MODIS DATA SYSTEM STUDY

TEAM PRESENTATION

January 13, 1989

AGENDA

1. Issues to be Addressed in the MIDACS Trade Studies

2. Outstanding Issues Progress Report

   --On-Board Processing
   --MODIS Data Coverage
   --Monitoring MODIS Data
   --Interface with the DHC
   --Non-MODIS Instrument Data Availability
   --DADS Issues
   --Implementation of Algorithms for Standard Product Processing
   --Joint Scheduling With Other Instruments
   --TMCF Specification

3. MODIS Instrument Operations Outline

4. Level-4 Processing Operations Concept Outline

5. Some Thoughts Regarding SPOT Data Proccessing

6. Science Requirements as Specified by Michael King

ISSUES TO BE ADDRESSED IN THE MIDACS TRADE STUDIES

1. Methods of Monitoring the MODIS Instument in the ICC
2. Standard Product Generation in the CDHF Versus the TMCF
3. Level of Analysis and Production Capabilities in the TMCF's
4. Electronic Versus Physical Distribution of MODIS Products
5. Choice of Distribution Media for MODIS Products
6. Complexity and Scope of Metadata
7. Preprocessing Versus On-Request Processing at the CDHF
8. Parallel Versus Serial CDHF Architecture
9. Similarity of CDHF and TMFC Architectures
10. On-Board Packetization by Channel
11. Advantages/Disadvantages of Different Types of Storage Media
12. Ability of DADS Processing to Flexibly Handle User Access

OUTSTANDING ISSUES PROGRESS REPORT

The outstanding MIDACS issues discussed briefly below share a common thread. The resolution of these issues and indeed, the specification of the performance requirements of the MIDACS, will be directly related to the state of our knowledge of the scientific requirements of the MIDACS. Currently, we have developed only three scientific data processing scenarios, and although these scenarios cover the three major geoscience areas and involve some of the major types of processing anticipated for MODIS data, many more scenarios are needed to accurately determine the scientific requirements of the system. Over 100 MODIS science data products are anticipated. Many of these products will have similar processing requirements, and in fact, we have attempted to estimate the overall MIDACS processing requirement by grouping these anticipated products according to their similarities with the three products for which we have data processing and data handling scenarios. As more scenarios are developed our grouping of data products will become more accurate, and therefore, our estimate of the system requirements will be improved. Eventually, there will be a detailed processing scenario for each anticipated MODIS data product, allowing a definitive resolution of these and other outstanding issues, and an accurate determination of the system performance specification.

Currently we are interacting with six scientists, mostly prospective MODIS science team members, to develop new scenarios for the following types of products: cloud climatology, Earth radiation budget, snow, ocean parameters, deforestation, land classification, and leaf water content. We anticipate the development of six to ten new processing scenarios by the time the science team is selected in early March, 1989. After interacting with the science team during and after the planned science team workshop in March, 1989, we will be able to refine the product list and begin the development of scenarios for each and every product anticipated by the team. These scenarios should be completed, in preliminary form, by May 1988 so that estimated requirements can be used in the non advocacy review process for Eos.

During the science team meeting in June 1989, these scenarios should be reviewed and revised allowing a more definitive set of requirements to be shared with the EosDIS developers. Also, the MODIS data system team will use these requirements to develop the specification of the MODIS specific components of the data system.

ON-BOARD PROCESSING

1.0 Introduction

One of the issues and uncertainties identified in the MODIS Data System Study relates to the nature and extent of processing support that will be available on-board the platform for MODIS use and the effect of such processing on the data system interfaces that the ground portion of the MODIS data system must accommodate. This document is a first attempt to identify design trade-offs that occur as required processing capabilities are moved between an on-board processor and appropriate sections of the MODIS ground data system.

Section 2 reviews general limitations that affect any satellite-borne processor and describes the general nature of improvements that could be provided by on-board processing, as seen from the perspective of the ground system designer. Section 3 addresses the minimum on-board processing capability required to provide basic MODIS service, and section 4 considers several potential expansions of on-board processing capability above minimum requirements and the design trade-offs associated with each. Since the experience and expertise of the data team relates to the design of the MODIS data system and not to the design of hardware components for on-board processing, this review is necessarily biased towards the needs of the data system. The material presented here should be considered along with complementary design and trade-off information developed by MODIS flight hardware engineers.

2.0 General Considerations

2.1 Negative Factors

Several unfavorable aspects of on-board processing from a satellite platform are immediately apparent. One obvious limitation is that physical access to the system is very limited after launch. The potential consequences of this obvious fact include the inability to provide timely repairs in the event of hardware failure and the inability to readily upgrade system software and hardware as new technology becomes available or as new processing needs become apparent. On-board processing uses the limited resources of an orbiting platform and takes place in a generally hostile environment that must be modified using special facilities such as temperature-controlled chambers, etc. Flight hardware must be specially qualified, and construction standards are rigorous and generally burdensome to meet.

2.2 Positive Factors

From the perspective of the MODIS data system designer, the potential advantages of on-board processing include improved data routing in the ground data system and potential improvements in instrument responsiveness. MODIS instrument data is routed to several points within the MIDACS, and a modest on-board processing capability would allow the creation of instrument data packets specifically tailored to the needs of the recipient system, i.e. containing only the specific data that is needed at the destination, in a format that allows easy identification and use of the data. If MODIS data packets are generated taking data in exact order as it is generated by instrument sensing elements, extensive on-the-ground data sorting will be required to create special-purpose blocks of data required at a number of locations within the MIDACS.

A substantial on-board processing capability could potentially support adaptive sensing and improved instrument reliability. Adaptive sensing would alter the instrument sensing routine or detector gains according to the target being observed, i.e. sensing routine or detector gains could be altered for land or ocean observations, cloud cover, etc. Also, extensive on-board processing could provide MODIS Built-in-Test capability, direct on-board monitoring of instrument status, automatically declare alarm status, and automatically initiate instrument safe-store sequences.

3.0 Minimum Required On-Board Processing Capabilities

Certain minimum on-board processing capabilities are required to obtain a functioning MODIS instrument. The essentials include on-board storage of command sequences, time-based retrieval of commands as they are needed for execution, the generation of MODIS data packets, and the buffering of observation data as it flows between the detector and the on-board LAN.

3.1 Command Storage and Retrieval

Instrument commands must be stored on the orbiting platform to provide instrument operation during the time periods when direct TDRSS contact is not available. Advance storage of commands also provides increased security in the event of equipment failure or other contingency affecting the command uplink. Time tags associated with each command specify the time when the stored command should be executed. Since failure in the command link could profoundly affect instrument operation, special precautions to assure data integrity may be needed in this portion of the data system.

3.2 On-Board Data Packet Generation

MODIS data packets must be fitted with descriptive header and footer information that describes the contents of the packet, provides error control (optional), and specifies the relative location of the packet in the sequence of all data packets returned by the instrument. While this procedure does not demand extensive hardware or software capability, the formatting of MODIS data packets is a MODIS responsibility and the MODIS instruments must certainly be equipped with at least enough processing capability to generate the required packets. Since individual items of observation data are perhaps not as critical to overall mission success as items of command data, this processing may perhaps be implemented in a separate system that operates independently of the command processing system and does not have the same provisions to insure data system integrity that are required in the command system.

3.3 Data Buffering

If the MODIS instruments use a rotating mirror scanning procedure, data will appear in bursts at the output of the detectors as the scan sweeps across the Earth observation area. Only a few, or perhaps no observations will be made in the "dead" region when the scan is outside the area of observational interest. [Certain calibration measurements may be made while the mirror is in a non-observing position]. Given the "burstiness" of the observation data, the processor must either accommodate the peak rate of the data burst or the data must be stored in a buffer that smoothes out data peaks and valleys and allows the processor to proceed at the average rate of data flow through the instrument. Efficient use of resources usually dictates the latter procedure. It will be assumed in this analysis that data buffering will be provided at the interface between the detectors and the MODIS on-board processor.

Similar considerations dictate that buffering must be provided between the output of the MODIS processors and the on-board LAN. The peak overall data storage requirement for the instrument data system occurs at the end of an observational sweep. During the sweep, data has accumulated at a much faster rate than it is absorbed by the LAN. Buffering allows the LAN to receive MODIS data packets at the average data throughput rate rather than the peak rate. Buffering at this interface is also required to accommodate those time periods when the LAN is not available for MODIS use because it is accommodating other Eos or platform instruments.

4.0 Candidate On-Board MODIS Processing Capabilities and Discussion of Trade-Offs

4.1 Construction of Data Packets to Exact MODIS Requirements

In this section we will discuss the likely MODIS data packet formats if the formats are chosen to minimize the on-board processing requirement. We will then consider the optimal data packet format for MIDACS on-the-ground use, and we will conclude with a comparison of the hardware resources that are required to implement each of these alternatives.

It would appear that the packet format that would make the least demands on the on-board processor would insert the data into packets directly as the data is generated at the detectors, in the same order as the data becomes available. This would allow the system to work with the construction of a single packet of observation data at a given time; the single packet would contain observation data from many spectral channels. Since data packets must be buffered at the interface with the LAN, the processor memory would contain many finished data packets awaiting receipt by the LAN and a single incomplete data packet that is receiving the latest data.

Because of the need to construct sample images to monitor instrument function at the ICC, the optimal data packets for MIDACS use would contain only data for a single spectral channel. Sample images at the ICC are constructed using data from only a few spectral channels; up to four spectral channels may be selected for each MODIS instrument (maximum of eight). If MODIS data packets contain only data for a specific spectral channel, only those packets containing required data need to be routed to the ICC. Otherwise, required data may be distributed throughout the instrument data stream, and the entire stream of data returned by the instrument may need to be routed to the ICC.

The construction of data packets containing only data for a given spectral channel requires the on-board data system to simultaneously construct many data packets - one for each spectral channel being observed. Data for each spectral channel appears at a different detector, so that data for a given detector can be routed directly to the appropriate packet. It appears that the additional processing capability required to route the data directly to the appropriate packet rather than to a single data packet that receives all spectral channels would be insignificant. Memory requirements to simultaneously generate many data packets may be somewhat greater than the corresponding requirements if only a single packet is constructed at a given time. However, because of the need to buffer data packets at the interface with the LAN, overall memory requirements are roughly comparable. If it is assumed that the maximum packet buffering requirement occurs at the end of an observation scan, and if it is also assumed that data packet generation is synchronized with the scan period and that all data packets are filled at the end of the scan, then the memory requirements of the two alternatives are exactly equal.

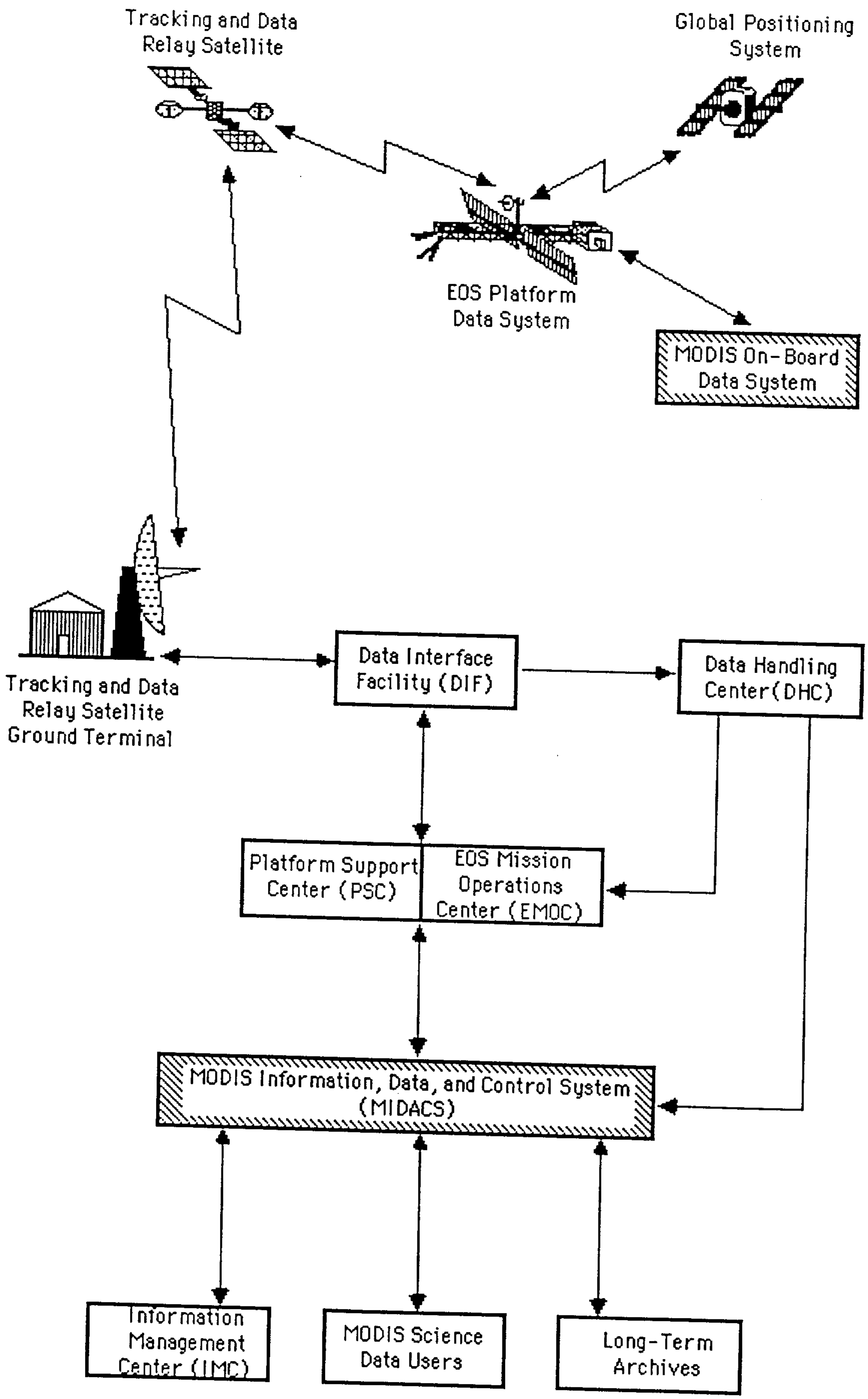
4.2 Real-Time Instrument Control

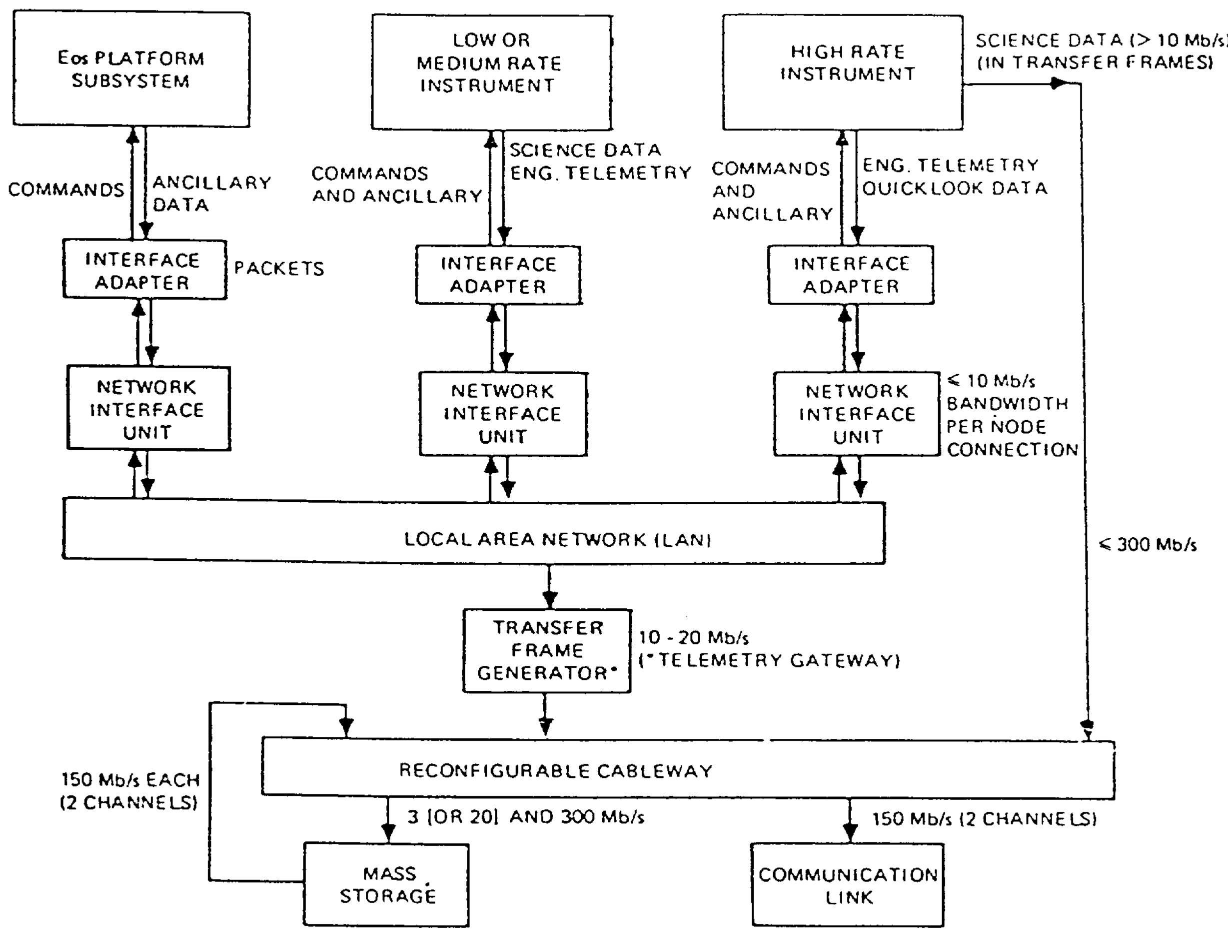
4.2.1 Adaptive Sensing

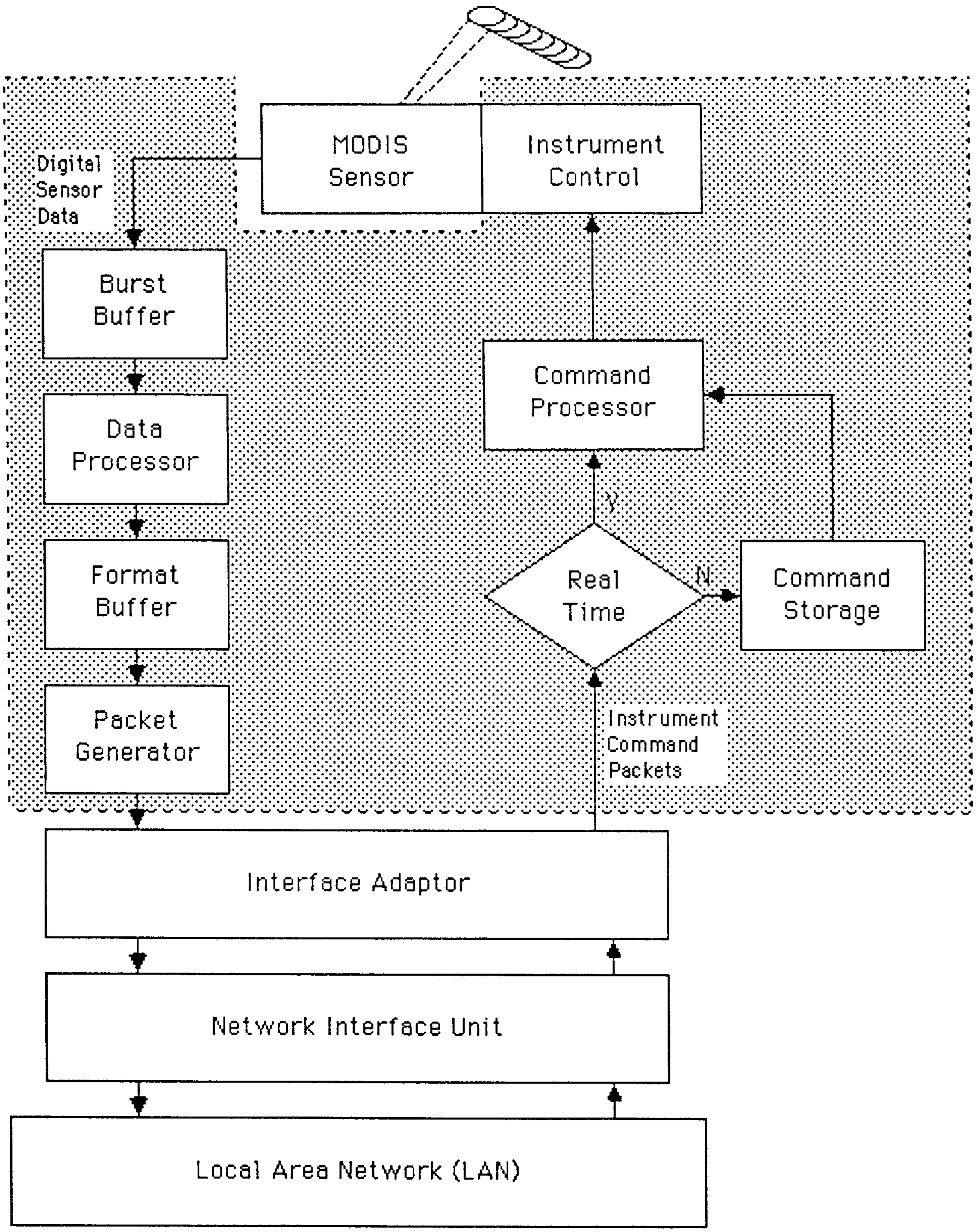
[This section will be expanded to include a discussion of possible adaptive sensing routines and the general resources that might be required to implement the alternatives.]

4.3 Lossless and Lossy Data Compression

[It is expected that all compression of basic science data will be lossless. This section will consider the possible application of lossy data compression to monitoring data and the system trade-offs that would result.]


Tracking and Data
Relay Satellite
Global Positioning
System
EOS Platform
Data System
MODIS On-Board
Data System
Tracking and Data
Relay Satellite
Ground Terminal
Data Interface
Facility (DIF)
Data Handling
Center(DHC)
Platform Support
Center (PSC)
EOS Mission
Operations
Center (EMOC)
MODIS Information, Data, and Control System
(MIDACS)
Information
Management
Center (IMC)
MODIS Science
Data Users
Long-Term
Archives


Eos PLATFORM SUBSYSTEM
LOW OR MEDIUM RATE INSTRUMENT
HIGH RATE INSTRUMENT
SCIENCE DATA (> 10 Mb/s) (IN TRANSFER FRAMES)
COMMANDS
ANCILLARY DATA
COMMANDS AND ANCILLARY
SCIENCE DATA ENG. TELEMETRY
COMMANDS AND ANCILLARY
ENG. TELEMETRY QUICKLOOK DATA
INTERFACE ADAPTER
PACKETS
INTERFACE ADAPTER
INTERFACE ADAPTER
NETWORK INTERFACE UNIT
NETWORK INTERFACE UNIT
NETWORK INTERFACE UNIT
≤ 10 Mb/s BANDWIDTH PER NODE CONNECTION
LOCAL AREA NETWORK (LAN)
≤ 300 Mb/s
TRANSFER FRAME GENERATOR*
10 - 20 Mb/s (*TELEMETRY GATEWAY)
150 Mb/s EACH (2 CHANNELS)
RECONFIGURABLE CABLEWAY
3 (OR 20) AND 300 Mb/s
150 Mb/s (2 CHANNELS)
MASS STORAGE
COMMUNICATION LINK



Possible Block Structure of On-Board MODIS Data System

MODIS DATA COVERAGE

The scientific requirements for MODIS data coverage are not yet formally specified. As such, this area remains an outstanding issue for the MODIS data system. The issue arises both as a requirement on the MODIS data system, and also on other aspects of the downlink data path, commencing with the MODIS instrument operations and ending with delivery to the long-term archives.

Domains of Consideration

Other than considering the data path as a whole, five individual components may be considered for specifying data coverage and data loss. These segments are : (1) the MODIS operations; (2) the platform LAN and tape recorder; (3) the TDRSS link; (4) CDOS; and (5) the MODIS ground data system.

1. Specific MODIS operations may result in data loss. This may be true for internal calibrations exercises, for specific tilt requests for the MODIS-T, for platform servicing, and for data rate conflicts with other instruments on the platform.

2. Data may be lost within the LAN, and may be recorded but never recovered from the platform tape recorders.

3. The TDRSS link may cause data loss through either the lack of access to any TDRS due to other missions (e.g., manned) competing for the same resources, due to errors in transmission, or due to TDRSS down-time.

4. Raw data may be lost or damaged between the DIF at White Sands and the DHC at GSFC. In addition, certain raw data may never be converted to Level-0 due to timeliness considerations or uncorrectable errors.

5. The MODIS ground data system may not process certain data for reasons not yet specified.

Allowable Levels of Data Loss

The MODIS instrument is capable of operating simultaneously in two modes. These have been termed the "survey instrument" mode and the "observatory instrument" mode. The survey instrument takes continuous observations and regularly observes the entire Earth. The observatory instrument acquires data only in response to a user's data acquisition request (DAR). Each of these modes will have a different level of allowable data loss.

1. Operating in the survey mode, the MODIS instrument will continuously produce geophysical parameters with global coverage. The entire Earth will be viewed every 1-2 days, depending on the latitude. For many users, long-term time series of MODIS-derived parameters will be required, with the parameters themselves averaged into daily, monthly, and annual

estimates. The accuracies of these averages will depend of the averaging interval, the spatial resolution of the grid, the quality of the measurements and algorithms, and the sampling characteristics of the MODIS instrument. The standard error of a time average will decrease by the square root of the number of independent samples. Assuming that every day is independent, equatorial monthly averages of measurements will be four times more accurate than the daily estimates. If one day of data is lost, the accuracy of the estimate will be degraded by several percent.

2. Operating in an observatory mode, the MODIS instrument and data system will be responsive to specific DARs, including those supporting near-real-time field experiments. If the requested data is either not taken or lost, then the DAR will not be satisfied. This will result in a 100% loss of accuracy, unlike the 3% impact of the previous example.

General Requirement for Data Coverage

When data collection is not dictated by a response to a DAR, the extent of lost MODIS data will be driven by the science requirements on the accuracy of the geophysical parameters (including the radiances). These requirements have not yet been formally stated, and will no doubt vary from parameter to parameter. It may be necessary to conduct system simulation studies to assess the impact of data gaps on the product accuracies. However, it is clear that no spatially systematic gaps in data coverage will be tolerable. Only non-systematic (random) data losses will be allowable. The exact specification on data coverage is uncertain. The specification may be 100%, 99.9%, 99%, or even 95%. It is worthwhile considering each of these in turn.

1. At a 100% coverage requirement, no MODIS data may be lost. The MODIS instrument must operate continuously and all data must be ultimately received by the ground data system and utilized in generating the archive data products. This requirement will drive the operation of the platform tape recorders, the TDRSS communications links, and even the level of service for error correction.

2. At a 99.9% coverage requirement, on the order of 1.5 minutes of MODIS data will be lost daily, corresponding to about 10^9 bits, 10^8 observations, and an equivalent swath along the track (if the error were systematic) of about 5° (600 km).

3. At a 99% coverage requirement, on the order of 15 minutes of MODIS data will be lost daily (1/7 orbit), corresponding to about 10^{10} bits, 10^9 observations, and an equivalent swath along the track (if the data loss occurred in one segment) of about 50° (6,000 km).

4. At a 95% coverage requirement, on the order of 70 minutes of MODIS data will be lost daily (3/4 orbit). Many regions of

the Earth will experience data loss more than once in a month. Furthermore, global daily analyses will include large (or many small) data gaps, which may require substantial interpolation.

As a check on the reasonableness of these limits, consider that on many instruments (e.g., ERBE) internal calibrations cause the loss of perhaps two orbits of terrestrial data per month. This is equivalent to 0.5%, yielding data coverage at the 99.5% level. From the point of view of the user, it makes little difference whether the accuracy of a parameter is degraded due to data loss for any reason. It may be reasonable to require that the general MODIS data coverage requirement for the end-to-end data path be 99.5% of all possible observations.

Requirement for Data Coverage in Response to a DAR

When data collection is dictated by a response to a DAR, there is a specific requirement for the data. The need for the data may be critical (e.g., supporting and directing aircraft flights), or alternate data may be acceptable (e.g., from two days later). It must be assumed that the 100% coverage requirement applies to the DAR, and that none of the requested MODIS data may be lost. Should a conflict arise that will result in the loss of the data covered by the DAR, then the MODIS science team leader must be involved in the resolution of the conflict.

MONITORING MODIS DATA

Introduction

Engineering and science data taken by the MODIS instruments, as well as selected platform ancillary data, must be monitored in the MIDACS' ICC. The primary downlink will be through the TDRSS. Science and engineering data will be stored onboard the platform for playback and downlink at the scheduled TDRSS contact.

Since it will not be possible to monitor the instrument in real-time with a 100% duty cycle, priority playback processing at the DHC will provide selected portions of recorded science data for monitoring. Engineering and ancillary data will be packetized separately from the science data and will automatically be routed to the ICC. Real-time processing now refers only to the processing of science data as it becomes available from the DHC.

Priority Playback Data

The procedures and operations for selecting and transmitting priority playback data are unclear at this time. A meeting should be held with personnel designing and studying the DHC requirements.

Processing of Data

Discussed below are several ways which data can be processed for monitoring of instrument performance.

1. Onboard processing of data for the formation of packets is still unclear. This will be discussed as a separate issue.

2. The DHC will perform level 0 processing of priority playback data to an agreed upon level. This processing may take the full advantage of level 0 processing capabilities of the DHC or may only include error checking, time ordering, and buffering. Some buffering is expected at the DHC for the time ordering of the packets unless the onboard processor accomplishes this task.

3. The ICC will request and process only a selected portion of the priority playback data for monitoring purposes. This may cover only approximately 5 % of all the playback data. The ICC will select a total of 8 channels out of this data and process it to level 1B or level 2.

Real-Time verses Priority Playback data - TBD

INTERFACE WITH THE DHC

The following notes are paraphrased from a telephone conversation with Gene Smith who is directing the Ancillary Data Support Service Study Team. The notes may help describe the interface between the MIDACS and the DHC.

Monitoring of science data for evaluation of instrument performance will require the receipt of priority playback data from the DHC. The following summarizes the status of uncertainties for the receipt of this data. The term, priority playback data is a DHC term and denotes a quick turnaround in the receipt and distribution of recorded data.

Priority Playback Data

The original purpose of the priority playback was to provide a small percentage of science data to the users. It was never anticipated that all the priority playback data would be passed to the users, MIDACS in this case. A small percentage of the data, approximately 5 %, would be distributed. Identification of the priority playback data or the subset of this data is currently unclear. The criteria for this identification is something that should be defined between the CDOS and unique EosDIS systems such as MIDACS.

Processing/Buffering of Priority Playback Data

The amount and level of processing of priority playback data by the DHC is unclear at this time. If the playback data is needed in a "real-time" sense, buffering may be held to a minimum, although the format of the playback data structure may make the buffering of playback data a necessity. That is, the playback data format is unknown at this time. Virtual channels may pass packets containing many observations from different instruments or may consist of packets having only one instrument associated with it. In any case, it is unlikely that the playback data will be sent to the MIDACS at the input data rate due to the need for buffering from packet selection, error checking, time ordering, etc. The communication lines and interface for the transmission of this data to MIDACS is dependent upon the NASCOM agreements.

Another impact is the design of the on-board processor that formats the packets and/or transfer frames and how data is placed on and retrieved from the tape recorder.

NON-MODIS INSTRUMENT DATA AVAILABILITY

Level 2 geophysical parameters can be classified as those requiring only MODIS data and those requiring a combination of MODIS and other satellite or non-satellite data. The performance requirements and other aspects of processing may differ considerably for the two types of data products. Some issues are:

Data Required

As of this time we do not know what data will be required for MODIS data processing. Probable data that will be required include AIRS/AMSU Level 2 data. What other instrument data will be required?

Performance Considerations

Level 2 data products are supposed to be generated within 72 hours of data acquisition. Some Level 2 data products may require data from other instruments and this data may not be available for 72 hours. Will the 72 hour performance requirement be relaxed for these data products and other interdisciplinary data products?

Non-Eos Sources

Currently several archive centers exist for remotely sensed data, and archives at both USGS and NOAA are being considered as archive sites for Eos data. The need for information from these data sets will be determined by team members on a product by product basis. Because of the timeliness and volume of MODIS data, it is not anticipated that Non-Eos data sources will rarely be used in the generation of standard data products. However, some special data products and some Level-4 products may require Non-Eos data sets.

Eos Sources

Because of the commonality of all Eos instrument data systems imposed by EosDIS, the merging of data sets from different Eos instruments as well as from different Eos platforms should be manageable task. It is anticipated the synergistic nature of many of the Eos instruments will require occasional or perhaps frequent merge of data sets for the production of standard products. It is through the careful interviewing of prospective team members along with the completion of in-depth data product scenarios that the amount and type of Non-MODIS (Eos and Non-Eos) data set use and availability will be determined.

DADS ISSUES

The following seven issues are DADS issues to the extent that their solutions will be implemented in the DADS. Many of these issues will be addressed in trade studies. Their eventual resolution will reflect these studies' findings, the Science community's priorities, and EosDIS policies and procedures.

Possible Resource Limitations on Processing Individual Queries

- User Account Balance
- Estimated Level of Remaining Resources

The DADS is the MIDACS element that processes the queries the users generate in the IMC. The various TMCFs are able to directly access the DADS for query processing and other Science support activities. The DADS is in effect reactive to requests originating in other MIDACS facilities. The MODIS functional requirements contain some query response times, and some projections have been made for the processor power necessary to support the expected 100 simultaneous users. Circumstances may arise that limit the DADS ability to respond to a given user's query or data request.

Before a query is executed, the DADS determines the charges to be made to the user's account, and verifies there is a sufficient balance. Should the balance be insufficient, the user is notified. The user can modify the query or take other steps to have the account balance modified either in terms of amount or as a factor in processing a given query. In this example the DADS resources necessary to process the query are available, but the user's ability to command them is limited.

In another possible scenario a user with a sufficient account balance may initiate a data request that would require an "excessive" amount of DADS resources. For example, a user may request every MODIS image of sediment build-up in the Mississippi delta for a five year period. This involves mounting, scanning, reading, and dismounting thousands of tapes. As a consequence one or more DADS tape drives could be indefinitely unavailable to other users, with the expected reduction in overall throughput. In this example the DADS resources necessary to process the query are available, the user's account is sufficient to cover the charges, but the effects on other users may have to be considered in honoring the query.

Possible solution areas include more closely specifying the role of user account balances; predicating the resources committable to a query as a function of user classes such as TMCFs, non-NASA investigators, and the general public; and having users specify (very) low priorities for queries that use excessive resources.

Advantages/Disadvantages of Different Types of DADS Storage Media

- Access Time Versus Capacity
- Expandability

The datasets stored in the DADS are accessed as a function of user query or request. As there probably is no sequence to the science community's daily requests, a random access device capable of storing all MODIS datasets may appear to be the solution. The DADS is expected to receive approximately 1.8 terabytes per day of MODIS products, requiring two optical tapes. An optical random access device such as a disk has approximately two orders of magnitude less capacity than an optical tape. Each day's MODIS data would require approximately 180 optical disks. The DADS will store two years of MODIS data, requiring 1,460 optical tapes or 131,400 optical disks.

The expected access time for information stored on optical tape is 23 seconds, plus however many minutes are necessary for physically retrieving and mounting the tape. Optical disks may have faster access times, but would require similar physical retrieval and mounting times. Even if disks were combined 20 per pack, nine packs per day (6,570 packs per year) would be required with several having to be retrieved and mounted for accessing parts of a day's data.

Expandability refers to the relative ease and/or difficulty experienced in increasing the DADS' storage capacity. Both tape and disk require storage space for the media. Additional drives will be necessary as the daily workload increases as a function of more data being stored, and possibly as a function of increased use of older data. Footprints and power requirements differ for tape and disk drives.

The issues to be resolved include determining dataset usage patterns and growth rates, the ability to meet any specified turnaround times, incremental relationships that may exist between storage media and drives, and any DADS physical constraints.

Electronic Versus Physical Distribution of MODIS Products

- Dataset Quantity
- Turnaround Time

The two methods of receiving MODIS datasets are electronically or on off-line media. Electronic receiving results in comparatively fast transmission of images. For example, an image defined as having 1,024 X 1,024 pixels with 12 bits/pixel contains 12,582,912 bits of information. A T1 line (1.544 Mbits/second) can transmit this image in approximately eight seconds. A slower line such as 9600 bps would require approximately 1,311 seconds or 21.8 minutes. Upon completion of transmission the MODIS datasets could be immediately available for processing. Off-line media would involve at least one day's transportation. The user would need to have the media mounted on a drive for the processor being used.

The issues to be resolved are determining the trade-offs between possible extensive use of electronic transmission and in effect batching requested datasets for transporting on media. Using electronic transmission requires the IMC/DADS to provide a quantity of various modems and/or other connection devices, while shipping off-line media requires stocking the media and staffing a mailroom.

Choice of Off-line Distribution Media for MODIS Products

- Computer-Readable
- Hardcopy

Non-image MODIS data is expected to be shipped to users on off-line media. Computer-readable media such as magnetic tape, optical tape, optical disk, and CD-ROM are possible choices or alternatives. Each would require the DADS to provide media and I/O devices. Hardcopy would require a printer.

The issues to be resolved include determining the quantities of non-image data users can be expected to request, as well as the anticipated uses of this data.

Additional Degrees of DADS Processing Support

- Minimal--Use Another Processor
- Support as Authorized
- Open Ended--Support as Needed

The science needs of single discipline investigators will be largely satisfied with the standard MODIS/HIRIS Levels 1-4 products. User requests for special processing or support of proposed new products will require additional processing of MODIS/HIRIS data. Multidisciplinary investigators are expected to generate requests for either non-standard products or for special processing of the data underlying existing products.

The facility where this additional processing will take place has to be identified. The TMCFs will be expected to support algorithm testing and validation, and other processing a Science team member may require. As the DADS processors provide the necessary datasets, performing any additional processing within the DADS may be a simple solution. Other alternative processing facilities are whatever the requester has available, one or more of the TMCFs, or the CDHF (as it has the most powerful EosDIS computing facility), will be a candidate for satisfying additional processing requirements.

The DADS processor suite has yet to be specified, meaning its raw power and excess capacity are unknowns. The DADS' primary function is the locating and retrieving of selected datasets. Depending on workload the DADS may not be able to provide meaningful levels of additional processing. As a function of processor availability, exceptions could be made on a case-by-case basis. Depending on their size one or more TMCFs may be able to support this processing, particularly if a Science team member makes the request. The individual requestor may or may not have access to a suitable facility. If processing requests become repetitive and involved, a new data product may become recognized with processing performed in the CDHF. Another alternative is for the DADS to honor all requests, assign them a low priority, and execute them in the background.

The issues to resolve involve determining how MODIS/HIRIS or perhaps EosDIS will respond to special processing requests, how the need for a new or modified data product is recognized and satisfied, and what if any boundaries are to be set for the individual MODIS/HIRIS processing centers.

MODIS Data Interchange

- Standing Orders
- Intelligent System

On recurring bases and for specific areas of interest, individual MODIS dataset users may require 100 percent of another EosDIS instruments' data, for example, HIRIS. Standing orders whose processing package is capable of recognizing the availability of the necessary data, will be one processing alternative. An intelligent or knowledge-based system that would recognize the existence of required data, will be another alternative. The present archive (NSSDC) environment can support the first alternative in the form of batch-processed standing orders capable of determining the presence of the desired datasets.

When initiated and subsequently executed by the system, standing orders would be capable of retrieving existing MODIS/HIRIS data. As Level 2 and above HIRIS products are generated only when requested, these standing orders would, in effect, initiate the production of the required HIRIS datasets that have yet to be produced. Depending on the number of HIRIS datasets produced in this fashion, HIRIS processing requirements could increase.

The issues to be resolved involve determining the more effective alternative for generating the paired datasets, and the possible impacts on HIRIS processing facilities.

Complexity and Scope of Metadata

- Standardized by Product Type
- Predicated on Descriptive or Derived Values
- Open Ended--Elements/Values Provided/Modified as Needed

Each MODIS/HIRIS image will be described in terms of specific attributes. The ranges of values for these attributes will be the basis for the queries submitted by users. For MODIS these attributes will include MODIS N/T sensor ID, product sequence number, version number, processing date, calibration algorithm ID, product start/stop time, orbit number(s), geographical boundaries, data quality, calibration quality, land/ocean, degree of cloud cover, instrument tilt (MODIS-T), scan numbers, solar and satellite zenith angles, platform ephemeris, time of observation, calibration coefficients, Level 2 product descriptors and quality flags, and Level 3 products with resolutions and domains. With other metadata elements and values to be defined later, there is the possibility of an image's descriptive data requiring more storage than the image.

When the metadata elements and values are initially assigned, they will reflect the initial Science types and their expected uses. As EosDIS matures new products may be added, with new metadata elements being described. Changes in MODIS/HIRIS dataset usage may result in changes to existing metadata element values/ranges, affecting not only future data products but possibly previous ones. Examples of the latter are an identifier for all products reflecting the immediate after-effects of a major earthquake, and an identifier for products along a given path for a specified time period.

For new products the metadata element values will be produced by the CDHF and stored in the DADS, with a copy available in the IMC for query processing. These would result exercising the EosDIS new product mechanism. The CDHF could also produce new metadata values for reprocessed datasets. Changes affecting current metadata element values, while possibly implemented on the DADS processor, will be the result of the then existing metadata policy. This policy can range from relatively flexible, meaning element values can be readily modified, to relatively inflexible, meaning the element values cannot be readily modified.

The issues to be addressed are ones of EosDIS policy and processing load. Modifying several years' datasets' metadata could prove to be a non-trivial ADP task, with impacts on the user community's ability to access the affected datasets. A flexible policy could result in metadata that reflects the needs of a relatively few but very active users. An inflexible policy could result in metadata reflecting past user needs with decreasing support for present and future needs.

IMPLEMENTATION OF ALGORITHMS FOR STANDARD PRODUCT PROCESSING

There are many unresolved issues at this point which can be characterized as standards issues, development issues, implementation issues, and data production issues.

Software and operating systems standards issues

1. Who decides what the EosDIS software standards are? Must these standards be rigidly adhered to or will the Eos Project Office allow exceptions if they are justified?

2. Each TMCF may have a different operating system. How will compatibility be maintained?

Algorithm development issues

1. Will science team members be allocated time on the CDHF computers for algorithm development? Will they have disk space at the CDHF?

2. The CDHF and TMCF computer architectures may differ meaning that code developed by CDHF will need to be modified to work on the CDHF computers or be modified to be more efficient. Portable code may work on many machines but be inefficient on some machines because of their differing architectures. Which is more important: portable code or efficient code?

3. Several science team members may want to derive the same geophysical parameter. Do we devote computational effort to deriving the geophysical parameter several times or do we select one derivation and have all science team members use it? If a single geophysical parameter is calculated several ways and different values result, what will the science team do? Will they rank them according to perceived accuracy? Will they stop processing until the differences are resolved? If, on the other hand, one algorithm is chosen to calculate a specific parameter which is used by several scientists, will each of these scientists modify their code or will a group of software developers mesh the separate codes together for increased efficiency?

4. At some point all Level 2 and 3 data products will need to be re-imaged on to a standard grid. Who develops this software? Who decides what the standard grid will be? Will the chosen grid be compatible with other measurements so that comparisons can be made easily?

Algorithm implementation issues

1. What criteria will be used to certify that an algorithm is working properly and producing useful data?

Data production issues

1. Are all Level 2 geophysical parameters equal in importance? Will different priorities be assigned to their generation? Can some Level 2 parameters be derived long after the observations are made?

2. Will all geophysical parameters be carried through to Level 3 or only selected high demand parameters? What spatial resolution will be used for Level 3 data products?

JOINT SCHEDULING WITH OTHER INSTRUMENTS

Joint scheduling of satellite observations raises issues associated with how often they are made, who is responsible for the coordination, how the ICC's interact, and how users obtain the data. Some unresolved questions are:

1. When will coordinated observations be made?:

   a. Whenever there are observations of the same Earth target?

   b. For calibration?

   c. For near real time experiments with in-situ measurements?

   d. For near real time experiments without in-situ measurements?

2. Will the science team leaders do the coordination?

3. How will the two ICC's interact?

4. How will the users obtain the data? Always through the IMC from their respective DADS? Through some special arrangement?

5. What data media will be used? What media are supported by MODIS? By HIRIS? By other instruments?

TMCF SPECIFICATION

The organization of the TMCF is still evolving. At present the following architecture for the TMCF is proposed:

1. Team Leader Computing Facility (TLCF)
    a. Calibration Support Team (CST)
    b. Science Support Team (SST)
2. TMCF for Ocean Sciences (TMCF-OS)
3. TMCF for Land Sciences (TMCF-LS)
4. TMCF for Atmospheric Sciences (TMCF-AS)
5. Domestic TMCF's
6. Foreign TMCF's

The special TMCF's (2-4) and SST are recent additions to the TMCF and have not yet been fully treated in the preliminary MODIS data system documents.

Several questions are raised by this organizational structure:

1. Is this organizational structure consistent with the Science Team Member proposals?
2. What data products are generated by CDHF and the various TMCF's?
3. Will a TMCF-OS which is located remote from the CDHF require its own duplicate data storage? How will data be transferred back and forth? Can the timing performance requirements still be met? TMCF-LS and TMCF-AS raise identical questions.
4. The special TMCF's seem to do far more than algorithm development and validation of data since they are envisioned as major data product producers. Is this allowed under EosDIS functional requirements? Under Level I requirements?
5. Will the special TMCF's acquire all their computing resources from the MODIS project?

MODIS INSTRUMENT OPERATIONS OUTLINE

I. INTRODUCTION

A. Purpose: Basis for understanding instrument operations

B. Scope: Covers instrument use and capabilities and the ground operations that support it

C. Assumptions: Instrument specifications per references and Barker and Salomonson, Science monitoring using playback data, Engineering and ancillary data using separate downlink, MODIS and HIRIS on same platform

D. References and Applicable documents

II. INSTRUMENT OPERATIONS OVERVIEW

A. Instrument Design Overview: Overview of instrument design (T/N) and use, current understanding of instrument operations to collect data

B. MODIS Flight Environment: Platform and LAN overview

C. MODIS Ground Environment: MIDACS Overview, introduction of ICC, IST, TMCF and the operation and science teams

D. Instrument Operations: What/How will be controlled by the operation team at the ICC

1. Modes Of Operations; Modes of data collection, operation envelopes, duty cycles, calibration ...

2. Operational constraints: Platform constraints of envelopes, power,...

3. Synergism: With HIRIS and others, Commonality

III. INSTRUMENT OPERATIONS MANAGEMENT: How team operates MODIS, controllable items...

A. Planning and Scheduling: DARs, Simulations, scheduling of data collection....

B. Command and Control: Routine commanding and verification of instrument operations

C. Safing and Emergencies: Non-routine operations, autonomous control, real-time commanding

IV. INSTRUMENT PERFORMANCE EVALUATION: Methods of analyzing correct operation of instrument, Displays

A. State-of-Health Monitoring: Engineering and ancillary data analysis

B. Data Collection Monitoring: Science data analysis using playback science data

C. Performance Analysis: Long-term analysis, reports, plans

V. INSTRUMENT CALIBRATION: Calibration techniques and responses in regard to instrument operations only

A. Scheduling: Unique and routine calibration

B. Commanding: Unique and routine

VI. INTERFACES

A. Traffic Analysis: Amount of expected data

B. Science Data Rates and Volumes: Expected data rates and volumes

VII. ISSUES

A. Monitoring of Science Data: Real-time, Playback, channel selection, timeliness

B. DHC involvement: Data transmission, Level-0 processing, data request, timeliness

LEVEL-4 PROCESSING OPERATIONS CONCEPT OUTLINE

I. PRODUCT DEFINITION

   A. STANDARD LEVEL-4 PRODUCTS

   B. SPECIAL LEVEL-4 PRODUCTS

II. INPUT DATA PRODUCT REQUIREMENTS

   A. LOWER LEVEL MODIS DATA

   B. OTHER EOS INSTRUMENTS

   C. NON-EOS INSTRUMENTS

   D. NON-EOS DATA ARCHIVES

   E. OTHER CORRELATIVE DATA

III. PROCESSING REQUIREMENTS

   A. TIMELINESS OF PRODUCT GENERATION

   B. DATA HANDLING AND STORAGE REQUIREMENTS

IV. LEVEL-4 PRODUCT EXAMPLES

   A. GLOBAL WEATHER FORECAST MODEL (J. SUSSKIND)

   B. GLOBAL VEGETATION INDEX PHENOLOGY (C. JUSTICE)

   C. GLOBAL OCEAN CARBON FLUX (W. ESAIAS)

   D. CLIMATE FORECASTS AND MODELING

   E. LAND COVER CLASSIFICATION (J. TUCKER)

   F. RADIATIVE TRANSFER MODEL COMPARISONS

Some Thoughts Regarding SPOT Data Processing

At present, processing of image data taken by the Système Probatoire d'Observation de la Terre (SPOT) is performed on three Digital Equipment Corporation computers (a VAX 8530, 11/750, and 11/785). It is worth reviewing the SPOT processing system for comparison to the MODIS data system performance requirements. The SPOT imaging system operates simultaneously in two modes (monochromatic and panchromatic). The monochromatic mode has a nadir-viewing ground resolution of 10 meters (0.51 to 0.73 microns). The panchromatic mode views three spectral regions, green, red, and near-IR (0.50 to 0.59, 0.61 to 0.68, and 0.79 to 0.89 microns) at 20 meter resolution.

The ground swath width of SPOT for either mode is 60 km, requiring 6,000 detectors for the panchromatic band, and 3,000 detectors each for each of the three 20 meter spectral bands. At a ground velocity of about 6.5 km/sec, the two sets of detectors will be sampled 650 and 325 times/sec, respectively. The resultant data rate is about (6,000 x 650 + 3 x 3,000 x 325) = seven million observations/sec. At eight-bit dynamic resolution, this sampling rate is recorded at 50 megabits/sec.

A SPOT scene is about 60 km square, or 6,000 lines x 6,000 pixels for the 10 m channel. Each scene requires 27 to 50 megabytes of storage on computer tape depending on the number of channels ordered (panchromatic versus all). About ten minutes of processing is required to generate a single SPOT scene, which consists of calibrated radiance data (equivalent to Level-1) acquired in eight seconds. The processing system is a factor of 75 slower than the data acquisition rate.

By contrast, the orbitally averaged MODIS data rate will be about one million observations/sec and ten megabits/sec. All MODIS data will be processed through Level-1. Because reprocessing may be done simultaneously (at twice the data rate), and due to other considerations, the data system must be sized for 100% utilization at an effective MODIS data rate six times that expected. This is on the order of the SPOT data rate, and suggests that certain aspects of the MODIS data system must be sized 75 times larger than the SPOT data system.

Cloud Optical Thickness, Thermodynamic Phase, and Effective Particle Radius of Cloud Particles (M. King, GSFC)

a) Input Data Volume: Channel 14′ (0.754 micron) of MODIS-N will be utilized to derive the optical thickness of clouds by comparison of the reflection function with asymptotic expressions for the reflection function of optically thick layers. Channels 21 (1.640 micron) and 23 (2.130 micron) of MODIS-N along with the derived optical thickness will be used to determine the thermodynamic phase of the clouds as well as the effective particle radius of the cloud droplets (or ice particles). An input volume of three MODIS-N channels will be used with an input data volume of

1.2 x 0.4 x 0.5 x 0.74 Gpix/day x 3 channels x 10 bits/chan-pix = 6 Gb/day.

where the products are derived only for the daytime portion of each orbit (40 percent) and for cloud-filled pixels (50 percent). The volume increase factor of 1.2 arises from Level-0 (10%) and Level 1B (10%) processing.

b) Processing Requirement: Currently, cloud optical thickness effective cloud particle radius, and thermodynamic phase of clouds are being derived from aircraft and Landsat TM data. The computer used is an IBM 3081, which requires a CPU time of 20 ms/pixel to determine these parameters. Thus the processing time for one day of data at 1 km resolution would be

0.4 x 0.5 x 20 ms/pix x 0.74 Gpix/day = 3 Ms or 35 days.

Assuming an effective rating of 5 MFLOPS for the IBM 3081, the computing capacity to process a day of data in a day would be

35 x 5 MFLOPS = 175 MFLOPS

c) Output Data volume: The total amount of data generated in the production of these two products will be

2 parameters/pix x 0.4 x 0.5 x 0.74 Gpix/day x 2B/parameter =0.6 GB/dAY

Aerosol Characteristics (Michael King, GSFC)

Input Data Volume: MODIS-N Channels 5-13 (0.435 - 0.865 microns) and the polarization channels 17'-19' at 0.7 microns will be used to derive aerosol optical thickness, size distribution, index of refraction, and single- scattering albedo for cloud-free pixels over the ocean. The input volume of these channels is

1.2 x 0.4 x 0.5 x 0.7 x 0.74 Gpix/day x 12 channels
x 10 bits/pix-channel = 15 Gb/day

where the products are determined for cloud-free pixels (50%) over the ocean (70%) and only for the daytime portion of each orbit (40%). The factor of 1.2 accounts for an increase in volume from Level-0 (10%) and Level-1B (10%) processing.

b) Processing Requirement: Current processing of aircraft or Landsat TM data requires a CPU time of 60 ms/pixel on an IBM 3081 to derive the 4 parameters. Processing one day of MODIS data at 1 km resolution would take

0.4 x 0.5 x 0.7 x 60 ms/pix x 0.74 Gpix/day
= 6.3 Ms = 72 days.

If a 5 MFLOP rating is assumed for the IBM 3081, the computing capacity to process a day of data in a day would be

72 x 5 MFLOPS = 360 MFLOPS

c) Output Data Volume: The volume of the 4 output data products will be

4 parameters/pix x 0.4 x 0.5 x 0.7 x 0.74 Gpix/day
x 2B/parameter = 0.9 GB/day

PROSPECTIVE SCENARIOS IN DEVELOPMENT

DR. ROBERT EVANS- U. MIAMI- CHLOROPHYLL

DR. JIM TUCKER - GSFC- GLOBAL LAND COVER CLASSIFICATION

DR. ROBERT SCHOWENGERDT U.ARIZONA- AGRICULTURE PRODUCTS

DR. STEVE RUNNING- U. MONTANA- FOREST PRODUCTS

DR. RICHARD WARING- OREGON STATE U.-SEASONAL PHOTOSYNTHESIS

DR. ALEX GOETZ- U. COLORADO-REMOTE SENSING MINERAL EXPLORATION

DR. DAVID LANGREBE- PURDUE- GEOGRAPHIC INFORMATION SYSTEMS

DR. SIEGFRIED GERSTL- LOS ALAMOS-

DR. JEFF DOZIER- U.C.S.B.- SNOW

DR. ROBERT GURNEY- GSFC

DR. HERBERT JACOBOWITZ- NOAA- ERB

DR. VERN VANDERBUILT-NASA AMES- ATMOSPHERIC CORRECTIONS