160-TP-003-001

A Science User's Guide to the EOSDIS Core System (ECS) Development Process

Technical Paper

Technical Paper—Not intended for formal review or government approval.

February 1995

Prepared Under Contract NAS5-60000

RESPONSIBLE ENGINEER

Thomas Dopplick /s/ 2/14/95

Thomas Dopplick, Science Office February 14, 1995
EOSDIS Core System Project

SUBMITTED BY

Joy Colucci /s/ 2/14/95

Joy Colucci, Science Office Manager February 14, 1995
EOSDIS Core System Project

Hughes Applied Information Systems
Landover, Maryland

This page intentionally left blank.

Contents

1. Introduction

1.1 Purpose 1-1

1.2 Organization 1-1

1.3 Acknowledgments 1-1

1.4 Review and Approval 1-1

2. EOSDIS Core System Context

2.1 Introduction 2-1

2.2 New, Comprehensive Earth Science Measurements 2-1

2.3 Unique Mission Support Capabilities 2-3

3. Key Features for Science Users and Data Providers

3.1 Services-Based Architecture 3-1

3.2 Object-Oriented Design 3-2

3.3 Cohesive Information Model 3-3

3.4 Focus on Data Quality 3-5

3.5 Emphasis on Calibration and Validation 3-5

3.6 Options for Distributing Local Science Products 3-6

3.7 Robust Infrastructure Components 3-7

3.8 Operational Resiliency 3-8

4. EOSDIS Implementation

4.1 Capabilities Developed as Needed for Mission Support 4-1

4.2 Specialized Skills and Teamwork 4-2

4.3 Multi-Track Development 4-3
4.4 Key ECS Development Challenges 4-4
4.5 Current Access to Selected NASA Data Sets 4-5
4.6 Evolution of ECS 4-5

5. Science Community Involvement

5.1 User Involvement at Technical Levels 5-1
5.2 Influence of User Feedback on ECS 5-2

Figures

2-1 Examples of New Measurements Available from EOS Instruments 2-3
2-2 EOS Instrument Data Flow 2-4
3-1 High-Level Services Provided by ECS Infrastructure 3-1
3-2 ECS Services-Based Architecture 3-2
3-3 Examples of Subsystem Design Features 3-3
3-4 EOSDIS Data Pyramid 3-4
3-5 Key ECS Automated Resource Components 3-7
4-1 ECS Development Schedule with Example EOS Missions 4-1
4-2a Technology Projection of Cost Per MFLOP (Processing) 4-2
4-2b Technology Projection of Cost Per Gbyte (Disk Storage) 4-2

Tables

2-1 The EOS Rebaselined Satellite Series 2-2
2-2 Locations and Scientific Expertise of the EOSDIS DAACs 2-5
4-1 Examples of EOSDIS Functional Components and Required Skills 4-2
4-2 Highlights of ECS Architectural Features that Facilitate Evolution 4-6

1. Introduction

1.1 Purpose

During the last series of User Working Group meetings, a recurring theme was the lack of understanding about the Earth Observing System (EOS) Data and Information System (EOSDIS) Core System (ECS) by the EOS science community. We heard there are too many acronyms, too many engineering documents, and too few summaries suitable for end users. So we organized an effort to try to summarize selected ECS material. This Technical Paper summarizes user-related development issues and how users can participate in the ECS development process.

We know that each of you is very busy with many other tasks, but we'd really like to hear about your reaction to this paper. Does the paper help you better understand the ECS development process? Are there any issues which we missed in the paper? Are there still too many acronyms? We tried diligently to reduce the number of acronyms - did we succeed?

1.2 Organization

This paper is organized into the following broad categories:

Chapter 2: ECS context within EOSDIS.

Chapter 3: Key features of ECS

Chapter 4. Implementation strategy

Chapter 5. User involvement

1.3 Acknowledgments

Drs. Karl Cox and Celeste Jarvis of the ECS Science Office also contributed material to this Technical Paper. Special thanks are due Dr. Stephen Wharton, NASA EOSDIS Project Scientist, for his encouragement and subsequent review of the Technical Paper.

1.4 Review and Approval

This Technical Paper is an informal document approved at the Office Manager level. It does not require formal Government review or approval. Questions regarding technical information contained within this Paper should be addressed to Tom Dopplick, 301-925-0333, tom@eos.hitc.com.

Questions concerning distribution or control of this document should be addressed to:

Data Management Office
The ECS Project Office
Hughes Applied Information Systems
1616A McCormick Dr.
Landover, MD 20785

This page intentionally left blank.

2. EOSDIS Core System Context

2.1 Introduction

NASA historically has built separate and centralized data and information systems for each science mission because of past focus on discipline-oriented research and the difficulties in building open, distributed systems. Modern technology now allows broad coupling of expanded services with advanced mission data collection to support interdisciplinary Earth science investigations, a necessary condition for resolving issues associated with global change. This revolution in science research is the fundamental reason for beginning the Earth Observing System (EOS), the centerpiece of Mission to Planet Earth which is NASA's contribution to understanding global change. The EOS Data and Information System (EOSDIS) provides end-to-end services from EOS instrument data collection to science data processing to full access to EOS and other Earth science data holdings. EOSDIS' infrastructure, the EOSDIS Core System (ECS), provides EOS and other U.S. and international scientists a broad range of desk top services from 9 science data centers, the Distributed Active Archive Centers (DAACs, see Table 2-2 for locations), operated by NASA and other agencies. The ECS infrastructure also supports exchange of data and research results within the science community, across multiple agencies and internationally. ECS is the evolutionary base for accelerating the pace of Earth science research.

2.2 New, Comprehensive Earth Science Measurements

Earth science research needs to broaden the use of remote sensing observations in order to complement *in situ* as well as theoretical studies owing to the complexity and spatial extent of the global Earth system. However, most geophysical variables cannot be directly measured by on-orbit measurements; rather, science algorithms applied to space measurements progressively convert raw instrument data to calibrated data to geophysical variables. A major strength of the EOS program is the breadth of geophysical variables for the global Earth system that can be derived from the numerous EOS instruments, such as variables for the land surface, oceans and atmosphere, as well as related biogeochemical variables. Table 2-1 shows examples of this breath as viewed through science mission objectives of the rebaselined EOS satellite series. The EOS satellite series will provide continuous, long term observations; whereas, other EOS instruments will fly on available non-EOS satellites, such as on the Tropical Rainfall Measurement Mission (TRMM) and international partner satellites. The number of derived variables and resultant new science products is large, with over 200 new science products from the EOS instruments on TRMM and the first EOS AM satellite.

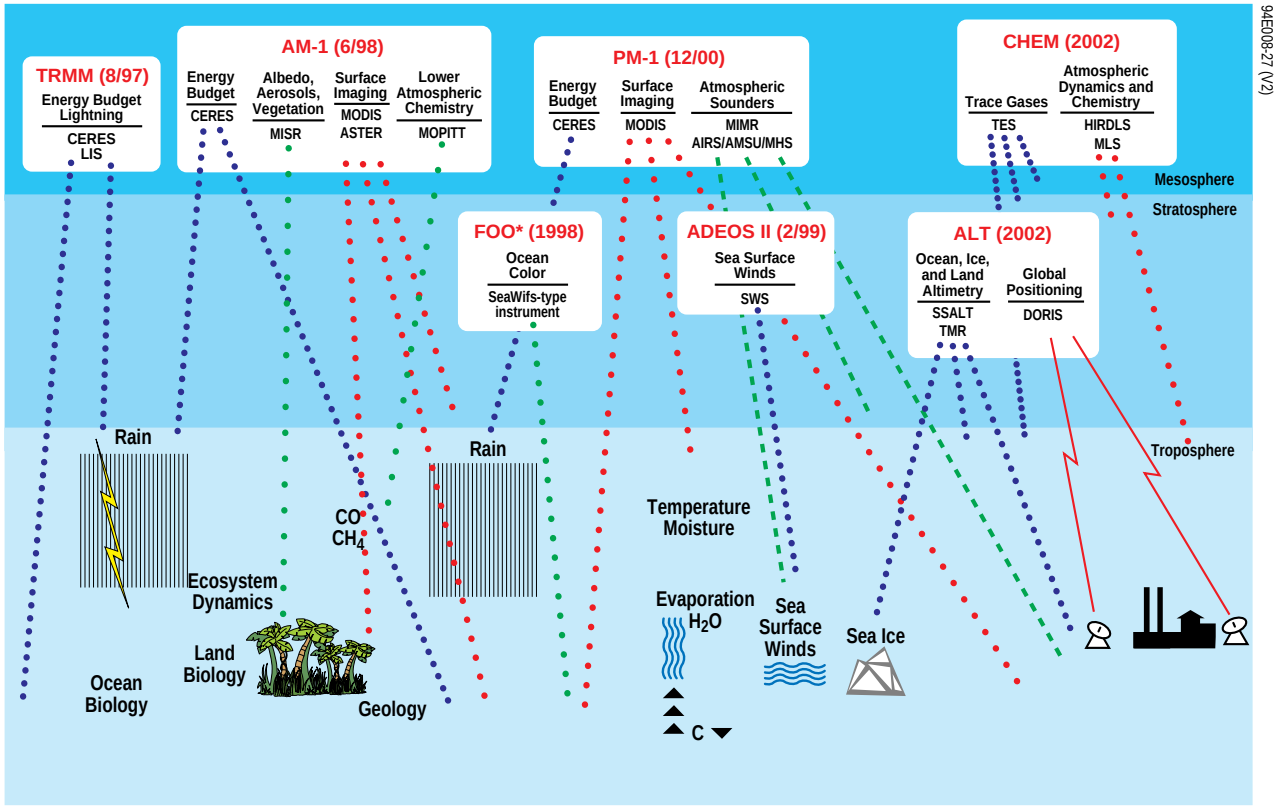
Table 2-1. The EOS Rebaselined Satellite Series

Satellites	Mission Objectives
EOS AM	Clouds, aerosols, and radiation balance, characterization of the terrestrial ecosystem; land use, soils, terrestrial energy/moisture, tropospheric chemical composition; contribution of volcanoes to climate; and ocean primary production productivity
EOS PM	Cloud formation, precipitation, and radiative properties; atmospheric temperature and moisture profiles; air-sea fluxes of energy and moisture; sea-ice extent; and soil moisture and snow over land
EOS ALT	Ocean circulation and ice-sheet mass balance
EOS CHEM	Atmospheric chemical composition and dynamics; chemistry-climate interactions; air-sea exchange of chemicals and energy

EOS investigators will analyze integrated, long-term earth observations from a variety of new remote sensing instruments in characterizing the components of global climate change and their interactions. Several small- and intermediate-sized platforms will be flown (see examples in Figure 2-1), with each platform carrying a suite of sensors selected to address key global change issues.

The view of Earth through these new sensors will be different from previous views; advances in remote sensing technology enable much higher resolution, spatially and spectrally, than has previously been possible on a global basis. For example, analysis of the 1 km data of the Moderate-Resolution Imaging Spectroradiometer (MODIS, see AM-1 in Figure 2-1), will enhance knowledge of global dynamics and processes on the surface of the earth and in the lower atmosphere. The 36 wavelength bands that MODIS will simultaneously measure have been selected based on prior experience to provide the necessary information to separate convolved surface and atmospheric effects. Similarly, the high spectral resolution of the Atmospheric Infrared Sounder (AIRS, see PM-1 in Figure 2-1), with over 2000 spectral channels, will yield much improved vertical resolution of atmospheric temperature, water vapor and ozone measurements. In addition, AIRS will allow much improved characterization of the spectral properties of clouds and land, and a more accurate determination of the height of cloud tops.

Each of the remaining instruments has much to offer individually but together they offer additional, synergistic benefit. Simultaneous observations of the same phenomena by complementary sensors allow much improved corrections to be developed for atmospheric uncertainties. For example, many derived geophysical variables, such as surface leaving radiances, require correcting for atmospheric uncertainties in order to produce an accurate science product. The synergism of these simultaneous measurements is very likely to lead to further improvement in the characterization of the land and ocean properties, clouds, aerosols, atmospheric processes, and radiative balance. The *EOS Reference Handbook* (Asrar and Dokken, 1993) contains details on all of the EOS instruments and is readily available via the World Wide Web using the URL provided in the reference at end of this chapter.



*FOO = Flight of Opportunity

Figure 2-1. Examples of New Measurements Available from EOS Instruments.

2.3 Unique Mission Support Capabilities

EOSDIS is an integrated system that supports multiple satellites and instruments, not a unique system for each. EOS includes instruments on satellites to be launched by NASA, the European Space Agency (ESA), and the Japanese National Space Agency (NASDA). Figure 2-2 provides a schematic view of the flow of data from the EOS instruments to the users. The ECS consists of the shaded portions of Figure 2-2, plus facilities for operation of the NASA EOS satellites and instruments and NASA EOS instruments on International Partner satellites.

The data from each EOS instrument will be sent to a Distributed Active Archive Center (DAAC) responsible for processing, archiving, and distributing EOS and related data. These data centers will house the ECS computing facilities and operational staff needed to produce EOS Standard Products and to manage, store, and distribute EOSDIS data, as well as the associated metadata and browse data, that allow effective use of the data holdings. The DAACs will exchange data via dedicated EOSDIS networks to support processing at one DAAC which requires data from another DAAC. NASA selected the DAACs based on their expertise in specific science disciplines (indicated in Table 2-2) and demonstrated long-term commitments to the corresponding user communities. An additional DAAC, the Socio-Economic Data and Applications Center acts as a link between the EOS Program and the socio-economic and educational community.

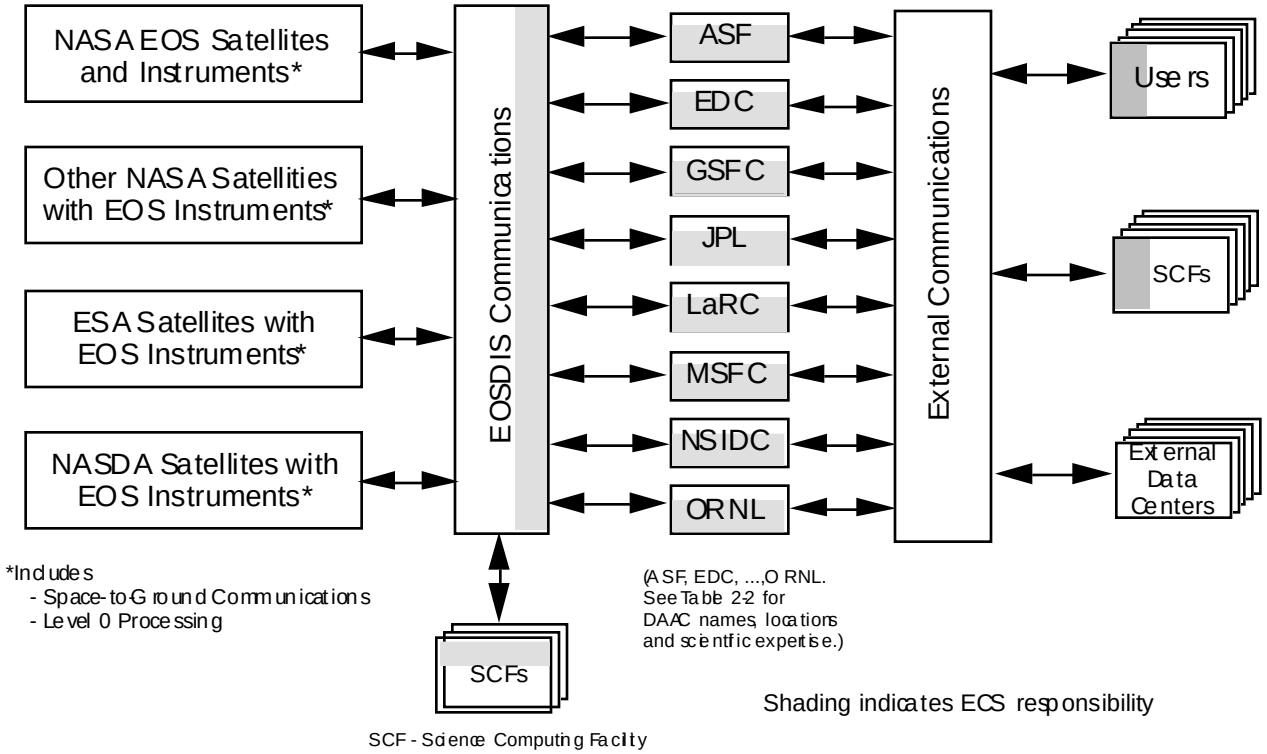


Figure 2-2. EOS Instrument Data Flow

The DAACs also house systems for processing and/or storage of non-EOS Earth science data. For example, the Alaskan SAR Facility currently provides systems for receiving, processing, and archiving data from Synthetic Aperture Radars on International Partner platforms.

Open access to the data by all members of the science community distinguishes EOS from previous research satellite projects, where selected investigators had proprietary data rights for a number of years after data acquisition. This open data policy will lead to greater utilization of EOS data products, for global change research and other applications. The EOS program is also distinguished by the large number of funded investigators (over 500), who provide expertise across the broad range of scientific disciplines in Earth system science.

Science Computing Facilities (SCFs), at EOS investigators' home institutions, are used to develop and maintain algorithms (for both Standard and Special Products), calibrate the EOS instruments, validate data and algorithms, generate Special Products, provide data and services to other investigators, and analyze EOS and other data in pursuit of the overall science objectives. The SCFs may range from single workstations to large supercomputer data centers. While the SCFs will be developed and acquired directly by the EOS investigators, ECS will provide software toolkits to the SCFs and other users to facilitate data access, transformation and visualization, and for science algorithm development. Some SCFs will play an operational role in quality control of the EOS Standard Products.

Table 2-2. Locations and Scientific Expertise of the EOSDIS DAACs

DAAC	Location	Expertise
Alaska SAR Facility (ASF), University of Alaska	Fairbanks, AK	Sea Ice and Polar Processes Imagery
EROS Data Center (EDC), U.S. Geological Survey	Sioux Falls, SD	Land Processes Imagery
Goddard Space Flight Center (GSFC), NASA	Greenbelt, MD	Upper Atmosphere, Atmospheric Dynamics, Global Biosphere, and Geophysics
Jet Propulsion Laboratory (JPL), NASA	Pasadena, CA	Ocean Circulation and Air-Sea Interaction
Langley Research Center (LaRC), NASA	Hampton, VA	Radiation Budget, Aerosols and Tropospheric Chemistry
Marshall Space Flight Center (MSFC), NASA	Huntsville, AL	Hydrology
National Snow and Ice Data Center (NSIDC), University of Colorado	Boulder, CO	Cryosphere
Oak Ridge National Laboratory (ORNL), Department of Energy	Oak Ridge, TN	Ground-based data relating to Biogeochemical Dynamics
Socio-Economic Data and Applications Center (Consortium for International Earth Science Information Network - CIESIN)	Saginaw, MI	Socio-Economic Applications

For more details on EOSDIS and ECS, a suggested reading list is provided at the end of each chapter. Several of the suggested documents are readily available via the World Wide Web at referenced URLs.

Suggested Reading

Asrar, G. and Dokken, D., 1993: 1993 EOS Reference Handbook. NASA Headquarters. URL: http://spso.gsfc.nasa.gov/eos_reference/TOC.html

Asrar, G. and Dozier, J, 1994: Science Strategy for the Earth Observing System. NASA Headquarters.

EOS Instrument Principal Investigators, 1994: Algorithm Theoretical Basis Documents. URL: http://spso.gsfc.nasa.gov/atbd/pg1.html

James, M. 1995: EOSDIS Data Set Reference Handbook. In preparation, to be published in Spring 1995 by NASA/GSFC.

Unninayar, S. and Bergman, K., 1993: Modeling of the Earth System in the Mission to Planet Earth Era. NASA Headquarters.

This page intentionally left blank.

3. Key Features for Science Users and Data Providers

3.1 Services-Based Architecture

The science community and the National Research Council reviewed earlier architectural concepts for ECS and strongly recommended changes to make ECS more extensible and open. Their recommendations were incorporated and presented at the ECS System Design Review in June 1994. After receiving approval to proceed, the revised system design was baselined for ECS. From a science user's perspective, the ECS infrastructure appears as services, and interfaces to services, which appear on the scientist's desktop. This perspective is similar to the view of the World Wide Web (WWW) servers as seen through a local client such as Mosaic or Netscape. When scientists invoke ECS services, they are actually exercising the client environment running in their local workstation (Figure 3-1).

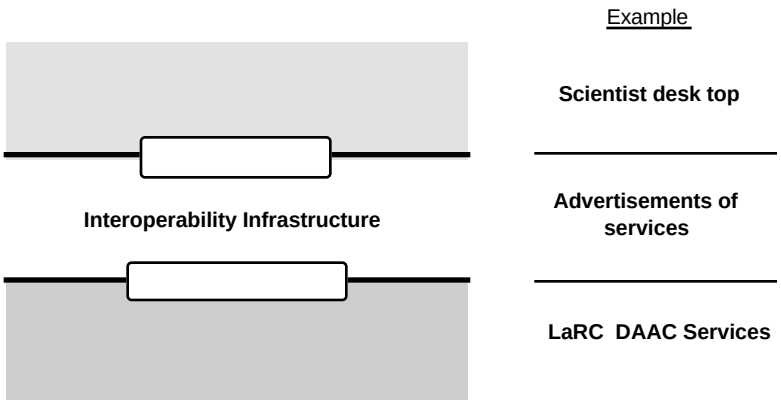


Figure 3-1. High-level Services Provided by ECS Infrastructure

The ECS services-based architecture (Figure 3-2) serves a wide range of user needs and allows scientists to focus on global change research rather than computer science details. The Client Services Layer allows users to easily explore and locate provider services that are advertised through the Interoperability Layer. Advertisements provide complementary, coupled views of services, data sets, and providers. After identifying a service of interest, users can immediately invoke the provider service, or the reference to the provider service can be saved on the desktop for later use. For example, users can locate and invoke an EOSDIS-wide service, such as cross-DAAC searching or choose local DAAC services, such as search of local data collections and DAAC-unique subsetting. A user can also query directly, using science-oriented forms or free text, for specific provider services. Providers are not limited to the 9 DAACs; providers can be other data centers or other scientists who adopt ECS interfaces and protocols.

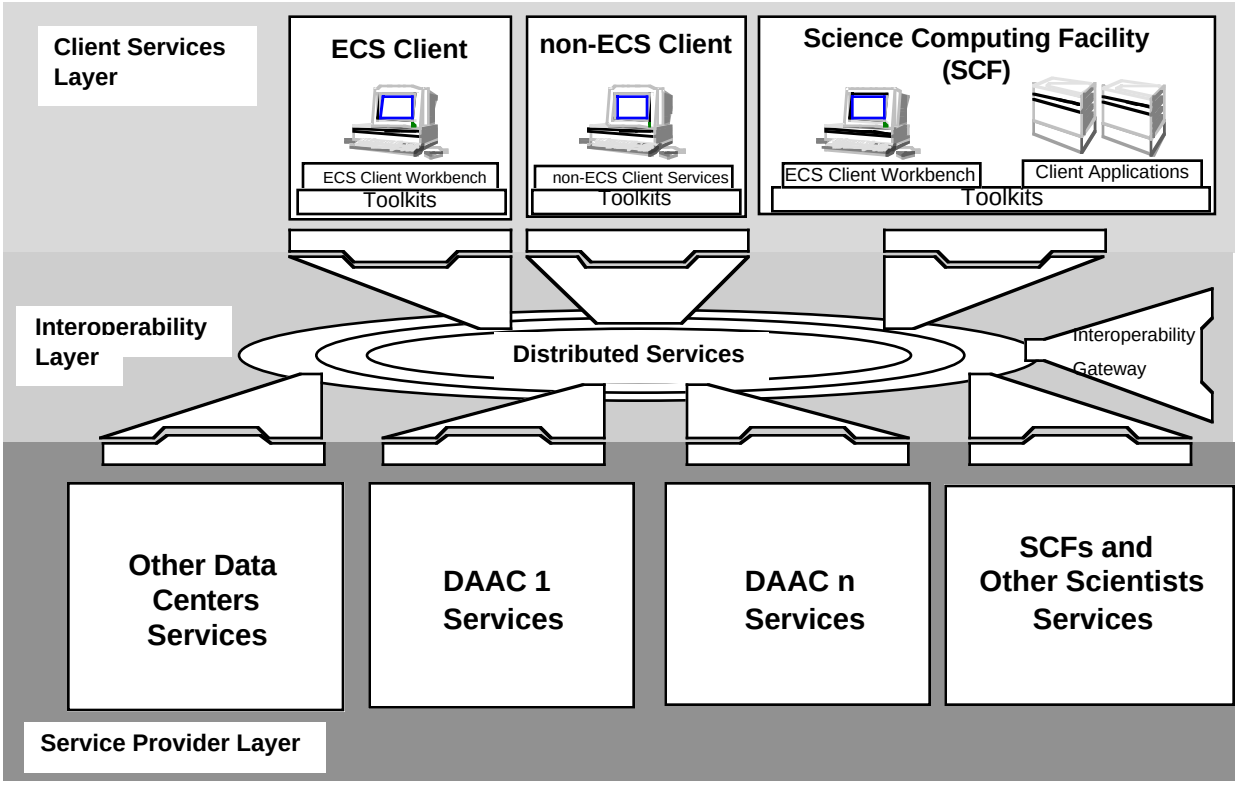


Figure 3-2. ECS Services-Based Architecture

ECS has chosen an open architecture that allows provider autonomy and independent, evolutionary development of components to improve services offered to users. Because of logical distribution, users and providers can tailor their components to their environments and still operate together as an extended distributed system such as found on the World Wide Web.

3.2 Object-Oriented Design

The engineering process that transforms the architecture to design is defined and described in numerous ECS technical documents and the results presented in formal reviews throughout the development life cycle. Using an object-oriented design methodology, the ECS design is iteratively refined with progressively more definition of subsystems, lower-level objects, and associated interfaces. In Figure 3-3 we highlight some of subsystem design features of importance to science users and data providers. For example, the client subsystem provides users a desktop workbench for accessing a broad range of services such as search services. Other subsystems, such as the planning subsystem, allow the DAACs to plan for instrument data processing in a data driven mode as well as respond to requests for on demand processing. Evolutionary upgrades will be implemented depending on technology and cost considerations. For example, content-based searching of EOS products is not practical with current technology and is a candidate for a future evolutionary upgrade.

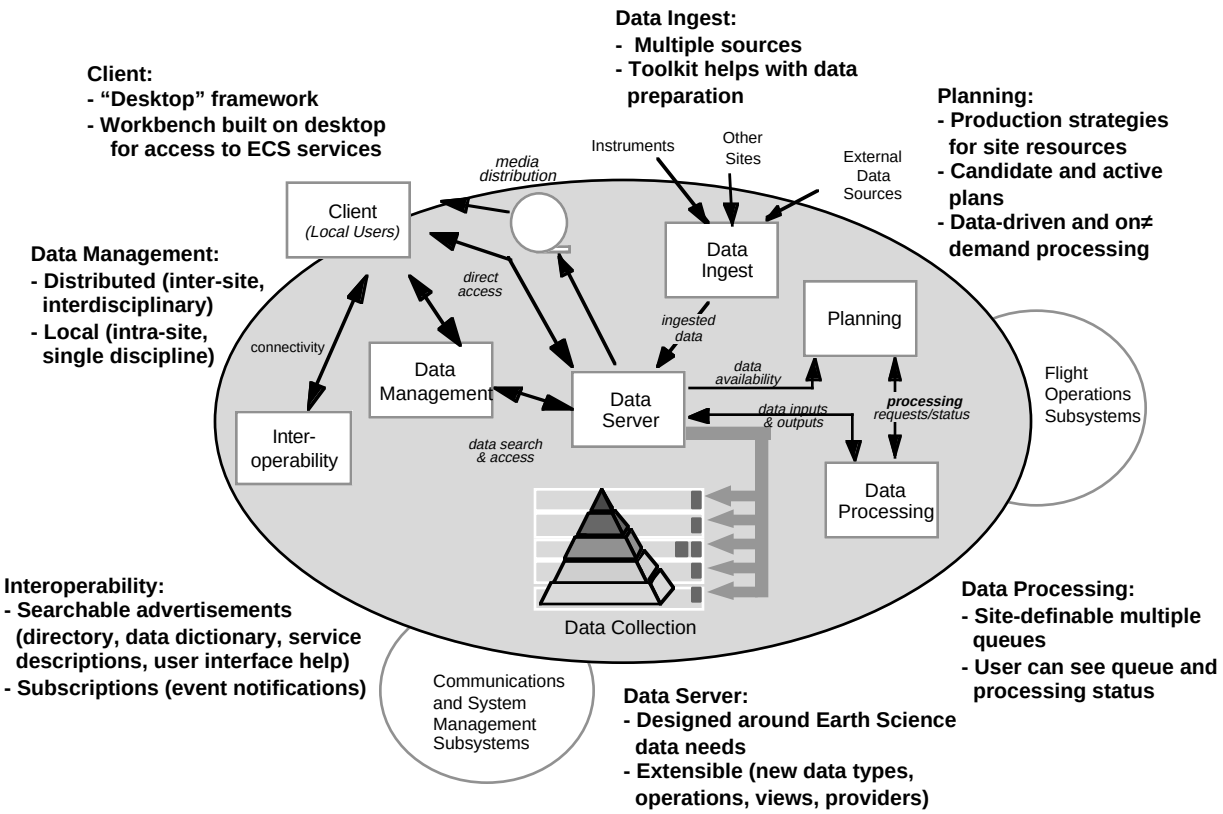


Figure 3-3. Examples of Subsystem Design Features

The ECS Team adopts commercial-off-the-shelf (COTS) design solutions wherever possible in order to leverage the technological revolution underway in areas such as open systems and distributed computing. ECS draws heavily on the R&D technology of the entire commercial marketplace, resulting in a design that uses 100% COTS hardware and extensive COTS software for long term affordability and supportability. COTS hardware is selected based on projections of performance versus cost to maximize return on evolution of commercial technology. The use of "just-in-time" COTS hardware procurement is discussed in Section 4.1 below.

3.3 Cohesive Information Model

To implement a services-based architecture, the ECS Team must identify, analyze, and capture the context and relationships of the science data and information holdings within EOSDIS, i.e. develop an information model. For example, a search service for sea surface temperature (SST) assumes an information model exists within ECS that models the relationship of the geophysical parameter SST with other SST-related variables such as instrument data attributes, space and time variability, and the science algorithm that converted the instrument data to SST. Thus, from product generation to product access, the context and relationships of science data and related variables must be modeled so users can request and apply services to the data. The choice of an information model is not arbitrary, rather it is based on careful examination of how data are used within EOSDIS by EOS scientists. Science data usage in EOSDIS is closely tied to data

products and the uses of those products, which implies a need to generate, store, and access varying collections and levels of data across distributed providers.

The starting point for developing the ECS information model was to characterize Earth science data into broad, multi-layered data categories as represented in the data pyramid (Figure 3-4). The data pyramid identifies data categories; it does not identify the context or relationships between the data categories. To understand relationships, extensive analysis is required to identify and relate the complex interactions between and among the different layers of the pyramid. Logical collections of data, based on their expected relationships, are developed to capture the variability in remote sensing instruments, science disciplines, and other characteristics of the Earth science community. For example, some EOS products have related properties (e.g., cloud type and cloud drop size) while other products are dissimilar (e.g., land vegetation indices and ocean productivity) which suggest certain logical groupings. Characteristics are often similar across a particular science discipline; often similar across products generated from a given instrument; but often different between provider sites because of differing science discipline focus, as well as organizational autonomy.

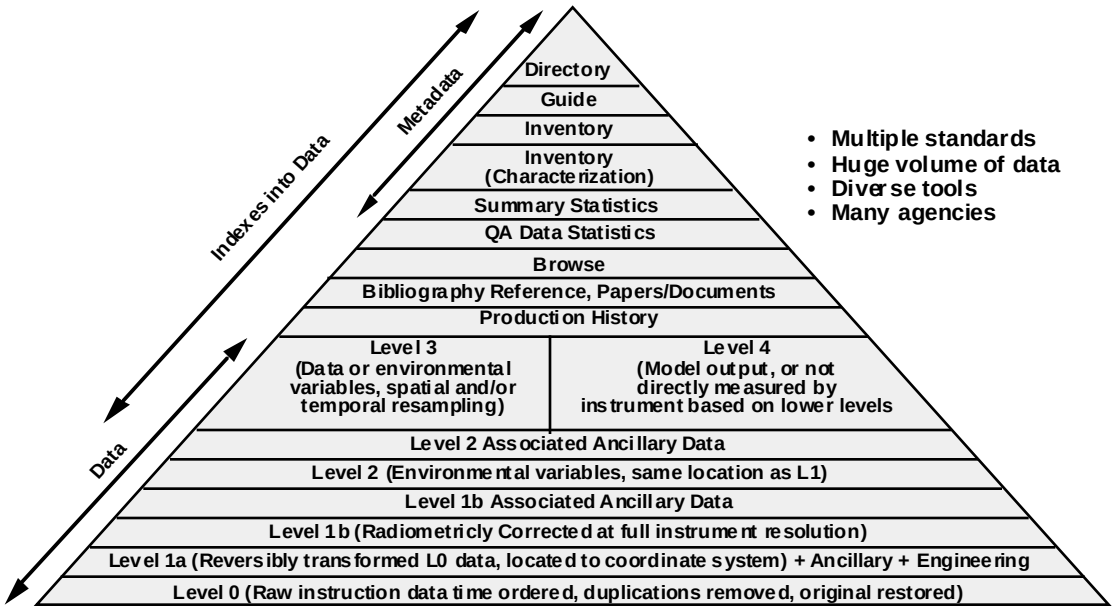


Figure 3-4. EOSDIS Data Pyramid.

Defining logical data collections typically begins with a systematic study of possible logical groupings with the goal of identifying similar characteristics. The resulting model is then examined from an applications perspective such as the expected access pattern to a logical collection. This process produces an information model which is capable of describing data in context. Logical collections are the basis for populating the ECS Data Servers at the distributed DAACs.

A significant benefit of the ECS logical information model is the ability to present the science user an Earth-science view instead of a computer-science view of data. Also, by making the ECS

information model available to researchers analyzing data, EOS scientists developing algorithm software, and scientists developing their own products, individual development can proceed within an overall information management framework that results in a consistent system view across widely distributed components of EOSDIS.

3.4 Focus on Data Quality

Interdisciplinary scientists will be using a wide range of EOS data from direct instrument measurements to geophysical variables derived using complex science algorithms. These scientists need not have detailed knowledge of the intricacies and uncertainties associated with all the instruments and the science algorithms. Rather, access to data quality heritage will promote correct usage and application of the EOS products in interdisciplinary investigations. Three Quality Assurance (QA) processes are applied to the production of standard products, two at the DAACs and the third at the Science Computing Facilities (SCFs).

First, startup QA is performed when production of a specific product is halted in order to assess the quality of intermediate output(s). Startup QA will be used during initial algorithm assessment, rather than during routine processing operations when operational algorithms are expected to perform internal in-line QA.

Off-line QA involves DAAC operations performing quality checks of output from algorithm processing after completion of an algorithm process. The checks might be visual checks requested by the instrument team as part of their operations concept for the algorithm or processing consistency checks on product formats, metadata content etc. DAAC operations perform QA using core services or with specialized software developed by the science user providing the algorithm. When QA is complete, QA results (inventory attributes, QA summary report, etc.) are automatically updated in the product processing history which becomes available to any user of the product.

SCF QA is similar to the off-line QA except that it is performed at the appropriate SCF. The QA in this case is performed using whatever manual/automated procedures the SCF team deem appropriate. The results from the QA may include updated metadata, annotations to the data set, and QA products added to the processing history. Again, the SCF QA results are available to any user of the product.

3.5 Emphasis on Calibration & Validation

NASA has placed strong emphasis on pre-launch and post-launch calibration of EOS instruments as well as validation of the EOS standard products so that interdisciplinary scientists need not have detailed knowledge of the instruments and the science algorithms . To ensure delivery of accurate and reliable measurements, on-orbit and ground resources have been specifically allocated to conduct routine calibration such as periodic measurements in orbit of on-board calibrations sources, as well as celestial sources such as cold space, the moon, or the sun. Field calibration exercises will also provide ground-truth calibration of some instruments. Maintaining accuracy and calibration over long time periods is crucial to studies of climate and global change. Lack of calibration has limited the use of operational satellite data in global change studies and

has spawned a joint NASA and NOAA retrospective effort, the Pathfinder Project, to develop stable calibration of historical raw data and consistent intercalibration among instruments in a series. Consistent intercalibration allows data to be chained together, for instruments in a series, to create longer time series.

Validation of the EOS science algorithms begins with careful analysis of the theoretical basis for the algorithms. Each of the EOS instruments on TRMM and EOS-AM1 produced an Algorithm Theoretical Basis Document which was formally reviewed by NASA and the science community. These documents provide a wealth of information about the science algorithms including expected accuracy of the derived geophysical variables. Each EOS instrument team is planning post-launch validation of their science algorithms through techniques such as comparisons with in situ measurements, field measurements, and intercomparisons with other EOS and non-EOS derived products.

3.6 Options for Distribution of Local Science Products

In addition to providing facilities and resources for generating standard and special products, ECS will provide facilities and resources for scientists to migrate their data to the DAACs for archiving and distribution. Special Data Products (generated as part of a research investigation using EOS data for a limited region or time period, or products that are not accepted as standard by EOS Investigator Working Group and NASA Headquarters), will normally be generated at investigator Science Computing Facilities (SCFs).

Scientists who create their own local products have two options for distributing their results. They can: 1) return these Special Products to a DAAC where they will be ingested, archived, and made available to the general science community, or 2) establish a local provider site and advertise services available at their site via ECS. Migration to a DAAC will involve a peer review process in which science issues, demand and allocation of scarce resources are key factors. A limited amount of storage has been reserved at the DAACs for migration of Special Products from the interdisciplinary teams. Requests from other scientists will be resolved through the peer review process. By adhering to published ECS interfaces and protocols, scientists can use ECS core services, such as ingest, to transfer their product to a DAAC.

Another option for distributing locally produced products is to establish a local provider site and advertise the availability of a new data collection and related services, such as search and retrieval services, and also the associated protocols. ECS will make available reusable components to be rehosted and adapted at provider sites. However, reusable components will contain both commercial products and custom code. ECS is not presently funded to provide commercial software or hardware products for establishment of a local provider site. Funding for commercial software or hardware would be the responsibility of the SCF or scientist. After installation of the local provider site, any software client with compatible interface protocols can access the new services or the user can download the client user interface and interoperability components from ECS. A new provider site can join the data network by providing only an information search and retrieval interface and advertising its services. No other part of the ECS design need exist at the site.

3.7 Robust infrastructure components

End-to-end EOSDIS services depend on ECS providing a robust infrastructure with some components having high reliability, high throughput or large storage capacity. Certain mission critical components must be highly reliable to support launches and to ensure that data are not lost. Examples of mission critical components include: 1) command and control of EOS spacecraft and instruments, and 2) maintenance of reliable, long-term data archives for global change research. Loss of either space assets or long-term data would seriously impair the EOS mission. High reliability is designed into mission critical components only where necessary since high reliability is a significant cost driver.

Other ECS components, i.e. the EOS data processing components, provide high throughput in order to ingest, process, and archive the high data rates from EOS instruments. Capturing the raw EOS instrument data (~200 gigabytes/day in mid-1999) and processing it to the level required to confirm data validity are mission critical functions. However, downstream processing of higher level products is important but not mission critical since recovery from processing errors or loss of data products can be accomplished by reprocessing from lower level input data. Processing of standard products is being reevaluated by NASA and the instrument teams, and will likely result in phasing of products into EOSDIS after launch of EOS spacecraft. Nevertheless, processing demands will be high with an expectation of 10's of GFLOPs needed to process the suite of instruments on each EOS spacecraft. Although processing power is important, other ECS components are equally important in building a robust infrastructure (see Figure 3-5).

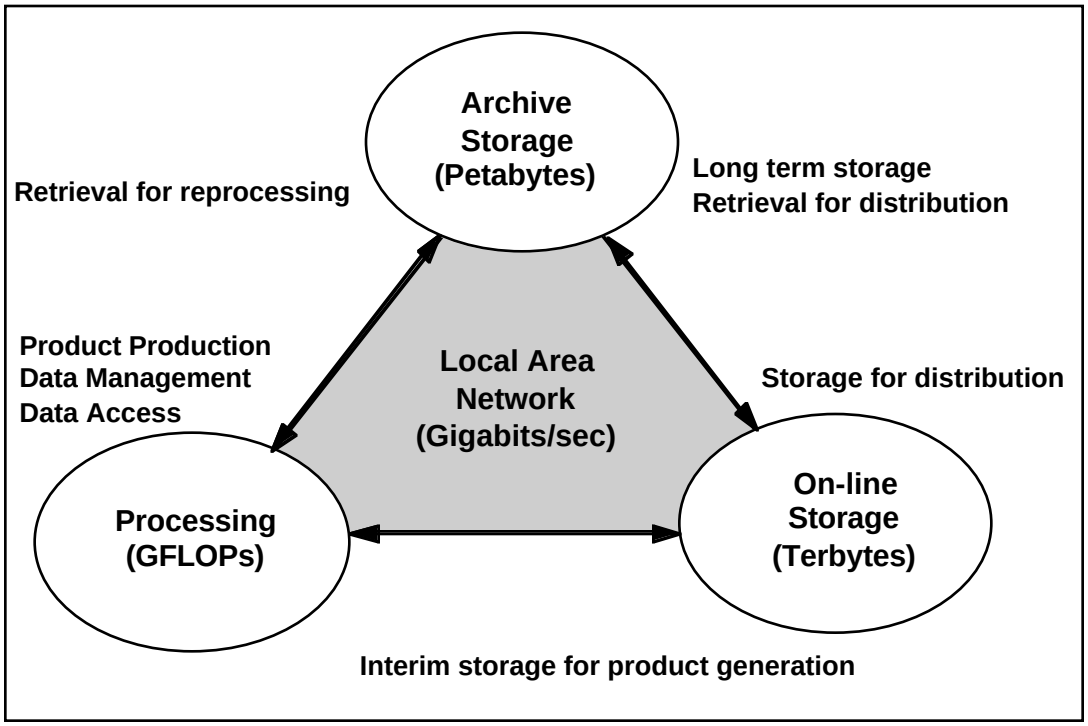


Figure 3-5. Key ECS Automated Resource Components

As a result of the high processing throughput and resultant product generation, ECS also must provide archive (near-line) and on-line storage on a scale not seen before by NASA data centers. For example, projected cumulative archive storage of all DAAC archives, for other NASA data, is estimated to be ~ 29 terabytes by mid-1999; whereas, cumulative storage for archiving EOS data is expected to be ~ 500 terabytes by mid-1999, growing to multiple petabytes[1] during the lifetime of EOS. Petabyte archive storage alone requires high network bandwidth (gigabits/sec) for data transfer to/from the archive. Further, to accommodate distribution, terabytes of on-line storage will be needed for timely distribution of data to local user sites.

Processing power (GFLOPs), archive storage (petabytes), on-line storage (terabytes), and high bandwidth of local area networks (gigabits/sec), must be addressed collectively as a system to achieve a viable ECS infrastructure. Individual components at each DAAC are chosen based on detailed modeling that seeks to maximize service to end users while choosing the most cost effective components at each DAAC and across EOSDIS. In addition to the initial costs of ECS components, maintenance of a viable infrastructure throughout the EOS life cycle involves recurring costs for routine maintenance and operations, as well as periodic upgrades to avoid technology obsolescence.

3.8 Operational Resiliency

ECS has many design features that improve operational effectiveness of user-related services. For example, partitioning of major functions is widely used to avoid resource competition such as separating flight operations processing from science data processing. Also, extensive error prevention and recovery in data handling will ensure end-to-end data integrity. In addition, ECS is being designed to minimize the amount of human interaction required for routine operations. For example, automated monitoring of performance and usage will allow problems to be anticipated, and solutions applied, before they adversely impact user services.

ECS operations will automatically track histories of search access to data collections. Low demand by science users may indicate that the collection does not adequately meet their needs and should be further analyzed for possible reorganization. Detailed analysis may indicate that an isolated subset of the collection is accessed, and that creation of a new logical view or fragmenting into smaller collections may better serve the needs of that particular community. High demand may portend a performance bottleneck, in which case, replication would be appropriate. Also, ECS operations will track user access patterns to determine if users ad hoc needs are becoming more routine and established, indicating the need to create another ‘view’ or context within the Data Dictionary for existing data and services, as well as for potential new collections.

---

[1] petabyte = 1,000 terabytes

Suggested Reading

Case, L., 1994: JU9405V1 SDPS Prototyping Plan White Paper.
URL: http://edhs1.gsfc.nasa.gov

ECS, 1994: ECS Prototyping and Studies Plan. URL: http://edhs1.gsfc.nasa.gov

ECS, 1994: Summary of the ECS System Design Specification. URL: http://edhs1.gsfc.nasa.gov

Elkington, M., Meyer, R. and McConaughy, G., 1994: Defining the Architecture of EOSDIS to Facilitate Extension to a Wider Data Information System. Proceedings of the ISPRS'94, Ottawa.

Endal, A., 1994: FB9402V2 ECS Science Requirements Summary - Version 2.
URL: http://edhs1.gsfc.nasa.gov

Moxon, B., 1994: FB9401V2 EOSDIS Core System Science Information Architecture.
URL: http://edhs1.gsfc.nasa.gov

Moxon, B. 1993: 19300611 Science-based System Architecture Drivers for the ECS Project.
URL: http://edhs1.gsfc.nasa.gov.

National Research Council, 1994: Panel to Review EOSDIS Plans, Final Report. National Academy Press.

This page intentionally left blank.

4. EOSDIS Implementation

4.1 Capabilities Developed as Needed for Mission Support

Step-wise implementation of ECS was chosen to accommodate a compressed development schedule and leverage evolutionary technology. Initial capability for the Tropical Rainfall Measurement Mission (TRMM) will be delivered by December 1996 with EOS-AM1 launch-ready capability delivered by September 1997 (Figure 4-1).

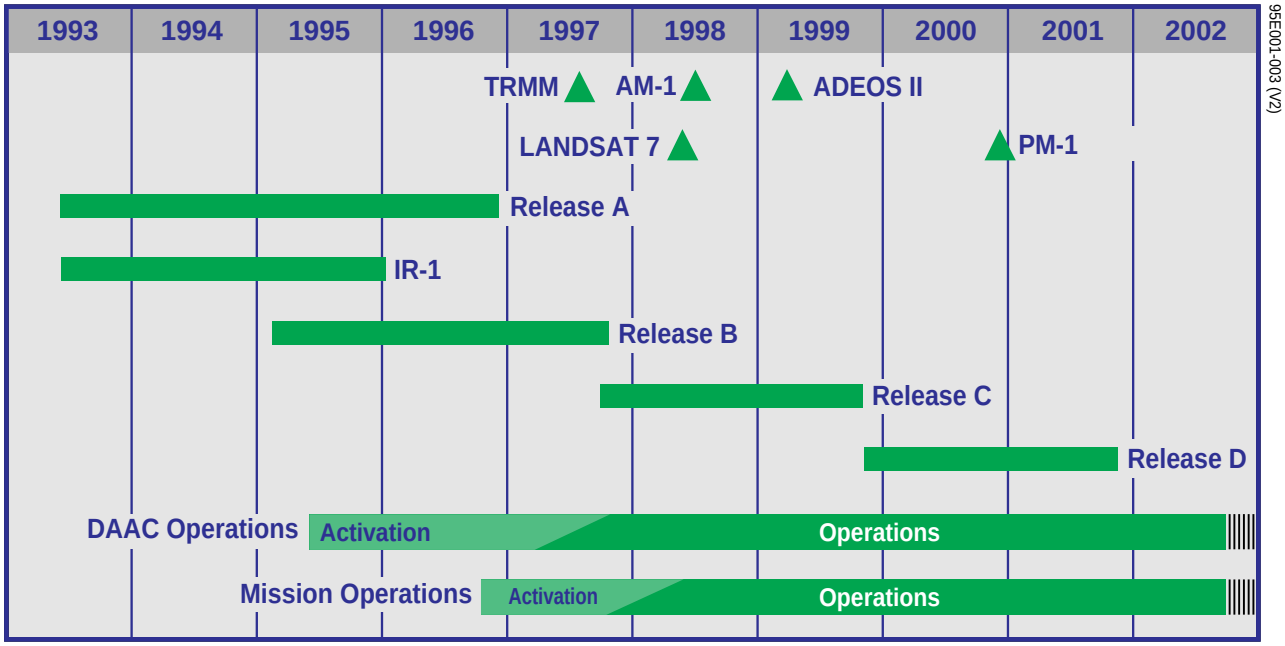


Figure 4-1. ECS Development Schedule with Example EOS Missions

Release A will support data functions (processing, archiving, data search and access, and distribution) for TRMM in addition to providing for flight operations interface testing and end-to-end data flow testing for EOS AM-1. Release A includes the algorithm development toolkit to support transition of algorithm software developed by the science community into the ECS DAACs. Release A will also include an Interim Release (IR-1) to support early interface testing and initial algorithm integration and test.

Release B will provide flight operations for EOS AM-1 and data functions for EOS AM-1, Landsat 7, and ADEOS II. Releases C and D will support future EOS missions, such as EOS PM-1, and will incorporate evolutionary changes such as new processing and storage technologies. Successive releases will provide expanded and increasingly enhanced data search and access, based on feedback from the science community.

Computer technology projections indicate significant benefit can be achieved by delaying hardware purchases until needed for mission support. For example, Figures 4-2a and 4-2b show ECS projections of cost per theoretical MFLOP (processing) and cost per gigabyte (on-line disk storage) over the next 5 years. These projections are based on information provided by major vendors of computer hardware and are updated by ECS as better information becomes available. Such technology projections suggest a "just-in-time" purchasing strategy to maximize computer resources for available funding. Also, some commercial software technologies are not expected to be mature enough for use in early EOS missions and will be incorporated as evolutionary improvements in later EOS missions.

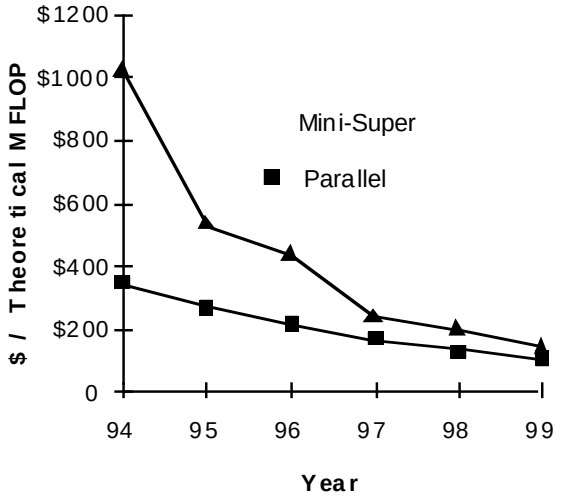


Figure 4-2a. Technology Projection of Cost Per MFLOP (Processing) as of April 1994.

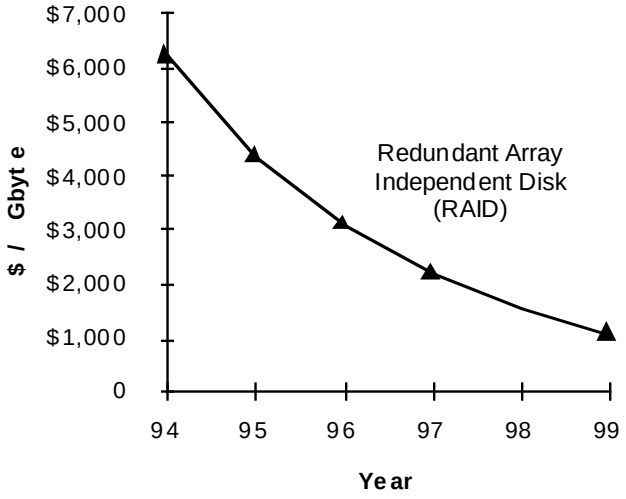


Figure 4-2b. Technology Projection of Cost Per Gbyte (Disk Storage) as of April 1994

4.2 Specialized Skills and Teamwork

ECS is to be developed and operated as a science-driven, albeit cost-constrained, infrastructure to service the EOS science community. To be effective, communication must flow among NASA, the ECS team, and the science community who possess highly specialized skills in diverse disciplines. Examples of EOSDIS functional components and required skills are shown in Table 4-1.

Table 4-1. Examples of EOSDIS Functional Components and Required Skills.

EOSDIS Functional Components	Required skills
Develop and operate science-based data and information system	Expertise in science information systems and research applications
Develop and operate EOS instrument processing	Expertise in algorithm Integration and routine product generation.
Develop and operate distributed Communications	Expertise in local area and wide area network technologies
Develop and operate network management and system coordination	Expertise in network and system coordination technologies
Develop and operate spacecraft and instrument control systems	Expertise in on-orbit flight operations

The ECS team is principally populated by engineers with experience in developing large-scale information and flight operation systems. Members of the science community use computer technology as a tool in their science research but generally have not developed or operated end-to-end systems that integrate information processing and flight operations. To build a bridge across the engineering and science disciplines, a science office exists within ECS to facilitate the bi-directional transfer of information between ECS and the science community. NASA, as project manager, has both engineering and scientific expertise to provide requisite guidance and direction. Effective teamwork among all groups will ensure that science requirements are successfully translated into correct engineering implementation and the implications of engineering decisions are understood by the ultimate end user, the scientist. The process for involving the science community in ECS is discussed in Chapter 5 - Science Community Involvement.

4.3 Multi-Track Development

The ECS Team is developing ECS using a multi-track development approach that includes 1) development of a portion of ECS on an incremental track and 2) parallel development of the remainder of ECS on a formal track using the traditional waterfall development methodology. The two primary drivers for development on the incremental track are volatility of user-sensitive requirements and commercial-off-the-shelf (COTS) intensive integration. To accelerate and accommodate early user feedback, a delivery mechanism called an Evaluation Package (EP) was devised to put incremental developments and selected prototypes in the hands of distributed users for evaluation and design iteration significantly in advance of formal track releases. Another purpose of the Evaluation Package is early integration of COTS hardware and software in order to evaluate advertised capabilities of commercial vendors.

The key to successful development on the incremental track is to provide structure without creating an administration overload that removes the freedom to react to objectives and design changes dictated by emerging circumstances, such as programmatic changes or technology evolution. To meet this challenge, we have adopted an Evaluation Package life cycle that merges selected practices from more traditional engineering methods with rapid prototyping methodology. For example, an objectives review is held with NASA and science community representatives at the beginning of each Evaluation Package to establish common understanding of design and evaluation goals. Other reviews with NASA and science community representatives include design, test readiness, consent to ship, and final readiness reviews. At each review, status and lessons learned are discussed and changes incorporated based on feedback by review attendees. Mockups, early prototyping, and in-process demonstrations give reviewers progressively better insight into the planned Evaluation Package functionality.

After final integration and testing, the Evaluation Package software is then distributed to the DAACs and other evaluator sites for a multi-week evaluation. Also, science community representatives are invited to participate in structured usability testing at the ECS development facility in Landover, MD. Usability testing is an efficient and low cost method of testing and

quantifying ease of use. Experience to date indicates that the minimum time to produce meaningful content in an Evaluation Package is about 6 months, and that evaluation will require an additional 2 months including time for data analysis and results sharing.

4.4 Key ECS Development Challenges

Currently, there are six key categories of ECS development challenges that could impact user services: 1) programmatic, 2) user interaction, 3) processing and storage demand, 4) database technology, 5) communications technology, and 6) storage technology. A successful ECS development program requires a balanced approach that addresses all of these challenges with carefully planned mitigation strategies.

1) Programmatic: The key programmatic challenge is compression of the ECS development schedule based on delayed start of ECS contract, early launches and extensive rework of the ECS architecture based on science community and NRC recommendations. Mitigation includes reallocation of functionality to releases, detailed definition of release content, attention to hiring and training, and taking advantage of incremental development and integration.

2) User interaction: The challenge is to model the number and activity of users. Accurate prediction of user activity is difficult for many reasons including the anticipated evolution of technology and user behavior. Mitigation includes refining the model of user activities, improving feedback from the science community, and design studies to reduce the sensitivity of the design to variations in the actual number of users.

3) Processing and storage demand: Challenges include processing and storage demand for standard products. Accurate understanding of science algorithm processing and storage requirements is needed to anticipate planning constraints and properly size data processing and storage resources at the DAACs. Mitigation includes working with algorithm developers, designing for scalability, and prototyping of alternative algorithm processing architectures.

4) Database technology. The key challenge is relational database management technology. The issue with existing technology is that current relational database systems are predicted to be inadequate for spatial, temporal & coincident searching. Mitigation includes ongoing assessment of current, and forecast, vendor products; prototyping; and monitoring of community research in database management technologies.

5) Communications technology: Challenges include the maturity, breadth, performance and evolution of the framework for distributed computing and the resultant integration and interoperability of commercial software products. Mitigation plans include prototyping, evolution planning to isolate applications from potential infrastructure changes, and keeping abreast of, and influencing, market place development through participation in various standards organizations.

6) Storage technology: Challenges span three areas of storage technology: cost effective storage; hierarchical storage management for ECS-class supercomputing applications; and scalability and maintainability of archives.

The issue with cost effective storage is the choice of archive components, particularly media and form factor for petabyte storage. Mitigation includes prototyping as well as keeping informed of developments in the storage technology marketplace and large volume applications such as found in DOD, Oil, and other science (e.g. CERN) applications.

The issue with hierarchical storage management (HSM) is that current systems are projected to be bottlenecks when scaled to EOSDIS-size applications. Mitigation includes partitioning the storage design into scalable, low-risk components that avoid bottlenecks for the archive and using non-HSM file management techniques

The main issue associated with scalability is that a single computer cannot handle an arbitrary I/O load especially in support of user-related services. A major mitigation is the investigation of network-attached storage. For maintainability, mitigation includes using proven robotics technology and pursuing the use of non-contact media technologies.

4.5 Current Access to Selected NASA Data Sets

EOSDIS Version 0 is the initial implementation of EOSDIS as a working prototype system that allows users to search for, and order data from, several DAACs in a single session. Through interconnection of the existing DAAC information management systems, Version 0 serves as a functional prototype of selected EOSDIS services. As a prototype, it does not have all the capabilities, fault tolerance, or reliability of later versions; however, EOSDIS Version 0 supports use by the scientific community in day-to-day research activities. Such use tests existing services to determine what additional or alternative capabilities are required of the full EOSDIS. The current Version 0 functions are data search and order, geographic coverage of selected data on an orthographic projection of the Earth, reduced resolution images for browsing the data, and detailed information on the selected data to be placed for order processing.

Although EOS data products are not yet available, data from the Pathfinder program are currently being used by researchers. The Pathfinder data are research-quality global change data sets available to Earth scientists through Version 0. Large remote-sensing data sets applicable to global change research have been developed from existing global and/or regional data sets. Higher level geophysical products are derived from peer-reviewed algorithms. The Pathfinder data sets have met stringent quality assurance and access requirements such as stable calibration of the raw data, consistent intercalibration among different instruments and any necessary archiving to a more accessible medium. Version 0 will evolve towards next generation EOSDIS by ECS taking maximum advantage of existing experience and by ensuring that no disruption occurs in services to current users.

4.6 Evolution of EOSDIS

Over the EOSDIS lifetime (at least two decades beyond the launch of the first EOS spacecraft), evolution will come from at least three separate sources: 1) Scientific needs will change as Earth system science matures and new applications of the data emerge; 2) Information system technologies must be refreshed as maintaining older technologies becomes more difficult and new technologies displace them; and 3) Changes in the information infrastructure (e.g., high

bandwidth networking) will lead to migration of functions to take full advantage of these capabilities. A key ECS goal is to provide a highly adaptable infrastructure that is responsive to the evolving needs of the Earth science community. Table 4-2 highlights ECS architectural features and associated benefits that facilitate evolution of EOSDIS.

Table 4-2. Highlights of ECS Architectural Features that Facilitate Evolution.

ECS Architectural Features	Benefits that Facilitate Evolution
Distributed search and access - Inter- and intra- DAAC services - Integration of additional DAAC-unique and user supplied search services	One-stop data search and access
Advertisements of new services - New and modified product sets - Submitted by producers	Extensible product set
Logical collections - Data modeled on Earth Science data taxonomies - Context and relationships captured for items within a collection	Information-rich logical data collections
Integration of investigator tools - Interoperability via standard data and control exchange protocols	Integration of independent investigator tools
Transparent access to system-wide resources: - ECS Client applications - Advertising and Subscriptions Services	Transparency and location independence
Search and access across multiple heterogeneous servers - Multiple protocols, including WAIS, and WWW	State-of-the-art protocols
Users can combine ECS services in arbitrary ways - Defined interfaces and associated protocols	User-tailorable services
Interoperability with heterogeneous systems - Standard or negotiated browse and data retrieval formats and protocols	International interoperability
Layered services - Minimize impact of technology insertion - Graceful evolution	Facilitate technology upgrades
User access to core system services - API toolkit - Customize user interface	Build favorite user interface
Toolkits - Software libraries, documentation, configuration tools, and APIs - Develop, insert and test new methods at their local site	Local development and testing of new methods

To facilitate evolution a broad range of strategies were used in defining the ECS architecture such as isolating changes to ‘affordable’ component replacements or modifications and separating fast changing components, such as user interface browsers, from slowly changing components, such as the archive storage. Other architecture features that support evolution

include location and access transparency, as well as avoidance of dependence on vendor computer hardware and system software. Use of layered, hierarchical services allow the application of state-of-the-art client/server topologies. Application Programming Interfaces (API) are a set of library routines between ECS components that allow program-wide access to particular ECS services. APIs are used to accommodate development of DAAC-unique interfaces such as special subsetting capabilities and new search techniques which can be advertised as value added services. APIs allow scientists to apply user-supplied methods to manipulate ECS data products from within an ECS query or executed from a user's own computing facility as a machine-to-machine process. The above architectural features, in addition to object-oriented design methodology, result in a design that is state-of-the-art in the use of technology and adaptable to evolutionary change.

Suggested Reading

ECS, 1995: 301-CD-002-003, System Implementation Plan for the ECS Project (1/95).
URL: http://edhs1.gsfc.nasa.gov

Elkington, M., 1994: 19400131 Defining the Architectural Development of EOSDIS.
URL: http://edhs1.gsfc.nasa.gov

NASA, 1994: EOSDIS VO IMS Home Page.
URL: http://harp.gsfc.nasa.gov:1729/eosdis_documents/eosdis_home.html/

Schwaller, M. 1993: Science Data Plan for EOSDIS Covering EOSDIS Version 0 and Beyond.
NASA/GSFC. Anonymous ftp host: eos.nasa.gov.
Change directory to /EosDis/Daacs/Docs/ScienceDataPlan.

This page intentionally left blank.

5. Science Community Involvement

Development of science-based services must have science user involvement and users have directly influenced development of ECS through several processes. Interaction with users occurs across a broad range of EOS organizations including the Investigator Working Group, the EOSDIS Data Panel, ECS science advisors, technical working groups, DAAC User Working Groups, and the general science community. Users participate in major design reviews; critique design considerations; collaborate in working group sessions; and submit on-line suggestions. The following sections describe how users are involved in ECS development at various technical levels, and how user feedback and suggestions have influenced the direction of ECS.

5.1 Users Involvement at Technical Levels

A broad spectrum of the user community provides guidance at all technical levels from high level policy decisions to prototyping. The highest level of guidance is provided by the Investigator Working Group and the associated panels, particularly the EOSDIS Data Panel which is composed of prominent scientists with strong data management backgrounds. The Data Panel meets quarterly to review EOSDIS policy, plans, and design, and provides recommendations on where EOSDIS should focus in the near- and long-term. Recommendations are reviewed, analyzed, and incorporated into EOSDIS planning under NASA technical direction.

Guidance at the technical level is also provided by the ECS Science Advisors. These science community representatives, sometimes referred to as ECS "tire-kickers", participate in reviews, evaluate ECS prototypes and evaluation packages and provide comments on a technical level for various issues as requested by NASA. The Science Advisors produce recommendations, evaluations and feedback about details of the technical design and implementation that are incorporated into future design changes.

Technical working groups are made up of members of the science community who work closely with NASA and ECS team members to address specific technical challenges. For example, the Ad Hoc Working Group on Production was formed to address issues associated with data processing, storage, and distribution capacities of EOSDIS. The Data Modeling Working Group was formed to address issues on information architecture and modeling. Technical working groups are formed as needed and include NASA, science community and ECS representatives.

The DAAC User Working Groups consist of science users with expertise or interaction with a specific DAAC. They provide guidance on DAAC plans, budgets and operations. The DAAC Scientist is the primary interface between ECS and the DAAC User Working Groups.

A broad spectrum of the science community is represented at the ECS reviews, where guidance is provided at all technical levels. Reviews focus on everything from EOS policy to specific implementation of design; Earth science community members are invited to review and comment on issues presented. Critiques are submitted by science community members, and ECS

responds to each critique by stating how the issue will be handled in the future. This process helps to guide work at all technical levels, and requires that ECS be responsive to the science user.

5.2 Influence of User Feedback on EOSDIS

The direction and design of EOSDIS has been strongly influenced by the user community beginning with Phase B design studies. Recommendations by the EOSDIS Data Panel during Phase B led to the adoption of a distributed architecture with multiple DAAC providers. Critique of the ECS architecture presented at the System Requirements Review in September 1993, by the National Research Council and the EOSDIS data panel, resulted in a revised architecture that is now more open and evolutionary. Specific technical characteristics of ECS have been influenced through ongoing technical interactions, disposition of review and documentation critiques, and the general user suggestion process.

Working groups, for instance, have improved design aspects of ECS data processing, storage and distribution. They also helped define metadata standards and browse packaging. Over 2,000 critiques have been received from members of the science community and others who attended major design reviews, many of which have been incorporated into technical level plans. Because ECS developers answer critiques while they are developing technical plans for the next level of design, many of the suggestions are incorporated into their plans. For instance, suggestions that ECS find an efficient means of serving the non-science community led to the development of the value added service provider concept.

ECS also provides scientists with an 'on-line' suggestion box for collecting, reviewing, and dispositioning user recommendations for the design of the ECS system. To date, over 600 recommendations have been entered into the system: 65% either match existing requirements or are being factored into ECS design considerations, 31% are in evaluation, and 4% rejected because of technical or cost considerations. These entries have varied from questions and comments about the ECS design to feedback on prototyping to recommendations on new design requirements.

An example of the way users have affected the ECS design is evident in the addition of a new requirement to support coincident searching. After review, assessment , and approval ECS now has a requirement to "search across multiple data sets for coincident occurrences of data in space and/or time and any other attribute(s) of metadata."

Currently, user recommendations are being evaluated for incorporation into the ECS design as part of the lower-level requirements definition process. Many of the recommendations are suggestions on how the user community would like to use the system. They illustrate typical operations the user will need, enhancement of Version 0, or special options which would ease the user's job of analyzing the science data. This feedback is proving to be an invaluable resource for optimizing the ECS design.

The many ways that users have to influence the design of ECS have resulted in a system design that has evolved from user feedback. As design and development proceed, these processes will continue to influence the development a system that will meet user needs.

Suggested Reading

Corbell, M., 1994: MR9407V3 User Environment Definition for the ECS Project.
URL: http://edhs1.gsfc.nasa.gov

Tyahla, L. 1994: 19400313 ECS User Characterization Methodology and Results.
URL: http://edhs1.gsfc.nasa.gov

Tyahla, L. 1994: 19400548 User Scenario Functional Analysis. URL: http://edhs1.gsfc.nasa.gov

West, S., 1994: 19400549 ECS Scientist User Survey (ESUS). URL: http://edhs1.gsfc.nasa.gov

Wingo, T. 1994: 19400312 User Characterization and Requirements Analysis.
URL: http://edhs1.gsfc.nasa.gov