1986

Earth Observing System, volume IIa: Data and Information System, Report of the EOS Data Panel

Raymond Arvidson
Washington University

Frederick Billingsley
Jet Propulsion Labortory

Robert Chase
Woods Hole Oceanographic Institution

Pat Chavez Jr.
United States Geological Service

Michael Devirian
NASA Headquarters

See next page for additional authors

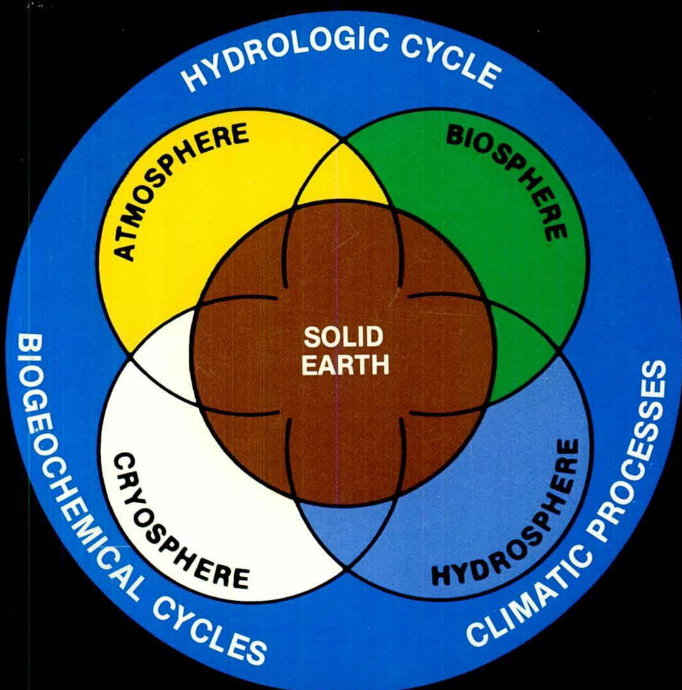
Authors
Raymond Arvidson, Frederick Billingsley, Robert Chase, Pat Chavez Jr., Michael Devirian, Frederick Mosher, and et al.

NASA-TM-87777 19860021622

Volume II a

EARTH OBSERVING SYSTEM

DATA AND INFORMATION SYSTEM



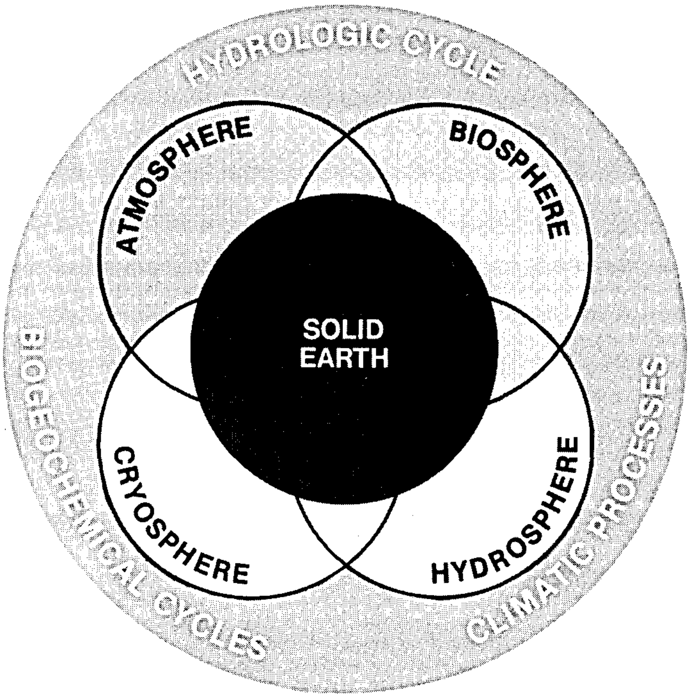
REPORT OF THE EOS DATA PANEL

NASA
National Aeronautics and
Space Administration

Volume IIa

EARTH OBSERVING SYSTEM

DATA AND INFORMATION SYSTEM



REPORT OF THE EOS DATA PANEL

NASA
National Aeronautics and
Space Administration
1986

Technical Memorandum 87777

EARTH OBSERVING SYSTEM REPORTS

Volume I	Science and Mission Requirements Working Group Report
Volume II	From Pattern to Process: The Strategy of the Earth Observing System Science Steering Committee Report
Volume IIa	Data and Information System Data Panel Report
Volume IIb	MODIS Moderate-Resolution Imaging Spectrometer Instrument Panel Report
Volume IIc	HIRIS & SISEX High-Resolution Imaging Spectrometry: Science Opportunities for the 1990s Instrument Panel Report
Volume IId	LASA Lidar Atmospheric Sounder and Altimeter Instrument Panel Report
Volume IIe	HMMR High-Resolution Multifrequency Microwave Radiometer Instrument Panel Report
Volume IIf	SAR Synthetic Aperture Radar Instrument Panel Report
Volume IIg	LAWS Laser Atmospheric Wind Sounder Instrument Panel Report
Volume IIh	Altimetric System Panel Report

DATA PANEL FOR THE EARTH OBSERVING SYSTEM

Raymond Arvidson, Washington University
Frederick Billingsley, Jet Propulsion Laboratory
Robert Chase, Woods Hole Oceanographic Institution
Pat Chavez, Jr., USGS Flagstaff
Michael Devirian, NASA Headquarters
John Estes, University of California at Santa Barbara
Gregory Hunolt, NOAA Satellite Data Services Division
J. Charles Klose, Jet Propulsion Laboratory
George Ludwig, Consultant
Frederick Mosher, NOAA National Severe Storms Forecast Center
William Rossow, Goddard Institute for Space Studies

EXECUTIVE SUMMARY

The purpose of this report is to provide NASA with a rationale and recommendations for planning, implementing, and operating an Earth Observing System data and information system that can evolve to meet the Earth Observing System's needs in the 1990s. The Earth Observing System (Eos), defined by the Eos Science and Mission Requirements Working Group[1], consists of a suite of instruments in low Earth orbit acquiring measurements of the Earth's atmosphere, surface, and interior; an information system to support scientific research; and a vigorous program of scientific research, stressing study of global-scale processes that shape and influence the Earth as a system. The Eos data and information system is conceived as a *complete* research information system that would transcend the traditional mission data system, and include additional capabilities such as maintaining long-term, time-series data bases and providing access by Eos researchers to relevant non-Eos data. The Working Group recommends that the Eos data and information system be initiated now, with existing data, and that the system evolve into one that can meet the intensive research and data needs that will exist when Eos spacecraft are returning data in the 1990s.

Eos SCIENTIFIC AND OPERATIONAL ENVIRONMENT

Meeting the Earth science objectives of the 1990s, delineated by the Eos Science and Mission Requirements Working Group, requires in many cases access to data from a number of instruments, from *in situ* observations, and to data that have been placed in time-series data bases that extend over a decade or more. In that sense, Eos as a program will clearly transcend the traditional NASA flight-project model, which stresses spacecraft observations over a selected, and usually relatively short, period of time, and includes initial analyses of the acquired data. Thus, major issues for Eos and its data and information system include how Eos will operate both as a flight project and as a long-term scientific research program.

We envision Eos instrument teams that will guide instrument and algorithm development and will be involved in mission operations and scientific analyses of the data. We also envision that consortia of researchers, many stressing multidisciplinary analyses, will form and intensely work with Eos data for extended, although not indefinite, periods of time. In some cases, these research groups will request that special sets of observations from a variety of sources be acquired to meet their research needs. Many Eos researchers will need efficient access to relevant non-Eos data bases and to a variety of models to conduct their research. We envision many Eos researchers being supported under ongoing NASA Earth science programs associated both with Eos and with other global Earth science programs.

We therefore recommend that the Eos data and information system serve both the Eos project and the more general NASA Earth Science and Applications program needs, providing needed continuity to receive data from other sources (e.g., the Upper Atmospheric Research Satellite, the Ocean Topography Experiment spacecraft) and from the Eos spacecraft, to deliver these data to temporary repositories for conversion to physical units, and to support long-term archives where the data will be accessible by and available to the greater scientific community.

Mission operations and standard data processing tasks should be tailored to acquire and produce viable scientific data for a given instrument. The initial data should be housed in temporary mission or instrument repositories for primary analyses. These data should then be provided to data centers for long-term maintenance and access by scientific researchers. In a number of cases, subsets of Eos and relevant non-Eos data should be maintained at active data base sites to support specific research tasks (active data bases are subsets of data that are being routinely used in research and are under direct control of a given research group). The resultant, highly processed data sets and associated documentation should migrate back to the data centers once the specific, chartered research tasks of an active data base have been completed. The core of the Eos data and information system should be an electronic information network, allowing access to the full suite of Eos system capabilities. This network should be flexible, providing access to mission operations, archives, selected active data bases, and to, for example, large mainframe computers for certain, very intensive, computational activities (e.g., modeling) needed to support Eos data analyses.

There will be at least four major types of Eos data and information system users:

1. The first type includes instrument team members and support personnel, associated with Eos instrument or mission operations centers. They will need to monitor continually a sampling of data in near-real time for quality assurance, error detection, and instrument malfunction assessment. They should have the capability to reconfigure observational sequences when malfunctions or special events occur. We do not expect that all personnel will need to be resident at the centers, since the Eos information network is envisioned to provide remote, electronic access to the centers' functions.

2. A second group of users will consist of researchers, instrument team members or operations-oriented personnel who need instrument-specific, near-real time, or even real-time data processing, delivery, and display capabilities. In some cases (e.g., NOAA or DoD) large data volumes will likely be required.

3. The third user group will consist of researchers who will need to interrogate directories and catalogs of Eos and other relevant data on an instrument, geographic location, and time of acquisition basis. They will also need to order data, and in some cases they will request that particular observational sequences be acquired with Eos instruments.

4. A fourth group of researchers also will need to interrogate directories and catalogs of Eos and pertinent non-Eos data but is distinguished from the third group by a need to browse Eos data visually via attributes, or preferably via expert systems (to find particular features, attributes, or special cases).

In many circumstances, ranging from pipe-line data processing activities to researchers dealing with Eos-data archives, access to non-Eos data will be crucial to meeting processing and analysis requirements. Thus, the Eos data and information system should provide a directory to the catalogs of pertinent non-Eos data and, in some cases, to the catalogs, data, and processing algorithms, per se.

RECOMMENDATIONS FOR THE 1990s

The Eos data and information system must be designed to meet the challenges of Eos mission operations, transparent data access, transmission, processing, and maintenance, as well as those imposed by the needs of scientists for non-Eos data sets. The system should also accommodate "operational" users (e.g., NOAA) if their activities do not detrimentally impact the conduct of Eos scientific tasks. If operational uses are to be made of Eos data, and if these users significantly affect the information system, then those operational users or agencies should be expected to provide the system enhancements needed to meet their specific requirements.

We delineate within this report a number of functions that an Eos data and information system of the 1990s should fulfill. These functions do not constrain physical location of personnel, data, or processing capabilities. In the 1990s, we expect that local processing capabilities, combined with modern network capabilities will allow a geographically distributed information system to become a reality. In fact, the goal of the data and information system should be to provide the scientific community with remote electronic access to the variety of capabilities and services that Eos will provide.

We recommend that the Eos data and information system of the 1990s be designed to accommodate a suite of required functions, including:

1. Eos flight system functions, including the recommended functions and characteristics of both remote sensing instrumentation and onboard data systems.

Flight systems functions are presumably more closely akin to those dealt with in previous missions. There are, however, new requirements being placed into this category, requirements consistent with both evolving research methodologies and developing technologies. Specifically, given the large data volumes and rates envisioned, there will be a need for expert systems, automated command and control, transparency in command control, rapid response, onboard monitoring, significant onboard buffers, etc. Technically, the flight systems aspects of Eos may be quite challenging.

2. User functions, which embrace the likely ways in which researchers, operations personnel, instrument scientists and teams, and other individuals requiring access to Eos services will operate.

User functions as a characteristic grouping are unique to Eos. In general terms, we envision the users of Eos and its data and information system playing a significantly greater role than in the past. Researchers should be tasked with the responsibility of ensuring the overall functional capability of the system, presumably in an oversight and advisory capacity. Scientists should be obligated to return to the archives reduced or derived data sets resulting from their research. Thus, the communications of data and information will be bidirectional and not merely a one-way conduit with the researcher at the end of the system.

3. Operational functions, including those inherent in spacecraft and data and information system operation.

Operational functions, like flight systems functions, will bear some resemblances to previous missions. Since we anticipate that Eos will operate during the same timeframe as the Space Station, it is likely that some subset of the operational requirements may be met by activities not under the direct control of Eos, the research project. As an example, the bidirectional communications link between spacecraft and ground facilities will undoubtedly be operated and managed by non-project personnel. There are a host of other operational requirements that will not be provided by other activities and must be satisfied within the confines of Eos per se. These requirements include quick-look data production for quality assurance purposes, command management, quality assurance, access services, near-real-time processing, etc.

4. Eos information services functions, encompassing the suite of network services (both space- and ground-based) that are required.

Information network services functions are envisioned to be far more extensive and comprehensive than those employed in previous experiments. System transparency while accessing geographically dispersed, heterogeneous data archives will be a fundamental requirement. We believe the primary goal of the system should be remote, electronic access to the host of services and capabilities afforded by Eos. These include mission and instrument operational planning; access to directory, catalogs, and data bases; a spectrum of communication services; management of inter- and intra-system information flow; access security, etc. Provisions must be made to ensure maximum flexibility in data management and communications network design, accommodating needs of future researchers.

5. The functional requirements characteristic of advanced data base management practices.

The requirements for advanced data base management functions are probably more extensive than for the other functional groupings. They may be subdivided into electronic directory and catalog, browse file, data ordering, documentation, data archives, and standards requirements and functions. The prime purpose of incorporating extensive advanced data base management capabilities is to maximize scientific efforts and exploit Eos capabilities to their fullest. The requisite functions within this group are designed to allow scientists to focus on research, rather than on the details of accessing and preparing data for analysis.

6. Eos data processing functions, tailored to meeting a spectrum of requirements imposed by the four major types of Eos data and information system users.

Data processing requirements will be highly varied. They range from quick-look processing for instrument and mission operational personnel, to routine preprocessing to Levels 0 and 1A, to higher level data reductions, to mainframe modeling capabilities. Near-real and real-time processing will be required for operational purposes and may be of value for interactive browse activities. Processing capacity and performance should be capable of flexible enhancement to meet evolving needs defined by instrument teams. Data reduction, grid overlay, standard projection data set production, and data set merger will be required by the majority of Eos researchers. Throughout these data processing activities, self-documenting software will be needed to produce inventory, catalog, and directory entries.

7. Functional requirements for non-Eos data bases consistent with overall Eos scientific objectives.

The principal function associated with non-Eos data bases is straightforward, transparent access. Ideally, this would provide the researcher with a "one-stop shopping" capability. Through close collaboration of archival personnel and advanced network capabilities, a researcher could explore data catalogs, browse data sets, select pertinent data, and order these data, all from the same facilities that he customarily uses for similar services from Eos archives. Although the major difficulties that will be encountered are political, every effort should be made to provide these functional capabilities since the success of Eos will be largely dependent upon a researcher's ability to utilize all relevant data in his work.

Eos will be a unique space-research program; unlike many, the functional elements of its data and information system are highly interdependent. Responsibilities residing within one element or category may require contributions from one or more additional elements for completion. *Thus, for Eos to be successful, all of the requirements and systems' attributes delineated within this report (and briefly summarized in an appendix) must be provided.* Furthermore, to ensure that the resultant system can meet researcher's needs, the architectural design should feature and utilize two fundamental principles or contemporary design techniques throughout. These principles are "layering" (the technique of dividing and conquering that produces modularity) and "standardization" (of formats and protocols, which promotes data autonomy, hence transparency). Together, they can be used to create a data and information system that is adaptively flexible, transparent to the user, and robust, providing the foundation for diverse and evolving data processing and archival needs of Eos researchers and operational users.

RECOMMENDATIONS FOR THE 1980s

The key to successfully implementing the recommendations for the 1990s rests squarely with the care, planning, and attention devoted to initiating an Eos data and information system during the 1980s. The requisite information system should be evolutionary, supporting ongoing and near-term NASA programs in the Earth sciences with existing data. We recommend that the Eos data and information system be built upon the experience gained from NASA's existing pilot data system projects (i.e., Pilot climate, Pilot Ocean, and Pilot Land Data Systems) with a focus toward Eos data and information system objectives. We further recommend close collaboration and coordination with the Unidata initiative of the University Corporation for Atmospheric Research (UCAR), and utilization of the Global Resource Information System philosophy as a unifying concept.

The pilot projects and Unidata address discipline-unique problems. On the other hand, the Global Resource Information System concept stresses access to data from each of these pilots and from other relevant NASA and non-NASA data bases. In essence, a major task that can and should begin now is facilitating access to the geographically dispersed, heterogeneous data bases that already exist within these pilot projects and at a number of university and agency locations.

Development of the information system should be done in close collaboration with scientists supported under NASA's Earth Science and Applications Division programs. In doing so, researchers would use the network and access, for example, the pilot data systems, thereby exercising the overall embryonic system and allowing NASA and its research community to gain experience appropriate to establishing a data and information system well suited to the Eos spacecraft era. We therefore recommend that the Earth Science and Applications Division fund a limited number of multidisciplinary studies and discipline-oriented research teams (requiring multiple data sources) that can begin to address specific Eos scientific objectives in biogeochemistry, hydrology, and climatic research, as well as in specific disciplines requiring multiple data sources. These teams will provide scientific focus for developing the Eos data and information system, including developing detailed requirements for needed data sets, for the data base management and network environment, and for evaluations of whether the evolving information system adequately meets scientific needs. In developing the information system, particular emphasis should be placed upon technology associated with flexible and transparent access to dispersed, heterogeneous data bases, to advanced data base management techniques (including search capabilities employing expert systems), and to cost-effective electronic network approaches (including direct broadcast of data).

The final, and perhaps pivotal recommendation of this panel is to initiate the planning and implementation of an evolutionary Eos data and information system without delay. A functional system providing the means through which Eos data will be fully exploited cannot be built in a matter of a few years; it must be encouraged and allowed to evolve in concert with the ever increasing knowledge base of the Earth sciences and with the requisite expertise to manage this data base. There are two imperatives or guiding principles which should be closely followed throughout this evolutionary process. They are (i) involve the scientific research community at the outset and throughout all subsequent activities, since the data will be acquired, transmitted, and processed for scientific research, and (ii) provide a representative group of active researchers with an oversight and review responsibility, since the most successful examples of data base management involve user oversight.

PREFACE

The sciences, by their very nature, are evolutionary. The questions posed and information needed to address specific problems (as well as the problems themselves) change through time as we learn more about a particular phenomenon or process; this is scientific progress. Similarly, the computer and communications industries have been and continue to experience significant technological progress year after year. With this comes inevitable change: change that when capitalized upon will lead to better and more efficient ways of conducting scientific research. A report such as this, which deals with scientific data and information systems, should not constrain a future system to today's innovations but rather should leave as its legacy a firm set of guiding principles and recommendations that will stand and remain valid long after its authors have accepted new challenges and well beyond the time when a specific piece of hardware or software has become obsolete. It will provide an approach and methodology for designing and building a flexible data and information system that is consistent with the evolving character of the sciences it seeks to serve. As well, its flexibility will allow new technological advances that loom on the horizon to be readily incorporated into its design. The Eos Data Panel has, I believe, succeeded in creating a report of this genre.

Clearly, a report of this breadth must draw quite heavily upon the expertise, experience, and knowledge of a large quorum of dedicated professionals. On behalf of the Eos Data Panel, I would particularly like to thank those individuals who graciously gave of their time and talent to contribute to this report. Alphabetically, they include:

Mark Abbott, Scripps Institution of Oceanography
Dixon Butler, NASA Headquarters
Jim Dodge, NASA Headquarters
John Dutton, Pennsylvania State University
Ed Greenberg, Jet Propulsion Laboratory
Dick Hartle, NASA Goddard Space Flight Center
Ed Hurley, NASA Goddard Space Flight Center
Tom Karras, NASA Goddard Space Flight Center
Ron Muller, NASA Goddard Space Flight Center
Rick Pomphrey, Jet Propulsion Laboratory
Stan Sobieski, NASA Goddard Space Flight Center
Jeff Star, University of California at Santa Barbara
Mike Ward, NASA Goddard Space Flight Center
Jim Weiss, Jet Propulsion Laboratory

To the host of other individuals who reviewed and commented upon various drafts of this report, our sincere thanks.

Robert R.P. Chase
Chairman, Eos Data Panel
Woods Hole, Massachusetts
October 1985

CONTENTS

I. PURPOSE AND SCOPE OF REPORT

The Earth Observing System Science and Mission Requirements Working Group was formed as an ad hoc advisory body to NASA's Earth Science and Applications Division and was charged with evaluating: (i) the potential for increasing our understanding of the Earth and how it works by utilizing systematic, multi-instrument observations from low Earth orbits, (ii) the role of *in situ* measurements, and (iii) requirements for programs of theory and data analysis to support scientific interpretations of these data. The Working Group delineated a number of high-priority, global-scale Earth science questions that should be addressed during the next decade. They then considered synergistic groupings of instruments to best meet these scientific objectives, the role of *in situ* measurements, and the need for a computation and data management environment that would facilitate research in an Eos era[1].

The Working Group pointed out that the data variety, complexity, and volume to be produced by Eos instruments would present major challenges in mission operations, data transmission, processing, and long-term maintenance of data that will ultimately constitute a portion of the time-series data bases. In fact, the Working Group considered that Eos as a whole should be thought of as an information system, whose development should begin now, with existing data. Through use of sound management principles and appropriate technologies, this information system could evolve to meet the needs of Eos during the 1990s.

The Working Group was succeeded by an Eos Science Steering Committee. This Committee provides a broad scientific oversight for the Eos Program and Project. The Steering Committee has, in turn, established a number of panels, one of which is the Eos Data Panel.

The Data Panel was charged with examining an Earth Observing System data and information system in more detail. This report, resulting from deliberations by the Eos Data Panel, is intended to provide guidance to NASA in developing the requisite Eos data and information system to meet the needs of the Eos research community throughout the next decade, including the era when Eos instruments will be acquiring data.

The Eos Data Panel began its deliberations by creating realistic scientific research scenarios (Appendix I) that closely parallel the scientific objectives set forth in the Eos Science and Mission Requirements Working Group Report. The Panel then reviewed the current state of and likely developments within the computer and telecommunications industries. This information, together with the lessons of history (embodied to a great extent within the National Academy of Science Committee on Data Management and Computation reports), forms the basis for developing the data and information system's scientific and operational environment (Chapter II). Based upon the requirements, attributes, and characteristics of this environment, information system design and architectural considerations (Chapter III) were determined.

Throughout this process, the Panel encountered many features that make Eos unique among flight projects. These unique features must be given careful thought and consideration by NASA management (Chapter IV) if Eos is to be successfully implemented. Finally, the Panel summarized its major findings and delineated an approach or implementation strategy for developing the Eos data and information system (Chapter V).

II. Eos SCIENTIFIC AND OPERATIONAL ENVIRONMENT

The Eos Data Panel developed requirements for an Eos data and information system operating in the 1990s by considering a number of possible research scenarios. Several of these scenarios are presented for reference in Appendix I. There are clearly stringent requirements, based on the scientific user scenarios, for determining what Eos, other NASA, and relevant non-NASA data exist, acquiring these data in suitable form, and having the capability to analyze the data. In addition, there are challenging requirements for mission operations, especially when operational or rapid response functions are considered. In the following pages we discuss requirements for an Eos data and information system designed to address those needs.

Eos DATA USERS

In determining the characteristics and attributes of an Eos data and information system, the intended user community must be identified, its access patterns anticipated and delineated, and the projected volume of requests estimated. We have identified four major classes of Eos data users:

1. Eos Operations Centers Users

In order to ensure the collection of complete data sets, free of instrument errors, timely delivery of data to Eos mission or instrument operations centers will be necessary. A sampling of data from all instruments should be processed and delivered to the centers in a format that is easily displayed and understood. These data will need to be continually monitored to ensure reliable collection. If problems occur, a center will need to have the capability to trace the data flow back through the processing system to the instrument, aiding in the isolation and correction of problems.

2. Instrument Specific, Real-Time Data Users

During the Eos spacecraft era, an instrument specific, real-time data user will require online data processing and display capabilities. Within this user group there are several subgroups. Operational users, such as NOAA or DoD, who will require entire, real-time data sets from specific instruments, constitute one subgroup. Their processing requirements will range from raw data to fully processed scientific data.

A second subgroup is composed of instrument teams or research groups selected from an Announcement of Opportunity. They may require control over the capabilities of various instruments to satisfy scientific requirements. In general, these groups will require quick-look processing of data from a specific set of instruments, over specific regions, at specific times. The processed information should be rapidly transmitted to these user teams so that it can be used in deciding, for example, the relocation of mobile, *in situ* measurement systems (e.g., aircraft, ships, autonomous platforms).

A third subgroup can be loosely termed the spectacular event monitoring group. This group responds rapidly to major scientific events, such as volcanic eruptions. It differs in function from the previous group in that the lead time for planning and coordination is greatly reduced. Consequently, it will require rapid decision-making capabilities in an Eos operations center and an established real-time processing and display capability that can be used to monitor an event.

The potential size of these groups can be estimated using existing organizations and funding as a guide. It has been assumed that there will be no major increase in the number of scientists or funding for the Earth sciences in general, except for the direct funding necessary to proceed with Eos. The only real-time operational users of Eos data requiring full, instrument data streams will probably be NOAA and possibly DoD. At any given time, there probably will be several groups (order 20) requiring quick-look data and a few (perhaps one or two) requesting instrument command control.

3. Users of Archival Data, Requirements Known

This group of users has known requirements for instruments, geographic areas, and times of data acquisition. It will require a catalog of data availability, with quality attributes of the data sets. The size of this group could be as large as 1,000 to 10,000 users. Of this group, approximately 50 to 200 will be active users at any given time. Each group member would be ordering 5 to 10 computer tapes per month. A few (1 to 10) of the active users will be ordering larger volumes of data. With the anticipated improvement in computer technology in the 1990s, it is expected that these users would want to order and process approximately 10 times as much data as they do currently.

4. Users of Archival Data, Attributes Known

This user group differs from the others in that it addresses scientific problems that require data with certain attributes (e.g., a particular type of cloud, a vegetation feature with certain characteristics). This will require a browse capability to select appropriate data.

There are two different types of browse requirements. One is a visual presentation of the data along with catalog listings of attributes. Well defined, standard data attributes within a discipline should be selected automatically and extracted for catalog entry during data processing. The attribute file should also be expandable, enabling user-identified attributes to be added. Similarly, as new algorithms are developed for identification of attributes, these algorithms should become part of the attribute file processing system (but old data should not be reprocessed to include these attributes without substantial peer and management review). The user community for these browse catalogues could be as large as 10,000 individuals.

The second browse capability required is a reduced volume data set suitable for custom processing by researchers. This data set will need to be available at a number of processing levels. Most users of these browse data sets will want highly processed versions of the data. The user community for this type of browse capability could be as large as 500 to 5,000.

DATA ACQUISITION

The concept of interactive, transparent system access should guide the creation of an environment in which scientists will operate when using the Eos data and information system. This concept includes the complete data acquisition (including archival data as well as new observational sequences) and analysis cycle, since it addresses both the need to transform a request into action (causing a remote system element to acquire data) and the need to deliver processed results to the requester. A researcher's system interface device may be as simple as a smart terminal or as complex as a large, mainframe computer, with the majority of scientists depending upon commercially available, microcomputer-based workstations. Access to in-flight subsystems should be provided through a coordination and control subsystem for all commands; those not requiring more than the allotted amount of critical system resources should pass through this subsystem transparently. Within a downlink telemetry access period, execution times for any flight system operating within the Eos data and information system should be in seconds, and those for archival data should be within minutes for small volumes, hours for medium volumes, and days for large volumes, depending upon the priority of the request. Thus, transparent access and rapid and timely responses are not only expected of the system, but also are a prerequisite.

Uplink Activities - Data Requests

Requesting observations from a flight instrument will involve decisions based on preset priorities and the cooperation of many simultaneous researchers and operational users. This necessitates the use of a master schedule that should be updated every orbit. The schedule should contain logistical information needed to plan an observation. Access to the schedule and the subsequent planning could be done electronically via a terminal, by formal proposal, or by agreement between collaborators. The system should utilize prompting algorithms for requests made electronically.

Uplink Activities - Activity Planning

Eos will utilize many strategies for data acquisition. Planning, of necessity, should be conducted interactively, utilizing project-provided software tools for design, integration and constraint assessment, and current resource summaries. These tools should allow command sequence design and integration, from simple requests to complex operations that require the coordination of several subsystems and instruments for multidisciplinary investigations. With these tools an experimenter could develop integrated command sequences that may be executed over long time periods, performing routine observations or providing for complex, critical observations in response to unique opportunities. The development of these plans could be interactively performed in concert with other investigators (via teleconferences) and with access to central archives of reference data.

Uplink Activities - Command

We envision several options available to the user, depending upon the desired operational scenario. Commands should be separated into categories that are noninteractive, interactive, and critical. Command execution could be in real time, near-real time, or preplanned. Interactive commands that require critical onboard resources should be processed at a ground-based control center, while noninteractive commands should be transparently transmitted from the user to the onboard system for final assessment and forwarding to the appropriate instrument or subsystem. Commands to be used in an emergency situation should have prior validation and be centrally stored for immediate delivery. Eos should utilize all of these operational capabilities.

Uplink Activities - Monitor and Control

During particular operations, necessary engineering and quick-look data should be provided to researchers for monitoring purposes. These data should be delivered in near-real time with delays of order seconds when the spacecraft is within a downlink telemetry reception period. For emergency commands, loop delays should not be more than 30 seconds between command delivery and received response. Investigators should have the capability to monitor their instrument from their home institution, and under certain conditions they may need command access. Access security must therefore be maintained at appropriate levels at all times.

Onboard Processing - Command

Onboard systems should be used for decisions when rapid response is needed. These situations may include changes of gain settings, filters, or data acquisition rates during anomalous conditions. These capabilities can be based on expert systems or software algorithms and could be used to plan interdisciplinary investigations. Systems of this genre could respond to short-term events in a coordinated, priority fashion, allowing scientists to take advantage of measurements that otherwise would not be acquired.

Onboard Processing - Function Monitoring

Flight instruments will require a certain level of onboard monitoring to ensure their functional capability and that of various subsystems. The areas of greatest importance are fault identification, isolation, and protection. Systems used for this purpose could be a combination of software algorithms and symbolics code. Many of the ground-based monitoring functions could be automated, tested on the ground, and eventually integrated with the spacecraft for onboard system control.

Onboard Processing - Data

Onboard data processing provides an opportunity to reduce the volume of information transmitted to the ground. Systems and algorithms that make decisions involving data acquisition strategies and record content should be used, thereby helping to maintain a reasonable downlink telemetry load. Techniques for data compression, editing, filtering, and refining data should be used for operational data sets. Certain levels of data reduction from onboard decisions may become a reality.

Onboard Processing - Telemetry

The onboard processing system should have the capability to format, error code, and buffer telemetry data, as well as the capability to merge critical ancillary data needed for quick-look and other analyses. This assumes that data collection would be performed by a scheme that does not require filler data to produce a standard record length, or set interrogation times.

Onboard Processing - Direct Broadcast

Direct broadcast data from NOAA environmental satellites have proven to be valuable for many users. Therefore, it is desirable that this kind of operation be continued from Eos platforms carrying operational instruments. It should be possible to receive directly transmitted data at any modestly equipped ground station that is within line-of-sight of an Eos platform. This is also an important capability for high data rate research instruments (e.g., the High Resolution Imaging Spectrometer or the Synthetic Aperture Radar), whose duty cycle may be constrained by the satellite's data recording system capacity and downlink availability. Most direct broadcast data should be produced in a raster format that can be readily interpreted.

Data Acquisition - Real-Time Monitoring

The Eos data system should have a real-time telemetry data processing system to monitor satellite and instrument status, and to confirm responses to uplink commands. Data for this purpose should be received in real time, perhaps by direct broadcasts from the satellite. Eos must maintain a mission operations center where this real-time data is processed and available to operations personnel. We envision that at least 95 percent of the real-time data broadcast by the satellite could be received, preprocessed, and displayed within 60 seconds from the time of transmission.

Data Acquisition - Data Records

The communications link between the Eos platform(s) and ground-based data reception facilities should be very stable. As a design goal, for any given orbit, all of the data transmitted by the space platform should be received in a form suitable for processing (e.g., as standard formatted data units) by the ground system, within 90 minutes of acquisition. To achieve this high percentage of data acquisition, it may be necessary to develop recall mechanisms to recover data lost during downlink transmissions. An accurate accounting of data received will be required.

Data sequence and encode peculiarities should be removed by the data reception system. Data that is transmitted from a recorder, in reverse order, should be rectified and presented in chronological sequence. Any encoding or data packet formation applied to data for downlink transmission must be removed (i.e., after the data have been preprocessed by the data reception system it should be in the same format that it was in when it left the instrument system).

For each unit of data (e.g., one orbit, one hour) preprocessed by the data receiving system, a catalog inventory record should be generated. This record should contain data record starting and ending times, data accountability information, etc. This information should be maintained by an Eos data catalog system.

Data Acquisition - Ancillary and Correlative Information

Both for research and quality assurance purposes, ancillary and correlative information will be needed. Much of this information will be derived directly from various spacecraft subsystems (e.g., attitude control, universal time) or through an advanced data collection and location system. In both cases, these data should be readily available to the Eos data and information system and provisions should be made to merge these data with Eos instrument data, as required.

Other types of ancillary information will be resident in non-Eos archives. Access to these data may be crucial to a particular research project or to understanding the characteristics and calibration of a given instrument. Consequently, provisions should be made to access this information as needed.

DATA PROCESSING

The full value of satellite-derived data cannot be realized unless careful thought is given to its processing, maintenance, and distribution. The Eos data processing system should:

1. provide near-real-time processing in support of field experiments, event monitoring, and quick-look scientific data analyses;

2. process data to a level that is readily usable, by applying algorithms that have been approved and validated by the research community; and

3. maintain and distribute data sets (i.e., both unprocessed and processed), ancillary information, and documentation that describes the processing procedures.

Eos data can be processed in the traditional sense by principal investigators or investigator teams at university-managed facilities, or by project teams at project-managed facilities. In either case the need for timely production of research-quality data sets is essential. All data processing requirements noted in this report are independent of specific organizational assignments and must be met by any facility (including commercially operated) that is producing Eos data.

Near Real-Time Processing

The Eos data processing system should support near-real-time data and information processing. These data will be used to support field experiments, monitor environmental events (volcanic eruptions, forest fires, floods, etc.), and provide the research community with an instrument quick-look capability. In general, quick-look information should be in the form of a graphics metafile or an image file that can be transmitted efficiently over medium-rate (9,600 bps) communications links and displayed on generally available graphics and image display terminals. To be of greatest utility, this information should be available within the time period of one orbit (nominally 90 minutes). Many of the near-real-time data sets will be selected prior to instrument deployment, but it should be possible to implement new data requests in response to evolving research needs and as new field experiment support requirements are identified.

Data Reduction - Processing Level Definitions

To facilitate the discussion of data processing, several processing "levels" are defined and used throughout the remainder of this report. These definitions are:

Level 0 - Reconstructed unprocessed instrument data at full resolutions.

Level 1A - Reconstructed unprocessed instrument data at full resolution, time referenced, and annotated with ancillary information, including radiometric and geometric calibration coefficients and georeferencing parameters (i.e., platform ephemeris) computed and appended but not applied to the Level 0 data.

Level 1B - Level 1A data that has been processed to sensor units (i.e., radar backscatter cross section, brightness temperature, etc.). Not all instruments will have a Level 1B equivalent.

Level 2 - Derived environmental variables (e.g., ocean wave height, soil moisture, ice concentration) at the same resolution and location as the Level 1 source data.

Level 3 - Variables mapped on uniform space-time grid scales, usually with some completeness and consistency properties (e.g., missing points interpolated, complete regions mosaicked together from multiple orbits).

Level 4 - Model output or results from analyses of lower-level data (i.e., variables that were not measured by the instruments but instead are derived from these measurements).

Level of Processing

A Level 1 data record is the most fundamental (i.e., highest reversible level) data record that has significant scientific utility, and is the foundation upon which all subsequent data sets are produced. Consequently, all

Eos data should be processed to at least Level 1. Given the volume of data involved and the likelihood that instrument teams and individual researchers will opt to process or have processed only select subsets of data, it is recommended that processing beyond Level 1A be handled on a request basis (except for the production of browse data sets). The basic processing should rectify the instrument data at full resolution, append calibration coefficients, geographic location parameters, etc. It is assumed that the accurate determination of calibration, Earth location, etc. will be performed by Eos instrument teams and will be the responsibility of instrument principal investigators. These tasks are critical to the successful utilization of Eos data and as such should receive significant attention to forestall any difficulties in producing data to Level 1A.

Level 2 is the first level that is directly usable for most scientific applications; its value is much greater than the lower levels. Level 2 data sets tend to be less voluminous than Level 1 data because they have been reduced temporally, spatially, or spectrally. Once verified, the value of Level 2 data remains high for a long period of time, declining only as newer data sets attract the interest of researchers.

Level 3 data sets are generally smaller than lower-level data sets and thus can be dealt with without incurring a great deal of data handling overhead. These data tend to be generally more useful for many applications. The regular spatial and temporal organization of Level 3 data sets makes it feasible to combine readily, data from different sources. The ability to produce combined or overlay data sets will greatly facilitate many scientific investigations. Therefore, it is recommended that any data from Eos instruments processed to Level 3 be produced in standard formats. Since the definition of a Level 3 data set may be application-dependent it may be necessary to produce several "standard" data sets or customized sets to meet particular needs.

In addition to Level 1A data, there will be components of the Eos data and information system that will be required to produce Level 2 and Level 3 data sets to meet scientific objectives established by Eos-sponsored scientific researchers or research teams. As each data set is processed, attribute and accountability information should be compiled (e.g., instrument name, data starting and ending times, processing level, algorithms used, number of records processed, percentage of data processed successfully) and appended. This information could also form the basis of an inventory record that should be maintained by an Eos data cataloging subsystem.

Browse Data Sets

It is recommended that a reduced volume data set be routinely processed to Level 2 for browse purposes. This data set should maintain the statistical properties of the original. A browse data set with a volume of one-tenth to one one-thousandth of the original set should provide these properties. The requirements of data set subsampling and spectral band sampling will need further study. However, it is suggested that several reduced-volume data sets be produced that are targeted for different scientific applications (e.g., from the moderate resolution imaging spectrometer), since the degradation of spectral resolution is discipline-dependent and in some cases instrument-specific. Spectral bands critical to each of the different scientific disciplines should be processed to identify specific data attributes (e.g., percent cloud cover, snow, soil, vegetation, water, scene average brightness, standard deviation, surface classification, time of day). Pattern recognition algorithms should be applied at this step to identify attributes that should be recorded in the browse file. One feature of these data sets is that they do not require external, ancillary data sources for processing. Finally, consideration might be given to producing browse files interactively, from a larger data set, as a researcher queries the system. Thus, it would be possible to add a variety of browse presentations to the total Eos holdings.

Instrument and Algorithm Performance

All Eos instrument parameters should be monitored through the periodic transmission of standard diagnostic data sets and by examining their performance statistically (e.g., compiling minimum, maximum, and mean values, and standard deviation for given measurements) or by generating an alarm if the value of a given measurement does not fall within a predicted range. This might best be done by examining Level 1 data and compiling instrument performance data for each unit of data processed. It is essential that the performance of each instrument be monitored, and that unexpected behavior be noted and evaluated. The history of an instrument's performance should be maintained in archives where it will be readily available to the research community.

Eos-derived environmental variables should be monitored in the same manner that instrument parameters are monitored. Statistical information should be compiled for each unit of Level 2 data processed. Continuous monitoring of algorithm performance is essential if Eos is to produce uniformly high-quality Level 2 data. The history of an algorithm's performance should be maintained in the archives.

Validation

All Eos data should be validated to the satisfaction of the research community. It is anticipated that this would be done by first conducting an intense validation effort lasting between three and six months, followed by an ongoing monitoring effort. During the initial validation period, Level 1 and 2 data would be substantiated by comparing data from Eos instruments with those of similar spaceborne instruments and with *in situ* observations. Once the initial validation has been accomplished it will be necessary to continue monitoring instrument and algorithm performance for departures from the norms established during the initial validation. It may be necessary to mount additional validation efforts if the instrument or algorithms do not perform as anticipated, for quality assurance purposes, or after known changes (e.g., replacement, refurbishment) of instruments. It is essential that the scientific community participate in both the initial and ongoing validation efforts and that pertinent feedback from any retrospective scientific analysis be utilized when appropriate.

Algorithm Enhancement and Data Reprocessing

It is anticipated that during the course of the Eos project, new or enhanced processing algorithms or procedures will be found that will substantially improve the utility of Eos data. When this occurs, the Eos data processing system should be sufficiently flexible, easily accepting new algorithms and procedures. Since algorithm changes are inevitable, an accurate processing history must be maintained for all Eos data. In some cases, algorithm changes will have such a significant impact that it will justify the reprocessing of previously processed data. Therefore the Eos data processing system must be able to efficiently accommodate new algorithms, access lower-level data that will be utilized in the reprocessing activity, and update the archives with the resultant, improved data.

Maintenance of Data Sets and Documents

The main function of the Eos archival subsystem is to preserve, maintain, and distribute data from Eos instruments for use by the research community. However, the term "archives" is broadly defined; it is not merely a repository for data. Rather, it is an information reservoir and conduit where both the quality and quantity of data are increasing and where access to and use of the data are encouraged. The Eos archives should be able to:

1. accept data from Eos instruments at all processing levels;
2. provide access to non-Eos data that are needed for processing of Eos data or by Eos investigators;
3. provide *in situ* data used in the processing and validation of Eos data sets, or as correlative data;
4. preserve, maintain, and distribute all Level 1A (produced by Eos platforms) and higher-level data holdings throughout the lifetime of the program;
5. accept higher-level data sets produced by investigators as part of their research work; and
6. reprocess Eos data when needed.

The archives should be able to store and maintain these data as well as provide a directory and catalogs for their retrieval. The directory should include information about all Earth sciences data deemed relevant by Eos-sponsored researchers. The catalog supporting the directory should include entries for all Eos data sets and extend to the holdings of other relevant archives, when needed. The archives should also maintain a complete inventory of relevant scientific and technical documents concerning the data and the archives. Finally, it is envisioned that the Eos archival subsystem will be geographically distributed with elements at both data centers and at the active data base sites. It is assumed that these elements will be connected via an electronic network and that uniform access to all system services will be provided.

User Interface

An Eos archival subsystem should be accessible from remote terminals or workstations via a prompting menu or natural language interface, supplemented by a free-form command language for experienced users. All capabilities and functions should be fully documented in a user handbook and in online help files. Consideration should be given to online tutorials that could be used to train novice system users. For those users who choose not to interact directly with the Eos archival subsystem, mail and telephone request should be considered.

Electronic Directory and Catalogs

An archival subsystem should provide an online catalog of data sets of interest to Eos investigators. The data set references in the catalog should be determined by an Eos scientific advisory group or by various user groups. Each data set should have a corresponding higher-level directory entry that contains general information about the data set (e.g., spatial and temporal coverage, parameters measured, data set name, archives location, availability, and contact person). References to pertinent literature should also be included. Those data sets being produced by the Eos data processing system should also have a lower-level inventory, containing more details (e.g., parameter measured, calibration information, processing levels, applicable algorithm descriptions, measurement precision and accuracy, documentation) about availability of data at specific times and places and appropriate access information (e.g., volume identifier for data sets on optical or magnetic storage media).

A user should have the ability to search the catalog by any and all of the following attributes: time, place, instrument or a related group of instruments, project or program, parameter(s), data set name, etc. The search should be limited to the top-level catalog, until a specific data set is selected by an investigator for further examination. At that point the user should be notified whether a detailed catalog exists, and, if so, what further search criteria are applicable. Detailed search criteria depend on information availability for each catalog, but generally should include at least time, place, instrument, and data attribute information.

An investigator should have the option of displaying a list of data sets found, or complete top-level catalog contents, in any of several information categories. If a detailed catalog search is requested, the results should contain all appropriate and available information for that particular catalog entry. Displays should be at the user's terminal, or on a central printer, at the user's option. Printed listings should include the user's mailing address, and should be shipped by operations personnel during the same shift.

A data catalog should contain references to pertinent data sets held by archives other than those within the Eos project. Methods for maintaining the timeliness and accuracy of these references must be established. It should be the responsibility of the Eos project to obtain information to be included in the catalog, and to verify the accuracy of the catalog contents.

Document Storage and Retrieval

The research community cannot effectively utilize Eos data unless they have access to both open and "gray" literature, containing pertinent technical and scientific information (e.g., project plans, instrument and algorithm documentation, validation results, data system documentation, data set descriptions, scientific papers). Therefore, the Eos archival subsystem should provide an online bibliography containing annotated references to all relevant published and unpublished literature derived from the project. Copies of unpublished project documentation should be retained as a part of the Eos data archives.

A researcher should be able to search the bibliography by specifying any number of attributes, including author, publication data, subject category or sub-categories, instrument, parameter, or original document number or citation. The bibliography subsystem should respond by displaying the number of entries found at each stage of specification. The user should then have the option of displaying a list of titles found, or complete citation, including abstracts. Displays should be at the user's terminal, or on a central printer, at the user's option. A user should have the capability to request a copy of any document found in the search. They should also be able to access the complete document text electronically for display or printing at their terminal or workstation.

Data Set Maintenance and Distribution

The primary function of the Eos archival subsystem is to store, maintain, and distribute data sets from Eos instruments, from other supporting spaceborne instruments, and from *in situ* measurement systems. The Eos archival subsystem should therefore provide the following services, all of which are considered to be of equal importance:

1. Ingest Eos data sets as they become available. This includes both scalar and raster satellite data, and *in situ* data sets. The archives should be able to absorb the large data sets of the 1990s and beyond on a continuing basis without any significant backlog. The key to performance and utility of the system is the way in which data sets are loaded into the archives. The data must be rapidly available, and of known quality. Considerable care must be taken in the loading process to remove duplication, to account for gaps in the data, to sort data into chronological order, to detect poor-quality data, and to annotate data whose veracity is questionable. The authority to purge any data from the archives should reside exclusively within the Eos-sponsored scientific research community.

2. With Eos scientific advisory committee approval, purge erroneous or outdated information or information with no significant value. The archives will be dynamic; just as they must efficiently load new data, so they must also efficiently deal with unwanted data.

3. Satellite swath data archives. Data coverage will be either regional (e.g., West Coast AVHRR and CZCS) or global. Swath data should be selectable by project, platform, instrument, level, version, parameter, attribute, time, and region.

4. Grid data archives. Coverage for Level 3 data sets could be regional (e.g., SSM/I polar grids) or global. Selection of Level 3 data should be by project, platform, instrument, level, version, parameter, grid type, time, and region.

5. *In situ* data archives. Pertinent data from both fixed and moving platforms should enter the Eos archives. Data from moving platforms may not be well organized spatially but should be stored in a systematic manner. The storage method for these data should depend on data volume and demand. Data selection should be by project, platform, instrument, level, version, parameter, time, and region.

6. Performance summaries archives. Data accounting summaries and other small non-spatial data sets should be maintained and distributed when requested. Summary data should be produced periodically. These data sets should have a fixed structure with no geographic dependence; thus, these data need only be referenced by the appropriate subset of project, platform, instrument, level, version, summary type, and time. The storage method for summary data will depend on data volume and demand.

7. Browse file archives. Browse files are composed of reduced resolution data sets, should be processed at least to Level 2, and have either regional or global coverage. Browse files are designed to provide a rapid response to a user wishing to locate specific data interactively. These files should be optimized to deal with communications and remote terminal limitations and should be referable by project, platform, instrument, level, version, parameter, time, and region.

8. Maintain data on media that provide long lifetime, rapid and random access, and economical storage. Currently, the primary candidate medium is high-performance digital optical disk. Optical disk storage media eventually could replace existing magnetic tape archives.

9. Distribute data sets on optical disk (or other appropriate media), or transmit data on a communications link. As the Eos data base grows, it may become necessary to develop a high-speed communications link to existing modeling facilities for the transfer of Level 3 (and possibly Level 4) data sets.

Data Set Support Categories

The Eos archival subsystem should provide several categories of support for archival data sets. The support category proposed for each data set should be selected by the Eos Project, based on guidance from the research community and cost/benefit considerations. The categories are:

1. Online support - subset selection by time, region, platform, instrument, and parameter from the archives; full range of output; full catalog support;

2. Offline support - subset selection by media volume only, from the archives copy; storage media output only; full catalog support;

3. Ordering support - catalog support; ability to place an order through the catalog subsystem; orders handled manually;

4. Catalog support only - limited information in the catalog; no support for obtaining data;

5. Directory support only - directs user to a specific non-Eos archival facility for further information.

All of these categories should be supported by the bibliography, but we expect that the level of support (i.e., the thoroughness with which relevant citations are selected, entered, or purged) will vary, according to the priority implied by this list.

Performance and Accounting

The Eos archival subsystem should meet suitable performance goals required by the research community, specified by the project, and agreed to by Eos investigators, at least during prime time (defined as 0800 Eastern to 1800 Pacific Time on weekdays). Performance goals should specify system availability, data loading time, extraction time, data set generation time, browse time (i.e., generation, transmission, and display), and system response time.

In order to monitor performance of the system and determine appropriate user charges, various accounting factors should be measured and reported. All required information should be recorded automatically and reported in weekly and monthly summaries. These data should include catalog usage, bibliography usage, archival loading, data set extraction, subset generation, and general resource usage. A detailed invoice should be prepared and sent to the user with each data shipment. Summary invoices should be prepared for each user on a monthly basis. In addition, online accounting algorithms should be available to allow researchers to estimate costs before requesting either data or data processing services.

TELECOMMUNICATIONS

The telecommunications industry is in a state of rapid change. Deregulation of the industry, the growth of satellite and fiber optics high-volume communication facilities, and the transfer of communications backbone switching and multiplexing equipment from analog to digital are causing major changes in communications capabilities. Networks for transfer of digital information are growing rapidly. Any requirements report relies on a background knowledge of what is feasible and readily available. Because of the uncertainties in the telecommunications industry, and the long lead time for Eos implementation, the telecommunications requirements should be periodically reevaluated and updated as new technologies become readily available.

Downlink Communications

Eos data rates are quite high in comparison to other Earth orbiting satellites. The expected average data rate is over 70 megabits per second, with peak rates of over 400 megabits per second. NASA communications commitments are for an average data rate of 70 megabits per second, but will not include continuous channel availability. Manned mission (i.e., the core Space Station) communications will have a higher priority. In order to buffer the high peak data rates and to guard against momentary downlink channel outages, an onboard data storage capability should be developed. This data storage subsystem should be sized to prevent loss of data. Data from one half orbit will be over 125 gigabytes. Onboard data storage needs to temporarily store these high data volumes, at a minimum. Data from onboard storage devices should be packed to take full advantage of broad-band downlink transmission capabilities.

Direct Broadcast Communications

NOAA and its international clientele will require separate data storage and relay systems that will "guarantee" data delivery from any NOAA operational instruments that may share the Eos polar platform. This system would likely function in a manner similar to current NOAA spacecraft, which have onboard storage and utilize burst transmissions over a receiving site. In addition, NOAA has international commitments to provide direct broadcast weather satellite data for remote regions. There are also a large number of international land-applications data users who are equipped to receive Landsat data. Consideration should be given to using these direct broadcast systems for transmission of high-resolution imaging spectrometer and synthetic aperture radar data when they are not being used to relay weather instrument data to NOAA facilities and clientele.

Repository Communications for Quick-Look Users

After relay by telemetry satellite, Eos data will first be delivered to NASA mission and instrument repositories. These data repositories are primarily to support NASA mission and instrument team operations, with the data subsequently being transferred to Eos data centers, active data bases, etc. The transfer of data to these facilities could be either through electronic communication channels, or through the physical transfer of disk, tape, or other storage media. However, while the data is still in an Eos repository, certain communication functions are still needed for mission operations, quick-look data sets, operational user data sets, and other principal investigator functions. The principal investigators will require quick-look data sets for instrument command decisions, and for preliminary scientific analysis.

Real-Time, Quick-Look Data Access

The users identified as requiring access to real-time, quick-look information need interactive image processing and display capabilities. These individuals may need some interactive image browse capabilities as well as the capability to reprocess further and display the data interactively. Further processing of the data may be mission dependent; consequently, applications software modules need to be easily added to the system. The user community requiring access to real-time interactive quick-look data will probably be limited to Eos instrument teams and selected active data base sites. Investigator processing requests should be completed within the timeframe of one orbit. Communication facilities are needed to link quick-look investigators with the repositories. Data rates between 56 kbps and 1.5 mbps will likely be required on this link. The link should also be capable of handling multiple users simultaneously.

Uplink Communications

Many mission operations activities will be planned and scheduled. However, there probably will be occasions when instrument command decisions will need to be made within the time period of one orbit. The requisite communications system needs to be capable of quickly providing scientists making these decisions with sufficient information. This information includes, in addition to quick-look data sets, trade-off information for various instrument configurations. Instrument team requests for configuration changes need to be included effectively into the command and control decision chain, and a mechanism should exist to arbitrate conflicting demands. The communications system needs to be functioning continuously, accommodating different user communities. These configuration decisions should be in addition to automated real-time selection (within a priority set of options) of measurement sites for select instruments.

Eos Information Network Services

Eos data centers will make Eos-derived data accessible by the greater scientific community. These centers should have archival data bases and advanced data base management capabilities in addition to demand processing facilities. The Eos information network should provide the required communications links to and within the scientific community. This network will require catalog and browse attribute file, browse image file, and archival data set access, as well as the quick-look access previously described. It should also provide for electronic communications between colleagues working on Eos-sponsored research.

Catalog and Browse Attribute File Access

Interactive electronic catalog and ordering functions should be available for Eos researchers with at least 9,600 bps dial-up capabilities. In the catalog, in addition to location, time, instrument status, and other available information, the attribute file should also be accessible for search purposes. An interactive catalog system should provide response to most search commands within 1 to 15 seconds. The system should be available on a continuous basis and be sized to handle at least 100 simultaneous investigators.

Browse Image File Access

In addition to the attribute browse capability, the system should allow for reduced-volume data browse. Because of data rate limitations of 9,600 bps communications links assumed for most users, electronic browse of image files may not be possible for many users. This would necessitate the publication of an image browse catalog. In this event, it could include both the catalog and attribute numerical files, which could be computer processed to locate images or sequences of images.

A select number of users (e.g., those associated with active data base sites) should have available higher data rate communications lines and they should be provided with interactive electronic image browse capabilities. Of particular importance are the data between the present and the last published image browse file. This interactive image browse capability could be used by quick-look investigators and research teams monitoring an instrument as a precheck of ordered data sets.

Access to Archival Data

Even in the 1990s, many researchers will probably receive high volumes of data through the mails. The majority of researchers can probably accept electronic interactive ordering (i.e., order placement and acknowledgment) with mail delivery. Turnaround time for orders should be consistent with mail service (i.e., a few days to a week at most).

The system should also allow for delivery of data electronically, since a few sites may have the high-speed communications capability needed to receive large-volume requests. Other researchers will have lower-speed communications, consistent with lower volumes of requests. The system should allow for all levels of archival retrievals to be sent over any communication line tied to the system. This would also include the use of low-speed links for interactive catalog interrogation.

Carrier Considerations

The Eos information network will require communications facilities that tie together a large number of researchers using various communication rates, ranging from dial-up lines to high-speed links of 1.5 mbps or more. Network protocols and standards may result from the adoption of standards set by Space Station communications. Users with different protocol capabilities will require translator boxes or packages at the gateways into the network. If extended to include local area network access, the NASA Program Support Communications Network (PSCN) would be one example of the type of capability that would be required. The PSCN was originally envisioned to link NASA researchers together with phone access, packet switched data (9.6 kbps), circuit switched data (56 kbps), and computer network subsystems (1.5 mbps to 6.3 mbps). While the PSCN will link NASA facilities together with a computer network subsystem, it is not clear if it will include voice, packet switched, and circuit switched access or if it will include extensions into the university and international research communities to serve Eos requirements. There are, however, numerous commercial carriers available that provide the full spectrum of communications requirements needed for an effective and efficient Eos information network.

DOCUMENTATION

Proper experimental documentation encompasses more than the description of the hardware and its initial calibration. Documentation should also include any information pertinent to scientific interpretation of the data record produced by the experiment. Since Eos is conceived as a global, multiyear Earth observational program, intercomparisons between measurements made at different times and locations is central to the scientific return. Consequently, accurate and complete documentation is imperative for the success of this program. Documentation is considered to be a vital part of the data record and should be stored and maintained with the same care as the data, per se. Some necessary elements of this documentation are briefly described below.

Trace of Design, Fabrication and Testing of Hardware

This information, which is routinely prepared by contractors, is rarely put into accessible form or kept as part of the data archives. Eos data volumes will be so large that digital storage of text materials can easily be supported. All key personnel should be identified and all reports collected in the archives. Much of this material may never be used, but some of it may prove crucial to understanding a particular set of critical measurements.

Hardware Description

Preparation of this document, including the above described information, should be included in the contractual obligations of instrument teams, instrument principal investigators, and facility-instrument Centers. It should provide a compact description of instrument specifications, including the nature of physical variables measured, noise characteristics, internal processing, and coding of telemetry. Sufficient detail should be included to allow another group or individual to convert raw telemetry signals into basic physical quantities, using ancillary calibration information contained within the transmitted data record.

Calibration

The procedures and results of all calibration experiments and tests should be reported in sufficient detail for other scientists to evaluate their thoroughness and accuracy. All approximations and instrument uncertainties must be identified. Instrument performance over a complete range of environmental conditions should be evaluated. This information should allow conversion of telemetry signals to standard physical units of measure. Calibration standards should be traceable to the National Bureau of Standards; this requires documenting the history of Eos data records.

Calibration History

In addition to documenting calibration tests, a specific time sequence data record should be created from information used to monitor instrument calibration over the course of an experiment. In doing so, drifts in instrument characteristics can be reconstructed. Both the calibration and calibration history files should have the provision to hold more than one calibration test result or more than one type of monitoring information; all sources of calibration information must be identified and placed in the archives.

Command History

Automated command sequence construction, together with the necessary computer involvement in routine spacecraft communications (i.e., the commands given every instrument together with its status, indicated in housekeeping telemetry) should be available in mission operations computers; this information is rarely retained. Provision should be made within Eos data archives to create a data record documenting the history and providing a trace of commands and instrument status throughout the lifetime of the project. Significant reconfigurations of instruments (e.g., gain setting changes) or observing sequences should be accompanied by log entries identifying the reasons for and the source of the requested change.

Bibliography

A list of all papers and reports published in the open literature and thought to be pertinent to the type of instrument or observation being made, calibration or sensitivity studies, and the precision and accuracy of the data should be included as part of the documentation. Consideration should be given to making the complete text of all unpublished documentation (i.e., the gray literature) electronically available.

Processing Algorithms

If an instrument data record is to contain quantities derived from measurements (those quantities obtained directly from the telemetry by application of calibration relations), then the methodology utilized must be thoroughly explained and all relevant references to literature concerning the methodology should be listed. All processing software should be preserved and documented, along with a report of design and testing rationale used by the software creator. This requirement also applies to processing software used to locate the observations and calculate observational geometry parameters.

Correlative Data Record

Often routine processing procedures utilize additional data or information to determine derived parameters. Although these data may describe simple or well-known quantities (e.g., topography), the particular version of the information used may differ slightly from other versions. Therefore, all such data should be considered part of the archives.

Data Usage Trace

Institutions responsible for maintaining Eos data archives should provide a trace of user access to the data, allowing other investigators to contact colleagues familiar with a particular data set's characteristics. The trace should contain specific information about the data ordered and the name and address of the individual accessing the data.

The above list of documentation is very exacting and will demand significant effort to acquire and maintain. Notwithstanding, given the requisite resources envisioned for Eos and the large investment of time and money it will require, this documentation effort appears trivial by comparison.

NON-Eos DATA AND MODELS

Most, if not all, of the research scenarios (vis. Appendix I) have some requirements for access to non-Eos data. In some cases, access to models (e.g., global circulation models), or modeling capabilities, will be needed to place Eos data in context with some physical or chemical process. While these external data, models, and modeling mainframes may not be a part of Eos as a space-platform mission, they are part of the Eos scientific requirements. Hence, a commitment by Eos to access external, non-project data sets and models should be made. Model results should be accessible and, in addition, researchers should be able to request model runs through an Eos data and information system.

There are six possible levels of commitment to external archives access. They range from accessing a directory of catalogs, to instant, remote access to the world's data bases. The six levels are:

1. Directory of catalogs
2. Information on specific catalog access routines
3. Ability to place an order through Eos
4. Direct access to other data bases through Eos
5. Direct orders and near-real-time access to external data bases through Eos
6. Direct order and real-time access to data banks through Eos.

It would be highly desirable if Eos could provide services through level 5. The user could locate a specific non-project data base, order that data through Eos, and have it arrive along with his Eos archival data in a compatible file format. A single software routine based on this common format would then be sufficient to input all of the data necessary for research.

STANDARDS

The Eos project will be operating within the same timeframe as the Space Station project, which is designing a Space Station Information System. It, in turn, will operate under the auspices of the NASA Office of Space Science and Applications, which is defining a Science Applications Information System. Each of these will impose some standards on its components. Space Station personnel have discussed standards for the operating system, programming language, software support environment, data base management systems, and the use of standard formatted data units. There are potential benefits to these interconnected systems if the various standards adopted by such top-level entities as the Space Station Information System are acceptable to, and accepted by, Eos in common or overlapping areas. This will require that Eos project personnel and researchers be familiar with, and preferably participate in, the definition of these standards.

Careful consideration must be given to the standards, conventions, or guidelines that Eos accepts or develops. Because Eos will be a primary researcher interface with the Space Station Program, its consistency of operation (i.e., the adherence to standards) and its flexibility (i.e., freedom from standards) will weigh heavily in its acceptance by the research community. Unless standards are perceived by the user to be beneficial, they will be viewed as overly restrictive and onerous. Thus, early decisions should be made, minimizing the set of standards to be invoked while maximizing their utility. Three factors should be considered before Eos adopts any particular standard:

1. Historical data from a suite of spaceborne and *in situ* instruments will be used for intercomparisons by Eos researchers. These data reside in numerous archives, are in a variety of formats, have varying resolutions, are of differing quality, and have a variety of appended ancillary information. Hence, these data will require resampling on space-time scales used for Eos data.

   Common data set organization (e.g., spatial resolution, map projection) and the use of logical catalog and data structures (allowing various computers to read the files) will minimize the reprocessing task. Common factors such as these should be studied to provide guidance in selecting standards that will minimize reprocessing at a future time.

2. Eos will be operating in a computer environment largely determined by developments in the commercial sector (i.e., by software and hardware manufacturers). We anticipate that Eos data and information system requirements will parallel those that focus commercial developments (e.g., nested catalog structures, advanced data base management capabilities, super micro- and mainframe computers). Since system costs would be minimized by adopting commercially available products, Eos requirements should be tailored and adapted to these product developments wherever feasible.

3. The computer and communications community is moving toward standard practices that will be both machine- and data-independent. These practices will allow data and algorithms to be readily interchanged even on future systems. Any standards adopted under the Eos data and information system auspices should take these practices into account.

Areas of Consideration

Areas of direct researcher and system interaction will include directory and catalog access, query and browse functions, archives access, algorithm development, and data interchange. Given the rapid rate of data base management developments, it is probably inappropriate to attempt to standardize specific data base structures, catalog organizations, query procedures, etc. Rather, a more productive effort will concentrate on standardizing the tools with which to describe and manipulate these data. Thus, appropriate items for standards include data definition languages (rather than a specific data structure) or data system model techniques (rather than defining a specific data model). This approach also recognizes the existence of numerous heterogeneous data bases and their catalogs, which must be dealt with by Eos. Several NASA Pilot Data Systems are considering the creation of uniform catalogs (i.e., uniform information presentation for the researcher) in contrast to standard catalogs.

Systems are usually described in block diagrams at various levels of detail. In such a description, the critical items are the functions to be performed by each block (i.e., its transfer function) and the characteristics of the paths between the block interfaces. Unless some overriding factor dictates standards within a block, the appropriate items needed for complete descriptions, and perhaps standards, are the functions and interfaces themselves.

Beyond functions and interfaces, careful consideration should be given to standard format structures for data interchange. While they may not explicitly cover data storage formats, there is a close relationship between storage, interchange, and functional formats. Standard formats could serve several purposes; these include:

1. Providing a data organizational structure that will accommodate all types of geographically related spatial and non-spatial data (i.e., scalar, polygon, and raster data), features and attributes, and necessary topological relations. While it may be desirable to define a single format, the diversity of data types suggests that this is inappropriate;

2. Providing a format volume record grouping capable of accommodating all necessary data in variously formatted records (i.e., feature identification, data quality, spatial data type, locational definitions, spatial relationships, and ancillary data). These format structures should be expandable to allow addition of future types of data;

3. Providing an interchange format that will allow a researcher to read the data set, determining the basic logical structure. A family of format structures necessitates conveying to the receiving computer the logical structure of the data being presented; this is the purpose of a data definition language;

4. Providing coding standards to accommodate a variety of information and attributes.

Glossary and Definitions

In the multiple-system, multidisciplinary environment of Eos, common understanding of terms is a necessity. This will require the adoption or development of a standard glossary containing operational definitions. This should include techniques as well as definitions, and include such items as the data definition language and data modeling techniques. It would also be appropriate for the glossary to list applicable external standards (e.g., Federal Information Processing Standards, American National Standards Institute, Consultative Committee for Space Data Systems, International Standards Organization).

Algorithm Interchange

Much has been said about sharing of algorithms among scientific researchers. This has been hampered in the past by the use of different operating systems, different programming languages, and different execution procedures. While it is unreasonable to expect the entire research community to standardize on systems and languages, it is reasonable to consider a standard interface executive. This would allow the independent development of algorithms and executives and provide a method for easy interchange between investigators.

Data Interchange Formats

An oft-voiced problem is that of data interchange, particularly in light of the infinite variety of formats that are in use today. Several studies are currently underway, aimed at providing a standard format structure:

1. Potential worldwide data transmission will be a feature of future systems. The International Standards Organization (ISO) has formulated a series of standards expected to facilitate international electronic message interchange. In addition, the Consultative Committee for Space Data Systems (CCSDS) has been developing a set of recommendations for special purpose techniques that are tailored to the unique environment of exchanging data through space data channels. ISO's DIS 8211 standard proposes a data definition language, and CCSDS proposes a number of data formatting standards that may be suitable for Eos consideration.

2. Two national committees are currently investigating the interchange format question, one convened under the U.S. Geological Survey (to unify

data representations from Federal agencies) and the other from the National Bureau of Standards via the U.S. Geological Survey and the American Congress on Surveying and Mapping (concerned with digital cartographic data standards). The National Committee on Digital Cartographic Data Standards is charged with producing a format structure recommendation for the cartographic community before 1986, with eventual consideration for inclusion within Federal Information Processing Standards. Careful consideration should be given to working with these national committees and subsequent adoption of the standards that they recommend.

3. NASA is currently sponsoring a study of standard formatted data units. A standard formatted data unit is a conceptual data object that could be transmitted between users. It will consist basically of a formatted and labeled data set; thus, it defines an interchange format. It will include a primary label that serves as a global identifier, a set of secondary labels that carry information about the data, and the data set itself. It currently focuses on the record level, but may also provide a nesting feature that would allow a given unit to be composed of multiple standard formatted data unit subsets.

Clearly, the area of standards and formats needs considerable attention by Eos and, in particular, consideration in concert with the research community who must deal with the resultant standards. Since we envision the necessity for an entire suite of format structures, system transparency to the researcher remains an overriding concern. This must be the guiding principle in establishing standards and format for the Earth Observing System data and information system.

SUMMARY

The foregoing discussion delineates a significant number of requirements, attributes, characteristics, and capabilities that we believe should be accommodated within the Eos data and information system. These factors can be conveniently grouped into seven interrelated and interdependent functional categories and are summarized in Appendix II.

These functional groupings are Eos flight systems, user functions, operational functions, information service, advanced data base management, data processing, and non-Eos data base functions. Eos is unique in that we expect that many of the requirements cannot be accommodated by any single functional element or group. Rather, we anticipate that many functional elements will need to coordinate their activities to accommodate a given task. The interrelationships and interdependence of the elements within the resulting data and information system is thus quite significant. Consequently, Eos information services functions will be central to ensuring straightforward, transparent intra- and intersystem interaction, a prime characteristic derived from these researcher-imposed dependencies.

III. ARCHITECTURAL CONSIDERATIONS

In this chapter, we propose appropriate design principles and consider an Eos data and information system for the 1990s, the Space Station era. We present an example of a functional architectural design, described in terms of elemental composition, top-level functions, and internal and external interfaces. This conceptual architecture illustrates one possible configuration that could satisfy the general data system requirements that are developed in this report.

To ensure that the designs considered were appropriate to the Eos data and information system, we categorized the requirements, characteristics, and attributes delineated in this report into seven functional groupings (cf. Appendix II). These groupings were then compared to each functional element of the architectural examples. We believe that the flexibility and modularity of the conceptual design presented in this chapter is well suited to these requirements and the needs of Eos.

Since the functional design of data transport facilities and services is a key and fundamental factor, we have selected an architectural example based on space data communication standards developed by the Consultative Committee for Space Data Systems and utilizing International Standards Organization general purpose protocols for transparent interchange between dissimilar system elements. We have examined this design in sufficient detail to ensure soundness of approach and methodology. As with any initial concept, this design is open to constructive criticism, revision, and improvement. A more detailed description of a system architecture is beyond the scope of this report.

DESIGN PRINCIPLES

Designing a data and information system for Eos is clearly a complex problem, particularly with respect to data handling. To ensure that the resultant system can meet scientific needs, the architectural concept should feature two fundamental principles or design techniques that we believe should be employed throughout: *layering* and *standardization*. Together, they can be used to create an information system that is flexible, transparent, and robust, providing the foundation for diverse data processing and archival needs within an Eos data and information system.

Layering and Modularity

The first architectural principle to be applied throughout the design of an Eos data and information system is layering, the technique of dividing and conquering. Layering breaks the diverse and complex system into easily comprehensible modules with clear "strata" in which common data-handling functions reside. Strata exchange data according to well-defined rules and are thus predictable in terms of their functional characteristics. Layering is essential for a data system that:

1. is understandable by a wide range of users and implementers;
2. can expand or contract easily to accommodate changing user requirements or technological developments;
3. can be tested in a modular fashion by maximizing the functional independence between processes residing in different system layers;
4. displays simple, well-defined system interfaces that are not constantly changing;
5. can adaptively respond to dynamic mission events;
6. facilitates recursive use of replicate hardware and software elements to lower system development and operational costs; and
7. avoids the need for major advances in technology by permitting complex tasks to be broken into pieces that can be handled within existing capabilities (e.g., parallel processing of very high-rate data streams).

Standard Data Structures and Data Autonomy

The second key architectural design principle is standardization of formats and protocols through which data are exchanged between distributed elements of the system. Standard structures promote "data autonomy," the transport of independent, user-defined units of data through the system. Using this principle, the internal format and content of the data are defined by the user and transparent to the communications system.

For the Eos data and information system to provide efficient, high-performance data services that handle a diverse combination of data types and requirements, it should be adaptively responsive to the data per se. The concept of data autonomy (encapsulation of variable data within standard, network-interpretable labels) provides a foundation for these adaptive data services. Data autonomy is achieved in this architectural example by employing standards that require each data source (e.g., engineering subsystem, instrument) to encapsulate its data messages into "source packets" having globally interpretable labels that define the source and destination, class of service, priorities and delivery conditions, and provide the information required for verification, validation, and accounting of data within the packet. Thus, while the packet format is a standard for the network, the format of its contents can be variable; the data exchange network need only transmit what it is given.

Adaptive Flexibility

We envision an Eos data and information system that is dynamic, evolving to meet the needs of its changing clientele and advantageously employing new technological developments. Layering, modularity, structures, and autonomy are hallmarks of contemporary system design. Together, they provide a means through which the resultant data system can evolve adaptively to meet these changing needs.

A data and information system design utilizing standard data structures affords a significant level of flexibility; its data services are readily adapted to changes in internal data format and data flow or routing without changing the system architecture. These structures should be deliberately designed for flexibility through the provision of features such as secondary headers, and they should have the flexibility of allowing future versions to be defined as needed.

DESIGN CONSIDERATIONS

The Eos data and information system must be designed, implemented, and operated affordably, in an environment of almost constant technological and researcher-initiated change. Some of the requirements and constraints imposed by this environment are now examined and their ramifications imposed upon subsequent designs.

Evolution

A key architectural consideration for the data system is its projected lifecycle, which will encompass the development and operation of a very complex, dynamic system that must accommodate growth and evolution of spacecraft instrumentation, technology, and user requirements. At the outset, performance requirements imposed on the system will push data handling technology to the limit.

Subsequently, data handling capabilities must evolve to support a maturing platform that carries an increasing number of instruments with growing requirements for data storage, processing, and transmission capacity. System technology will become obsolete quite rapidly unless the system design accommodates changes.

As the data system further evolves, spacecraft and ground-based data handling and processing operations may become more independent. The system's data communications and processing capabilities will need to expand in a controlled, evolutionary fashion to support growth. The evolving character of the environment in which the system operates requires that the architecture display a high degree of modularity and structure. Adaptive flexibility must become a cornerstone upon which the system evolves.

Distributed System Considerations

When the transmitted data stream is received at ground-based facilities, it will be distributed to a confederation of geographically dispersed data processing and archival facilities. We anticipate that these facilities will be interconnected by both private (e.g., NASCOM) and shared public (e.g., X.25-based) communications networks and private and public broadcast satellites.

System facilities may be distributed worldwide, and will provide specialized data handling, processing, and archival services in support of a large, and not yet fully defined, cadre of users. The breadth and diversity of the user community suggests that the Eos data and information system should feature a spectrum of format and protocol standards for data interchange. Application-oriented standards are essential to facilitate the early development of a dynamically adaptive, distributed data system that can sustain planned evolution.

The majority of researchers will also have changing requirements for routine data handling services such as accounting, merging, grid overlay, map projection, ancillary data processing, temporary storage, and retrieval. These are changes that the data and information system must efficiently accommodate.

Data Types and Rates

The data system will be required to transport and deliver a diversity of digital instrument and engineering data through bidirectional TDRSS (Telecommunications and Data Relay Satellite System) communications links. As a result, a spectrum of performance requirements will exist, including: (i) raster data that can tolerate delivery delays but not data outages (it is moderately tolerant of data errors when uncompressed, but intolerant when compressed); (ii) other digital data and data transport processes, including low- and medium-rate instrument telemetry, engineering and housekeeping data, ancillary data, data base transfers, command sequences, memory loads, and text and graphics, that have differing and often conflicting requirements (some, such as programs and data base transfers, are intolerant of any data errors or outages, while others are tolerant of even the poorest communications).

The wide diversity of user data requirements suggests that the Eos data and information system could utilize several different classes of data transport service, each class displaying a well-defined quality of service that includes clear specifications of data rate, error rate, delay, sequential character, and completeness. The data system must also provide certain value-added utility services for user data streams such as merging, sorting, storage, and remote access.

TDRSS Compatibility

All types of Eos digital data, each with its own service requirements, are merged into a composite data stream that flows across TDRSS transmission paths. TDRSS services include uplink and downlink transmission on S-band and K-band data channels. A 300 Mbps K-band single access service, using I and Q channels, will provide the dominant downlink data transport service for platform scientific and operational data. The S-band single access channel will primarily carry platform engineering and some real-time operational data.

Initially, data system downlink services will need to accommodate a composite data rate that periodically approaches (and even transiently exceeds) the present TDRSS limit of 300 Mbps. The uplink data service will be provided on the single-access S-band channel at 100 kbps when scheduled. It is important to note that TDRSS has the capability to communicate with the Eos platform at any time via the multi-access forward link, but realistic operational support considerations will probably limit the use of this uplink to emergency situations. These communications services must transparently support instantaneous, dynamic, and adaptive changes in the blend and volume of data that flows through the link.

Direct Downlink, Broadcast, and Uplink

It is possible that Eos research instrumentation and NOAA operational instruments will reside together on the same polar platform(s). NOAA operations will likely require a direct downlink independent of TDRSS. The data and information system should accommodate this enhanced capability with minimal system impact.

NOAA operations in support of international weather services require that some onboard processed data be broadcast from the platform to ground receiving stations. These transmissions would only contain real-time data and would not include previously recorded information. Relay broadcast services could also be accommodated for search and rescue activities.

Direct uplink capabilities would include only relay broadcast and data collection activities. Instrument and platform command functions would be handled through mission operations via the TDRSS uplink.

Data Interchange Standards

Within the framework of the layered concept of "Open System Interconnection," the International Standards Organization (ISO) is currently developing a broad spectrum of commercially supported general purpose data protocols. These protocols are designed to provide transparent interchange between dissimilar elements within an open, worldwide communications system. Emerging ISO standards will likely have direct application to the Eos data and information system, although some of the standards are relatively unattractive for specialized data exchange through either capacity- or bandwidth-limited space data channels.

The Consultative Committee for Space Data Systems (CCSDS) has been developing a set of recommendations for special purpose techniques that are tailored to the unique environment of exchanging data through space-based data channels. These recommendations supplement ISO standards in those specific areas that are unique to space missions. Many CCSDS standard data link protocols are directly applicable to the Eos data and information system, although some, particularly space-to-space links, may require adaptation.

In support of Eos scientific research, the data and information system will need to exchange heterogeneous data sets not only among its different elements but also with other systems both internal and external to NASA. The use of standard formatted data units will enable common services within and interchange of these data among various data systems.

Taken together, the ISO and CCSDS standard protocols form the basis for standardization, hence data autonomy, within the architectural example of an Eos data and information system considered here.

FUNCTIONAL ARCHITECTURE: TOP LEVEL

The Eos data and information system should provide its users with the services of a complete information system, including the bidirectional communications capabilities required for transparently transferring many diverse types of data between the various ground and space-borne elements of the system. Consequently, it will be necessary not only to procure and maintain certain physical elements of the system, but also to provide an interface with and use of facilities, utilities, and data provided by other NASA and non-NASA organizations in a manner that is transparent to the researcher.

Overview

A simplified functional diagram of a distributed data and information system concept is shown in

Figure 1. This architectural example has been examined in some detail to ensure that it meets the requirements, characteristics, and attributes developed in this report. It is important to note that this example describes a distributed data system in which the same functions (e.g., acquisition planning, analysis, storing, and cataloging of data) may be executed on different data sets at one location or the same data sets at different locations. These functions or processes and their results must be coordinated, either onboard or on the ground. Exactly where and how this coordination takes place will be a function of the level of intelligence that is built into the onboard and ground-based elements of the systems. No attempt has been made to detail all of the functional capabilities included within a given system element. Rather, a few key functions are highlighted for each and the reader is referred to Chapter II and Appendix II for a more detailed accounting.

Two key elements of the system are the Interface Unit and the Data Management and Communications Network. The Interface Unit includes a processor and memory, and provides the means together with telementry and telecommand capture and dispatch systems (imbedded within the Data Management and Communications Network), through which commands will be sent and data received from spacecraft instrumentation. Similarly, position and timing information from the Global Positioning System can be appended to the data streams via Interface Units.

On the Eos platform, Interface Units can play a key role in system resource management. With a large user complement, it is likely that conflicts will arise when users simultaneously request resources (i.e., power, thermal, pointing, communications, etc.) that exceed those available. Rather than attempting to check every request centrally, a distributed resource management system can be implemented using Interface Units. A set of software-controlled management services (network interfacing, actuator interlocking, power limiting, command verification and validation checking, status monitoring, fault containment and isolation procedures, etc.) could be downloaded to distributed Interface Units for execution.

With the exception of direct broadcast access to polar orbiting platforms by users with receiving stations, the user interface to the Eos data and information system will be the same, regardless of

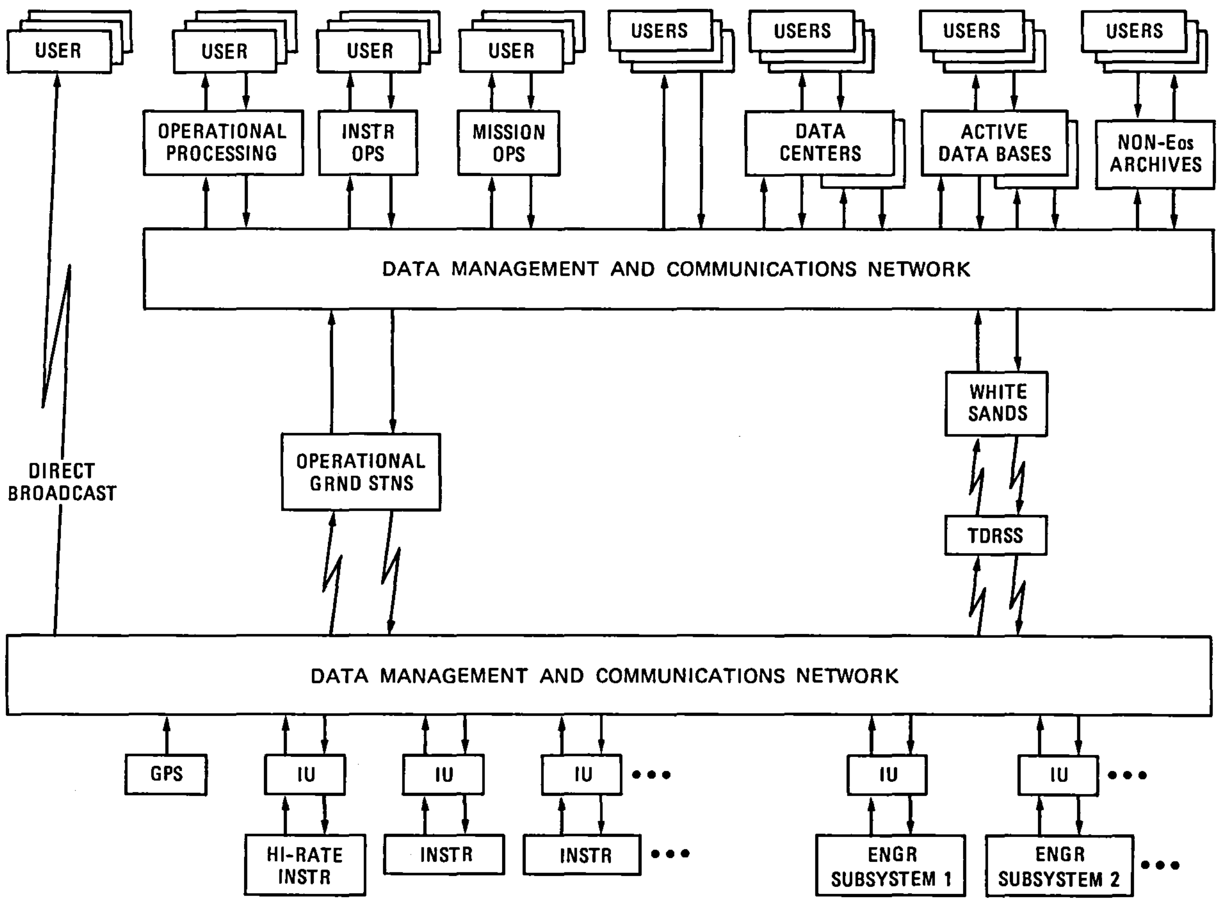


Figure 1. An example of a data and information system architecture suitable for Eos.

geographic location. Operations-oriented data will flow through its own ground station(s) and Eos data will flow through TDRSS. The types of data flowing to various Eos facilities and the processing capabilities resident at these locations are implementation questions requiring systems trade-off analysis. The generic architecture considered here does not constrain the physical location of personnel, data, or processing capabilities. It is important to note that the use of data standards enables ground data processing facilities to upgrade their capabilities and transfer functions to different elements without major perturbations to the entire system.

A user's low to medium data rate instrument or engineering subsystem will have an interface with the Eos data system Local Area Network via a standard Interface Unit, or via the operating system services of an onboard data processor that supports instrument operations. A source packet will be the standard data structure at the Interface Unit. The Interface Unit provides the physical interface to the local onboard data network and its protocol, a packet transfer frame. High data rate instruments (e.g., High Resolution Imaging Spectrometer, Synthetic Aperture Radar) will format their packet data directly into transfer frames, which can be switched directly into dedicated "virtual channels" in the downlink stream (a virtual channel is created when a broadband link is subdivided into two or more parallel paths operating quasiautonomously). These instruments can also exchange low-rate source packets via the Network for adaptively optimizing data acquisition.

At the downlink receiving site, the high-rate serial frame stream is immediately split into parallel virtual channels, and selected frames are routed to appropriate lower-rate data handlers such as decoders, packet extractors, data capture devices, etc. Closed-loop retransmission protocols will be implemented for certain virtual channels carrying critical data that cannot tolerate data outages. The elements of protocol required to execute closed-loop command operation procedures may be placed in a secondary header of the telemetry transfer frames assigned to those virtual channels that use retransmission techniques.

The use of virtual channels in a telemetry frame allows a single 300 Mbps data stream to be immediately split into many lower-rate parallel paths. Thus, processing functions such as decoding can be done at data rates that are well within the capabilities of existing technology. Bulk data streams can be readily sent to dedicated real-time processors or a storage medium for immediate reception and possibly non-electronic transfer to a user.

Data flowing through TDRSS will have Level 0 processing performed at a ground data handling center (presumably located at White Sands) before transmitting: (i) platform engineering data to the Mission Operations Center; (ii) subsets of the instrument and engineering data to the Mission Operations Center for quick-look analysis by investigators; (iii) high-rate instrument and engineering data and subsets of platform engineering data directly to dedicated Instrument Operations Center(s); (iv) moderate and low-rate instrument and engineering data and subsets of platform engineering data to appropriate Instrument Operations or Mission Operations Centers.

Operational Facilities

Operational Ground Stations can acquire data directly from the spacecraft in a manner similar to current NOAA practices. Assuming operational instruments are resident on the Eos platform, these Stations provide transparent user access to their instruments.

All data flowing through an Operational Ground Station will be transmitted to an Operational Processing Center for reduction and creation of directory and catalog entries. These data will be accessible to other system elements linked to the Network.

Instrument Operations Centers

There are two types of Instrument Operations Centers envisioned, (i) those that are dedicated to specific high data-rate instruments, and (ii) those designated for groupings of low to moderate data-rate instruments.

We anticipate that all unprocessed data would be stored at the White Sands receiving center for a two-day period. During this time period, the data would necessarily have to be transmitted, received, and acknowledged by the appropriate Instrument Operations Center.

Processing quick-look, Level 1A, and higher-level data, and generating catalog, directory, and browse file entries, would be performed at the appropriate receiving Instrument Operations Center. This data and information would then be forwarded to an appropriate Data Center for long-term archival storage, catalog and browse file maintenance, and dissemination.

Requests for observational sequences, commands, and command sequences generated by instrument teams or associated investigators would be transmitted to the Mission Operations Center maintaining a master schedule.

Mission Operations Center

The Mission Operations Center supports planning and execution of instrument operations, resulting in the introduction of new Eos data into the system. Through Mission Operations, a researcher's instrument request will be formulated into a command sequence and forwarded to the spacecraft. Should the need arise, the Mission Operations Center should provide rapid response for instrument teams or researchers requesting command control of an instrument from remote locations. The Mission Operations Center, like the Instrument Operations Centers, will maintain software-supported planning aids, employing menus or other "user-seductive" techniques. This software will assist researchers in planning and scheduling of observational sequences by Eos instruments.

Data Centers

Currently, it is anticipated that there will be a lead Data Center that is responsible for algorithm development and maintenance; coordination of overall data system activities; development, distribution, and maintenance of various technologies; network management; and maintenance of data archives, the directory, and its own catalog. Data directory and catalog entries will be forwarded to and maintained by this lead Data Center. The directory will be more comprehensive than the Eos data per se; it will include entries from other non-Eos data bases pertinent to Eos scientific objectives.

Since we recommend that processing beyond Level 1A be performed on demand, it is anticipated that this processing and eventual scientific analysis could take place in both Instrument Operations and Data Centers. The resulting data sets will have directory, catalog, and browse file entries developed at their processing location. The resultant data would be entered into an appropriate Data Center for archival storage, catalog, and browse file maintenance, and distribution.

Data Centers will be characterized by a staff having scientific expertise in one or more disciplines, the location of archives of Level 1A and higher-level data sets, the capability for scientific data processing and analysis, maintenance of data catalogs and browse file entries for all data resident at the Center, and above all, the ability to retrieve and distribute archival data and information.

Active Data Bases

Active Data Bases will be facilities where focused, extended scientific analyses are conducted with archival data derived from a Data or Instrument Operations Center. Value-added processing (e.g., data reduction) will be performed at these sites and the resulting data, together with directory, catalog, and browse file entries, will be transmitted to an appropriate Data Center for long-term storage and distribution. Additionally, directory and catalog entries for these data sets will be generated and transmitted to the lead Data Center.

Non-Eos Data Bases

Many Eos scientific objectives require access to non-Eos data bases. This architectural example provides for a complete bidirectional interface between any number of archives and the Data Management & Communications Network. Thus, researchers accessing the system through any other component will have transparent access to pertinent non-Eos holdings. Similarly, Eos researchers working through the auspices of these non-Eos Data Bases could have full access to all system services.

Data Management and Communications Network

The Data Management and Communications Network will link together all other system elements and subsystems, in a manner that is transparent to a user. Its goal is to allow researchers to conduct mission planning, request and receive data, interrogate directories and catalogs, and browse data sets remotely and interactively. The success of Eos is therefore directly tied to the effectiveness of this information Network.

The Network should be managed by Data Center personnel, thereby ensuring that the Data Centers are involved directly in this key activity. The performance of the Network should be evaluated periodically by personnel who lead activities of instrument teams and Active Data Bases. Thus, since the selection of Active Data Bases and instrument teams will be governed by the scientific community through the peer-review process, those selected will have a significant vested interest and will act to ensure the utility of Data Centers and the Data Management and Communications Network through direct oversight responsibilities.

Transparent Data Handling

The scientific and operational environment envisioned for the Eos era (viz. Chapter II) suggests quite strongly that transparency in data handling will be a key factor and fundamental attribute of an Eos data and information system. In this example, the functional configuration is based on standard protocols that support the integrated transmission of many types of user data (e.g., instrument and engineering data, computer memory exchange, text and graphics) through common, limited-capacity data channels. It is also compatible with commercially supported terrestrial standards for open system interconnection. These characteristics are achieved by using two CCSDS standard application data structures (the source packet, and the standard formatted data unit) and two CCSDS data link structures (the telemetry transfer frame and the Reed-Solomon telemetry codeblock) throughout the system.

Data entering the system utilizes either a source packet, or the standard format data unit (that contains either sets of raw packets, or processed results). The format for transmitting data bidirectionally through TDRSS links is a telemetry transfer frame, that may or may not be encoded using the Reed-Solomon error-correction algorithm.

Selection of a telemetry transfer frame as the standard data link structure for bidirectional use on data channels has significant ramifications. The frame is optimized for efficient channel utilization. It is a fixed-length data structure that has a "natural" size of 10,080 bits, a convenient quantity of data to be handling on links operating at multi-megabit rates. Additionally, it is organized around the concept of virtual channels, that provides a mechanism for segregating different types of data, transmitted serially through a common data channel, into several logically parallel paths. This permits a high-rate serial stream to be immediately split into many parallel, lower-rate data handling processes at a receiving facility (using relatively unsophisticated hardware). This significantly reduces requirements for developing high-rate processing technology.

The telemetry transfer frame is optimized to fit within a high-performance Reed-Solomon code-block structure. Because this is a block-oriented code, all of the Reed-Solomon parity bits are simply appended to the end of the frame. Therefore, some frames can be transmitted with the Reed-Solomon parity bits attached (permitting virtually perfect-quality frame data to be received after decoding), while others can be tranmitted without Reed-Solomon protection (in which case a frame will have the TDRSS channel bit error rate of $1x10^{-5}$). Thus by selectively encoding some frames, and transmitting others uncoded (saving 15 percent coding overhead), different data qualities can be provided for dissimilar data types being transmitted through a common channel.

FUNCTIONAL ARCHITECTURE: PEER LAYERS

Redrawn (Figure 2), the architectural schematic reveals the "peer" layers of data handling. The standard protocol data units, which provide the mechanism for data exchange across the layers, are shown

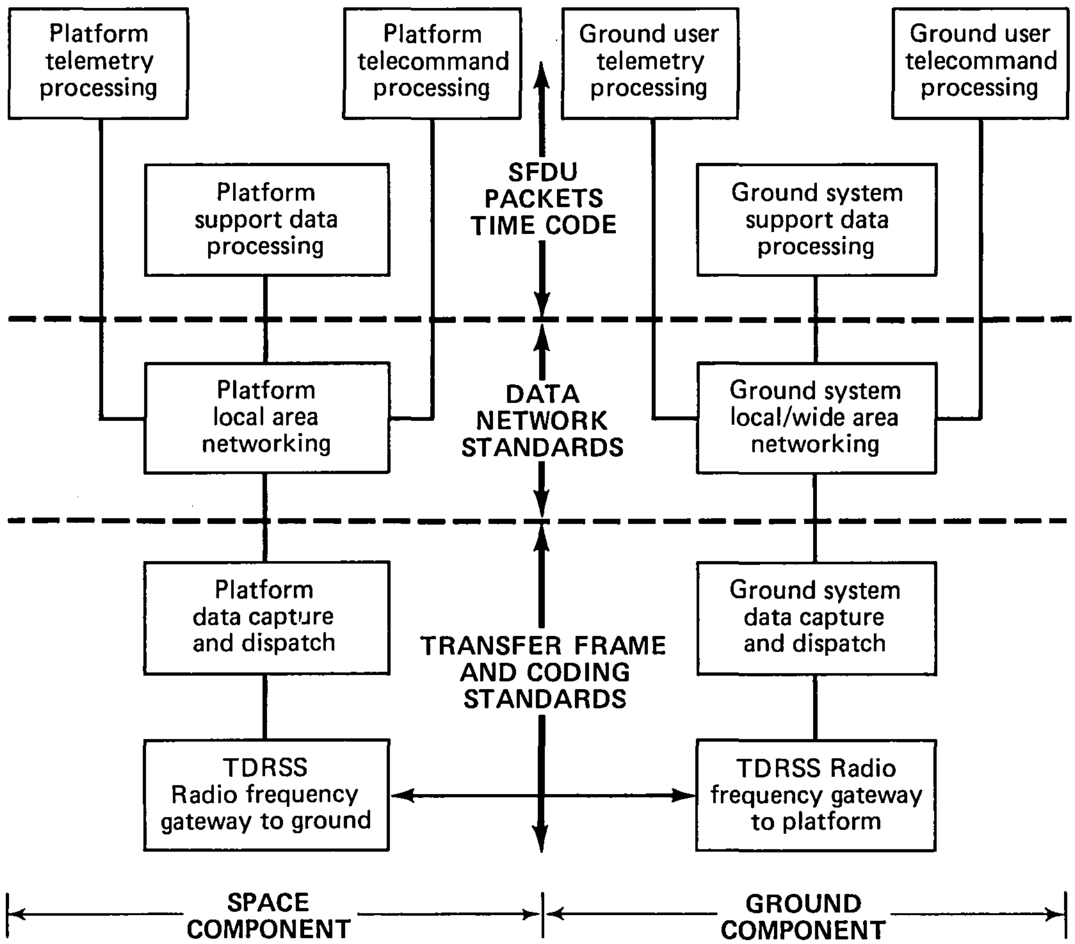


Figure 2. Peer protocols within the data system example.

as well. Although for brevity, the following discussion is limited to the space-to-ground data transfer link, it is important to note that identical, mirror-image techniques are used on the uplink path and are applicable to both sides of the ground-based network. This self-similarity in data handling methodology, employed throughout the architectural example, is easily seen.

The concept used here of transmitting bidirectional data streams through space-to-ground links is slightly different than conventional techniques used by other spacecraft. Within the Eos system, the uplink and downlink data streams are similar; both contain low- to medium-rate instrument and engineering data and computer-to-computer data exchange, and the physical characteristics of the uplink and downlink are virtually identical. Data retransmission protocols known as command operation procedures can be incorporated to provide fidelity and highly reliable delivery for both uplink and downlink data traffic that cannot tolerate data outages (e.g. data base transfers, memory loads).

IV. MANAGEMENT CONSIDERATIONS

There have been several recent reports published that deal with various aspects of data management problems in the space sciences. Three of these reports are particularly apropos and should be carefully studied and followed by NASA management as the Eos data and information system evolves. These reports are the National Academy of Sciences Space Science Board, Volume 1: Issues and Recommendations (1982)[2] and Volume 2: Space Science Data Management Units in the 1980s and 1990s[3]; and Space Research Data Management in the National Aeronautics and Space Administration[4].

These reports examine many features of past data management practices and lay a firm foundation for minimizing problems with future systems. Volume 1 from the Academy's Committee on Data Management and Computation develops a set of seven principles for successful scientific data management; this panel fully endorses these principles and urges NASA to closely follow them in the future. We therefore consider the recommendations contained within these reports to be appended to those contained herein.

We note that these reports consider a spectrum of past difficulties and delve into the future to the extent that projects in years to come will bear resemblances to those of the past. We envision an Eos that markedly differs in many ways from prior NASA projects. Consequently, rather than dwelling on past sins, we deal in this chapter with aspects of Eos that are unique and will require careful thought, consideration, and appropriate attention if the Eos data and information system, hence Eos per se, is to be a success.

The unique features of an Eos data and information system provide a suite of management issues that fall into two categories, those pertaining to program and project management, and those involving information systems management. We deal in this section with both classes of management issues. The distinction between the two categories is, in most cases, readily apparent.

LONGEVITY

A crucial difference between Eos and previous data collection and analysis projects, which requires procedural changes, is the intended length of this effort. Current practices for projects of limited duration (up to five years) are determined by and for a single group of engineers, programmers, and scientists who participate throughout the lifetime of the project. Many current efforts, having lifetimes of nearly 10 years, already show the difficulties of changing team members and lack the flexibility to embrace new technologies and methodologies. The fate of data sets acquired by these projects is generally to fade rapidly from memory. Reasons for this problem include:

1. Lack of useful documentation. Changes of personnel or dissolution of groups cause loss of knowledge of instruments and data characteristics, of calibration and processing procedures, and of data formats.
2. Lack of planning for hardware or personnel changes. Data quality monitoring procedures are often nonexistent and necessary software changes are not controlled or documented adequately.
3. Perception of data archives as passive storehouses that at best only dispense data. Little that is learned about the data is returned to the archives, hence data value can only degrade with time.

The success of Eos depends on changing existing practices in data collection and analysis, overcoming these problems.

Although Eos is essentially a large, long-term data collection, processing, and analysis project, its data are intended to serve as a dynamic resource for research on global phenomena. Hence, unlike data processing in many projects, where value resides in the final product, value in Eos is distributed over many stages of data processing, all of which must be retained by an Eos data and information system. Furthermore, retrieval of much information is a matter for experimentation; fixed algorithms are difficult to define. Yet, many steps in the processing of Eos data will be common to all analyses; these steps need to be identified and performed once. These characteristics of Eos suggest several new principles for the data system and its management that should be included in its design so that research access to the data is facilitated for years to come. These principles are:

1. Archival activities should occur at several stages in the data processing system, representing "raw" data, processed observables, and inferred or derived quantities. The definition of these stages depends on the nature of the observations and processing algorithms. If calibration and navigation information are accurate and readily available, "raw" data retention and maintenance may be avoided. Processed observables refer to raw data converted to the physical quantity directly measured by the instrument (i.e., spectral radiance). Inferred quantities are those that are derived from directly measured quantities through application of auxiliary data or processing models.

2. Data should be arranged into stable, predictable units to allow for automated cataloging and easy user access to very large data volumes.

3. Processed observable data should include complete calibration, navigation, and other ancillary information to facilitate use and to avoid repetitious processing. This and other pertinent information should be appended to the data with convenient frequency (e.g., per image).

4. Catalog documentation should include results of automated quality assessments and a trace to cross reference other studies that have learned something about the data.

Automated Data Processing and Management

Much of the documentation chore can be handled by the addition of appropriate software for data processing that generates and accesses summaries of data attributes (i.e., self-documenting software). There are three types of software needed: Eos-specific, non-Eos, and user processing software. The Eos data system should be automated to handle the large data volumes effectively. Proper quality control and data management require that the processing management software obtain and verify many attributes of the data; hence, recordkeeping and generation of summary reports should be added functions. The management software should determine the source and contents of data sets, record what processing has been performed, where the data are stored, and form a cross reference list of data coverage, time, etc. In other words, this software should prepare a catalog entry containing this information. Any "use" of the data should also be monitored and recorded by the management software. Acquisition of non-Eos data sets for Eos purposes requires processing of these data to allow equivalent catalog entries to be made.

Planning for Change

The longevity of Eos and its data and information system, together with practical considerations of equipment, computer hardware and software, and personnel, requires that the data system must develop by evolution and that changes to the system be planned rather than ignored. Design of the data processing and management software and the associated institutional arrangements should incorporate procedures for changing EVERY component, including the human element. Efficiency is not as important as flexibility and clarity of design.

1. Changes of personnel will occur. The documentation of the total system, especially when highly automated, should be thorough to allow ready training of new operations and management personnel. Knowledge of the system should be independent of the personnel.

2. Instrumentation will change (e.g., improve). Data quality and catalog software should be able to detect and record these changes. Data analyses that lead to change should also be part of the Eos archives.

3. Computer hardware should change when necessary to keep overhead low and to improve system capability. Software and procedural changes triggered by such hardware changes should be planned; benchmark tests and standard data sets for reprocessing should be defined to verify the operation of the whole system after such changes occur.

4. Software changes will occur to accommodate new data, new instruments, and improved analysis algorithms. Design of the data management system should acknowledge this likelihood. All of these changes should be controlled and documented. Management arrangements should ensure that operational software is secure, change procedures clearly defined, and changes are made using a self-documenting, monitoring system.

5. Human-machine interfaces in the data processing and management system should be designed to maintain proper oversight and quality control. Human tasks should be designed to be interesting (not too lengthy, not fatiguing, not too repetitive). Tasks should require problem solving and judgment. One good way to ensure this type of interaction is to include users in the management system such that their knowledge is effectively included in data quality assessment. The processing and archival institutions should have associated researchers with vested interests in the data per se.

6. Use of expert systems in the management software should not overlook the utility of incorporating various "scientific" analysis algorithms. Even though simple or standard algorithms may not be valid for all scientific problems, such algorithms can be usefully applied for auxiliary quality assurance purposes by assessing data patterns and characteristics that can be used to monitor data attributes. For example, a sea surface temperature retrieval method, which is inaccurate for climate research, may be sufficiently accurate to monitor data quality, since expectations about the behavior of a physical quantity like sea surface temperature (e.g., no rapid changes) can be utilized for quality control.

7. Data catalogs should consist of online, electronic data sets that can be amended to accommodate: new data sets, new attributes for existing data sets, new cross references among data sets, processing and use trace, and bibliography.

8. Data holdings can change in several ways other than addition of new data sets: (a) processing can provide an alternate form that has been sorted (mapped), edited, or labeled by analysis, (b) processing can produce associated, inferred quantity data, and (c) processing of several data sets can produce alternate, correlated data sets. Planning of self-documenting data catalogs and a directory and holding structures should accommodate this inevitable evolution.

Institutional Commitment

Considering the above points, in addition to the requirements for incorporating scientific results into the archives we conclude that the role of data processing and archival institutions will be more active than in past or current practices. These institutions need to be part of the research process so that interest in the data and vigilance over the system will remain high. Provision of the many services outlined here is not a simple or easy task; appropriate institutional rewards must be found. Finding a solution to this problem must take account of several characteristics of the current practice of science:

1. Publication of data and its documentation in refereed scientific literature is not possible, especially for the large satellite data sets. Hence, scientists do not spend much time on data structure or documentation.

2. Quality assurance and formal acceptance procedures for large data sets do not exist, and reliance on individual scientists is inadequate.

The usual tools of scientific quality management, namely peer and publication reviews, are inadequate for creating the Eos data base needed for future Earth sciences research. This suggests that the role of the processing, management, and archival institutions should be extended to include participation in research projects and the publication of data documentation. Consequently, some institutional publications might well be elevated to full journal status by instituting review procedures similar to those of the refereed literature.

SYNERGISTIC CHARACTERISTICS

Coordination will be the keystone of a successful Earth Observing System and the modus vivendi of the data and information system. Unlike many flight projects of the past, virtually every facet of Eos, hence its data and information system, will require coordinated interaction to preserve its synergistic characteristics. Synergy within Eos occurs at three levels, all pertinent to the data and information system. They are: the scientific, project, and program levels.

Scientific Level

On the scientific level, we anticipate system resource conflicts arising that will demand resolution within the confines of the data and information system. Multidisciplinary researchers and research teams will have needs for particular observational sequences, while disciplinary researchers may well have requirements for entirely different measurements. Both groups will be affected by spectacular environmental events and the pressures (both scientific and political) to respond.

These scientific conflicts could easily (and will likely) have an impact on the limited resources of the data and information system. While they may ramify and pervade the system, the areas most likely affected will include command management, instrument and mission operations, network services, onboard buffers, and data processing services. Since these services and facilities will have limited performance and capacity, a fast, efficient, and effective mechanism must be established to maximize researcher benefits when dissimilar scientific objectives collide.

Project Level

Resource conflicts at the scientific level directly translate into concerns at the project level. A major task of project personnel will be resource management, assuring continued, uninterrupted performance of the entire data and information system. Clearly, this will neither be a trivial task nor will the results of its performance be scientifically inconsequential.

There is an additional concern with resource management that heralds difficulties for the data and information system. This is synergistic instrument operation. While the Eos Science Steering Committee has discussed a number of synergistic instrument scenarios, we do not presuppose that these are totally inclusive. As the research community increases its knowledge base, new requirements for multiple-instrument observational sequences will be identified. Preexisting requirements may remain and circumstances be further compounded by newly deployed instruments that may well place greater demands on available resources. The data system design, implementation configuration, and operation must be based on preserving flexibility to the greatest extent possible, maximizing scientific productivity to its fullest measure.

Program Level

At the program level, resource conflicts take on a somewhat different character. To a large degree, the data and information system will be the key to ensuring Eos' scientific success. Consequently, fiscal resources must be identified and preserved to ensure its full and complete implementation, regardless of the inevitable pressures to do otherwise. This task will not terminate with the deployment of one or more Eos platforms. Rather, as new instruments are proposed (and subsequently deployed), and new synergistic groupings of instruments are identified, adequate funding must also be secured for enhancements to the data and information system, when required. In doing so, the fidelity and performance of the system may be preserved to the ultimate benefit of scientific research.

DATA VOLUME AND RATE

The Eos data and information system will be required to handle daily more data than any system ever conceived. In general terms, Eos will produce several orders of magnitude more data per day and is envisioned to have a duration exceeding any mission ever before proposed. Today, the only space mission at all close to projected Eos data rates, hence volume, is Landsat. While Landsat processing capabilities have been improving, the resultant data remain relatively inaccessible and are used by a rather small number of individuals compared to projections for Eos. Clearly, the operation of an Eos data and information system will create management problems of a magnitude that cannot even be fully appreciated at this time by either NASA management or the scientific research community who must cope with these data in their research.

There are numerous management issues associated with high data volumes and rates that can be clearly identified at this time. They involve virtually all of the recommended functional characteristics that the Eos data and information system should exhibit. Since the Eos data and information system will of necessity be a limited resource, it is reasonable to expect that the major issue associated with high Eos data rates and volumes will be scheduling, including scheduling of data acquisition, data set production, data access, and distribution timing.

The architectural example discussed in Chapter III of this report concentrates on data flow and transparency. We have recommended the use of expert systems to the extent possible to reduce delay times. Yet, we anticipate that even with these advanced technologies, limitations in telecommunications bandwidth, computational processing power, etc. will generate a significant number of scheduling problems. Further, it is likely that the synergistic nature of the overall mission will both solve what might otherwise be a complex scheduling problem, and create yet others, particularly if the Eos platform(s) host instruments of an operational genre together with a suite of individual, principal investigator, and facilities-class instruments.

We know that existing systems, such as Landsat, are inadequate, and we know from a management perspective that an Eos data and information system must be vastly superior; we do not know the specific solutions to the myriad management problems that these high data rates and volumes will create. We therefore recommend that the Eos data and information system be built in an evolutionary fashion, enabling solutions to evolve along with the technology and expertise to cope with this new era in space research.

ARCHIVAL DATA SETS

A new objective unique to Eos is to increase our knowledge base, represented by data archives, by persistently adding the results of scientific data analyses to the basic observational data holdings. This continuous evolution of information content requires redefinition of the roles of scientists and archival institutions as integral elements of this system (which is conceived as a network tying users, dispersed data processing, and archival functions together with a central data directory and network management).

Individual investigators or scientific teams have new obligations to accomplish this knowledge increase. In return for access to the data system, these investigators must return their results to the system. Whether this obligation is accomplished by providing system access to their individual holdings or by physical transfer of data sets to project archives, additional processing of the data resulting from an investigation will be required to provide:

1. a catalog entry containing descriptions (defined by Eos) of data sources, data properties, analysis methods, and attributes (e.g., location, time, wavelength);
2. a standard format to allow access from Eos software, and processing by Eos archival software;
3. documentation of data set contents, processing algorithms, instrument characteristics; and
4. an evaluation of the results, including error analyses and validation tests, as well as a relevant bibliography.

Archival institutions must be active participants in the data processing that supports scientific studies rather than passive storehouses. This new role requires data processing and distribution functions, in addition to the usual data acquisition and holding functions (i.e., storage, security, purge). Data processing will be necessary to:

1. put data sets into any Eos standard formats or develop software interfaces between data in non-Eos formats and any Eos standards;
2. collect simple statistics of the data that can be used to assess data quality;
3. examine data or apply simple algorithms to develop attribute lists used for search purposes;
4. compare data attribute lists producing cross reference entries in the data catalogs; and
5. prepare browse data sets by sampling the original data volume.

Data distribution functions will be facilitated by:

1. development and maintenance of data directories (for all pertinent data sets) and data catalogs (for data sets held by Eos);
2. production of browse data sets;
3. establishment of peer review procedures for data quality; and
4. publication of data documentation catalogs.

FLEXIBILITY IN DESIGN: FUTURE USERS

The Eos data and information system is being designed as part of the Eos project by evolution of a dispersed data processing network. Since the network is intended to provide flexible access to large data processing volumes by many project-related users, its design will have potential for other users entering the system at a later time. Thus the Eos data and information system is meant to continue the process of data analysis and research beyond the realm of the data collection phase of Eos alone. Provision should thus be made in the network design for expansion of data access to and for acquiring new data sets from two classes of users.

The smaller user class includes individual scientists or small groups investigating specific phenomena, incidents, or specific geographic locations. This type of user, as a group, will access data directories, data catalogs, and browse data very frequently but request only modest volumes of data. The network design must provide proper access security and usage accounting, allow remote access to directories, catalogs, and browse data, provide usage statistics for accounting, and allow ordering of data. The Eos data and information system design may allow a user to order non-Eos data that is subsequently reformatted by Eos before being transmitted to the user.

The larger user class is composed of consortia of investigators formed to undertake specific, large research projects external to Eos (e.g., a special field study may wish to acquire supporting data being collected by Eos or a retrospective study may be initiated to produce a climatological data set). This type of user will access data directories, catalogs, and browse files relatively infrequently but will request rapid or scheduled delivery of large data volumes, possibly for long periods of time. The Eos information network should be designed to allow expansion (paid for by the user) to support additional large demand uses. A key feature of the Eos information system, in this case, is access to data quality assurance and calibration information, in addition to the basic observation data, and provision of appropriate software interfaces within the network.

ECONOMIC FACTORS

Most academic research is funded with an assumption that data, once acquired, will be exchanged through publications and by other means where the cost of obtaining the data by other investigators is essentially negligible. In essence, scientists exchange data for personal recognition, which translates into professional growth, increased income through advancement, and peer recognition. Thus, grants for space research are often dominated by salaries of scientific personnel compounded with institutional overhead. Computer costs may also be substantial, but an approach being adopted more and more often is for a funding agency to provide a shared-use computer facility and allocate time to its investigators at no cost to the researcher. This tends to encourage full use of any large-scale machine, and the pattern is generally that new computers are fully subscribed within one to two years.

The traditional procedure for obtaining Earth remote sensing data for research purposes is somewhat analogous to the super computer situation. Investigators propose to participate in space missions and, if successful, receive access to data at no cost. Copies of archival data are made available at the marginal cost of reproduction. Under this system, the aggregate scientific community works to justify funding for a major space program or mission and then receives the resultant data at marginal cost. In essence, if the promise of future knowledge is sufficient, then present professional stature is exchanged for access to data at little direct cost. Full exploitation of the data is encouraged because the only barriers to obtaining data are successful peer evaluation leading to selection for participation in a mission or research program.

These considerations have direct bearing on the financial model that should be used in operating Eos. The current research system will continue to function if data are provided at costs not exceeding the marginal cost of making the data available. However, an alternate system, which requires those obtaining data for research to pay a proportionate share of the capital cost of acquiring the data (e.g., Landsat), will necessitate major changes in the way research is funded. A full-cost recovery system will work against the interests of maximizing use of Eos by discouraging access to data, and further is diametrically opposed to the unique feature of Eos that requires researchers to return reduced or derived value-added data sets to the archives.

The scientific community and its sponsors strongly support a system where the value of implementing a mission is judged in advance, and access to its data is governed by a peer review selection process. Under this system, institutional capabilities such as a data system or spacecraft are funded as part of the decision to proceed with the activity, not recovered directly from the researchers utilizing them. Consequently, it will be more straightforward to successfully implement Eos and ensure its maximum utilization if monetary charges for data are limited to the marginal cost of reproduction.

Although this economic model will work adequately for some investigators, it will not work for all. Many researchers will be contractually obligated to provide value-added data to the archives. If a researcher is required to pay even the marginal cost of reproducing the initial data he receives, then it is reasonable to expect that he should receive financial remuneration for processing, copying, and providing his derived or reduced data to the archives. Unless these costs are considered and included from the outset of proposal preparation, the task of returning these data to the archives will necessarily require the use of an individual scientist's research funds; hence, the task will be of low priority.

Thus, some adjustments to the existing economic system will need to be made. We anticipate that these changes will not require a major restructuring of the system, but rather note that the problem will likely arise and recommend that NASA management take all necessary steps to ensure that priority is given to the acquisition of value-added data from Eos researchers.

INTERACTION WITH OTHER GOVERNMENTAL ENTITIES

Unlike many other space research missions, Eos should provide a means for addressing a suite of multidisciplinary research problems. To do so requires access not only to data derived from Eos spacecraft, but also to other archives, many of which are operated and maintained by other governments and governmental agencies. Similarly, we anticipate that Eos will enable multiple data source, disciplinary research to be more effectively conducted. Here too, many of the requisite archival holdings are under the purview of other governmental entities. Clearly, to achieve the scientific objectives of Eos, researchers will need easy access to directories and catalogs, as well as to the data per se resident within these external archives.

The technical challenges of efficiently and transparently linking heterogeneous data bases are not beyond the capabilities that should be readily available during the 1990s. The problems arising from the multidisciplinary scientific objective will be managerial in nature. Project management's principle task will be to effectively ensure that the directories and catalogs that researchers access are updated on a very frequent (perhaps daily) basis. Thus, inter- and intra-governmental communications will likely be the deciding factor in determining whether a researcher's data needs are met.

On a program management level, the problems generated by this archival access issue become political in nature. We anticipate the need for multiple interagency agreements detailing access procedures, updating methodology, responsibilities, funding, etc. On the international, intergovernmental level, similar negotiations must take place and agreements be defined. In both cases, we anticipate that these external governmental entities will request reciprocity in archival access. Since many research scientists and colleagues are in the employ of these external organizations, we expect that many will likely be collaborating with project-sponsored investigators. Clearly then, reciprocal access will be neither a technical nor a scientific issue but, rather, political.

On this basis, we urge NASA to undertake appropriate negotiations with those governmental bodies responsible for the operation and maintenance of archives holding data pertinent to Eos scientific objectives. The resultant agreements should be consistent with the access requirements delineated within this report.

REWARDS AND PUNISHMENT: A PERSONNEL DILEMMA

If the scientific community is to have easy access to existing, planned, and future data sets, the Eos data and information system must fully utilize advances in data system software and hardware. In order to completely integrate these improvements into the Eos data and information system it will be necessary to greatly expand upon and upgrade existing human resources involved in the fields of data management, preparation, maintenance, and scientific validation. The success of the data and information system requires that these professionals be recognized, encouraged, and appropriately rewarded for their efforts.

Many traditionally trained scientists and engineers lack the inclination or ability to adequately perform the tasks necessary to manage, produce, and maintain data sets that are useful to the greater scientific community. These tasks include scientific quality assurance, validation, algorithm maintenance, and, of particular importance, documentation.

Employment in the data management sciences is currently rendered relatively unattractive because workers receive fewer promotions and other types of recognition than coworkers in the more traditional scientific disciplines. Consequently, the field has not attracted a large number of individuals of the type and caliber that Eos will most assuredly need. There are several reasons for these differences in recognition and rewards, but primarily they stem from the nature of the work; data management scientists do not produce the same type of publications as those engaged in traditional scientific endeavors. Data management publications are produced less often and do not generally appear in the refereed scientific literature. Indeed, the major derivative of data management is the data set itself, which is merely described by publications or, more often, user's guides and other types of similar documentation. Throughout the scientific community, recognition of professional accomplishment and excellence is primarily related to publications in refereed scientific journals. A major step toward encouraging highly qualified and motivated scientists to enter and remain in the field of data management would be to broaden the base of what is recognized as a truly significant scientific publication. This then remains the guide for measuring professional performance, but should include publications of technical memoranda (which describe data sets), user's guides, algorithm and validation study results, as well as the data sets, per se.

KEY ISSUES

Although there are many considerations, both technical and managerial, noted within this report, there are two prime issues that NASA management must address at this time. They are (i) the need for immediate action to plan, design, and implement an Eos data and information system, and (ii) the need to develop the experience, expertise, and technology necessary to ensure that the resultant system effectively meets the needs of the research community. One means of addressing these issues is to initiate a top-down functional design effort that is very closely coupled to and iterated with prototype experience to efficiently converge on an optimal system by the beginning of the next decade. Regardless of the approach taken in addressing these issues, there are two imperatives or guiding principles that should be followed throughout the evolutionary process. They are (i) scientific involvement and (ii) scientific oversight.

The prime objective of a scientific mission is to obtain new knowledge and understanding. Therefore, it is reasonable that since data are acquired for scientific research, scientists should play a significant role in determining what data are required, how they are acquired, how they are processed and disseminated, and the disposition of the data after they are collected. Scientists must be actively involved throughout the Eos data and information system evolutionary process to ensure production of and access to data sets of the highest quality. This involvement will maximize the return on investment and improve the quality of the resultant data.

Management of scientific data bases has been the responsibility of project personnel, principal investigators, and project archives (such as they are). Those scientists outside of the project per se who actually use the data for research purposes have not been actively involved in this process. These scientists should be involved at the outset via an oversight and advisory function, since the most successful examples of data base management involve user oversight.

V. RECOMMENDATIONS AND CONCLUSION

The goal of the Eos data and information system must be, by the time Eos spacecraft are returning data, to meet the challenges of Eos mission operations, data transport, processing, and data management, in addition to the challenges of access to information and data not under direct control of Eos or even NASA. An Eos data and information system must be a system that includes geographically distributed sites of varying capabilities and responsibilities. We expect that by the 1990s local processing capabilities, combined with network technologies, will allow such a geographically distributed system to become a reality. In fact, we envision the key objective of an Eos data and information system to be providing remote and interactive electronic access to the variety of capabilities and services that the system offers. We consider the management of this data and information system to be considerably more difficult to implement successfully than the technological aspects.

OPERATIONAL AND COMMERCIAL USERS

We recommend that the Eos data and information system be designed to meet research needs, but that it also be used to meet the needs of operational (e.g., NOAA, DoD) users, to the extent that such needs do not detrimentally impact use of the system for research. Commercial users, likewise, should be accommodated to the extent that their use does not deter Eos research and to the degree that commercial use costs can be recovered. If operational users significantly impact Eos operations, then the relevant operational entity should fund or provide the enhancements needed to meet their needs. This recommendation necessarily implies that the information system be designed in a manner that ensures flexibility, allowing enhancement without affecting operation of the remainder of the system. We believe that a modular, flexible architecture such as described in Chapter III of this report has the potential to meet this as well as many of the other requirements presented herein. Consequently, we urge NASA to explore the adoption of a modular, transparent architecture for the Eos data and information system.

FUNCTIONAL ELEMENTS: THE 1990s

Within this report, there are a number of recommendations, requirements, attributes, and characteristics pertaining to detailed functions and capabilities of the Eos data and information system. An annotated listing of requirements is presented in Appendix II and grouped conveniently into seven separate categories. Figure 3 schematically portrays these functional categories and their interconnections.

The seven functional categories or groupings include flight systems, user, operational, information services, advanced data base management, Eos data processing, and non-Eos data base functions. Flight systems include the functions and characteristics of both remote sensing instrumentation as well as the

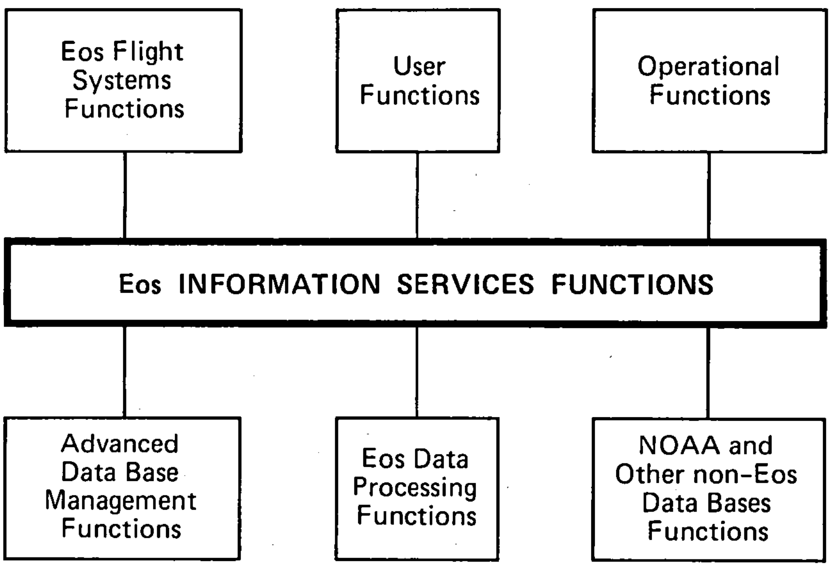


Figure 3. Eos data and information system - Generic functional groupings. The schematic interconnection of these elements indicates the dependencies of any one task upon other functional domains.

onboard data system. User functions embrace the envisioned modi operandi of researchers, operations personnel, instrument scientists and teams, and other individuals requiring access to Eos services. The functions and characteristics inherent in spacecraft and data and information system operation are included in operational functions. Eos information services functions include the suite of network services (both space- and ground-based) that we envision during the Eos era. Many of the new functional requirements for the data and information system are grouped under advanced data base management, while processing requirements for Eos data, per se, are grouped separately. Similarly, the needs for access to and data from non-Eos sources are encompassed within the final functional grouping.

Unlike many space research missions, the requirements levied upon Eos as a whole are such that these seven functional elements are very much interdependent upon one another. Tasks contained within one element may not be accommodated solely within the domain of that element but rather may require input from one, two, or perhaps an even greater number of other functional domains. Thus, for Eos to be successfully implemented, all of the data and information system functional requirements delineated in this report should be provided.

Consequently, we recommend that the Eos data and information system of the 1990s include the functional characteristics depicted here in elemental form, identified throughout this report, and summarized in Appendix II.

INFORMATION SYSTEM EVOLUTION: THE 1980s

Many existing scientific research projects and pilot data system studies are considering aspects of problems that must be solved to implement successfully a data and information system supporting the project and program needs of Eos. Therefore, progress toward an appropriate information system can be obtained by focusing existing efforts, stimulating new scientific uses of data, and coordinating an iterative learning process that is directed toward evolutionary or "test-bed" information system concepts. These activities (many already begun or planned) should be focused and used to generate experience with information management problems and their optimal solutions, and provide the nucleus of a functional Eos data and information system.

The conceptual structure for an Eos information system that underlies many current activities is that of a central coordination and information management function, together with access to data holdings located at a number of data archives. The recommendations noted below are meant to focus study on the problems of defining the functions and structure of an Eos data and information system by attempting to carry out these functions on a smaller scale. The key lies in trying several solutions to the various problems so that planning for the resultant system will be based on experience. To secure the enthusiastic participation of researchers and managers, the motivating factor for these information management studies should be real scientific investigations that demand solutions to many of the problems and functions outlined in this report.

Expansion and Focus of Pilot Data Systems

We recommend that NASA's current Earth science pilot data systems be continued and expanded (as appropriate) in their representative disciplines. Additionally, we recommend close collaboration with the UCAR Unidata initiative, which provides a similar focus for the atmospheric sciences. These efforts, each in a unique fashion, are addressing many of the technical and managerial questions that must be answered to design an Eos data and information system. Future efforts should address four areas: (i) develop data formats and interface software to allow access to heterogeneous data from any of the existing pilots; (ii) develop (with researchers) new analytical tools for multidisciplinary research; (iii) improve data sets in pilot archives by developing improved documentation and formats that cross traditional Earth science boundaries; and (iv) produce an electronic data directory for all NASA Earth science holdings, and similarly electronic catalogs for all Earth science pilot project holdings.

Some or all of these objectives have already been identified by individual pilot efforts, but they should be focused with an Eos information system thrust, leading to progress toward Eos. A key feature must be strong interaction and collaboration with ongoing research efforts that emphasize the collection and analysis of multiple data sets. The evolutionary Eos data and information system should work with researchers to facilitate archival access, and the researchers should return reduced and value-added data and analytic results to the archives. This interaction will allow the information system, the pilot systems, and analysis efforts to evolve with a new level of sophistication. Developed software should be considered part of the archives. The Eos data and information system as well as the pilots can be structured to properly manage information while stimulating research.

We further recommend that the Global Resource Information System concept be focused on Eos objectives, then utilized for developing the information network aspects of the Eos data and information system. Particular attention should be given to developing and utilizing technology to allow transparent access to heterogeneous data bases, to advanced data base management software with intelligent (expert systems) search capabilities, and to cost-effective network capabilities, including direct data broadcast. To the extent applicable, guidance from the Space Physics Analysis Network activities may be both desirable and worthwhile.

Science Projects to Focus System Evolution

We recommend that the Earth Science and Applications Division fund a limited number of multidisciplinary research teams to investigate a select number of crucial scientific objectives in hydrology, biogeochemistry, or climatology, using currently available data sets. These research activities would foster multidisciplinary studies of Earth, studies that require access to multiple, diverse data sets. These research teams will be responsible for using the evolving Eos information system, collecting data from external archives, analyzing the data, and returning their results and experience to the embryonic information system. The link between multidisciplinary research projects and information management systems can provide the experiences necessary to move toward an operational Eos data and information system in the next decade. The key is selecting teams to do this type of research now, using current data and archival systems to define where many of the problems lie and what the solutions might be. Finally, these scientific projects can be utilized to identify key data sets that need to be retained, maintained, and organized in a time series data base format for comparison with Eos data in the 1990s (e.g., preservation of the Advanced Very High Resolution Radiometer archives).

CONCLUSION

The final and perhaps most significant recommendation of this panel is to initiate the planning and implentation of an evolutionary Eos data and information system without delay. A functional system providing the means through which Eos data can be most fully utilized will not be built in a matter of a few years; it must be allowed to evolve along with the increasing knowledge bases of the Earth and data management sciences.

There are two fundamental principles that should be followed throughout the Eos data and information system evolutionary process. They are: (i) Involve the scientific research community at the outset and throughout all subsequent activities, since the data will be acquired, transmitted, processed, and delivered for scientific research purposes; and (ii) provide the researcher with an oversight and review responsibility, since the most successful examples of data base management rely on the active involvement of scientists.

REFERENCES

1. Science and Mission Requirements Working Group for the Earth Observing System, "Earth Observing System," Volume 1, Part 1, NASA TM 86129, 1984.

2. National Academy of Sciences Space Science Board, "Data Management and Computation," Volume 1: Issues and Recommendations," National Academy Press, Washington, DC, 1982.

3. National Academy of Sciences Space Science Board, "Data Management and Computation, Volume 2: Space Science Data Management Units in the 1980s and 1990s," National Academy Press, Washington, DC (in press).

4. Ludwig, G.H., "Space Research Data Management in the National Aeronautics and Space Administration," Unpublished consultants letter, 1985.

APPENDIX I - RESEARCH SCENARIOS

It is useful to consider how scientists in various disciplines would utilize Eos and its data to address research problems. In doing so, the Data Panel developed requirements and system attributes by considering examples of research activities, most of which parallel the scientific objectives delineated within the Eos Science and Mission Requirements Working Group report. Necessarily, these scenarios focus on current activities and methodologies. However, we do not envision major variations in the ways Eos investigators will approach problems in a functional sense. Therefore, these scenarios provide a background for developing Eos data and information system requirements.

CLIMATIC RESEARCH

This scenario portrays information system requirements for climate research, with a focus on atmospheric phenomena. In climate research, not all of the quantities to be derived from the data are known beforehand. Indeed, a main purpose of the data analysis is to find and diagnose patterns of behavior; the knowledge of climate therefore grows by accumulations of multiple descriptions of the same data with special emphasis on correlations between quantities. Summary statistics obtained from the data are compared with climate model performance to improve understanding and simulation capability.

Data Characteristics

The primary characteristics of a climate data set are its global coverage and temporal record length, but to be useful the coverage and length must have uniform properties. Research objectives, namely diagnosis of climate processes, also require sufficient space and time resolution (i.e., detail) to observe key atmospheric phenomena. Therefore, remote sensing climate data can be described as multispectral images with space and time resolution of 10 to 50 km and 1 to 12 hr., respectively. The dominant length scale of atmospheric motions argues for image scene sizes greater than 1,000 km ("imagery" means data taken rapidly over large areas and includes such data as temperature sounding data). Useful record lengths are from a few years to decades. The importance of radiation processes in climate is such that research interest in measured radiances will be equal to the interest in quantities derived from the radiances.

Data Access Patterns

Three interrelated types of data access are often used to attain climate research objectives: (1) case studies, (2) routine statistical analyses, and (3) correlation studies.

Case Studies

Case studies are analyses of specific, high-detail data sets to diagnose climate processes, to validate remote sensing data analysis techniques by comparison to other measurements, or to validate the physical interpretation of statistical patterns inferred from coarser data analysis. Case study data sets are generally limited in geographic and temporal coverage, but still of large volume because of the need for high resolution. Access to raw data is usually needed. To generalize results from case studies, an ensemble of similar data must be examined. Thus, interaction of the research scientist with the data archives involves catalog search to find data sets for specified geographic regions and times, catalog searches of many different data sets to find those cases with adequate data for intercomparison and validation studies, and browsing of reduced-volume versions of the data to find other examples of particular phenomena. Data must be cataloged according to place and time, instrument and satellite, and physical quantities measured to allow for correlation searches. Browsing reduced-volume data to search for specific kinds of events can be facilitated by an interactive expert system (i.e., a computer system that "learns" to sort data from other investigators who have already sorted the data). This system is possible only if the results of previous case studies and statistical surveys are made part of the data archives. Updating the catalog and archives and the menu of sorting criteria allows the use of data to become more sophisticated with time. More attention should probably be paid to implementation of this approach than to providing processing software to investigators.

Statistical Analyses

Routine statistical analysis of the data demands repeated access to long, continuous time series of measurements. These analyses can be considered as progressive sorting of data, producing statistical characterizations of content. Access must be to both original measured quantities and to derived quantities. Information on calibration, calibration history, and documentation of techniques used to derive quantities are necessary parts of the data record. Although most users would perform their primary processing locally, several routine utilities would be used if available: mapping, statistics, and sorting and classification. The first refers to provision of the data in a standard map grid so that statistics may easily be obtained. A mapping utility should allow not only a selection of the grid and its resolution, but also control of the procedures (accumulation, sampling, replication, interpolation) used to reproject the data. Simple statistical utilities would be useful to reduce the data volume sent to a researcher. For example, an investigator interested in interannual variability of synoptic cloud features might request data averaged over the diurnal cycle. Samples of the data may also be desirable. Sorting and classification utilities make it possible for a scientist to select from a long data record those events of interest (determined by simple physical criteria). Updates of these utilities based on previous research results allow for more sophisticated analyses subsequently.

Correlation Studies

Correlation studies can be either case studies or statistical analyses; however, the focus is on the correlation between many variables. This kind of study is most demanding of resources because it requires all of the capabilities discussed above as well as accurate time and space collocation of multiple-source data. It will also require greater access to other data sets outside of those produced directly by satellite missions. Sorting of multiple data sets into collocated time sequences requires significant computer resources and detailed catalog and inventory information for each data set. Multidisciplinary research (e.g., air-sea interaction, land-atmosphere processes) will make extensive use of this type of data analysis. Access to multiple archives containing results of previous data analyses will also greatly facilitate this type of work.

Needed Utilities

Three types of software utilities need to be available to researchers either online or as documented software that can be used locally. The first examines data documentation to determine physical variables (in SI units), to allow specification of measurement geometry and conditions, and to provide time and location of the measurement. The second allows mapping of data on a standard grid, time selection, and data volume reduction by a variety of simple processes: sampling, averaging, sorting, classification, and calculation of simple statistics. The third updates catalogs, inventories, and selection-software menus to reflect previous analysis results.

Mission Operations

The constraints placed on mission operations by climate researcher data requirements are primarily those required by the maintenance of long, consistent measurement records. Changes of observational strategy generally would not require rapid system response. If space, time, and spectral resolutions of the instruments can be altered on command, climate research needs could be met by ready access to command history information (for short time record interruptions) and by data processing to produce the defined climate observations. Examples of this last strategy include continued production of lower-resolution climate observations by processing of higher-resolution data collected for some other purpose, or continued production of spectral images by processing spectrometer data to simulate broader band images. The key requirement is to obtain a long, uniform data record.

Data Management and Control

The underlying structure of the data base required to support the type of research discussed above is that of a centrally coordinated network of dispersed data archives. The central coordination activity involves maintenance of a master directory of all holdings and a library of software modules needed to interact with archives transparently. The interaction with specific archives includes examination of local (more detailed) catalogs, browsing, ordering, and addition of data. Software modules to read (and write) data files should be provided, either by the central coordination institution or the local archives.

Key technical difficulties that need to be resolved (based on current archival practices) are:

1. improve documentation to provide complete experiment and calibration information;
2. improve navigation information and its accuracy;
3. resolve difficulties of multiple data set access caused by widely varying data formats;
4. provide data browse capability in systems that can "learn" progressively more sophisticated sorting;
5. provide for more remote data manipulation to allow preliminary (simple) studies that lead to more focused data selections.

Scale of Activity

Analysis of very large data sets will continue to be limited to a small number (~10) of groups, usually associated with large models, that have sufficient computer resources. The interactive nature of archives is therefore crucial because the results of an analysis are usually smaller in volume than the original data and of more interest to a larger number of researchers. Consequently, these few, large groups should probably have component archives of the total network; however, these groups will not be very interested in providing services to others. The role of a coordinating function will be to provide services (i.e., format documentation, catalogs). The more numerous (~100) users of smaller data sets will also be more interactive and use more processing power to access data.

CLIMATIC MONITORING

This scenario represents the type of observations that might be needed in the future to support seasonal forecasting or to target anomalous events for intensive study. Data collection methodology is focused on uniform, global coverage; however, near-real-time analysis is required to produce a reduced resolution version of the data along with statistical summaries. In addition, this type of scenario requires an ability to switch rapidly from a low level to a high level of effort. Thus, routine examination of summary statistics may lead, in quick sequence, to a request for more detailed data already in the archives and to a change in instrument configuration. This capability would support study of the time evolution of major climate anomalies.

Data Characteristics

The basic data taken by specific instruments represent a global description of the thermodynamic state of the atmosphere. These data would be equivalent to current routine weather observations, but a version of the data with somewhat lower spatial ($\sim$500 km) and temporal ($\sim$1 day) resolution would be prepared. Climate monitoring would entail near-real-time calculation of a number of statistical quantities (e.g., anomaly maps) presented in a standard map-projection or grid format. Appearance of a "significant anomaly" would trigger an examination of the full resolution data being acquired.

Data Access Patterns

Routine access to the archives for examining reduced resolution data and statistical summaries might be limited to a very few ($\sim$5) institutions in near-real time (weekly or monthly). Other research institutions may access these same records at other times to test forecast models, which would require a browse facility to locate "interesting cases." If seasonal forecasts are being attempted, then rapid access to the higher-resolution (unprocessed) data will also be needed. If intensive observations are requested or some change in observational strategy is called for, then data characteristics will be similar to those for case studies or field studies discussed above; but access will need to be more rapid. Part of the decision process in changing instrument configuration would be examination of "current" data in mapped or image form.

Needed Utilities

The basic utilities needed to provide near-real-time access to statistical summary data are processing algorithms required to produce statistical quantities (including a climatology data set for anomaly calculations) and a mapping facility. This type of access may take the form of electronic transfer of small data sets (maps) to a remote location for display. More rapid response to data requests (based on a browse data set) or a method for remote display of higher-resolution data sets may be necessary.

Mission Operations

Unlike climate research, this type of study requires not only more timely access to the data stream (within one week) and near-real-time processing on a routine basis, but also the ability to change observational strategies quickly to get better information about some particular location. Since the research supported by this capability is directed toward developing monitoring and prediction techniques, some evolution of the observational strategy over the course of the mission is expected.

Data Management and Control

Not only the instrument configuration, but also the data processing system will evolve as more is learned about the controlling factors for climate variations. Consequently, proper quality assurance and documentation of observations and processing software is imperative. Results of experimental anomaly predictions, whether retrospective, or near-real time, should be stored in some fashion for model intercomparisons. A trace of studies focused upon particular anomalous events can benefit subsequent analyses of climatic variations.

Scale of Activities

Routine access to the archives is likely to be limited to a few institutions with climate-scale Global Circulation Models and large computer resources; however, access to higher-resolution data sets for certain case studies (e.g., El Niño, volcanoes) will be more widespread.

LAND SURFACE CLIMATOLOGY

The objective of a land surface climatology project is to develop a better understanding of the processes occurring within, and the interactions among, the Earth's biospheric, edaphic, hydrologic, and atmospheric systems, and to determine their role in influencing climate over land surfaces. To accomplish this objective, principal experimental efforts would be organized and conducted as three parallel activities, each of which depends on the development of a supporting data base and data analysis system.

The first activity would be to conduct an analysis of existing remote sensing data in an attempt to select climatically representative study regions. The purpose is to determine the extent to which changes in the land surface can be determined and measured, and to assess the relative sensitivity of climate to various land processes. The second activity would be to prepare and validate comprehensive global data sets derived from operational satellites. The validation would be performed to document the current state of the Earth's land surface with respect to a number of select environmental parameters. The third activity would involve pilot experiments on specific regional or continental land masses to correlate remote sensing measurements with climate sensitivity parameters and to validate or modify land-atmosphere interchange models for these study sites.

These studies could include:

1. Vegetation: Fluctuations in green leaf biomass (monthly, seasonally, and annually) would be related to precipitation and surface temperature on a continental scale. Biomass would also be related to ecological units (i.e., Holdridge Life Zones) and to processes such as desertification, deforestation, and habitat destruction.

2. Soils: Regional measurements of soil moisture would be related to remote sensing surface signatures to determine if a broad range of relative differences in surface moisture conditions could be delineated using remote sensing instruments.

3. Hydrology: The application of remote sensing techniques to provide better estimates of surface water areal extent and volume, evapotranspiration, precipitation, snow cover and volume, and soil moisture would be explored on a global or regional basis and used if proven feasible.

4. Near-Surface Atmosphere: The capability to remotely detect and quantify climatic variations in the land surface record and to relate these changes to climate process models would be also assessed. Measurements would include multi-stage sampling of the land surface for albedo, vegetation cover, surface roughness, insolation, ground temperature, precipitation, etc.

Approach

Historical land remote sensing data would be examined to determine if in regions known to have experienced significant variations in their climate, these climatic changes can be detected through an analysis of changes in land cover. The approach would be to assemble land surface data obtained from satellites into a data base, including developing procedures to gain rapid, easy access to heterogeneous data in geographically dispersed archives. These data would need to be preprocessed and intercompared. They would also be compared to collateral data sets in the form of tabular meteorological records; digital topographic data; polygonal land cover and soil designations; and intensive point, area, and transect data from field measurements. This would necessitate integrating selected ground reference data, such as soil and land use maps, into the data base. A georeference structure would be used for relating these data sets.

While analysis of retrospective data would provide some indication of the influence of past land surface changes on climate, the present physical and biological state of the land surface should also be described. The best current sources of this information on a global scale are data sets derived from operational satellites. These satellite and ancillary data sets would be assembled into a global data base. Specific study sites, ranging in size from 10 to 500 km^2 would be selected to represent different climatic regimes, and a data base would be assembled. Pilot experiments that entail collection of field data and remote sensing data would be conducted. These experiments would be designed to determine if specific processes can be detected through remote sensing of earth surface features, and would use data from a variety of instruments.

Land-atmosphere process models representing the exchange of mass, energy, and momentum between the land and atmosphere systems would be developed. Detailed data sets acquired over the study sites would be used to initialize or parameterize models in simulation runs, and to verify and validate the results of the models. The response of the land surface in terms of biomass productivity and water budget would be modeled given climate forcing functions (precipitation and insolation) and terrestrial system properties (vegetation cover and surface roughness).

Data Needs

Satellite remote sensing data to be examined would include visible, near infrared, thermal infrared, and microwave radiances acquired at spatial resolutions ranging from 30 m to 30 km, and temporal resolutions from less than 1 day to 18 days.

In addition to these data, observations from a variety of spectrometers, radiometers, and cameras, flown on aircraft, would be used. Ground-based measurements of vegetation and soils would also be collected at selected study sites. These measurements would include physical sampling of vegetation biomass and soil moisture for laboratory analysis. Other measurements would include reflectance in the visible and infrared portions of the spectrum using hand-held and truck-mounted radiometers, and standard meteorological parameters, such as temperature and relative humidity, from meteorological stations.

Participants

Institutions participating in this type of project would fall into two categories, those that would conduct scientific research and those that could provide data. The number of scientific investigators involved would probably increase from perhaps 10 initially to as many as 50 in subsequent years. They could be geographically located at a total of perhaps 15 domestic and foreign government agencies, universities, and research institutes. All of their facilities would have the hardware and software expertise for digitally processing remote sensing data. Perhaps as many as 10 national and international institutions would be involved in providing data, including satellite, aircraft, and field measurements

Information System Considerations

Project needs would include locating and retrieving appropriate data, providing data to co-investigators, and transferring data to computer systems appropriate for each processing step. These steps would typically involve reformatting, preprocessing, processing and information extraction, and developing or verifying process models.

This type of project would need an information system that provides access to remote data bases and processing services to accomplish the following tasks: data search, browse, data storage, data processing (reformatting, registration, etc.), geographic overlay of dissimilar data, and data base management. An additional requirement would be to provide these services quickly and in a straightforward and easy-to-use manner. This could involve the use of techniques being developed in the field of artificial intelligence. The focus of the information systems needs discussed here allows scientists to focus on the research, and not waste their efforts on the details of accessing and preparing data for analysis.

GEOLOGIC MAPPING

Digitally merging and interpreting data collected by different remote sensing instruments with map and field data is rapidly becoming a major task of scientists in many different disciplines. For example, Landsat Multispectral Scanner (MSS) data have been merged with both Seasat and SIR-A radar data for geological analyses. These data have also been combined with geophysical, geochemical, and soils data that were digitized from existing map sources and then converted into images. One can envision projects that utilize Landsat thematic mapper, synthetic aperture radar, digitized topography, soils maps, and geophysical and geochemical data, together with Eos data. A main objective of these projects would be to extract new information on the types of soils and rocks in given areas.

One of the first tasks would be to identify the data that are available in a locale of interest. If multiple-image data sets exist, the researcher must be able to access information that will help identify the best data sets for a particular project (e.g., if there are 20 Landsat images over the area, which one should be used). Information needed to select an image includes data quality, cloud cover, sun elevation, and weather information (e.g., visibility, wind speed, water vapor content). When using data from multiple sources one of the most critical requirements is that the various data be geometrically coregistered in the desired projection. Thus, once the various data sets have been collected and are in digital form the next step would be geometric correction. Without good registration, the analysis and information extraction process can generate unsatisfactory results. Rectification is currently one of the most time-consuming steps in data reduction, and most researchers would prefer not to deal with this complex task. An Eos data and information system should provide standard services that include projection changes, image coregistration, vector-to-raster conversion, and generation of data sets with various grid point settings.

Another critical task is radiometric calibration. The data and information system should be able to produce quantitative digital-image data representing some physical unit (e.g., albedo for Landsat data; backscatter for radar data). For imaging spectrometer data, radiometric calibration should include both instrument corrections (gains, offsets, noise removal) and, in some cases, correction for atmospheric contributions and variations induced by changing incidence, emission, and phase angles. Corrections such as these will allow data collected at different times or by different satellites to be used and compared, because the sensor values will represent the same physical units relative to ground materials or cover.

The processing stage comprises geometric and radiometric calibration, and once complete, the researcher can begin the data analysis and information extraction phase. One of the first problems will be determining which set of "group" of data should be used to extract the desired information. Because of the large volume of data that is present when various data sets are merged (e.g., imaging spectrometer has dozens of image planes; thematic mapper has three visible, three near-infrared, and one thermal band; synthetic aperture radar with one microwave band; plus topography, digitized soil data, and geophysical data), the researcher must consider ays of grouping the data into subsets. The choice can be a subset of three bands, which has the most information (with the least amount of duplication) for color composite representation or digital classification; or bands where data with high correlation can be grouped together. Statistical methods such as principal components analysis, "selective" principal components analysis, and optimum index factors can be used to accomplish data plane reduction and help extract information from the data. These techniques must be used because of the large number of combinations that can be generated (e.g., even Landsat thematic mapper data has 15 independent ratios, which can be grouped into 455 three-ratio combinations). Alternatively, deterministic approaches can be used, such as expert systems for identification of particular rock types from imaging spectrometer data.

METEOROLOGICAL USES

Meteorology has a very active component of operational activities involved in daily weather forecasting. The operational activities are strongly tied to numerical weather prediction models. These models use global coverage of various atmospheric state parameters (including pressure, temperature, moisture, and winds throughout the depth of the atmosphere) and runs are made every 12 hours. Horizontal resolution on the order of 100 km and vertical resolution on the order of 1 km are required in the troposphere. Boundary conditions such as sea surface temperature, soil moisture, snow cover, cloud cover, and topography are required by some of the more advanced models. Current data assimilation models use conventional radiosonde and surface observations, satellite sounding radiometer data, cloud drift wind data, aircraft reports, and satellite imagery.

Most of the recent advances in weather forecasting skill have resulted from either improved observing systems or better numeric models. Because Eos instruments will provide global coverage of meteorological parameters that could improve operational weather forecasting, there will be a very strong desire by the operational user community to gain real-time access to selected Eos data. In particular, the moderate resolution imaging spectrometer, the high resolution multi-frequency microwave radiometer, the laser atmospheric sounder, the scatterometer, and the Doppler lidar would be useful for forecast models.

Data access patterns for operational users of Eos data will require the full data stream processed to include calibration, earth location, and scientific units in near real-time (within three hours of observations). Previous experience with new data sources has shown that the general sequence of events includes theoretical studies that show the potential impact of the new data, data systems tests of one to three months when the data first become available, followed by eventual operational utilization. The theoretical studies generally precede the data systems tests by 1 to 10 years. The data systems tests are usually performed within the first year of data availability. The processing system is set up to provide data in near real-time for a one to three month test period. It is set up as a "bare bones" system without the redundancy, complete documentation, etc., required of an operational system. New data are fed into a model running in parallel to an operational model, and the forecasts are compared to assess the impact of the new data. These data systems tests are generally run in real-time with a post facto analysis and evaluation period. If the data show a positive impact, they are then brought into operational use with the necessary processing equipment and software being procured to provide reliable service. This generally requires one to three years after the data systems tests. If the satellite system providing new data will not be available for a number of years, the sequence stops at the data systems test until an operational satellite system is being procured. This generally requires five to eight years after the initial data availability on a prototype satellite system.

Field Studies

There have been a number of large meteorological field studies (BOMEX 1968, GATE 1975, SESAME 1979, FGGE 1979, etc.) that last for several months and focus on a particular phenomenon, such as tropical convection. These studies use a combination of *in situ* instruments and remote sensing gadgetry. An operations center directs the placement of aircraft and dictates instrument schedules to maximize information on the mechanisms under study. In order to do this, the operations center requires real-time, quick-look data from all instrument systems, including satellite instrumentation. In previous field experiments, satellite data have either been received and processed directly at a field site, or arrangements have been made to have the data received and processed at a central facility and then transmitted to the field site.

Following the field exercise, there generally is a two to five year period where research groups do intensive studies with the data. Quick-look field processing of the data has already established a general framework of locations, times, etc., that are most promising for further study. Hence, the catalog requirements are fairly straightforward, including lists of times, dates, etc., of data availability, data quality indicators, etc. Prior experience with archives of satellite data used within field experiments has shown that approximately 20 to 100 separate groups will request data. Data requests are generally for 1 to 10 computer tapes, with an occasional user requesting 100 or more tapes. For a two-month experiment, frequently two to five days will be selected by most groups for intensive study with a few groups requesting data throughout the entire time period. The data requests generally reach a peak about one to two years after the field experiment. The studies generally have a requirement for merged data sets from a large number of different data sources. Hence, accurate location information is extremely important in these studies. Calibration and conversion to scientific units is also required for these studies.

The actual instruments of interest for a given field of study will vary according to the purposes of the experiment. The Eos Surface Imaging and Sounding Package (SISP), Sensing with Active Microwave (SAM) package, and the Atmospheric Physical and Chemical Monitors (APACAM) would have instruments of interest, including very limited data sets of the very high-resolution imaging spectrometers and synthetic aperture radar instruments.

Case Studies

Meteorological case studies generally focus on a specific time and place where a particular meteorological phenomena is occurring. The investigators are generally of two types, those having specific times and places already established from other data sources (these investigators are generally like field study investigators in that all they require are lists of times and locations of data availability, quality indicators, etc.) and those with only a specific phenomena in mind (but no time or location). These researchers require extensive browse capabilities to locate likely data sets. The most successful browse capability includes sampled data sets in an image or graphical presentation. Books, movies, video discs, etc., of these-sample data sets can be rapidly examined to locate a place and time that shows the desired phenomena. Then, the listings of data times, availability, etc., are used to order data.

VEGETATION BIOMASS

The purpose of this type of research is to gain a better understanding of the spatial distribution of vegetation characteristics and processes, including biophysical factors (leaf area index, biomass, net primary productivity, canopy temperature, and albedo), and plant physiological processes (evapotranspiration, photosynthesis, and respiration).

Objectives include: developing methods to measure (by remote sensing) biomass and net primary production of terrestrial vegetation, and to employ satellite images for assessing and improving the current representational accuracy of continental-scale land cover information.

Approach

Two approaches would be employed in this type of research. The ability to infer key vegetation characteristics from remote sensing data is central to the economy of large-scale research. Therefore, in the first approach, spectral signatures of vegetation would be collected and correlated with laboratory measurements such as leaf reflectance. These data would be used to interpret measurements from aircraft and spacecraft where atmospheric conditions attenuate and distort the characteristics of the signatures.

The second approach would employ both manual interpretation and machine classification of satellite data to stratify vegetation and other surface features into broad, physiognomic categories (based on vegetation structure) suitable for global comparison. Aerial photographs, field reconnaissance and other data sources would be used in this analysis. Maps would be derived on scales consistent with existing small-scale vegetation maps. This approach would provide both a comparison for current information sources, and an assessment of the methodology of very large-area vegetation mapping. However, because of the resources that would be required to process data for the entire land surface of the Earth, an appropriate strategy would include the use of coarser resolution data for primary stratification in a multistage sampling approach. Even using coarse-resolution data and statistical sampling, assembling the required data would be a significant challenge.

Information System Considerations

In support of this research, a data and information system should provide computation and information management support for a highly dynamic and diverse set of processing and data requirements that result from biospheric research needs. The services of the system should include data acquisition, search, browse, access, preprocessing, image processing and display, data management (physical and electronic), and operational management functions.

1. Data Access: Online catalogs of and browse access to archival data would be beneficial. In addition, the ability for direct ordering would also be of value. The capability to browse large-image data bases before ordering images would be an important function.

2. Data Input: Direct transmission of *in situ* data (both vegetation and radiometric) between field sites and processing centers would reduce delays by an order of magnitude (from months to a few days). Entry or conversion of ancillary data (topographic, soils, climatic) to an acceptable format would add significantly.

3. Preprocessing: Registration (band-to-band and sensor-to-sensor) and common data formatting would be of tremendous value and high priority. Also of value would be the

capacity to digitize photographs with interactive input from investigators at remote locations.

4. Analysis: Efficiency of analysis could be increased if real-time interaction between centers and remote investigators were possible.
5. Archival Storage and Catalog: An active directory with up-to-date documentation of parallel and ancillary data sets held within and external to Eos would be of great value and high priority.
6. Distribution and Network: Access to geographically dispersed data bases and the ability to overlay them in common format is of high priority. Data, besides being in compatible file format, must carry documentation of quality and type. Time scales for such access should be on the order of a few days. Networks of computer processors would be valuable.
7. All of the above require that an advanced data management and control system be a part of the overall Eos concept. This system should facilitate researcher access, processing, and analysis in a manner essentially transparent to the scientist. It should in effect permit scientists to function as scientists, not librarians, communications experts, image processing specialists, or computer scientists.

TEMPORAL INFORMATION EXTRACTION

Following are several short scenarios that demonstrate the use of temporal information extracted from data collected by remote sensing instruments.

Mapping Lava Flows

To understand volcanic processes and to gauge the extent of potential hazards, radar images collected every four to six hours over active volcanic fields are used to map the spatial distribution, flow direction, and volume of each new flow. Data must be collected during the eruption regardless of the weather conditions. Because of high cloud cover probabilities, an imaging radar system would have to be used. If radar stereo images were collected, topographic information before and after the various eruptions could be produced and used to calculate the volume of magma erupted. Topographic data could also be used to compute slope, which is used to predict the direction and possible speed of any new flow. Eos and aircraft data could be used to meet the temporal coverage requirements.

Vegetation Mapping

Along certain areas of the Colorado River, three to four major crops are grown with irrigation water from the river. The amount of water that can be used is strictly controlled, and information on its use is gathered by the Water Resources Division of the United States Geological Survey. Information gathered with stream gauges along the river is used to compute water usage. A study was conducted to see if Landsat multispectral scanner data could be used to predict water usage along the river. The method used a ratio of band 4 (chlorophyll) to 7 (vegetation cell structure) to map the crops in an area. Then, known water usage for each crop was used to predict the amount of water removed from the river and the data sets correlated. This is clearly a time-dependent problem.

There are a number of problems where temporal changes (man-made or naturally occurring) are of interest. There are several simple statistical parameters that can be computed for images and used to monitor temporal changes of interest. Statistical information that can be used for this purpose includes the average, standard deviation, and skew coefficient, plus the correlation coefficient between images and a mean image for the area. Principal components of the images would also be useful to identify changes that have occurred.

With Eos data bases, differences between the averages and standard deviations of images can be used to check for low frequency and total contrast changes. The correlation coefficient between the mean and any given image can be used to select images that warrant examination. Temporal monitoring and comparisons should be made against the current scene average. Monitoring and change detection capabilities will allow the system to do automatic-browse and search of data. These derived data would be of value in a number of other research areas.

Updating Digital Line Graph Data

The Mapping Division of the United States Geologic Survey generates Digital Line Graph data by digitizing most of the information present on either 1:250K or 1:24K topographic maps (e.g., roads, railroads, drainage features). They are interested in using data collected by remote sensing instruments to update their Line Graph files. High spatial-resolution image data that have been geometrically and radiometrically calibrated is required. Digital processing techniques for pattern recognition and classification should be used (e.g., mapping of roads) and the registered Line Graphs should be periodically updated. Image data with at least 10-meter spatial resolution recorded in one to three spectral bands will be needed. Useful bands will be similar to thematic mapper bands 2, 4, and

5. A single band with five-meter spatial resolution may be preferred because of the very fine detail that will be needed for identifying roads and other high-frequency patterns (spatial information may be more important than spectral information for this application). Although a high spatial-resolution imager is not included on the current Eos payloads, one can imagine that an Eos Information System could provide directory and catalog information for this type of data.

APPENDIX II - REQUIREMENTS SYNOPSIS

This report presents more than 100 requirements and recommendations concerning the functional characteristics and attributes of a requisite data and information system for Eos. We include this Appendix as a means for assessing the interdependence of functional system elements and as a convenient reference for analysis of the architectural example presented in Chapter III. This listing is not complete, nor is it intended to be the sole reference for individuals seeking a definitive accounting of requirements and recommendations set forth by the Eos Data Panel. Rather, it is meant to show the relationship between functional elements of an Eos data and information system and to allow the reader to quickly obtain an overview of the requirements contained within this report. Eos Data Panel recommendations should not be taken out of context, and therefore the reader is referred to the main body of this report for a complete accounting.

The requirements listed below are grouped into seven generic classifications: Eos flight systems, user, operational, information services, advanced data base management, data processing, and non-Eos data base functions. An eighth category, management considerations, has been included and contains recommendations dealing with both implementation and the unique characteristics of the data and information system. The seven classes of system requirements are subdivided as appropriate and references to text indicated. No priority is indicated by the relative order in which these recommendations or requirements are listed.

I. FLIGHT SYSTEMS FUNCTIONS

1. Execution times for any system operation within an Eos data and information system should be in seconds, and those for archival data should be within minutes for small volumes, and days for large volumes, depending on the priority of the request.
2. The Eos spacecraft should include a flexible flight information subsystem for controlling onboard processing of both uplink and downlink information.
3. Interactive commands that require critical onboard resources should be processed centrally at a ground-based control center, while non-interactive commands should be directly transmitted from the user to the onboard system for final assessment and forwarding to the appropriate instrument or subsystem.
4. Onboard systems should be used for decisions when rapid response is needed (e.g., gain settings or filter changes, or data acquisition rates during anomalous conditions).
5. Flight instruments will require a certain level of onboard monitoring to ensure their functional capability and to format, error code, and buffer telemetry data, as well as the merging of critical ancillary data needed for quick-look and other analyses.
6. The flight data system should be capable of acquiring and merging ancillary (e.g., time and position) and correlative (e.g., *in situ*) data into instrument data streams.
7. It should be possible to receive directly transmitted data at any modestly equipped ground station that is within line of sight of an Eos platform.
8. At least 95 percent of the real-time data broadcast by the satellite should be received, preprocessed, and displayed within 60 seconds from the time of transmission.
9. In order to buffer the high-peak data rates and to guard against momentary TDRSS channel outages, an onboard data storage capability should be developed.

II. USER FUNCTIONS

1. Capabilities of data centers should be periodically evaluated by key personnel who lead activities of instrument teams and active data base sites.
2. All Eos data should be validated to the satisfaction of the research community.
3. Investigators will require quick-look data sets for instrument command decisions, and for preliminary scientific analysis.
4. Users will submit instrument requests to an instrument operations center for execution. This interaction will most likely occur over an electronic information network.
5. One of the unique features of Eos is the requirement that a researcher return to the archives the results of his research, in the form of reduced or derived data (i.e., value-added data).
6. Instrument teams will guide instrument and algorithm development and will be involved in mission operations.

III. OPERATIONAL FUNCTIONS

A. Mission

1. A master schedule should be updated every orbit and it should contain logistical information needed to plan an observation.

2. A mission operations center is the focus for real-time command and control of the space platform. Instrument operations center(s) are the focus for Eos instrument operations.

3. Subsystems of the information system will support mission operations, including relevant network functions, acquisition and delivery of data to mission and instrument repositories, quick-look data production, and the large-scale processing needs of instrument teams.

4. Instrument teams will guide instrument and algorithm development and will be involved in mission operations.

5. A mission operations center will need to continually monitor a sampling of data in near-real time for quality control, error detection, and instrument assessment.

6. The capability to reconfigure observational sequences when malfunctions or special events occur is needed.

7. If problems occur, control centers need to have the capability to trace the data flow back through the processing system to the instrument, aiding in the isolation and correction of these problems.

8. The capability should exist to integrate commands and create command sequences from simple requests to complex operations that require the coordination of multiple subsystems and instruments.

9. Access security must be maintained at appropriate levels at all times.

10. All Eos instrument parameters should be monitored by examining their performance statistically.

11. Automated command sequence construction, together with the necessary computer involvement in routine spacecraft communications, should be available in mission operations computers.

12. In some cases, instrument command decisions will have to be made within the time period of one orbit.

B. Instrument

1. A mission operations center is the focus for real-time command and control of the space platform. Instrument operations centers are the focus for Eos instrument operations.

2. Access to spacecraft subsystems should be channeled transparently through any control system for non-interactive commands.

3. Commands should be categorized as non-interactive, interactive, and critical.

4. Research groups should have the capability to request acquisition of special observational sequences from one or more instruments.

5. Instrument teams may require command control over various instruments to satisfy scientific objectives.

6. Researchers will require near-real-time processed data to decide, for example, appropriate locations for mobile, *in situ* instrument systems (e.g., aircraft, ships).

7. Instrument specialists and operations personnel will require rapid decision-making capabilities. This requirement includes establishing real-time processing and display capabilities that can be utilized to monitor events.

8. All Eos instrument parameters should be monitored by examining their performance statistically.

9. If problems occur, control centers need to have the capability to trace the data flow back through the processing system to the instrument, aiding in the isolation and correction of these problems.

10. Users will submit instrument requests to the instrument operations centers for execution. This interaction will most likely occur over an electronic information network.

IV. INFORMATION SERVICES FUNCTIONS

1. Required functions and capabilities should not constrain physical location of personnel, data, or processing capabilities.

2. A major goal of the system should be to provide remote electronic access for the scientific community to the variety of capabilities and services that Eos will provide.

3. Particular emphasis should be placed on technology associated with transparent access to dispersed, heterogeneous data bases.

4. The system will require cost-effective network capabilities, including direct broadcast access.
5. The network should provide access to mission operations, archives, selected active data bases, and to large mainframe computers.
6. Data center personnel should manage the network, ensuring that the data centers are involved directly in many key aspects of the project.
7. Because of rapid advancements in the telecommunications industry, and the long lead time for Eos implementation, the telecommunications requirements should be periodically reevaluated and updated as new technologies become more readily available.
8. Communications linking quick-look users with data repositories will be required. Data rates between 56K baud and 1.5M baud will likely be required on this link.
9. The system should provide for a spectrum of electronic data delivery rates, ranging from 9,600 baud to 6.3M baud.
10. Access security should be maintained at appropriate levels at all times.
11. Provisions should be made within the network design for future expansion to accommodate new user's access to the services and capabilities afforded.
12. NOAA and its international clientele will require separate data storage and relay systems for data delivery if they deploy operational instruments on an Eos spacecraft.
13. Subsystems of the information system will support mission operations, including relevant network functions, acquisition and delivery of data to mission and instrument repositories, quick-look data production, and the large-scale processing needs of instrument teams.
14. Users will submit instrument requests to the instrument operations control centers for execution. This interaction will most likely occur over an electronic information network.
15. Access security must be maintained at appropriate levels at all times.

V. ADVANCED DATA BASE MANAGEMENT FUNCTIONS

A. Electronic Directory

1. Eos-sponsored multidisciplinary and multiple data source, disciplinary-oriented researchers will require an electronically accessible directory of pertinent Earth sciences data.
2. Self-documenting software will be required to create directory entries for newly acquired data.
3. The Earth science data directory requires maintenance and periodic updating as project-sponsored researchers identify new data sets and as new Eos data are acquired.
4. The master directory should include information on non-Eos data deemed pertinent by Eos-sponsored researchers.
5. A researcher should be able to search the directory by project, platform, instrument, data processing level, version, parameter, time, location, or any combination of these attributes.
6. The directory should include information about supporting catalog access.

B. Electronic Catalogs

1. Eos-sponsored multidisciplinary and multiple data source, disciplinary-oriented researchers will require electronically accessible catalogs of Earth sciences data.
2. Self-documenting software will be required to maintain and update the catalogs as new data are acquired.
3. The Eos data catalog should be accessible by project, platform, instrument, data processing level, version, parameter, time, geographic location, or any combination of the above.
4. The Eos catalog should contain a use trace including user name and address for any particular data set.
5. To the extent possible, the Eos project should endeavor to create, maintain, and update similar catalog services for pertinent data held external to the project.

C. Browse Files

1. The Eos project will be required to create, maintain, and update browse files within the archival system.
2. Browse files should be updated in real-time for non-image data.
3. Near-real-time updates will be required for some critical image data. It is anticipated

that these data will be of reduced resolution (either spectrally or spatially).

4. Browse files should be accessible on an instrument, time, geographic location, or any combination of these factors.
5. Browse files should be searchable visually via attributes and by expert systems.
6. Data attributes (e.g., cloud type and cover, vegetation type and cover, snow cover, data quality) should be appended to an inventory record within the browse files.
7. Attribute files should be expandable, enabling researcher-identified attributes to be added subsequently.
8. Researchers will require processing of reduced-volume data sets to Level 2 for browse purposes.
9. Pattern recognition algorithms should be developed and applied to browse data for attribute identification.
10. Online browse capabilities may not be possible for many users; this will necessitate publication of an image browse catalog.
11. Consideration should be given to creating special purpose browse files interactively.

D. Data Ordering

1. The system should support online, electronic ordering of all Eos data.
2. To the extent possible, the system should support the online, electronic ordering of pertinent non-Eos data sets.
3. Interactive ordering capabilities should be available with at least 9,600 bps dial-up links.
4. The system should be available on a continuous basis and be sized to handle at least 100 simultaneous users.

E. Documentation

1. Documentation is considered to be a vital part of the data record and should be stored and maintained with the same care as the data per se.
2. The documentation should include a concise description of instrument specifications, including the nature of physical variables measured, noise characteristics, internal and external processing history, and coding of telemetry.
3. All sources of calibration information, procedures, and results must be identified and placed in the archival documentation.
4. An annotated bibliography covering all pertinent papers and reports published in the refereed literature should be included as a part of the project's documentation (e.g., instrument descriptions, observations and sequences, calibration and sensitivity studies, precision and accuracy of measurements).
5. Documentation should also include any information pertinent to scientific interpretation of the data produced by various experiments.
6. Consideration should be given to providing online, electronic access to technical reports and memoranda (i.e., the gray literature).

F. Archives

1. The archival system should be able to assimilate data at all processing levels.
2. Access to non-Eos data that are needed for processing of Eos data or by Eos investigators will be required.
3. *In situ* data used in the processing or validation of Eos data should be easily accessible and maintained within the archives.
4. The archives must be able to assimilate, maintain, and distribute higher level or reduced data sets produced by either active data base sites or by individual investigators in their research work.
5. Directory and catalog entries for all data stored and maintained within the archives will be required.
6. The archives will require a fast, efficient, and cost-effective retrieval system for all of its data holdings.
7. A history of both instrument and algorithm performance should be maintained and accessible via the archives.
8. The Eos archival subsystem should be accessible from remote terminals and workstations via a prompting menu or natural language interface.
9. All capabilities and functions of the archival subsystem should be fully documented in a user handbook and in online help files.

10. The archives should maintain a complete inventory of relevant scientific and technical documents concerning the data and the archives.

11. Processed data should be retained in accessible form, along with associated documentation, to avoid redundant processing.

G. Standards and Formats

1. Considerable and careful attention must be given to the adoption of standards and formats; the research community should be intimately involved in this process.

2. We envision the necessity for an entire suite of format structures and, consequently, system transparency remains an overriding concern.

3. Consideration should be given to various standards adopted by the Space Station Program in areas of overlapping concern if they are acceptable to Eos.

4. Careful consideration should be given to working with both established national and international committees addressing standards and possible adoption of the standards that they recommend.

5. A common, standard glossary containing operational definitions used within Eos should be created and maintained by the project.

VI. DATA PROCESSING FUNCTIONS

1. Processing requirements within the system will range from raw data to fully processed scientific data.

2. Instrument-specific, near-real time, or even real-time data processing, delivery, and display capabilities will be required.

3. Algorithms to be used for processing data should be approved and fully validated by the research community.

4. Requirements for processing include: Level 0, all data (for temporary repositories); Level 1, all data (for long-term archives); Levels 2, 3, and 4, only upon request.

5. The requisite data processing capability must support some near-real-time data and information processing. These results should be available to the requester within the time period of one orbit.

6. Data processing subsystems should be capable of enhancement for special purposes during fixed periods of time (e.g., scientific processing for particular instrument team activities).

7. Access security must be maintained at appropriate levels at all times.

8. Users identified as requiring access to real-time, quick-look information need interactive image processing and display capabilities.

9. Processing facilities will be needed to support requested higher-level data processing for scientific research purposes.

10. Data reduction, grid overlay, standard projection, data set merger, etc., will be required by Eos researchers.

11. Distributed data processing facilities will be required to perform similar higher-level data processing functions.

12. All data processing requirements are independent of specific organizational assignments and must be met by any facility (including commercially operated) that is producing Eos data.

13. Access to large mainframe computer facilities will be required for Eos researchers engaged in model development, evaluation, simulation, and experimentation.

14. Data sequence and encoding peculiarities must be removed by the data reception system.

VII. NON-Eos DATA BASES FUNCTIONS

1. Many Eos-sponsored research tasks will require access to operational data bases such as NOAA's. The project should make every effort to provide this access in a user transparent fashion.

2. International data bases such as the holdings of the World Data Centers will be required by Eos researchers. Access to these data sets should be through the Eos data and information system.

3. National archives such as those of the National Climate Center, National Oceanographic Data Center, etc. should be online to Eos researchers via the data and information system.

4. Reduced data sets produced from data derived from other archives, as well as the original data, should be maintained within the information system.

5. Directory entries covering pertinent, non-Eos data sets should be obtained and maintained within the system.

6. The data and information system should accomodate catalog entries for non-Eos data sets.

VIII. MANAGEMENT CONSIDERATIONS

1. The scientific community must be involved during the development and evolution of the Eos data and information system.

2. Active researchers should have an oversight and review responsibility for the data and information system.

3. NASA management should pay particular attention to existing National Academy of Sciences reports dealing with space research data systems.

4. A crucial difference between Eos and previous flight projects that requires procedural changes is the length of the effort.

5. An Eos data and information system must evolve and be a system that includes geographically distributed sites of varying capability and responsibility.

6. A critical factor in developing the Eos data and information system will be strong interaction and close collaboration with ongoing research efforts that emphasize the collection and analysis of multiple data sets.

7. We recommend that the Global Resource Information System concept be focused on Eos objectives, then utilized for developing the network services aspects of the Eos data and information system.

8. We recommend that the Earth Science and Applications Division fund a limited number of multidisciplinary research teams to investigate a select number of objectives in hydrology, biogeochemistry, and climatology using existing data sets.

9. We recommend that the Eos data and information system be built on the experience gained with existing Earth science pilot projects, that these projects be focused toward Eos objectives, and that close collaboration be maintained with the UCAR Unidata initiative.

10. We recommend that planning and evolutionary implementation of an Eos data and information system begin without delay, enabling the resultant system to reflect the experience and expertise of those people who must manage and utilize its capabilities.

11. Eos should accommodate operational users if their activities do not detrimentally impact the conduct of Eos-sponsored research.

12. If operational uses are made of Eos data, and if these uses significantly affect the information system, then those operational agencies should provide enhancements to the system to meet their needs.

BIBLIOGRAPHIC DATA SHEET

1. Report No.
NASA TM-87777

2. Government Accession No.

3. Recipient's Catalog No.

4. Title and Subtitle
EARTH OBSERVING SYSTEM
DATA AND INFORMATION SYSTEM
REPORT OF THE EOS DATA PANEL - VOLUME IIa

5. Report Date
June 1986

6. Performing Organization Code
610

7. Author(s)
See Supplementary Notes

8. Performing Organization Report No.
86F5001

9. Performing Organization Name and Address
Laboratory for Atmospheres
Goddard Space Flight Center
Greenbelt, Maryland

10. Work Unit No.

11. Contract or Grant No.

12. Sponsoring Agency Name and Address
National Aeronautics and Space Administration
Washington, D.C.

13. Type of Report and Period Covered
Technical Memorandum

14. Sponsoring Agency Code

15. Supplementary Notes
R. Arvidson, Washington Univ.; F. Billingsley, Jet Propulsion Lab; R. Chase, Woods Hole Oceanographic Inst.; P. Chavez, USGS Flagstaff; M. Devirian, NASA Headquarters; J. Estes, Univ. of California at Santa Barbara; G. Hunolt, NOAA Satellite Data Services Div.; J. Klose, Jet Propulsion Lab; G. Ludwig, Consultant; F. Mosher, NOAA National Severe Storms Forecast Center; W. Rossow, Goddard Institute for Space Studies.

16. Abstract

The purpose of this report is to provide NASA with a rationale and recommendations for planning, implementing, and operating an Earth Observing System data and information system that can evolve to meet the Earth Observing System's needs in the 1990s. The Earth Observing System (Eos), defined by the Eos Science and Mission Requirements Working Group, consists of a suite of instruments in low Earth orbit acquiring measurements of the Earth's atmosphere, surface, and interior; an information system to support scientific research; and a vigorous program of scientific research, stressing study of global-scale processes that shape and influence the Earth as a system. The Eos data and information system is conceived as a complete research information system that would transcend the traditional mission data system, and include additional capabilities such as maintaining long-term, time-series data bases and providing access by Eos researchers to relevant non-Eos data. The Working Group recommends that the Eos data and information system be initiated now, with existing data, and that the system evolve into one that can meet the intensive research and data needs that will exist when Eos spacecraft are returning data in the 1990s.

17. Key Words (Selected by Author(s))
Eos Data Users, Data Acquisition, Data Processing, Telecommunications, System Architecture, Data Management

18. Distribution Statement
Unlimited - Unclassified

STAR Category 42

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price*
Unclassified	Unclassified	62	A04

*For sale by the National Technical Information Service, Springfield, Virginia 22161

GSFC 25-44 (10/77)

End of Document