# General Disclaimer

One or more of the Following Statements may affect this Document

- This document has been reproduced from the best copy furnished by the organizational source. It is being released in the interest of making available as much information as possible.

- This document may contain data, which exceeds the sheet parameters. It was furnished in this condition by the organizational source and is the best copy available.

- This document may contain tone-on-tone or color graphs, charts and/or pictures, which have been reproduced in black and white.

- This document is paginated as submitted by the original source.

- Portions of this document are not fully legible due to the historical nature of some of the material. However, it is the best reproduction available from the original submission.
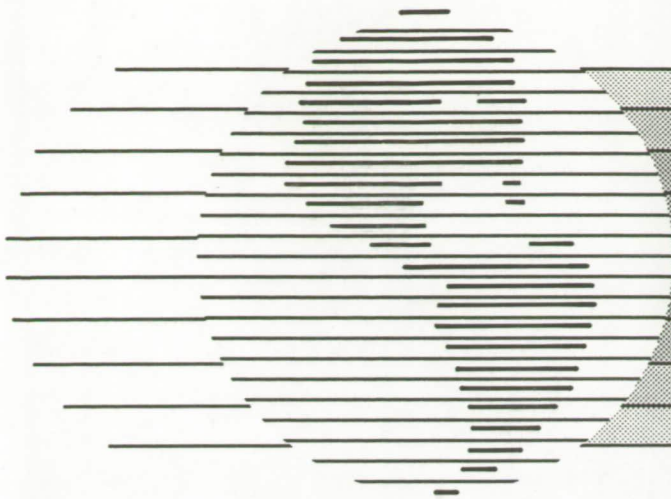
**Earth
Observing
System**

INITIAL SCIENTIFIC ASSESSMENT
OF THE EOS DATA AND INFORMATION SYSTEM (EosDIS)

SCIENCE ADVISORY PANEL FOR
EOS DATA AND INFORMATION

Eos-89-1

NASA

National Aeronautics and
Space Administration

Goddard Space Flight Center

Preface

"The emphasis here is on those aspects of the system that will make it responsive to the needs of science and scientists, rather than being a data archive in the form of the write-only memory that has been all too frequently encountered in the Earth sciences. In fact, . . . a successful Eos Data and Information System must do much more: it must be designed and implemented so it will engender powerful new modes of research, foster synergistic interactions between observation and simulation with models, and promote thought about the Earth and its processes at higher levels of abstraction."

—John A. Dutton
IEEE Trans. Geosci. Remote Sens.
March 1989

This report from the Science Advisory Panel for Eos Data and Information summarizes our view of the progress of EosDIS, some issues that need to be addressed, and some planned actions of the Panel. It is written after the Preliminary Requirements Review and the Systems Architecture Review by the two contractors, Hughes and TRW, who are participating in studies of the Definition and Design Phase ("Phase B") for EosDIS, but before the selections for the single contractor for the implementation and execution phase ("Phase C/D"). These Phase B studies will be completed in the Spring of 1990, and the Request-for-Proposal for Phase C/D will be issued in the early Fall, with a response deadline of 60 days. In the Summer of 1991 the winner of the Phase C/D contract will be announced.

We now seek comments and criticisms from the rest of the Eos community—not only Eos Investigators, but other scientists who plan to use Eos data for investigation of global environmental processes and change. We feel that EosDIS will surely fail unless a large fraction of the science community feels personally involved in its conception and design. It is crucial that EosDIS succeed. The goals of Eos depend not only on its creative scientific investigations and innovative instruments, but also on how well Eos and EosDIS create and integrate large-scale data sets and geophysical measurements of established quality and reliability. A report like this one is never "finished." Concerns and opinions

about EosDIS will persist well past its implementation, but now is a good time to get some important issues on the table and to resolve or at least identify differences between EosDIS, the Eos Program, and the scientific community.

Please, therefore, bring to our attention any omissions, misinterpretations, or other misdemeanors that are in this report. Comments may be directed to the Panel Chairman or to any member. Addresses are listed in the Appendix.

Science Advisory Panel for Eos Data and Information

Jeff Dozier, Chair

John Barker	Robert H. Evans
Bruce R. Barkstrom	David Halpern
Roger G. Barry	James F. Kibler
Francis Bretherton	David A. Landgrebe
Josef Cihlar	Mark R. Schoeberl
John Curlander	Bob E. Schutz
Lee Elson	Ronald Welch
G. David Emmitt	Al Zobrist

Eos Program/Project Contributors

Graham Bothwell	Eni Njoku
Dixon Butler	H. Ramapriyan
Al Fleig	Gerald L. Soffen
Daesoo Han	Daniel Wenkert
Strat Laios	Darrel L. Williams
Gary Metz	W. Stanley Wilson
David Nichols	

Contents

LIST OF FIGURES

LIST OF TABLES

Acronyms

AIRS	Advanced Infrared Sounder
AMSR	Advanced Microwave Scanning Radiometer
AMSU	Advanced Microwave Sounding Unit
ASAS	Advanced Solid-State Array Spectroradiometer
ASF	Alaska SAR Facility
AVHRR	Advanced Very High Resolution Radiometer
AVIRIS	Advanced Visible and Infrared Imaging Spectrometer
CCSDS	Consultative Committee for Space Data Standards
CD-ROM	Compact Disk—Read-Only Memory
CDF	Common Data Format
CDHF	Centralized Data Handling Facility
CDOS	Customer Data and Operations System
CERES	Clouds and the Earth's Radiant Energy System
DLS	Dynamics Limb Sounder
DMSP	Defense Meteorological Satellite Program
Eos	Earth Observing System
EosDIS	Eos Data and Information System
ERBE	Earth Radiation Budget Experiment
ERS-1	(European) Earth Resources Satellite–1
FFT	Fast Fourier Transform
GFLOP	10^9 floating-point operations
GSFC	Goddard Space Flight Center
HIRIS	High-Resolution Imaging Spectrometer
HIRRLS	High-Resolution Research Limb Sounder
HIRS	High-Resolution Infrared Sounder
ISLSCP	International Satellite Land-Surface Climatology Program
IWG	Investigators' Working Group
J-ERS-I	Japanese Earth Resources Satellite–1
JPL	Jet Propulsion Laboratory
MISR	Multi-angle Imaging Spectro-Radiometer
MLS	Microwave Limb Sounder
MODIS	Moderate-Resolution Imaging Spectrometer
MOPITT	Measurements of Pollution in the Troposphere
NAR	Non-Advocacy Review
NASA	National Aeronautics and Space Administration
NMC	National Meteorological Center
NOAA	National Oceanic and Atmospheric Administration
NSIDC	National Snow and Ice Data Center

PRR	Preliminary Requirements Review
RFP	Request for Proposal
SAFIRE	Spectroscopy of the Atmosphere Using Far Infrared Emission
SAGE	Stratospheric Aerosol and Gas Experiment
SAR	Synthetic Aperture Radar
SPOT	Systeme Probatoire d'Observation de la Terre
SQL	Scientific Query Language
SSM/I	Special Sensor Microwave/Imager
SWIRLS	Stratospheric Wind Infrared Limb Sounder
TDRSS	Tracking Data-Relay Satellite System
TES	Tropospheric Emission Spectrometer
TOGA	Tropical Oceans and Global Atmospheres
TOMS	Total Ozone Mapping Spectrometer
TOPEX	The Ocean Topography Experiment
TRACER	Tropospheric Radiometer for Atmospheric Chemistry and Environmental Research
UARS	Upper Atmospheric Research Satellite
USGS	U.S. Geological Survey
WOCE	World Ocean Circulation Experiment

Page intentionally left blank

Initial Scientific Assessment of the Eos Data and Information System (EosDIS)

Executive Summary

Crucial to the success of the Earth Observing System (Eos) is the Eos Data and Information System (EosDIS). The goals of Eos depend not only on its instruments and science investigations, but also on how well EosDIS helps scientists integrate reliable, large-scale data sets of geophysical and biological measurements made from Eos data, and on how successfully Eos scientists interact with other investigations in Earth System Science. Current progress in the use of remote sensing for science is hampered by requirements that the scientist understand in detail the instrument, the electromagnetic properties of the surface, and a suite of arcane tape formats, and by the immaturity of some of the techniques for estimating geophysical and biological variables from remote sensing data. These shortcomings must be transcended if remote sensing data are to be used by a much wider population of scientists who study environmental change at regional and global scales.

EOSDIS AND EARTH SYSTEM SCIENCE

Eos is a pivotal part of the U.S. Global Change Research Program, and hence of the international effort to understand how the Earth functions as a complete system. Earth System Science objectives require a data and information system that will facilitate and encourage multidisciplinary and interdisciplinary investigations. Data from Eos platforms must be combined with data from other agencies and nations—including remote sensing data from other satellites or aircraft and *in situ* operational and experimental data. These data sets must be accessible and integrated with scientists' computing facilities and models of environmental processes and global change.

Recommendations

The Eos Program at NASA Headquarters must:

- provide an environment of incentives and challenges that persuade scientists to devote their efforts to those tasks in EosDIS for which their creative talents are essential;

- establish dependable linkages, that really function as dependable partnerships, with other agencies and nations that collect, archive, and analyze Earth science data; and

- establish data and information interfaces between EosDIS and other national and international active data archive centers for multidisciplinary and interdisciplinary scientific data analysis.

SCIENTIFIC INFORMATION FROM EOSDIS

What distinguishes EosDIS from current remote sensing data systems is the commitment to provide usable scientific information to the geophysical, biogeochemical, ecological, and interdisciplinary communities. Eos data products will be used by a wide spectrum of scientists and the public during the 15-year life of the Mission and for decades afterward. Standard, reliable data products, essential to distinguish natural and anthropogenic variations, will give the community access to independent measurements to validate and drive models of processes at local, regional, and global scales. The characterization will include algorithms for generating the products and descriptors of data quality, and each data set will include the identity of the responsible scientists. Algorithms for standard products are certified by a peer-review process, and the products will be available through EosDIS for all cases for which the appropriate input data exist. Moreover, specialized products created by Eos scientists in their own computing facilities are to be archived and distributed by EosDIS.

Recommendations

The Eos IWG, assisted by the Eos Project, must:

- develop peer-reviewed algorithms for standard products and methods for scientific quality control and assessment;

- provide mechanisms to resolve and support inter-instrument product dependencies; and

- compare data products expected from Eos instruments with requirements established by

Eos interdisciplinary investigations to identify gaps, so that we can work toward realistic expectations.

The Eos Program must:

- nurture the creative investigators who begin to develop the appropriate data sets; and

- help the Eos investigators and Project find the best ways to move scientific algorithms from local interactive facilities to wider distribution.

CONTINUITY WITH PRESENT SYSTEMS

Eos data will continue existing measurements, some of which currently extend for more than a decade. Eos Science begins now, not with the launch of the first Eos platform in 1997. Starting immediately, EosDIS must develop current and previous data sets, measurements, and provide them to other investigators, so that the Eos community can gain experience with data processing, archive, and distribution centers. A few current centers where remote sensing data are intensively and routinely analyzed into scientific products provide the heritage for design and prototyping of Eos data processing and distribution, especially for data sets consisting of scientific interpretations rather than satellite-level radiance measurements.

Recommendations

The Eos Program and Project must:

- identify appropriate current and previous data and promote rapid development of them so that they will be compatible with anticipated Eos data;

- assist the EosDIS designers to acquire experience with data processing centers and archives that are currently active; and

- begin the development of data and information interfaces between EosDIS and other national and international active data archive centers for multidisciplinary and interdisciplinary scientific data analysis.

EVOLVING DESIGN AND ARCHITECTURE

The important questions that drive the design and architecture of EosDIS involve the procedures by which scientific products will be created and distributed, as well as how the styles of interaction with EosDIS will change as Eos science matures. An Eos data processing and distribution system, including visualization and browse capability for both image and non-image data and information derived from Eos sensors, requires an evolutionary, distributed design because of flexible research utilization of the data and because of rapid developments occurring in computer hardware and software. As Eos matures, specific algorithms will be formulated, mutual product dependencies resolved, and interdisciplinary data requirements defined. The process of producing and analyzing data will continue to lead to new methods of producing scientific products and new computing requirements. The system architecture must accommodate data from different kinds of sensors, changes in available computer hardware, software, and communications, different levels of human involvement in the creation of standard products, and different centers of expertise. A system that can address the changing nature of both its tasks and the available hardware and software inherently must be designed for easy, graceful evolution, both before and after launch of the Eos platforms.

Recommendations

The Eos Project, assisted by the Eos IWG, must:

- develop scenarios, for processing of Eos data into scientific products, that will address the style, volume, and human interaction needed by the processing facilities;

- promote diverse user facilities that take advantage of existing and anticipated scientific expertise; and

- develop realistic expectations for rapid browsing and visualization of large data sets.

EosDIS must:

- adhere to a flexible, distributed, portable, evolutionary design;

- distribute data products by appropriate high-bandwidth communication or other media; and

- operate prototypes in a changing experimental environment.

PARTICIPATION OF EOS SCIENTISTS

The responsibility of EosDIS to provide scientific information as well as raw data has significant implications for the time and requirements for support of many participating scientists. An Eos investigator who is responsible for a standard data product has a commitment for

the life of the mission—including maintenance of the algorithm and code as the instrument and environment changes, as well as quality control and validation of the product. The investigator must also provide additional information (meta-data, browse products, documentation) that will enable others to identify and use the scientific products. The scientist must also be flexible about generating information the community needs even if not promised in the original proposal. For algorithms to be maintainable, the code must be readable and portable, and therefore written and documented using good scientific coding standards.

Recommendations

The IWG scientists must participate in:

- development and review of standards and procedures for software, graphics, images, and operating systems;

- definition of products, formats, units, and coordinate systems;

- peer-review of documentation of algorithms; and

- crediting and citing in the resulting research papers the scientists who developed the data sets.

CRITERIA FOR SUCCESS OF EOSDIS

The scientific community will judge EosDIS by how well it supplies reliable and significant data products and promotes scientific interaction with those products. Indeed, EosDIS will have to supply the data and the tools that will allow the community to establish the long-term quality of Eos data. The community will also use EosDIS to derive and validate Earth System Science models and to document global change.

Assertion

The most important criterion for judging EosDIS is its ability to improve researchers' productivity in Earth System Science investigations. A completely objective measurement of productivity will prove too elusive to define, but the Eos Mission and EosDIS will be judged by the quality, compelling results, and creative ideas in Eos scientists' publications.

I. A Conceptual Framework for EosDIS

This introduction is intended to help focus thinking about policy issues and functionality toward Eos data and information management, in particular on how to approach the question "What is EosDIS?" It articulates a generic model for the scientific enterprise, which may help to identify general characteristics relevant to system design and evolution as well as gaps in present thinking about information management. This model, however, is certainly oversimplified and should not be applied as a template in individual cases or construed as a constraint. Likewise, the framework incorporates, in summary form, some programmatic and conceptual developments that post-date the existing documents for Eos science and for EosDIS. It should not be regarded as superceding these documents on substantive issues, but instead as a spur to their review and amplification.

THE CHALLENGE FOR EOSDIS

Investigation of the causes and magnitudes of environmental change, especially on large regional and global scales, depends on future excellence at tasks in which the record of the scientific establishment is not encouraging—the integration of data sets and geophysical and biological products of established quality and reliability. If EosDIS is to succeed, the participating scientists must have a strong role in its development and a sense of ownership in that success. The Science Advisory Panel for Eos Data and Information represents the Eos Investigators' Working Group and non-Eos scientists to help ensure that EosDIS meets the goals of constituencies interested in global change.

EOS SCIENCE GOALS

The key scientific questions that Eos addresses are:

- How does the Earth function as an integrated system?

- At what rate is it changing? How are the changes in the system forced? What is the envelope of variability?

- Can the effects of human activities on the Earth system be unambiguously detected?

- Can the cumulative effects of human activity be predicted decades or centuries in advance?

- Can local manifestations that accompany global change be detected and predicted?

In addition to their relevance to Eos, these questions are crucial to the scientific objectives of the U.S. Global Change Research Program: to monitor, understand, and ultimately predict global change, both natural phenomena and the effects of human activity. Eos activities include both monitoring and quantitative modeling. The tools include:

- conceptual models—highly simplified but susceptible to easy experimentation to focus questions that are subsequently addressed in greater depth;

- process models—designed to treat some restricted aspect of the Earth System with as much fidelity as practicable; and

- comprehensive system models, yet to be fully developed, capable of describing global changes on time scales of decades to centuries.[1]

Such comprehensive system models are based on a set of state variables and algorithms that can, with prescribed external forcing functions, compute the evolution of this set with time. State variables are typically global scale fields such as temperature or nitrate concentration. For simplicity, intermediate constructs that have geophysical or biological significance, such as average water vapor flux, will also be included in this category. Comprehensive system models are used for a variety of purposes, each with distinct requirements for input information, as summarized in Table 1.

Associated with meeting these needs of models for information are three different types of scientific activity within Eos:

Sustained Global Measurements

Sustained measurements are needed that result in globally analyzed products of established accuracy. These are needed to document changes that are really occurring. The emphasis

---

1. Earth System Sciences Committee, NASA Advisory Council, *Earth System Science: A Closer View*, Figure 2.4.2 and Chapter 6, National Aeronautics and Space Administration, Washington, DC, 1988.

Table 1. Uses of Comprehensive System Models

Purpose	Input Information
Encapsulate existing knowledge:	Understanding of relevant state variables. Process algorithms.
Analyze model sensitivity to uncertain parameters:	Plausible ranges for critical parameters. Comparisons with process models.
Assimilate observations:	Measurements of global state variables.
Validate model performance:	Records of natural variability, including past states. Measurements of prominent phenomena, e.g. the annual cycle. Case studies of isolatable subsystems, including measurements of interface and key state variables.
Monitor change:	Sustained global measurements of selected state variables.
Analyze cause and effect:	Plausible scenarios for changing external forcing or internal processes, including both anthropogenic and natural perturbations.
Predict the future and assess uncertainty:	Best available input scenarios. Results of model sensitivity experiments. Analyses of what may have been left out.

here is on sustained products of established accuracy. A bench mark for our performance will be the quality and reliability of our measurements and our analyses, as seen from a future perspective: What will our successors think, 20 years from now, when they are making similar measurements and trying to determine whether the apparent changes are real or are artifacts of instrument design, calibration, sampling, or treatment of the data? For almost every variable the required measurement system is composite, involving a blend of satellite and *in situ* observations, many of them not under the direct control of research scientists. Few of the existing measurement and analysis systems are adequate, because of the dearth of information on long-term accuracy and lack of identified responsibility for reviewing or ensuring end-to-end performance.

Process Studies

Process studies concentrate on critical processes or phenomena, and resulting in better algorithms or measurement techniques. They are typically team projects of limited duration, guided by a science steering group, and drawing on a mix of intensive measurements from ongoing satellite and *in situ* systems and special-purpose instruments. Data management and quality control are typically handled within the project. Alas, just when this unique collection of data is beginning to affect process models and algorithm development for comprehensive models, project funding typically runs out and the data become inaccessible and unavailable, except in fragmented form, to individual project investigators and their students. Procedures for taking care of the broader community interest in this data set, as opposed to the highest priorities of the project, are often ineffective, even if they exist.

Data Archeology

Data archeology assembles and interprets existing records to reconstruct the history of individual variables or sets of variables. These studies are essential to determine the background of natural variability in the Earth system, and to broaden the range of independent data sets for validating system models. Major problems typically encountered are the lack of accessibility of existing data, the incompleteness of ancillary records, such as station histories or calibrations, and the difficulty of raising funds for demanding but unglamorous tasks.

It should be noted that each of these three classes of activity requires investment of high quality scientific talent in activities that do not directly result in scientific papers clearly attributable to individual authors. A major determinant to the effectiveness of the whole endeavor is likely to be the inability of management to foster

an environment in which individual scientists are motivated to apply their talents where they are most needed.

PROGRAMMATIC ENVIRONMENT

Eos is a NASA flight project, with substantial participation by USGS and NOAA, and an integral part of the U.S. Global Change Research Program,[2] which, in turn, contributes to internationally coordinated activities such as the International Geosphere Biosphere Program and the World Climate Research Program. Eos also includes basic research on selected topics in Earth System Science that are not directly related to high priority thrusts of the Global Change Research Program, but for which Eos capabilities are particularly advantageous. There are many direct participants in Eos from other nations, as well as data links with other U.S. and foreign measurement programs, both from satellite and *in situ*.

Within Eos, EosDIS provides a hardware and software system so that investigators can access data from the Eos instruments and products derived from these data. The term Eos information management is used here in a broader sense than EosDIS, a process that involves systematic consideration of everything that is required to turn the bit stream from Eos instruments and other sources into information that is really used for global change and other scientific objectives. Although principally composed of U.S. investigators, Eos is an international program, and EosDIS should therefore serve, stimulate, and strengthen international scientific collaboration.

EOS INFORMATION MANAGEMENT

Eos is a space flight mission with associated ground based information handling and science. Although a part of the U.S. Global Change Program, it has other objectives as well. Thus it cannot and should not take responsibility for all that is necessary to meet the needs of that program. Nevertheless, the magnitude and scope of Eos are such that it contains within itself a microcosm

of all these issues, as well as having a myriad of interfaces to the broader program both within and outside NASA that involve similar problems. It is clearly the flagship endeavor, and unless it has an effective strategy for approaching the human issues of information management, little progress will be made in the program as a whole.

Such a strategy must concentrate on incrementally building confidence in a partnership between Eos and the community that turns data into scientific information. For example, it should:

- Involve Eos investigators and, where necessary, non-Eos scientists, using existing data sets NOW to establish role models for effective data management and production of integrated products that really re-direct the progress of science. This will also involve funding participation in pre-launch process studies such as WOCE, TOGA, and ISLSCP using data from *in situ* and heritage instruments.

- Undertake analyses of the end-to-end performance of the existing global measurement system for selected variables for which sustained measurements are essential and Eos is expected to contribute significantly to the analyzed product; examples are stratospheric temperature, cloud amount, type and height, and land cover.

- Articulate the present status and required actions on instrument calibration, validation, combination of *in situ* and satellite data, quality control, data editing and analysis procedures, availability of independent measurements, and documentation, including potential system trade-offs.

- Focus from the outset on the mix of management options and hardware and software system design that can ease the critical transition within Eos from algorithm development and product definition to routine production. In particular it is vital to retain the committed involvement of key scientists in developing tools to monitor the production process, including quality control, establishing the strengths and weaknesses of the product, and demonstrating its utility in scientific applications. Depending on the variable, this transition will take years of patient nurturing, not a few months of checkout just after launch.

2. *Our Changing Planet: The FY 1990 Research Plan,* The U.S. Global Change Research Program, Office of Science and Technology Policy, Committee on Earth Sciences, July 1989.

- Establish mechanisms that ensure that Eos data dependencies on or obligations to other agencies or nations are treated as collaborative partnerships, not we/they divisions.

- Interact with the management of Eos science to ensure that a seamless support structure of carrots, sticks, hardware and software tools, and institutional arrangements is systematically fostering the involvement of scientific talent where it is most needed for complete program objectives, including all aspects of information management.

EOSDIS ARCHITECTURES

Within the context of a clear management focus, yet to be defined, on Eos information management in the broad sense, there is a range of architectures possible for the hardware and software system known as EosDIS. The appropriate compromise is under study now, but it may be worth reiterating two extremes or end members.

1. One architectural extreme concentrates on the data processing implications of high bit-rates from the Eos instruments, and envisages one or a few centralized production facilities institutionally that are distinct from and typically physically remote from the scientists who have developed the algorithms for deriving these products. These facilities would come on line "in toto" at or shortly after launch, and the main criterion for their success would be cost-effective throughput. This approach could be parodied as the "monolithic" alternative.

2. The alternative extreme approaches EosDIS as an evolutionary process, starting now with the selection and encouragement of independent activities, using existing data from various sources to derive immediately useful products under the hands-on control of individual scientists. As the more successful of these develop experience, and as others are added in the areas where there are gaps, these prototypes become, by aggregation, EosDIS. The launch of more capable instrumentation would be an incident in this evolution. The main criterion for their success would be the demonstrated application of data outputs in process studies and numerical models in a variety of areas of Eos science. This could be parodied as the "prototypical chaos" approach.

In reality, major compromises between these end members will be necessary. Handling large quantities of data can be expensive, but rewriting software or failure to reach program objectives can be even more so. As emphasized above, the people dynamics of EosDIS are probably even more important than the hardware and software dynamics. Experience shows that monolithic systems tend to be institutionally inflexible, and work against the "try it and see" approach and close personal involvement of most productive scientists. Likewise, close analysis of the prototype approach shows that it presumes at the least some standards for ready communication of data and data products between independent centers. Without constraining the internal representation of data structures within a database, it is critical to have a common interface language that both preserves the information content of data relationships and is reasonably efficient. This must be backed up by a centralized data directory and catalog function. Also needed is a dictionary of terms. For example, the word "temperature" is used in many different senses within the Earth Sciences community, and someone has to establish and maintain system-unique names and aliases that can be used on-line to resolve ambiguities and conflicts. Apart from the cost of bulk data shipments in near real time, the problems that have to be addressed include version control for multiply-distributed data sets and conflicting priorities within many different groups. Probably the only thing that can now be said with confidence is that the initial specifications of EosDIS will prove inadequate, and will need to be changed. .

CRITERIA FOR SUCCESS

How might the success of Eos data management be judged? There are three distinct constituencies, each of which must be reasonably satisfied.

- The first is the typical Eos investigator, asking primarily how EosDIS has helped with his or her everyday data tasks.

- The second is the Earth system modeling community, asking where are the reliable global analyses and integrated data products that can be used to validate and run models, and whether observed trends signify anthropogenically-induced global change or just natural fluctuations.

•, The third is typified by a member of the congressional staff, asking: "What has Eos done for the nation, and, incidentally, what was the contribution of EosDIS?"

Clearly, satisfying all three constituencies will require leadership and imagination of the highest order. The challenge is to fashion approaches, institutional arrangements, and reward structures that aid rather than obstruct this task.

Page intentionally left blank

II. Charter and Activities of the EosDIS Advisory Panel

FUNCTIONS

The Science Advisory Panel for Eos Data and Information (the EosDIS Advisory Panel) is chartered by the Eos Investigators' Working Group (IWG) to fulfill the following functions:

1. To review EosDIS on behalf of its users—the Eos science community and non-Eos scientists who will want to use Eos data—and the Eos Program and Project for responsiveness to scientists' needs and incorporation of scientists' suggestions. Such reviews shall include both Project-originated reviews, such as the System Architecture and System Design Reviews, as well as Panel-initiated reviews. The Panel will report the results of their reviews and make appropriate recommendations to the Eos Investigators' Working Group (IWG) and individual investigators, to the Eos Program at NASA Headquarters, and to the Eos Project at NASA Goddard Space Flight Center.

2. To represent the scientific community in matters relating to Eos data production and science interfaces, including:

- assuring access to Eos data by the scientific community, including non-Eos scientists;

- advising EosDIS developers on requirements and priorities of Eos scientists;

- suggesting approaches to EosDIS developers that might improve the system or save costs or schedule; and

- assisting in disseminating EosDIS status and approaches to the scientific community.

3. To assist the EosDIS developers in areas of panel expertise, including

- reviewing, advising, and consenting to project recommendations on standards, formats, and nomenclature for data and software;

- providing advice regarding scientific requirements and science interfaces;

- reviewing illustrative scenarios for data production, EosDIS operations, and data validation;

- participating in prototyping and test evaluations of EosDIS; and

- suggesting priorities for EosDIS development.

4. To assist in the development of appropriate data policies including:

- availability of Eos data and derived products with no period of exclusive use;

- approval of algorithms;

- reprocessing of products; and

- data use.

INTERFACES

The EosDIS Advisory Panel shall:

- provide an interface between EosDIS and the Eos investigators on questions relating to EosDIS and Eos data; and

- provide an interface between Eos scientists and the EosDIS system developers.

MEMBERSHIP

The Advisory Panel is composed of scientists or their identified representatives, chosen to represent the interests of EosDIS participants and scientists. Each of the Facility Instrument Teams shall have a representative, and there are also representatives from selected Instrument and Interdisciplinary Investigations. The Panel also includes scientists with no direct participation in the Eos Project, as well as experts in data systems, software development, space flight operations, and instrument control who have no direct involvement in the EosDIS design activity. The Panel selects its Chairman from members of the IWG who are either Eos Interdisciplinary Investigators, Instrument Investigators, or Facility Instrument Team Leaders and who have no implementation role in EosDIS.

The Advisory Panel for EosDIS is established by the Eos IWG, with advice from the Eos Project Scientists and the Eos Program Scientist. Ex-officio members of the committee include the Project Scientist for Data, the EosDIS Manager, the EosDIS Program Manager, the Eos Project Scientist, and the Eos Program Scientist.

IMPLEMENTATION

The EosDIS Advisory Panel will prepare and update a detailed list of meetings and reviews. The Panel will also prepare a detailed implementation plan for its activities, including necessary support. This plan will show how it will interact with EosDIS, with the Eos Program, and with the IWG and Eos investigators. The plan will be

updated semi-annually by the Chairman. The Panel will convene subpanels, possibly ad hoc, when necessary.

Advisory Panel funding shall be supplied by the Eos Program office at NASA Headquarters and the Eos Project at NASA/GSFC. Initially funding will include support for:

- panel members to review and experiment with recommended standards, guidelines, etc.;

- meeting-related travel expenses not covered by other funding;

- hardware and software for possible additional electronic data communication between members;

- entering, organizing, maintaining, and reviewing input and responses between panel members and the EosDIS Project; and

- prototyping and test-bed data collection, entry, and analysis by panel members.

PLANS OF THE ADVISORY PANEL

The Advisory Panel will review the work of the EosDIS project and the contractors and make suggestions for modifications or new directions. We believe the Advisory Panel should provide guidelines in those areas where science efforts could be significantly affected by adverse proposals. The Panel will review and propose changes as necessary to EosDIS-generated documentation. The prime areas of panel concern are likely to be: system environment, software implementation standards, data standards, coding standards, user interfaces to EosDIS, and the agreements on allocation of development and operational responsibilities between science team members and EosDIS. Panel coordination with CCSDS and other activities to develop standards is especially appropriate.

Suggested Subpanels

The Panel will convene and disband subpanels, as necessary, to address particular tasks.

Project Review and Design

- Attend presentations by Project and contractors and provide comments to Advisory Panel, EosDIS Project, Eos Program, and Eos IWG as appropriate.

- Evaluate guiding philosophy of Project, including the overall conceptual design, its

correspondence with scientific guidelines, and alternative architectures.

- Examine current analogs of EosDIS, so that the proper heritage is incorporated. Help establish a mechanism for resolving design problems as EosDIS evolves.

Geographic Information Systems

- Develop geo-reference and mapping recommendations.

- Develop topographic reference recommendations.

- Develop recommendations and review prototypes for geographic searching of EosDIS.

Information Access

- Identify crucial non-Eos and pre-Eos data with multiple investigator use.

- Assure that EosDIS includes appropriate non-Eos and pre-Eos data to provide the necessary long-term geophysical time series to evaluate global change.

- Develop recommendations to assure that data producers make data, including geophysical products, available in a timely fashion.

- Provide mechanisms for resolving disputes between scientists, producers, and EosDIS.

- Provide adequate definition of data browse requirements.

Software and Data Standards

- Review proposed software development procedures and standards and make appropriate recommendations to EosDIS Project and Eos Program. Included are standards that have already been developed or are under development if these can accommodate Eos requirements, such as the NASA Master Directory and the CCSDS.

- Experiment with prototypes for standards.

- Review proposed standards and guidelines for software documentation, coding standards, acceptance criteria, and configuration control procedures for algorithms.

- Review proposed and prototype standards for data exchange, including data dictionaries and lexicons, measurement descriptions, catalog contents, and input and output data structures.

- Review proposed configuration management.

- Review proposed scenarios for prototyping.

Scientists' Computing Facilities

- Review proposed interaction and networks between EosDIS and scientists' computing facilities at their home institutions.

- Evaluate potential roles of scientists' computing facilities in production of geophysical products.

- Review the suite of computational styles adopted by Eos scientists, and examine standardization needed to interface with EosDIS.

- Recommend guidelines for acquisition and operation of scientists' computing facilities.

Project Phasing

- Review and recommend scenario development regarding project post-launch phases, including initiation;
  optimization of sensors, mission, and products;
  cross-product experiments; and
  end-of-life experiments.

Support for Panel Activities

Meaningful professional recognition must be provided to the scientists who devote their limited time to activities of the EosDIS Advisory Panel. The current highly active members spend about 25% of their total working time on Advisory Panel activities. Time devoted to this work will diminish the time available for algorithm development and instrument development. Without arranging for proper participation, these key issues will not be addressed until too late to be effective.

The following steps seem to us to be essential:

- provide support for investigators and supporting skills on Advisory Panel;

- provide for paid staff to provide independent review of EosDIS and support for standards; and

- provide for professional recognition of Investigators who contribute to Science Advisory Panel for Eos Data and Information.

Table 2. Scheduled Activities of the EosDIS Advisory Panel

March 20-24, 1989	Eos "All-Hands" Meeting. EosDIS Advisory Panel Formed.
April 3-7, 1989	EosDIS Phase B Program Requirements Review. Advisory Panel Preliminary Comments.
April 14, 1989	Preliminary Panel Comments to Project.
April 17, 1989	Panel Meeting at GSFC. Discuss Charter and future work.
April 30, 1989	Panel Report on PRR to HQ and Project.
May 3-4, 1989	NAR Dry Run. No Panel participation.
June 28-29, 1989	Panel meeting. Formulate request to Eos Program for support for panel activities.
July 20, 1989	Eos IWG Science Executive Committee meeting. Panel Chairman reports on Panel activities.
July 24-28, 1989	EosDIS System Architecture Review. Panel Representatives attend and formulate comments to Panel, EosDIS Project, Eos Program, and Eos IWG.
September 14, 1989	Eos IWG Science Executive Committee meeting. Panel Chairman reports on Panel activities.
September 19, 1989	Panel meeting. Edit this report. Plan presentations for October Eos IWG meeting.
October 11-14, 1989	Eos IWG Meeting at JPL. Panel Chairman and members report on Panel activities. Comments about EosDIS received from IWG members.
October 30 - November 3, 1989	EosDIS Preliminary System Design Review. Panel Representatives attend and formulate comments to Panel, EosDIS Project, Eos Program, and Eos IWG.
February 12-16, 1990	EosDIS Final System Design Review. Panel Representatives attend and formulate comments to Panel, EosDIS Project, Eos Program, and Eos IWG.
April 1990	Contractors' Phase B reports due. Panel advises project on Phase B contractor activities.
September 1990	Release of EosDIS Phase C/D Request-for-Proposal. Panel reviews and approves RFP before it is released.
November 1990	Phase C/D Investigator proposals due. Panel assists in evaluating contractors' proposals.
August 1991	Eos Phase C/D Investigator selections made.

III. Conceptual Design and Implementation

Among the issues for EosDIS, the first one is to resolve its conceptual design. What do we want EosDIS to do for us? Included in this area are the data supplied by EosDIS, geophysical and biological products supplied, time required to produce various levels of products, and the style of interaction between the scientist and EosDIS.

DISTRIBUTED ARCHITECTURE

The architecture emerging from recent discussions with EosDIS Project and contractors seems much less centralized than the presentations at the All-Hands meeting suggested. We must make sure that a distributed system results, with the right mix of processing and development at facilities operated by investigators and those operated by EosDIS. The production facility for data products, the support for algorithm development, calibration and analysis, and the distribution of archived products are probably not best done in single entities. Normally architecture is not an issue for functional requirements, but instead one for implementation. We raise the issue here to make sure that the form of EosDIS is not finalized before the functional requirements are settled. It is essential that the various Memoranda of Understanding signed do not unduly constrain the design of EosDIS.

Some of the presentations given thus far show that centralized systems along the "Computer Center" model are being considered, and these will dictate the style of interactions between the users, the algorithm developers, and the various instruments. The highly distributed nature of the community and the availability of inexpensive computer technology at many levels of speed and size should be recognized with the allocation of function-dedicated processing facilities convenient to individual or small groups of users. The experience of the university community is that a set of smaller computers provides more power more readily available at a lower cost than large centralized facilities.

Our recommendation is that the scientific functions of EosDIS—command and control of the spacecraft, data acquisition, distribution of instrument Level 1 data to investigators, information management, interaction among investigators, creation of geophysical and biological products, and archiving and distribution of data and information—should be separately optimized.

Smooth interfaces are also important. Where a centralized system is appropriate, let us build one, but for those functions where more distributed processing is best, let us use EosDIS to provide networks, occasional access to supercomputers, standards, maintenance, and advice. Where processing can be routine, without continuous involvement of scientists, let us design a system to process efficiently, but where continuing scientific evaluation is needed in the creation of science products, such a centralized environment will hamper progress.

Most scientists' concept of their ideal computing environment involves:

- powerful workstations under their control;

- painless and affordable access to data from big archives or from other investigators' workstations;

- confidence that the geophysical and biological products they obtain are produced by knowledgeable people and have established quality and reliability;

- communications networks, for easy exchange of information, without the arcane knowledge currently required to go between systems; and

- occasional access to bigger and faster computers.

BROWSING OF EOS DATA

Browse capability is an essential feature of EosDIS, but it is potentially an unbounded cost driver. The Advisory Panel requires that EosDIS provide interactive electronic access to some rationally chosen subset of Eos data, and not just to "metadata"—summaries or catalogs of data. However, it is critical to get the perception of this rationally chosen subset bounded so that EosDIS and its contractors will accept it as a reasonable requirement. Interactive requests that require analysis of a large volume of data while applying conditional tests will consume enormous resources, but restricting browsing to catalogs and metadata will increase the volume of data that investigators have to order.

Thus we must work for an acceptable compromise. There are several reasons why access to actual data is important to Eos. Such capability is needed for:

- checking data for cloud cover or image quality before ordering;

- monitoring of instrument health and safety;

- data validation; and

- full interdisciplinary use of Eos.

Consider cloud cover. It is the community's common experience with Landsat or SPOT data that the cloud-cover statistics furnished as metadata, either in percent or tenths of cloud cover, are useful only when near either extreme. If half of the image is cloud-covered, one needs to look at it to determine if the areas of interest are covered.

Consider instrument health and safety in the context of an incident that threatens a failure of several instruments. We may expect the Eos instruments to send out many more pieces of housekeeping telemetry than will fit comfortably on one screen. Furthermore, the data rate for many of the instruments is high enough that their data cannot be conveniently plotted except for a few, preplanned scenarios. An instrument engineer may need to examine many different kinds of data with time-history plots to determine the cause of a problem and suggest appropriate action.

Consider next the period of data validation. Here we expect that the investigator will have discovered some anomaly in the data, but the cause of the anomaly is not clear. Usually, the investigator must sift through various measures to find the cause of a problem. On ERBE, for example, such an interactive database tool was essential to be able to determine that observations of the contamination covers were not appropriate to estimate scanner offset. The same tool was used to select appropriate data for the offset determination. A history of the individual data points had to be plotted before noisy data could be removed. Although the need for offset determination had been expected before launch, the values could not been found rapidly enough to produce useful data without this interactive data analysis tool.

Finally, consider the ability to access multiple data sets during an interactive investigation. Let us suppose we are interested in building up a statistical picture of how volcanic eruptions act on their environment. One strategy for such an investigation would be to select MODIS scenes that bracket an eruption. If we can examine a MODIS scene that surrounds the volcano, we can perhaps count pixels to estimate the area affected immediately after the eruption. By taking several such scenes and counting pixels, we could build a history of the area affected and its recovery. Confirming scenes from HIRIS could then be identified, screened for clouds, ordered, and subjected to more intensive analysis of the affected areas. At the same time, cloud identifications from CERES and temperature and humidity profiles from AIRS could be used to estimate changes in precipitation and runoff associated with the change in the volcano's surroundings.

In this last scenario, the plot of affected area and its correlation with runoff over several years' time might constitute the final result of the investigation. If electronic browsing allows the investigator to access the data, it is conceivable that it could be completed in one or two interactive sessions with EosDIS. A scenario that allows access only to metadata or highly summarized browse data-sets limits such interdisciplinary investigations. Indeed, without this interactive data access, Eos data use are almost certain to be confined to our current disciplinary structure. Interactive access will increase efficiency and productivity and foster interdisciplinary cross-fertilization and development of new concepts.

The alternatives are analogous to those posed by libraries with closed stacks as opposed to those with open stacks. In a closed-stack system, a researcher browses through the card catalog for book titles or authors of interest to him. He then requests these books. After these arrive, he must search them for the few pages of interest. The system is easy to monitor. It provides the library with a simple data access problem, although staff must be maintained to search for the appropriate book. In contrast, an open-stack library allows the user direct access to the few pages that he needs. Only these need to be copied. Furthermore, he can examine nearby books for similar titles and better explanations.

Costs of such access are likely to be high on initial examination. However, we must remember that there are hidden costs in the usual view of interactive searching through data. With browse only through metadata and predetermined browse data sets, there is less electronic data flow to the initiator. The catalog is small, making the search fast. However, there is a higher probability that the investigator will miss the signal

he is looking for. More important, the investigator will have to order large amounts of data and provide his own software for sorting through the data. For truly interdisciplinary investigations with multiple data sources, such software development is likely to become a larger drain on resources than is the search through the data.

CONTINUITY WITH PRESENT SYSTEMS

Access to selected non-Eos and pre-Eos data must be part of EosDIS. Earth Systems Science does not begin with the launch of the Eos spacecraft in 1997; instead, Eos must continue the available time series of geophysical measurements on a global scale. We recognize that this requirement, as stated, is open-ended. At the same time, it is clear to us that the way to prepare for Eos is to start now to produce, archive, and distribute geophysical products from available satellites and *in situ* data collection systems. The way to combat unease and disillusionment about large, futuristic data systems like EosDIS is to develop successful prototypes that allow us to investigate and solve today's problems. We need to examine the successful role models, improve on them, and use the vast existing archives of data, accessible now only to the *cognoscenti*. We need to start with selected end-to-end tests with existing data.

Data and products from previous Nimbus missions and from the NOAA polar-orbiting and geostationary satellites are an essential key to the strategy of smooth transitions between the past and the future, and between research and operational satellites. Although most NOAA operational instruments are no longer included on the Eos polar platforms, deletion of NOAA data from EosDIS should not follow as a consequence. EosDIS and the appropriate Eos investigators must identify critical NOAA operational data that are still required, and such data must be available to the investigators, either directly from NOAA or through EosDIS. NOAA operational data, particularly the gridded temperature, pressure, humidity, and wind fields, along with normalized vegetation indices, are needed by Eos investigators to provide synoptic views of the world for various data processing tasks. These include accounting for atmospheric effects and cloud determinations. We cannot be sure that the retrieved quantities from Eos will replace the gridded data, so the NOAA data would provide a useful backup function. In addition, the NOAA products integrate data from other sources that

EosDIS will probably not include routinely, such as radiosonde observations.

Data from many other missions should be included in EosDIS prototypes: SAR data from the Earth Resources Satellites (ERS-1, J-ERS-1, Radarsat), the SSM/I data products from DMSP polar orbiters, Earth probes, foreign Eos missions, plus appropriate *in situ* and airborne data.

NEAR REAL-TIME DATA

The two-day turn-around for Eos Level 1 data is well beyond previous experience for some communities, but is terribly slow for others. EosDIS should recognize the scientific justification for rapid availability of a few of the data, within hours of satellite overpass. There is at least two Eos interdisciplinary investigations—studies of volcanoes and sea ice—whose conduct requires rapid access to data.

What is not required, however, is an approach that forces all quick-look and near real-time data processing through the same central facility where data are routinely processed. EosDIS expects to be able to deliver priority data (unprocessed), at the cost of reproduction and distribution, via TDRSS and retransmission by communications satellite within hours of acquisition for up to 5% of the total data volume. Quick-look products should be available in the same time-frame. Does this place a special performance requirement on EosDIS that may not be needed in the general case? For small, quick-look data sets, it may be feasible and cost-effective to process by other means, including direct down-link. The likelihood of more timely delivery of data may also be enhanced in this scenario.

Among the other possible reasons for adding direct communications are:

- backup of TDRSS if that system is unavailable;

- transmission of research data to remote sites to support field experiments or operations;

- provision of data to international partners more rapidly and more directly than possible through TDRSS/EosDIS; and

- demonstration of next-generation direct broadcast service for prototype operational instruments.

Still at issue is the volume requirements for such a direct down-link. Possible for implementation are a 2–10 Mbits/sec X-band link, which

could serve both research and prototype operational users, and a 100 Mbits/sec S-band system that would be compatible with existing Landsat and SPOT ground stations. There is also interest in the possibility of a 56 kbits/sec UHF link to serve low-cost ground systems.

ON-BOARD PROCESSING

Some proposals for on-board processing are not traceable to real requirements. They appear to be driven by optimistic assumptions about the need for on-board control including modifications to sequences, command generation and validation and "generic" data processing. Common functions frequently do not have common implementations when more detailed analysis of requirements is performed. Even common functions frequently need separate copies dedicated to individual uses to avoid aggregating a simple problem into an unmanageable one. Rather than assume an extravagant menu of capabilities on a universal scale that is probably beyond the technology that will be available in the flight-qualified form, we should concentrate on the few most basic and plan for the ability to add functions on later platforms as the initial capabilities are proven.

SCENARIOS FOR EOS INVESTIGATIONS

The two competing contractors, Hughes and TRW, have developed scenarios to guide EosDIS planning. This is a positive and extremely useful step, and we believe that these scenarios need to be made independent of the contractor efforts, distributed to the Eos investigators, and iterated with the Project. We think that these scenarios are an opportunity for the project to produce enthusiastic investigators. They are specific enough to provide useful feedback in a short time. They can serve as a test-bed for operations planning and testing. They may be useful for bringing potential interactions between investigation teams to light. This could aid in conflict resolution and cooperation amongst investigators early in the project. To fulfill these goals requires that the set of scenarios be complete, comprehensive, and understandable. They must reflect the full spectrum of Eos investigations and the Eos concept, not just an interesting subset, and they must cover the full end-to-end set of interactions between scientist and instruments.

IMPLEMENTATION

Once we have a set of functional requirements, the question is: How do we design and build EosDIS? Included in this area are the choice of hardware and software, networks among investigators and EosDIS nodes, and staging of the system. Among the Eos IWG, we have a wide range of experience in creation of geophysical time series from large remote sensing data sets. Thus it is important that we learn from our past experiences about how EosDIS should be configured.

EOSDIS DEVELOPMENT

The Project must not aim for rigid requirements and architecture on EosDIS early in the life of the system. There is considerable experience in the data processing community that suggests that specification of firm, difficult-to-change requirements too early in the development is likely to produce a system that does not respond to what users need. For example, *Science* recently carried a story criticizing the development of the Space Telescope software.[3] Rigid adherence to requirements without providing for user interaction is quoted as a cause of a substantial increase in cost, as well as a system that does not work as desired.

The requirements will change during the design and construction of EosDIS and throughout its life. A perceived impediment to best designing and building EosDIS is the problem of phasing within the NASA procurement structure. It is difficult to write a specification for a system that one does not want to specify until one has tried out prototypes. Yet we must begin to put money and talent beyond what is available within the EosDIS project into the execution phase system. A potential solution is to design an evolutionary process into the EosDIS contract. Such a scenario would divide design and development into three phases:

* system engineering and integration services;

* firmly specified command and control functions—tightly coupled with spacecraft and instrument design and of a confined scope; and

3. Waldrop, M. M., "Will the Hubble Space Telescope compute?," *Science*, vol. 243, pp. 1437-1439, 17 March 1989.

• data processing and archiving and software development—to include prototypes, initial software-support environments, scientific products from existing data sets, hardware, and access to networks.

EosDIS needs to make specific provision for input and revision during the entire contract. Incremental development and testing are key concepts in modern methodologies for software-intensive systems. NOAA AVHRR data can be used now to develop methods to process and archive MODIS data and create geophysical products from them. Similarly aircraft data from AVIRIS and ASAS and satellite data from Landsat and SPOT can be used to develop strategies for HIRIS. Detailed methods are suggested in the later section on alternative architectures, our vision of how EosDIS development might proceed.

It is necessary to define external data interfaces as early as possible, to allow users to get on with development inside the system. It would also allow external suppliers and users to have a well-defined interface while development proceeds.

Continuing development of a complete set of user scenarios based on inputs from Eos investigators is necessary. These should reflect the variations in use of Eos data from each of the different types of investigators and should be updated as their plans mature. The users' scenarios are usually based on specific geophysical data. Most users do not want to have to know much about EosDIS as a whole. The Eos IWG, with help from the Project, should develop a set of data use scenarios, to help guide EosDIS design.

The Project would have a much more definite way to judge sizes of data products and timing of their availability. The scientific users would have specific suggestions of how their data would be used, and this involvement should increase their interest in the system, thereby making them more attentive to the details by which the system will live or die, as well as making them proponents of the system if they interact with it successfully.

Everyone would get training in how to plan and schedule for EosDIS. Scenarios that emphasize the roles of EosDIS and the various science contingents in planning, scheduling, and commanding would help to clarify an area that currently is not cleanly defined.

We need early prototyping of a functional skeleton for EosDIS. Modular design, identifying the most critical modules, and user evaluation of prototypes of critical elements are consistent with modern methodologies for large systems. Asking groups of users to respond to requirements before they have had any experience with the system is likely to lead to misleading answers. Users are busy and frequently do not speak the same language as the system designers; they are not likely to worry about abstractions for activities more than six months into the future. Questions such as "Does this software meet your requirements?" are not likely to elicit helpful responses. On the other hand, a more specific question—"Is this display of a simulated temperature profile from your instrument coming back to you quickly enough for you to build your climatology?"—is likely to get a more interested and thoughtful response.

Use of skeleton configurations augments formal inspections of software. It is a formidable task to wade through masses of documentation and to test programs adequately, but no one can make sure a program will do what it is supposed to solely by examining specifications. Most of us need real experience with how the system works with our own data. Details and the sequence of details matter!

We strongly urge the EosDIS Project to deliver working prototypes of parts of the system as early as possible, even if the prototypes lack the features that the full system will have. Early prototyping and experience with EosDIS is essential to provide adequate understanding of the effort required. Algorithms need several years of work before they can be trusted.

ALTERNATIVE ARCHITECTURES

The current concept of EosDIS is that two or more "Central Data Handling Facilities" will fulfill all processing needs except algorithm development and individual scientists' investigations. These will be complemented by "Data Archive and Distribution Systems" that will be accessed by an Information Management Center and that will distribute data sets to investigators. Explicit in this concept is that a single facility or generic class of facilities can best fulfill these functions. However, such a solution is not optimal because different products demand different expertise.

Some functions are indeed best handled by centralized, spacecraft-oriented or instrument-

oriented facilities. "Raw" Level 0 products involve the spacecraft and data-relay satellites. Level 1 products, i.e. radiative measurements at the spacecraft, require intimate knowledge of instrument characteristics and need to be done quickly.

For some functions, an instrument- or discipline-oriented facility would provide the most reliable products. The creation of geophysical, Level 2 products usually, but not always, requires data from only one instrument and involves intimate knowledge of its characteristics. Some scientific products use data from more than one instrument, and some products may be developed by scientists who are not on an instrument science team.

Higher-level (3 or 4) standard products may be most reliable if produced at a discipline-oriented facility or at the investigator's home institution. Many useful regional or global-scale scientific products are derived from a suite of instruments, and scientific expertise, cooperation, and interaction are needed.

Separation of product generation into different nodes within EosDIS is consistent with scientific goals and interdisciplinary research. The one facility that requires centralization is the Information Management Center, so that investigators can access products from a variety of sensors and disciplines. Therefore standards for data formats are crucial, and a single common catalog is needed. Any user should be able to investigate the availability and characteristics of all archived data, without having to use separate catalogs for different instruments or to learn new access techniques.

The current architectures proposed by the EosDIS contractors are based on their impressions of the needs of the Eos, and do not necessarily support the actual needs of the Eos investigators. Sizing and placement of computing resources should adapt to demonstrated needs. This would give each investigator the opportunity to consider and make known more realistic requirements.

INTERFACE DEFINITION DOCUMENT

Algorithm integration and testing should be regarded as similar to instrument integration and testing on a spacecraft. In other words, bringing an algorithm to a processing facility might be regarded as similar to bringing an instrument to be integrated on the spacecraft. This section of our report explores this metaphor and suggests some actions as a result of this view.

In dealing with instruments and spacecraft, project managers and engineers recognize that they must specify the interfaces between the spacecraft and the instruments long before hardware is delivered and connected. These interfaces are documented in an "Interface Definition Document." Such a specification assures that instruments know how much and what kind of power they will work with, the protocol for exchanging information with the spacecraft, the amount of flexure and location of center-of-gravity they are allowed, etc. This specification of interfaces allows the spacecraft engineers to proceed independently of the individual instrument engineers, at least in theory. Even with careful control of these interface specifications, projects still insist on testing that the instruments on a spacecraft do not interfere with each other after they have been bolted on and plugged in.

Software development for EosDIS will benefit from the same view. We are expecting the science investigators to develop their own algorithms. The software implementing some of these algorithms will then be brought to EosDIS to be incorporated into the system. EosDIS must therefore provide an interface definition, so that the algorithm developers will have the information needed to get their algorithms to work.

Accordingly, we would suggest that algorithm development proceed with interface definitions and controls being provided with documents similar to those for instruments supplied to spacecraft. In other words, we need an Interface Definition Document and an Interface Control Document. Plans and procedures for testing the integration of algorithms into EosDIS need to be defined as carefully as the plans for instrument integration into the platform.

IV. Scenarios for Scientific Investigations

The EosDIS project and the contractors have interviewed a big segment of the investigators, to obtain information needed to design EosDIS. However, much of the emphasis so far has addressed the size of the system. The presentations at the Eos All-Hands Meeting in March 1989 reinforced this perception, focusing on:

> How many terabytes per day?
> How many millions of tapes?
> Etc.

The data volume is huge, admittedly, but we believe that this should NOT be the major driver in the design of the system architecture. Instead, we should address the style in which scientists do their computing and how they expect these styles to change in the next decade. A particular concern is the amount of human involvement, especially of high-level scientists, needed to make "standard products." The view of the Project is that these are "routine" or "automated," and scientific involvement is more or less restricted to validating the algorithms and periodically checking the results.

We present a few scenarios in this section, to present an accurate picture of how a few particular scientists will operate, i.e. "One Day in the Life of . . ." Some of them are team-member activities, where the emphasis is on the production of Level 2 and 3 geophysical products. Some are interdisciplinary, where results of models that use such data are emphasized. Continuing development of such scenarios is needed, to

• generate interest in EosDIS;

• serve as a base for operational testing;

• provide specifications for tools and processes needed (e.g. browse products); and

• increase the likelihood that the system will function as we want it to a decade from now.

STRATOSPHERIC RESEARCH AT GSFC

Mark R. Schoeberl
NASA Goddard Space Flight Center

Research Overview

Both modeling and data analysis programs take place within our group. The modeling programs are at an early stage in comparison to weather-prediction models that are used at NOAA. Our models are of the same type, but they extend much higher into the atmosphere than the usual models; therefore we have a different suite of problems to be solved. The data analysis involves remote sounding data (from both the ground and from satellite) as well as aircraft data from the recent polar ozone missions and satellite data. We look at the data for their own sake and we compare the model calculations with the data. Thus much of our activity is simply trying to bring the various data sets to a common space- and time-resolution for intercomparison and examination. There is no unique or standard way this can be done. To work with the data we have to have a good knowledge of their resolution and their accuracy. That knowledge determines the way we will intercompare the data.

What Do We Do With Data?

The first type of activity involves what could be loosely called "Data Examination." We look at data sets and ask, "Do they make sense?" We plot them up in different ways. Is there anything interesting here? Then there is "Data Comparison." Do these data look like equivalent data taken by another instrument or taken the previous year? Often we rewrite the data to expedite the comparison. For example, early on we realized that the data volume produced by the Total Ozone Mapping Spectrometer (TOMS), one value for each 50 km × 50 km grid cell for each day, was too large to be handled quickly by our computer system. We rewrote the data by averaging it into 5° × 2° bins. Even this is a large data set (after 10 years of measurements) so we mostly use the monthly means. Here the data rewrite was motivated by the issue we were trying to address (not the computer system, although it sounds that way). We were looking at large-scale changes in ozone. For that activity a 2° × 5° resolution is fine enough. In another example, we extracted the highest resolution TOMS data, but over recent volcanic eruptions. It turns out that TOMS is good at sensing SO_2. Again the data set was reduced, but that reduction was scientifically motivated.

How Do We Work With the Data?

Our group activities involve writing many little programs in a high level language (we use IDL put out by Research Systems Inc. in Denver).

These programs extract and plot up data. We have tried menu systems and generally hate them—they are too constraining. In our experience, scientists do not want to be bound by a menu system; they invent their own analysis techniques as they go along. However, that is not to say a good package of analysis routines isn't helpful. For example, one common routine we use is a simple polynomial fit. Other routines like FFT's, eigenvalue generators, plotters of all types and mapping routines are all frequently used. These types of routines are part of the standard analysis "tool kit."

How Are the Data Stored?

We have written a group of generalized reader routines that grab chunks of data off disk as specified in the call list. These core routines, common to all users, also deliver back information on the local time, where the data might be bad, etc. These routines are then called by yet other routines that further process the data, or link the data sets (e.g. ozone data from one instrument, temperature from another). The data are stored in directories with logical names so that movement of the data from one physical unit to another is transparent to the user. The point is that we try to make the low-level data handling as invisible to the user as possible.

What Do We Do With Model Output?

Modeling output values are in many ways just like regular data, except they are usually of much lower quality. By this we mean that we are more likely to throw away a model run (because of a bug or a change) than we are to throw away data. Model output could also overwhelm a system more easily since one tends to save more things from a model than one can measure from an instrument and one tends to generate a lot more model values. Therefore we store the model output by changing some of the parameters to do sensitivity tests. Once we have some model output, we treat the values much like data. Everything about the analysis is about the same.

Drawbacks and Suggestions

Where are the bottlenecks?

1. We never have enough quickly accessible storage. We buy a new 600 Mbyte disk every 6 months and quickly run out of space again.

2. We spend a lot of time putting up data sets. Most people who give us data

- have incompatible systems or media;
- write data in ill-defined or inconsistent formats;
- do not give enough information about their data;
- do not even look at what they write out; or
- do not bother to check the values.

The storage problem is not as horrible as it sounds, we do swap the most timely data sets off disk when we are not using them. Storage constraints we can live with, but we balk at waiting for more than a day to get some important data.

Eos Specific Questions

Prelaunch Data

We are already working on prelaunch data products that could be put up on the EosDIS. This includes our own model output and aircraft and satellite data sets. These data sets are currently stored locally, but we could put them up on a prototype machine and make them available to the other investigators. NASA Headquarters has made some of these data sets available on optical read-only disks (CD-ROMs).

Most of the Eos constituent data we will select to analyze are the data sets that are similar to UARS (MLS, HIRRLS, DLS, SAGE III, SAFIRE, SWIRLS, AMSU) with the possibility of TES, MOPITT, and TRACER among the tropospheric instruments. We would also select NMC meteorological analysis and data assimilation analysis, which does not now exist but may be available during the UARS launch phase. A big problem for us will be mapping data sets at the same equivalent space and time.

We will be producing some assimilated data products of the same size as the meteorological analysis (global data on say a 2° by 3° grid, 20+ fields at 30 levels). These will be archived, and reprocessing with newer model versions will in all likelihood take place. We think, at least for the remote sounding instruments that analyze a portion of the spectrum, reprocessing is a reality. However, there is no need to keep the older (originally processed) data products on-line.

If it is feasible to run our data assimilation in real time, it will probably lag behind the satellite data collection by a week to two weeks.

What We Expect EosDIS Will Be for Us

We do not want to be running our own high level interdisciplinary analysis on EosDIS, but we

will build our own computer environment. We expect to connect to EosDIS through a network for data products or run programs on our machine that will access data on EosDIS. EosDIS should also route us to the data we may need to get from other investigator sites. EosDIS should also evaluate and develop high level software tools that can be used on our local machines. However, that activity should be closely scrutinized by and synchronized with the user base—too often computer centers become unresponsive to the users if not closely tied to the user base.

As far as our data assimilation is concerned, this has yet to be defined, but most likely we will store the output from our data assimilation in EosDIS for common use.

POLAR OCEAN ICE DATA AT NSICD

Roger G. Barry
National Snow and Ice Data Center
University of Colorado

The overall goal of this research project is to model polar surface fluxes of energy, ice, and carbon.

Pre-Launch Phase

Data
We use "historical" data sets (SMMR, SSM/I, buoy data of wind and temperature) and bulk-transfer formulations to compute flux maps for a low-resolution model of the Arctic Ocean ice-surface. Sea ice fluxes will require AVHRR thermal imagery. A major problem here is the absence of an archive of gridded thermal products commensurate with the SMMR-SSM/I grids available from NSIDC. Table 3 identifies the data requirements and potential sources of the data. EosDIS directory searching and ordering of such data will be necessary. Long period, hemispheric data sets suitable for analysis at large time and space scales are required.

Methods and Procedures

1. Test algorithm for snow on thin ice using passive microwave and SAR data.
   Output: Improved algorithm.

2. Test ice/open water algorithm using aircraft data for Bering Sea.
   Output: Improved algorithm.

3. Analyze ocean color and sea ice data.
   Output: Level 3 gridded products and time series for ice marginal areas.

4. Intercompare ocean analyses with ship/aircraft data.
   Output: Improved analyses.

Table 4 shows a schematic diagram on how seasonal ice cycles are modeled.

Post-Launch Phase

Goals
Updated polar ocean surface-flux model incorporating Eos and other new data and new models.

Data Required
Table 5 summarizes the changed data requirements following launch of Eos. Note that non-Eos data (ERS-1/Radarsat SAR, J-ERS-1, AMSR, and NOAA AMSU data) are required. Levels 1 and 3 data (orbital and gridded sensor radiances) and gridded geophysical parameters are needed within one month of acquisition. No real-time data requirement is anticipated.

An issue to be determined is whether selected medium-resolution MODIS and high-resolution HIRIS data will be required as unfiltered level 1 output, or whether a spatially-averaged cloud-free product will be preferable for the surface-flux calculations. In the former case, cloud filters or masks would be applied by us.

Data Delivered to EosDIS
Table 6 shows the anticipated products. Intermediate level data will include unaveraged values of upper ocean salinity, temperature and density from buoy locations. Guidance is requested on descriptors for the data products and related metadata summaries.
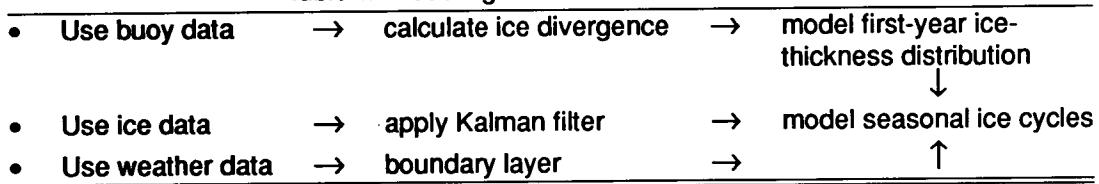
Resources
Computer resources for the project are being assessed. It is likely that several powerful scientific workstations will be required at University of Washington and one workstation at NSIDC, with commensurate magnetic and optical disk capacity.

Table 3. Data Rates for Polar Ocean Ice in Pre-Launch Phase

Data rates below 0.01 Gbits/yr are symbolized by a "—." All required data are Level 3.
Data needed over ice-free ocean are symbolized by (O), over ice-covered ocean by (I).

Variable	Sensor	Number of 8-bit variables or channels	Temporal resolution (days)	Spatial resolution (km)	Volume (Gbits/yr)	Source(s)
Surface wind and pressure	NSCATT (O)	2	7	50	0.01	?
	E-ERS-1 Scatterometer (O)	2	7	50	0.01	ESA
	Drifting buoys (I)	3	7	100	—	U. Wash, NSIDC
	WMO Surface Station Data	3	1	30 stns	—	NCAR, NCDC
Sea surface temperature	AVHRR IR (I)	1	7	4	0.8	NCDC?
	Drifting buoys (gridded)	1	3	100	—	U. Wash, NSIDC
Ice motion	Drifting buoys (gridded)	2	3	100	—	U. Wash, NSIDC
	(J-)ERS-1 SAR	2	3	10	0.6	ASF, ESA
	Radarsat SAR	2	3	10	0.6	Canada?
Ice type	SSM/I (and SMMR)	4	3	30	0.1	NSIDC
	(J-)ERS-1 SAR	4	3	10	0.6	ASF, ESA
	Radarsat SAR	4	3	10	0.6	Canada
	Airborne SAR	4	3	10	—	JPL
Chlorophyll	CZCS NASA ODAS ONR MARS Aircraft	1	7	10	0.1	NSSDC, NODS GSFC
Ocean parameters	Salargos buoys					J. Morrison, NODC?
TOTAL					2.4	

Table 4. Modeling of Ice in the Polar Oceans

• Use buoy data	→	calculate ice divergence	→	model first-year ice-thickness distribution ↓
• Use ice data	→	apply Kalman filter	→	model seasonal ice cycles
• Use weather data	→	boundary layer	→	↑

POLAR ICE PARAMETERS FROM SAR

John C. Curlander
Jet Propulsion Laboratory

A key scientific application of synthetic aperture radar (SAR) data is in observation of temporal changes in polar ice. The general requirement is for observations on several spatial and temporal scales depending on the phenomena to be observed. The key issues are:

• extent of polar ice cap as indicator of global warming;

• movement of ice pack and individual floes in the marginal zone as an indicator of large-scale polar circulation; and

• fractional concentration of ice type by age for evaluation of physical processes associated with heat fluxes between air and sea interface.

There are many detailed physical processes that also have important scientific implications that are not listed above. However, we confine the data flow scenario to these listed applications since they represent the primary classes of scientific studies. In addition, these applications have received a significant amount of consideration in support of the Alaska SAR Facility science team who will be working with the ERS-1, J-ERS-I and Radarsat satellite SAR data.

Table 5. Data Rates for Polar Ocean Ice in Post-Launch Phase

Data rates below 0.01 Gbits/yr are symbolized by a "—." All required data are Level 3.

Variable	Sensor	Number of 8-bit variables or channels	Temporal resolution (days)	Spatial resolution (km)	Volume (Gbits/yr)	Source(s)
Surface wind*	Scat-1,2	2	1	25	0.3	
Wind velocity*	LAWS	2	1	50	0.07	
Sea surface temperature	AMSR	1	7	30	0.01	
	MODIS	1	7	10	0.1	
Air temperature*	AMSU	1	2	30	0.05	NESDIS
	Drifting buoys	1	2	30	0.05	
Moisture*	AMSU	1	2	30	0.05	
Ice motion	SAR	2	3	10	0.6	
	Drifting buoys	2	2	100	—	
Ice type	AMSR	10	2	15	1.9	Japan
	SAR	4	3	5	1.2	
Chlorophyll	MODIS	1	7	10	0.1	
TOTAL					4.4	

* indicates a near-surface atmospheric value for establishing a
difference between surface and near surface.

Specifically, the geophysical parameters to be operationally measured at the ASF and plotted on a Polar Stereographic grid are:

- ice motion vectors on time scales of days to weeks; and

- fractional concentration over four ice types:
  (a) multi-year, (b) first year, (c) new ice, (d) open water.

The SAR data flow is shown in the functional block diagram (Figure 1) for the end-to-end processing chain. After ground reception and the standard Level 0 processing (to eliminate telemetry artifacts) in the raw SAR signal data, the data is staged on tape for Level 1 processing (image formation). Level 1 processing requires both pre-processing and post-processing steps. The pre-processing is a critical step since its accuracy determines the final quality of the image product (i.e. focus, signal-to-noise) and depends on the raw data quality and knowledge of the platform ephemeris. A poor result in estimating the processing parameters in this step may result in the need to repeat the Level 1 process. Acceptance of the Level 1 product is typically determined by visual inspection of the output

image. The main stage of the Level 1 processing is a two-dimensional matched filtering process using filter parameters determined during the pre-processing step. Level 1 processing is an extremely computationally intensive step typically requiring between 3-7 GFLOPs (depending on the radar imaging mode) for real time throughput. These high resolution (single-look) complex images are referred to as Level 1A data.

The post-processing step consists of two primary operations: radiometric correction and calibration and geometric correction and geo-coding. The radiometric correction process is to compensate for all system distortions and drifts about some baseline, such that all SAR image data numbers can be uniquely related to a radar back-scatter coefficient. Depending on the accuracy required, this may be a lengthy process requiring operator analysis over a several day period. Typically this analysis involves combining elements of pre-flight test data, on-orbit engineering telemetry and ground calibration site imagery to derive the processor correction factors. The geo-coding and terrain distortion corrections can be an automated procedure whose accuracy depends on knowledge of terrain elevation and

Table 6. Polar Ocean Data Products for EosDIS

Variable	Number of 8-bit Variables	Volume (Gbits/yr)
Fluxes—over ocean and sea ice		
Net short-wave radiation	1	0.05
Net long-wave radiation	1	0.05
Sensible heat	1	0.05
Latent heat	1	0.05
Salt flux to ocean	1	0.05
Moisture	1	0.05
Light to ice and ocean	1	0.05
Sea Ice		
Thickness distribution	1	0.05
Ice-type distribution	1	0.05
Atmosphere		
Surface wind field	2	0.1
Geostrophic wind field	2	0.1
Surface pressure field	1	0.05
Ocean		
Pigment concentration	1	0.05
Primary production	1	0.05
Total radiances	1	0.05
Diffuse attenuation coefficient	1	0.05
Temperature profiles	10	0.5
Salinity profiles	10	0.5
Profiles of horizontal velocity	20	1.0
TOTAL	60	2.9

AREA 30×10^6 km^2 of polar oceans
(Arctic 7×10^6 km^2; Antarctic 23×10^6 km^2)
Spatial resolution 25 km
Temporal resolution 3 days

22

JPL   FIGURE 1: FUNCTIONAL BLOCK DIAGRAM OF SAR LEVEL 0 TO 4 DATA FLOW FOR POLAR ICE STUDIES

JC-01-03-17
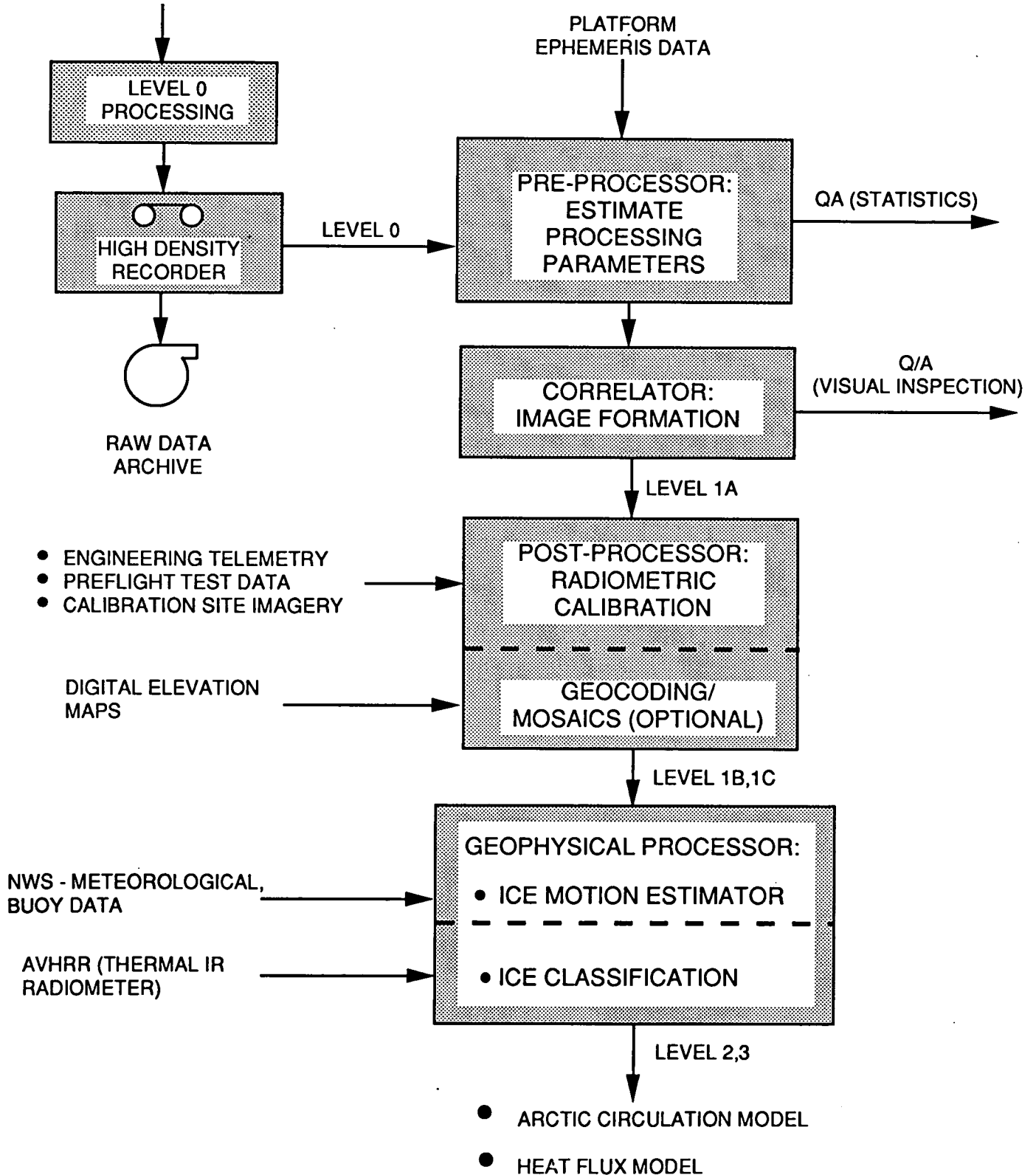
satellite position and velocity. Note that a digital elevation map is required to generate Level 1C products. Radiometrically detected images in ground plane format are termed Level 1B data products, while the geo-coded imagery (in a uniformly gridded, map-projected sample spacing) are Level 1C.

To date, there is no operational SAR processing system that routinely generates products beyond Level 1C and most stop at Level 1B. Therefore, discussion of an operational data flow beyond Level 1 to geophysical products (Levels 2 and 3) and large scale models (Level 4) is mostly conjecture and at best could be deemed experimental. However, a developmental effort is currently in process to routinely generate ice and ocean geophysical products for inclusion in an on-line database that can be browsed by the Alaska SAR Facility science team. The team will also have the capability to request specific processing from a remote site and evaluate new algorithms on this system as a local user.

These Level 2 and 3 processing algorithms, which are partially automated, require geo-coded low-resolution SAR images as input with correlative data from the SSM/I (microwave radiometer) and the National Weather Service for ice classification and determination of ice motion. An overview of the functional block diagram of the geophysical processor is provided in Figure 2. In the following section, a simplified scenario for the geophysical processor operation is described with a particular emphasis placed on the conditions under which the automated process will require manual intervention to ensure consistent product quality.

Consider first the application for tracking of sea ice motion in the Arctic Ocean. The goal is to build a database gridded by latitude/longitude or other suitable geographic reference system and time of motion vectors (magnitude, direction) for the Arctic region (i.e., 60°N to the pole). The primary physical forces for this motion are ocean currents and surface winds. Correlative data such as is available from drifting buoys and meteorological satellites are used to start the tracker by estimating the general ice motion, but this data is insufficient to produce a detailed motion map of the region. The estimated motion vectors from the model, using correlative data, is input to the database to determine the two most probable SAR image frames for matching. These images are then transferred from the on-line archive for processing. The matching process is complicated by the sea ice's rotation, deformation, and sometimes a change of state between multi-temporal passes. Therefore, the matching is based on defining a feature vector for a given piece of ice that is invariant under these change of conditions. In reality a feature vector cannot be constructed that is invariant under all possible conditions and that the success of the matching process reduces to a maximum likelihood problem. Thus, the probability of achieving many good matches depends on the quality of the input data used to establish the range of variation in image characteristics cased by the changing environmental conditions. This scenario forces the question of whether the operations can be fully automated. The answer appears to be that full automation can only be achieved at the cost of consistent product quality until a sufficiently large database can be built to account for nearly all ice and imaging conditions.

A more practical solution is to automate the systems to the extent required to do quality assurance checks that can detect (say to 95% probability) a bad match. These images are then set aside for manual (computer aided) matching. The good matches are then fed back into the model to improve the database and the performance of the ice motion predictor. This type of feedback system therefore, enhances the level of automation as the processing matures over the mission lifetime. A key element is that full automation is not desirable at the outset since it is at the cost of product reliability, which is much more damaging to most science studies then the delay associated with an operator assisted process.

The ice classification is a similar procedure except that it uses the microwave radiometer data (SSM/I) to measure brightness temperature. Initially the classification algorithm is being confined to derive only a four level classification map. Since the classification is based on the relative difference in backscatter coefficient between each class (as well as texture measures), radiometric calibration is critical to good performance. Also, since the existence of a water layer on top of the ice, which occurs during the melt season causes the backscatter to change dramatically, the SSM/I-derived surface temperature is also necessary for accurate classification.

In the processing scenario presented, the two SAR-derived geophysical products (motion and

JPL FIGURE 2: FUNCTIONAL BLOCK DIAGRAM OF LEVEL 2,3 DATA FLOW FOR ICE MOTION / CLASSIFICATION PRODUCTS

JC-01-03-18
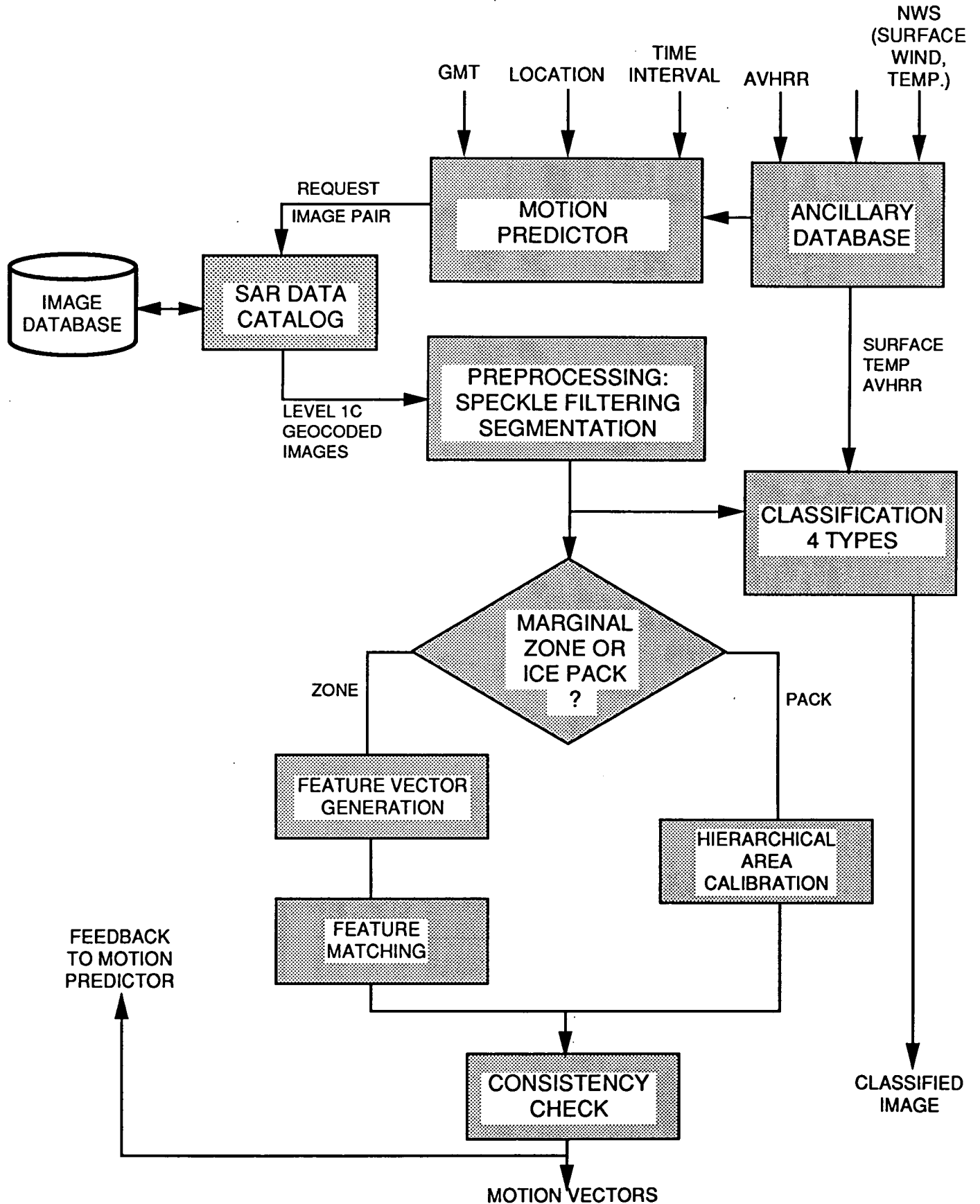
classification maps) could be generated simultaneously since they use the same image and the same resampled output grid. In addition, the result from the ice motion matching process can be used to improve the classification map. Consider for example that two images have been previously matched and the same ice floe is identified on both images. This flow should always be classified as the same ice type and therefore independent estimates from both images can be used to improve the classification.

In summary, the generation of geophysical products from the SAR image data is a multistage process that requires precise ephemeris, calibration analysis, target elevation data and selected correlative data from other remote sensors or ground measurements. Erroneous or inaccurate data from any of these sources will compromise the quality of the geophysical products. To ensure consistency in the data set, checks on quality are required throughout the processing with operator intervention when appropriate.

V. How to Evaluate "Success" of EosDIS

INTRODUCTION

In planning for the Eos Data and Information System, it is essential to examine how the system may affect researchers. We assert that the fundamental justification for EosDIS lies in improving the productivity of the researchers who use Eos data to investigate important questions in Earth System Science. EosDIS will be judged by the quality, compelling results, and creative ideas in Eos scientists' publications.

These attributes are difficult to quantitatively measure. Tenure and promotion committees in universities and laboratories have struggled for years to develop objective criteria, and it is their common experience that numbers of papers in refereed journals and numbers of citations are helpful, but not definitive. What is used instead are thoughtful reviews of scientists' work by their peers; we think this mechanism will serve equally well for Eos and EosDIS.

But it is still necessary to examine ways in which EosDIS will improve the efficiency and excellence of scientific research, and build such criteria into the design of EosDIS. We therefore suggest a model of how researchers spend their time interacting with data now, and we suggest some ways in which EosDIS can improve the efficiency of scientific research. We also suggest that a Level I Requirement for EosDIS is an assessment of the benefits that will accrue to its users.

We need to define who the users of EosDIS are. A reasonable definition is:

> "An EosDIS user is a scientist or group of collaborating scientists who will use data produced by the Earth Observing System to pursue a scientific investigation that results in at least one paper that describes the investigation, the Eos data used, and the results, and is published in the archival, peer-reviewed literature."

There will be other users of EosDIS. Scientists who build instruments and reduce their data will use EosDIS for data processing and archival. Weather forecasting agencies and mapping agencies may be users of EosDIS, as will land use planners and oil companies. However, the principal reason for Eos and EosDIS is to investigate Earth System Science, and success of the scientists is essential. The success of these other users is important but not essential.

There are two important correlaries:

1. EosDIS can be judged by the increased quality of refereed papers on the Earth that are widely cited. Numerical measures for quantifying this judgement will prove elusive. However, EosDIS should cause an increase in the number of papers in Earth System Science that are cited in appropriate references more than two years after they are initially published. Secondly, the number of investigators who publish research on the Earth and its environments should increase.

2. EosDIS can be judged by the improved efficiency with which scientists can interact with the data to reduce it to publishable results. Again, this efficiency is difficult to quantify, but it can be evaluated in terms of decreased time to obtain measures of the phenomena being studied.

CURRENT RESEARCH MODEL

One major challenge of the Earth Observing System is the collection and dissemination of data. So, our research model needs to describe the activities and times associated with various phases of conducting research and publishing the results for a paper involving data. While there are similar steps for theoretical investigations, we suggest that data-related investigations lie at the heart of what most users will do with EosDIS.

To us there seem to be eight important steps in producing a paper using data:

1. deciding on types of data to use, requiring browsing;

2. ordering and receiving data;

3. preliminary sorting through data;

4. preliminary relations with phenomena;

5. selecting and ordering additional or supplemental data;

6. re-selection and resorting of relevant data;

7. refinement of relations with phenomena; and

8. writing and publishing.

We have not included data collection or instrument-related work because we are seeking

a generic model involving users who may not be involved in actively collecting data. We want a research model that includes Interdisciplinary Investigators, Facility Team Members and Team Leaders, and Instrument Investigators, who will order their data from EosDIS. We expand on the steps below.

Deciding on Types of Data to Use

The first step in a data-related investigation is to find out what data are available for conducting the investigation. An investigation of oceanic heat transport might use *in situ* temperature and current measurements or radiation budget and atmospheric temperature and wind measurements. The investigator must first make up his mind what data are suitable for his work.

This first step is important. It is part of a researcher's professional training to shorten the time needed to locate particular kinds of data. The tools in this type of work include atlases, catalogues, review papers, and contacts with other colleagues, as well as browsing through Eos data.

Ordering and Receiving Data

Once he has identified relevant data types, a researcher must order and receive the relevant data. This part of the process includes preliminary identification of the time span and geographic location for relevant data. It also includes finding the ordering agency, the media on which the data are stored, and costing. Sometimes, data may be available in printed or electronic form. For larger data sets, the data are likely to be delivered in electronically readable form, such as optical disk.

Preliminary Sorting Through Data

Once the data are received, they must be assimilated. For most modern data investigations, this means that the data must be read into a computer. If the data are on magnetic or optical media, suitable formatting programs must be available. Often, this part of the investigation produces major headaches.

A second part of assimilating data is visualizing, editing, and selecting the portion of the original record that is appropriate for the investigation. Let us consider these steps in order. The first substep is to visualize the data that have been received. If we are dealing with images, we may want to see "what the data look like." In other cases, we may want to examine individual profiles of temperature or concentration, look at time series, or examine in other ways the contents of the received data. This visualization also includes establishing reliable estimates for expected maxima, minima, means, standard deviations, or other statistical properties of the data.

The next step is to edit the data for usefulness. By this we mean establishing criteria for what data will be selected. Often, this involves establishing criteria for noisiness, for geographic or temporal location, or thresholds. Only some of the data will be useful for the investigation. Then we select the useful data subset based on preliminary editing criteria. This step is important. It often determines the success or failure of the investigation.

The steps in this part of the investigation are almost certain to be repeated; they consume a substantial portion of the initial part of a data-related investigation.

Preliminary Relations With Phenomena

After the investigator has selected a preliminary portion of his or her data, the selected portion can be used to provide preliminary ideas about the phenomena the investigator wanted to examine. Occasionally, the data selection itself might provide the relationship. For example, simply producing the net radiation of the Earth over a year could produce the needed number. Usually however, additional work must be done.

We do not expect investigators to be done with this part of the investigation after the preliminary step. Often, they will have certain quantitative expectations that suggest the form of the relationship.

Reselection and Resorting of Relevant Data

We do not expect investigators to be blessed with a magic intuition that produces the appropriate selection of data the first time. Usually, outliers in the relationship inform the investigator of influences not expected at the beginning of the work, or they may reveal flaws in the investigator's model. Thus, these outliers must be examined in a more detailed way. Unusual data points will have to be selected and intensively analyzed.

This part of the investigation lies at the heart of working with data. It is a portion that cannot be shortened without astounding good luck. It is invariably iterative.

Refinement of Relations With Phenomena

After data have been more carefully selected, they can be used to refine the model parameters or to suggest a different model. Again, this work is iterative and cannot really be removed from a realistic model of how data-related research goes on.

Writing and Publishing

Writing and publishing are the final parts of an investigation. Good researchers will engage in the writing as they go along. We believe that this shortens the time devoted to writing at the end of the process. However, this practice does not lessen the burden of writing and having an investigation published.

Summary of Research Step Burdens

Table 7 summarizes the percent of time we expect researchers to spend in the research steps. We also include an estimate of the number of days that may be invested, derived by assuming that a university researcher who publishes two papers per year may spend 50% of his time in writing papers. This means that one paper requires 500 hours (about 60 days) of research work, broken down into the steps in Table 7. We recognize that many data-related investigations may take longer than the scenario.

INVESTMENTS IN CURRENT MODES

Any currently active researcher has adapted to his environment. Thus, an investigator is likely to have a significant investment in the current way of doing research. We believe these investments have four components: (1) knowledge base; (2) computer hardware; (3) computer software; and (4) procedures for investigating data. We examine these below.

Knowledge Base

Active researchers already know a good deal about what they need and want for conducting scientific research. Thus, they have knowledge about existing data sources and about the community's opinions regarding these sources. They will know how to order appropriate data. They will probably have a reasonable idea about

relative costs (both monetary and time) of various data sources.

Computer Hardware

Most investigators already have computer hardware that they use for data-related research. Such hardware will usually include tape readers, magnetic or optical disc readers, as well as display terminals with windows, display devices for color images, printers, and plotters. These may be combined as workstations or as interactive terminals connected to mini-computers or main-frames. Over the life of Eos, the current equipment will become obsolete, and active researchers will replace their hardware with more powerful equipment.

Computer Software

Investigators also have heavy investments in computer software, probably performing three functions: (1) data reading and interpreting; (2) data display; and (3) data-relationship generation. Data on fixed media, such as magnetic tape, must be read into the computer and placed into a more usable form. Although this is not in itself a difficult task, the variety of formats makes the job somewhat tedious, yet exacting. This part of the job is usually not regarded as "scientific," and is resented by most researchers.

Data display software includes simple statistical summary generators and complex three-dimensional plotting packages. This type of software is one of those areas where researchers have quasi-religious differences of opinion about which packages are best.

Software for generating data relationships includes everything from simple regression packages through radiative transfer codes and atmospheric circulation models, together with many intermediate tools.

Procedures For Investigating Data

The final area where researchers have a heavy investment in the current environment is in procedures for investigating data. By this we mean the order in which they apply their tools to data sets. A procedure is usually time consuming to set up. One author of this report recalls spending six months learning how to process ERBE flight calibration data to produce two numbers for each calibration. Once the procedures were set up, the numbers could be produced in a single, interactive three-hour session on the computer.

Table 7. Production of a Research Paper

step		fraction (percent)	time (days)
1.	Deciding on data to use	5	3
2.	Ordering and receiving data	10	6
3.	Preliminary sorting	15	9
4.	Preliminary relations	10	6
5.	Ordering supplemental data	5	3
6.	Reselection and resorting	25	15
7.	Refinement of relations	20	12
8.	Writing and publishing	10	6
	Total	100	60

A Note on the Influence of Current Investments on EosDIS

Because investigators have such a heavy investment in their current way of doing business, there is a significant cost to changing the way they work. If EosDIS plans to improve their efficiency, the reinvestment must be repaid. For EosDIS to be perceived as useful to the research community, it must prove to the researchers how they can become more efficient through this system's use.

IMPROVEMENTS THROUGH EOSDIS

EosDIS can intervene to help researchers in several ways:

- widen knowledge of data availability;

- shorten ordering and data receipt time;

- shorten data searching and selection time; and

- shorten time for developing preliminary relationships.

We have limited our suggestion for improvements through EosDIS for several reasons. First, we do not believe EosDIS will be useful in writing and publishing, except in providing bibliographic searches. Second, we believe that EosDIS probably cannot provide cost-effective help in the refinement of relationships. Investigators have such a diversity of opinions and selection procedures that EosDIS is unlikely to provide a completely satisfactory interface. We would not exclude the possibility that some EosDIS data examination tools could produce final results. However, we believe that the costs of providing such a diversity of tools and display functions would be beyond reasonable bounds. In other words, we believe that EosDIS is likely to be useful in the preliminary steps of an investigation. Let us consider the four modes of intervention.

Widen Knowledge of Data Availability

EosDIS can widen the spectrum of data that users might take advantage of. The basic tool for this purpose is likely to be interactive electronic access and display of data catalogs and metadata. This tool is similar in function to a library card (or computer) catalog.

Such a catalog search is mainly useful to the uninitiated. We expect that experienced investigators would use such catalog searches rarely, perhaps on a one-time basis, for checking on what kinds of data EosDIS holds. This is not the same as checking for the availability of a particular time and place, which we mention in the following two intervention modes. Instead, with catalog search, we expect the researcher will use the catalog a few times in a given investigation for generic data selection.

Shorten Ordering and Data Receipt Time

EosDIS will probably provide tools for ordering and shipping a particular data set. This tool might be used several times during an investigation. We would expect the amount of time a user would spend with this tool is small, perhaps as small as a few minutes every few months. The time saved and productivity enhancement with this tool depend on the amount of time the user's turnaround between ordering and receiving data were improved. We believe that the time saved is relatively small.

Shorten Data Searching and Selection Time

EosDIS might provide several aids in selecting data before they are ordered. These tools span a substantial range of requirements on EosDIS:

- browse of data availability;

- preselected (example) data browse;

- on-line data selection browse; and

- on-line product generation.

By browse of data availability we mean the ability to examine what data have been collected. Such browsing would include time and geographic location of data collected and processed. It does not include examining the data or reduced-resolution browse data.

By preselected or example data browse, we mean the ability to examine products prepared by data producers. In its simplest form, we might expect generic forms of data products. For example, a MODIS investigator might select one or two images in each spectral band from one or two days. The user could then examine these products to determine if they fit generic needs. More complex forms of this tool could include browse products generated from data processing. For example, each MODIS orbit could have a special reduced resolution browse product.

By on-line data selection browse we mean that the user can select how he or she wants to browse the data. As an example, users might specify the sampling resolution for display of images. They might be able to select the map projection or which spectral bands would be displayed. By on-line product generation we mean that the user can produce new types of data during a data examination session. This would require substantial non-production interactive power in a centralized facility, so we recommend instead that software be made available for investigators' computers. For example, a user might want to resample HIRIS data to a MODIS image resolution and produce a normalized difference image. Because this mode of intervention is likely to be expensive and open to a large variety of investigators' desires, we think the function is best transferred to their home facilities.

We believe that the possibility of electronic data examination and browse are high-leverage investments by the project for user satisfaction. Users are likely to use this type of facility frequently, and if they can do so easily their satisfaction with EosDIS will be enhanced.

Shorten Preliminary Relationship Development

This mode of intervention by EosDIS includes three more complex functions:

- data editing before ordering;

- cross-product visualization; and

- data transformation and correlation.

By data editing we mean the ability to order a product from EosDIS that includes only edited data, either by simple selection or user-generated editing rules. For example, we might be able to order only data with flags set to "GOOD" or only data whose radiances in a $1.6\mu m$ channel of MODIS were greater than a user-selected threshold.

By cross-product visualization we mean the ability of the EosDIS user interface to display relationships between various types of products. For example, EosDIS might display a MODIS image and a selected low-resolution HIRIS image side by side. In some investigations, it might help the user to create a new product that contained both images together.

By data transformation and correlation we mean that the user can create simple relationships between transformed data sets, including, for example, multiplying data by simple constants or producing regression coefficients. With one dimensional data sets, we might want to be able to correlate one variable against another (surface temperature against primary production for a particular geographic region). Again, this attribute places substantial non-production power in centralized facilities, and would best be served by delivering data and software to investigators' computers.

GAINING ACCEPTABILITY OF EOSDIS

In the previous pages, we have been trying to suggest how EosDIS might affect the working lives of its users. EosDIS users already have a significant investment in their current ways of doing business. Thus, to make these individuals aware of their options and to make EosDIS useful will require some thought on how to make the transition to the new system.

We believe that it is necessary to provide early indications of how EosDIS will affect the users. However, it is also true that users are busy and will tolerate few sessions with wasted effort. Thus, we need to develop a strategy that will make wide interaction with EosDIS as painless and as useful as possible. In other words, if we ask a research scientist to interact with a preliminary version of EosDIS, we need to have as high

31

a probability of producing a useful product during the interaction as possible. Thus the preliminary interactions should be with currently available data.

As a preliminary suggestion, we feel that it may be necessary to find some working researchers who can serve as volunteers for the community. There are difficulties with this approach. The volunteers may not represent the user community. The volunteers may not be able to devote enough time to the activity to provide a useful sounding board. The volunteers may not receive much professional recognition for their activities (particularly if EosDIS does not live up to its billing). We also recognize that users are a community. The reputation of EosDIS hangs on word-of-mouth communications.

Unfortunately, most scientists only want to discuss results, not how results are produced. Thus, papers that discuss EosDIS methods are not likely to provide the word-of-mouth reputation we want. This aspect of EosDIS intervention needs some care and systematic thought.

SUMMARY AND RECOMMENDATIONS

This section of the report provides a preliminary breakdown of the steps that go into a user's interaction with Eos data. It also provides an estimate of the time an Eos user will spend in working each step. Thus, we have provided a tool for estimating how EosDIS can improve a user's interaction with the Eos data. We have also provided a preliminary categorization of ways EosDIS might interact with the users.

Most of the benefits to users appear to come from data selection and transformation tools, which are best made available in investigators' home facilities. A stupendous catalog browse is likely to slightly affect the final users because they will use that kind of browse only sparingly. Just the ability to examine DATA before ordering is much more likely to increase user productivity.

It also appears that EosDIS is not likely to produce a 10× decrease in the time users must spend in producing papers. The reputation of Eos and EosDIS will instead depend on the quality of the research results.

The EosDIS designers should try to show how users will benefit from their interaction with the system. Thus, the EosDIS Advisory Panel recommends addition of the following Level I Requirement:

> "The EosDIS system design and planning must provide an assessment that demonstrates how use of the system will increase the quality of research by Eos scientists."

We hope that this preliminary model of EosDIS interactions with users will stimulate a more concrete and fruitful discussion with the system designers that will lead to an early chance to try out various features of EosDIS. The model we have suggested is tentative, and almost certainly needs tempering by a wider community of researchers. However, we believe that it carries the seeds of a more fruitful tool for the users of EosDIS.

Appendix
Science Advisory Panel for Eos Data and Information

MEMBERS

Name and Address	EosDIS Role	Telephone	Electronic Mail
Jeff Dozier University of California CSL/CRSEO Santa Barbara, CA 93106	Chair	805/961-2309	JDOZIER/NASA dozier@crseo.ucsb.edu jdozier@ucsbuxa.bitnet
John Barker NASA Goddard Space Flight Center Code 625 Greenbelt, MD 20771	MODIS	301/286-9498	JBARKER/GSFCMAIL
Bruce R. Barkstrom NASA Langley Research Center MS 420 Hampton, VA 23665-5225	INTD	804/864-5676	BRBARKSTROM/NASA
Roger G. Barry University of Colorado CIRES/NSIDC Boulder, CO 80309-0449	INTD	303/492-5488	NSIDC/OMNET rbarry%kryos@vaxf.colorado.edu
Francis P. Bretherton University of Wisconsin Space Science Center Madison, WI 53706	non-Eos	608/263-9824	F.BRETHERTON/OMNET FBretherton@vms.macc.wisc.edu
Josef Cihlar Canada Centre for Remote Sensing 1547 Merivale Road, 4th Floor Ottawa, Ontario Canada K1A 0Y7	INTD	613/952-2734	JCIHLAR/GSFCMAIL
John Curlander Jet Propulsion Laboratory MS 300-223 Pasadena, CA 91109	SAR	818/354-8262	JCURLANDER/J.P.L.
Lee Elson Jet Propulsion Laboratory MS 183-301 Pasadena, CA 91109	SWIRLS	818/354-4223	LELSON/J.P.L. elson@scn3.jpl.nasa.gov
G. David Emmitt Simpson Weather Associates, Inc. 809 East Jefferson Street Charlottesville, VA 22902	LAWS	804/979-3571	G.EMMITT/OMNET

Name and Address	EosDIS Role	Telephone	Electronic Mail
Robert H. Evans University of Miami, RSMAS/MPO 4600 Rickenbacker Causeway Miami, FL 33149-1098	MODIS	305/361-4018	R.EVANS/OMNET REVANS/GSFCMAIL bob@miami.miami.edu
David Halpern Jet Propulsion Laboratory MS 300-323 Pasadena, CA 91109	SCANSCAT	818/354-5327	D.HALPERN/OMNET
James F. Kibler NASA Langley Research Center MS 423 Hampton, VA 23665-5225	CERES	804/864-5386	JFKIBLER/NASA ASDVAX::JIM
David A. Landgrebe Purdue University School of Electrical Engineering West Lafayette, IN 47907	HIRIS	317/494-3486	DLANDGREBE/NASA landgreb@ee.ecn.purdue.edu
Mark R. Schoeberl NASA Goddard Space Flight Center Code 616 Greenbelt, MD 20771	INTD	301/286-5819	PACF::SCHOM
Bob Schutz University of Texas Center for Space Research, WRW 402 Austin, TX 78712-1085	GLRS	512/471-4267	R.SCHUTZ/OMNET BSCHUTZ/GSFCMAIL
Ronald Welch South Dakota School of Mines and Technology Institute of Atmospheric Sciences Rapid City, SD 47701-3995	HIRIS	605/394-2291	RWELCH/NASA
Al Zobrist The RAND Corp. P. O. Box 2138 Santa Monica, CA 90406-2138	non-Eos	213/393-0411	zobrist@rand.org

PROGRAM AND PROJECT REPRESENTATIVES

Name and Address	Eos	Telephone	Electronic Mail
Al Fleig NASA Goddard Space Flight Center Code 630.3 Greenbelt, MD 20771	EosDIS	301/286-7747	AFLEIG/GSFCMAIL
Strat Laios NASA Goddard Space Flight Center Code 501.0 Greenbelt, MD 20771	EosDIS	301/286-8611	SLAIOS/GSFCMAIL
Gary Metz USGS EROS Data Center Sioux Falls, SD 57198	EDC	605/594-6125	GMETZ/GSFCMAIL
David Nichols Jet Propulsion Laboratory MS 180-703 Pasadena, CA 91109	JPL	818/354-8912	DNICHOLS/J.P.L.
Eni Njoku NASA Headquarters Code EEC Washington, DC 20546	Program	202/453-1748	ENJOKU/NASA
Gerald Soffen NASA Goddard Space Flight Center Code 600 Greenbelt, MD 20711	Project	301/286-9690	GSOFFEN/GSFCMAIL
Darrel L. Williams NASA Goddard Space Flight Center Code 623 Greenbelt, MD 20711	Project	301/286-8860	DLWILLIAMS/GSFCMAIL
W. Stanley Wilson NASA Headquarters Code EEC Washington, DC 20546	Program	202/453-1725	S.WILSON.NASA/OMNET WSWILSON/NASA

Page intentionally left blank

Epilogue

After EosDIS Advisory Panel meetings it has become traditional for a few Panel members to go out to dinner to a Chinese restaurant. We leave the interpretation of the fortune cookies to our fellow Eos investigators, but this is what we've gotten so far, honest:

Your cheerful outlook is one of your assets.

Put the data you have uncovered to beneficial use.

Rely on your own good judgement to lead you to success.

The coming month shall bring winds of change in your life.

If you have confidence you will gain confidence in others.

Watch your relations with other people carefully; be reserved.

We cannot all do all things.

Your wishes will materialize with a little extra effort.

Failure teaches success.