions and is responsible for the enhancement of user and inter-DAAC connectivity.

The Version 0 activity is only the first step in the EOSDIS development process. The EOSDIS Core System Project will develop the Beta-DAAC and subsequent versions of the system through a series of phased, incremental releases. At the same time, the project will be supporting a series of prototypes and special studies that will focus on particular topics that address user-sensitive areas and emerging technologies. These prototypes will be performed by the ECS contractors, other industry participants, the DAACs, academic institutions, or

428

members of the in-house staff. The results of these activities will continuously feed into the implementation of the operational systems. The science community will be constantly involved in the specification and evaluation of the prototypes and releases.

Conclusions

The breadth and complexity of the Program goals and the size and diversity of the Earth science research community place unique data and information access requirements on the EOSDIS which impact all aspects of its design and development. The distributed architecture is a reflection of the data volumes to be supported and the broad scope of participants, in terms of both institutions and investigators. The flexibility of the access functions and the wide range of capabilities that they must provide are necessary because of the differing needs of the individual science users. However, the greatest impact is on the overall approach that is being used in the development of EOSDIS.

The EOSDIS conceptual architecture and functional requirements describe the current understanding of global change research and the role to be played by the data and information system. However, there are many aspects of this process that are not yet well understood and it is believed that the process will evolve over the course of the program. The EOSDIS development plan recognizes and reflects these factors. More direct interaction with the users is required to more accurately define the current requirements and is a primary justification of the Version 0 effort. Technology is rapidly changing in many ways that will enable the science community to better pursue its research and that will potentially alter the requirements and characteristics of its access to EOSDIS. The intention of the prototyping plan is to give the users the opportunity to use and evaluate the technological advances. Finally, the delivery of the EOSDIS through a series of releases will allow the systems to evolve in response to changes in user requirements and data systems technology.