

# The EOSDIS Data Server

Jeanne Behnke

Earth Science Data Information Systems/NASA GSFC

Code 423, Greenbelt MD 20771

+1-301-614-5326

[jeanne.behnke@gsfc.nasa.gov](mailto:jeanne.behnke@gsfc.nasa.gov)

Presented at the THIC Meeting at the Naval Surface Warfare  
Center Carderock

Bethesda MD

October 3, 2000

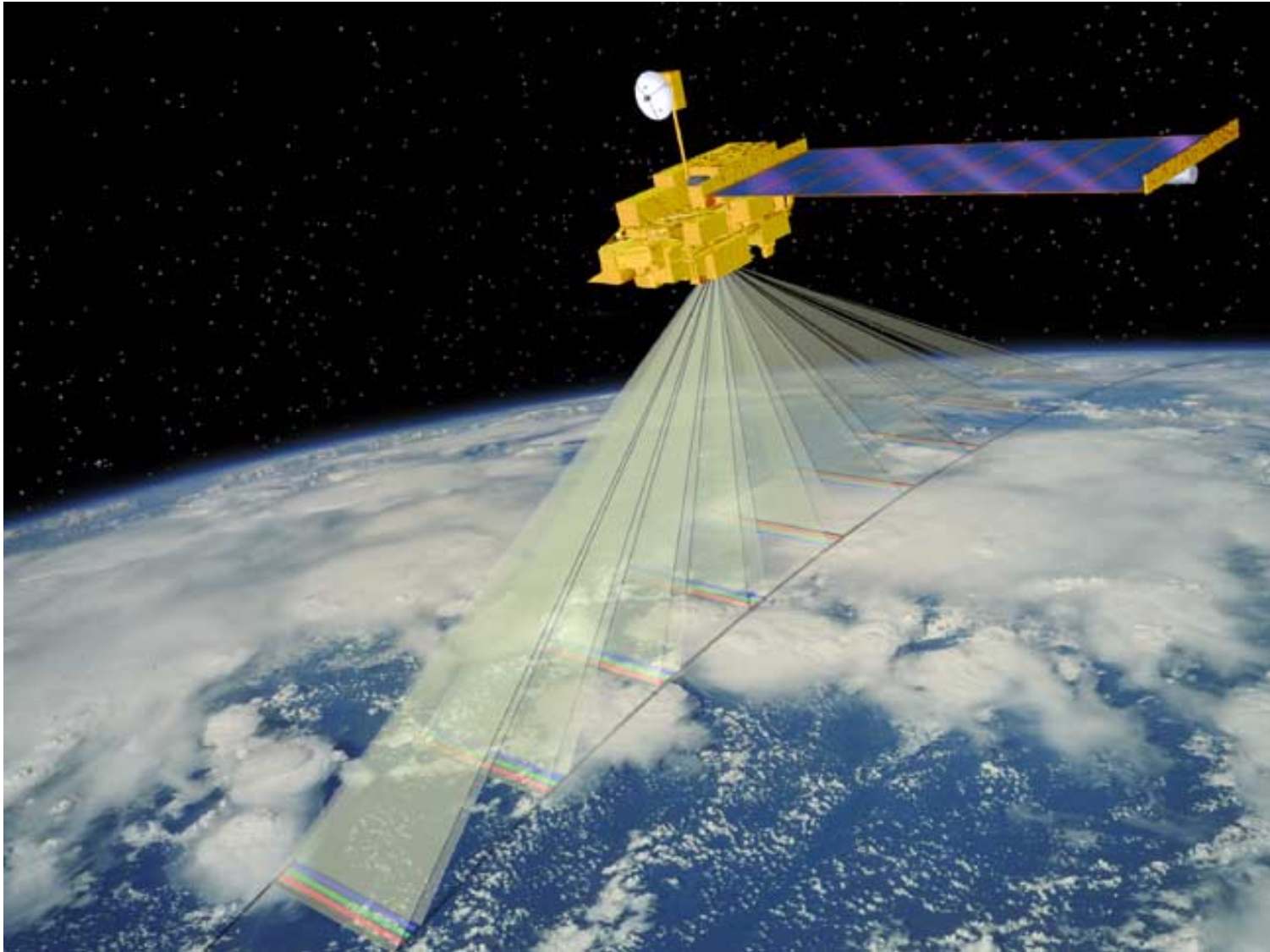


## A Big Mission!

- EOS = Earth Observing System
  - Centerpiece of NASA's Earth Science Enterprise
  - Collect Earth remote sensing data for a 15 year global change research program
- EOSDIS Data Information System
  - Software architecture is designed to receive, process, archive and distribute several terabytes of science data on a daily basis
  - User community consists of several thousands of science and non-science users
  - 7 major facilities across the US
    - Distributed Active Archive Centers (DAACs)

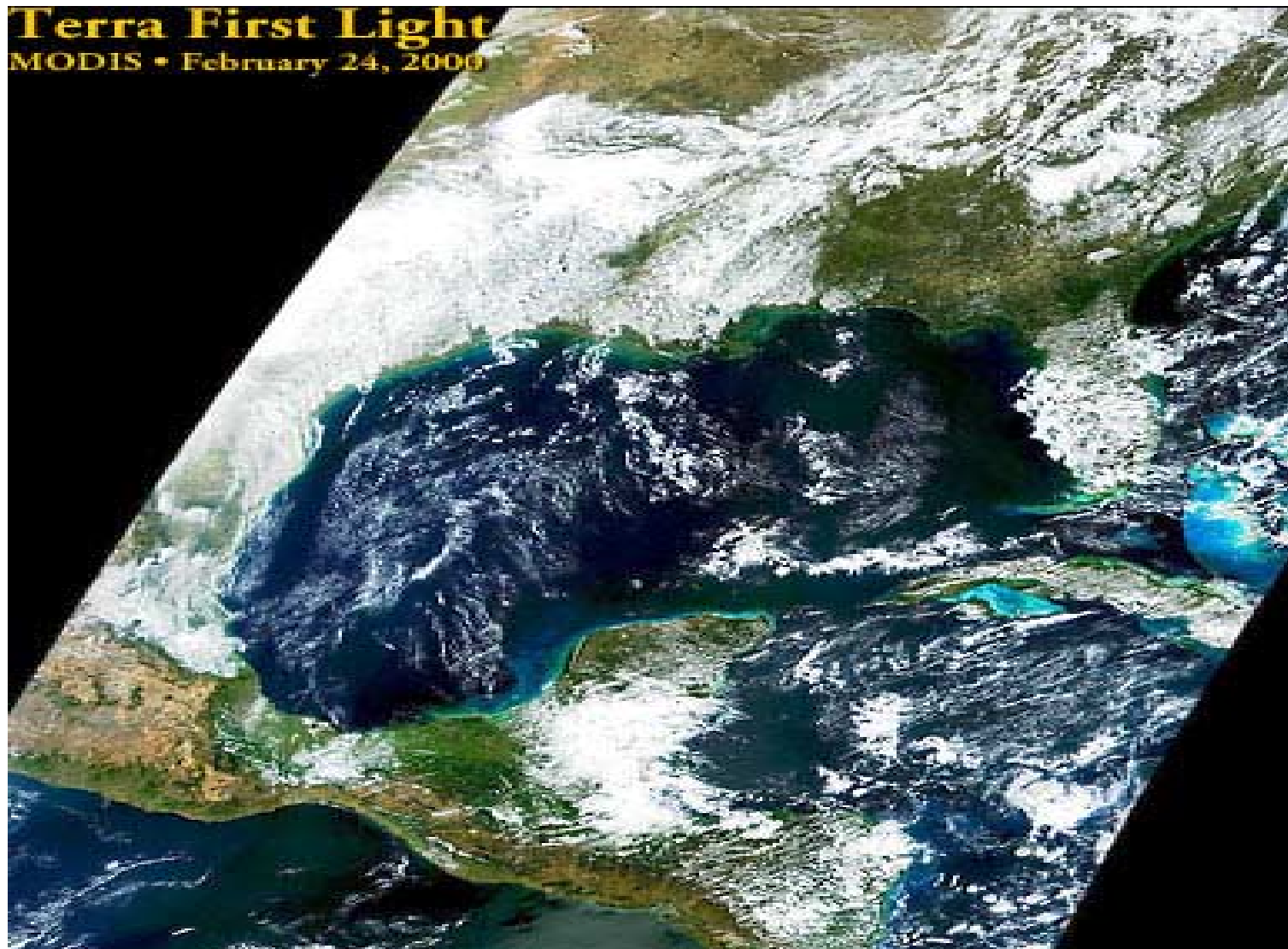


## EOS Terra in orbit



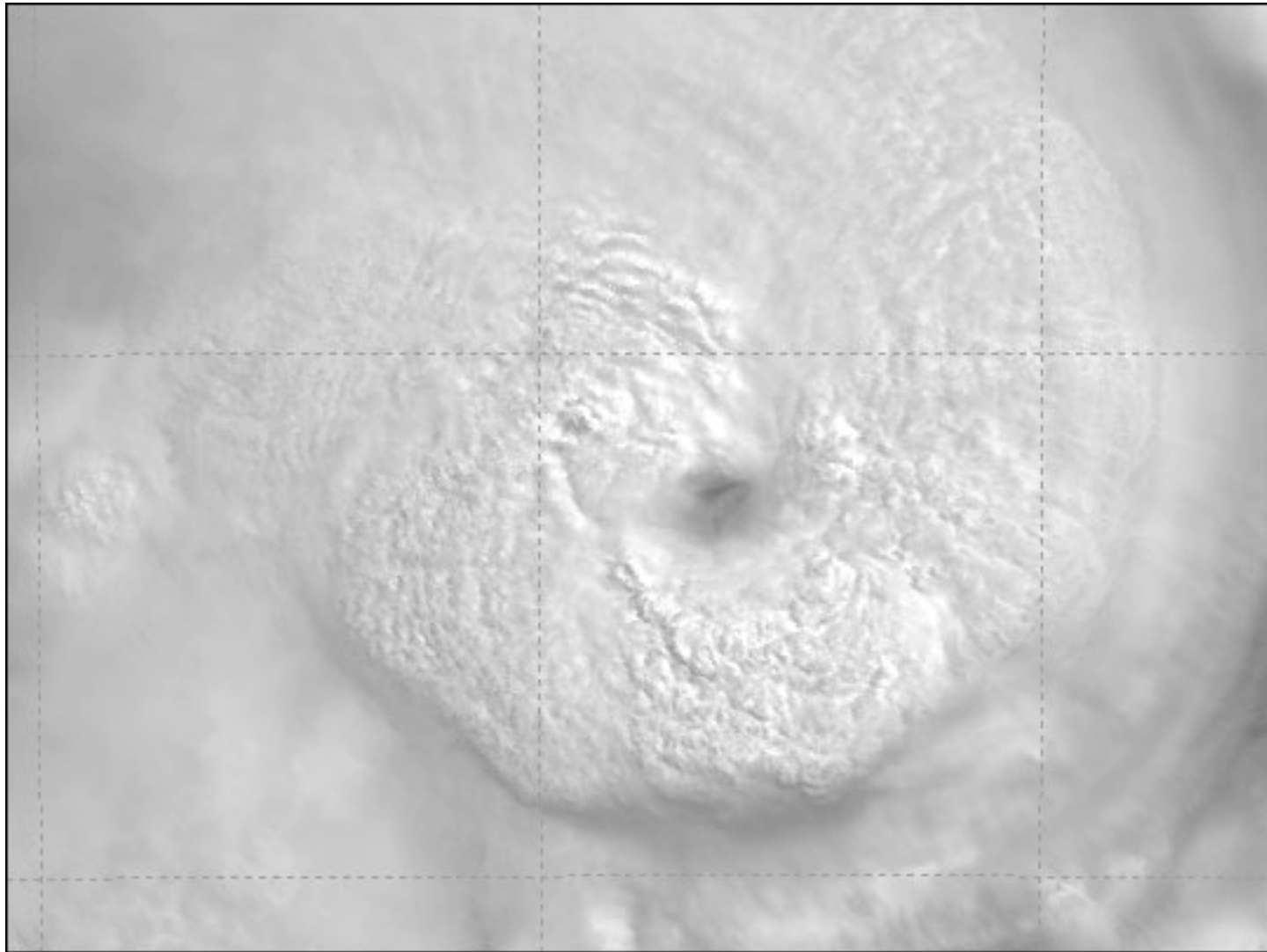


# Terra Image





# TERRA Image - MODIS



**Cyclone Hudah in Indian Ocean taken March 29, 2000**



# TERRA Image - ASTER



Reno, Nevada taken April 18, 2000

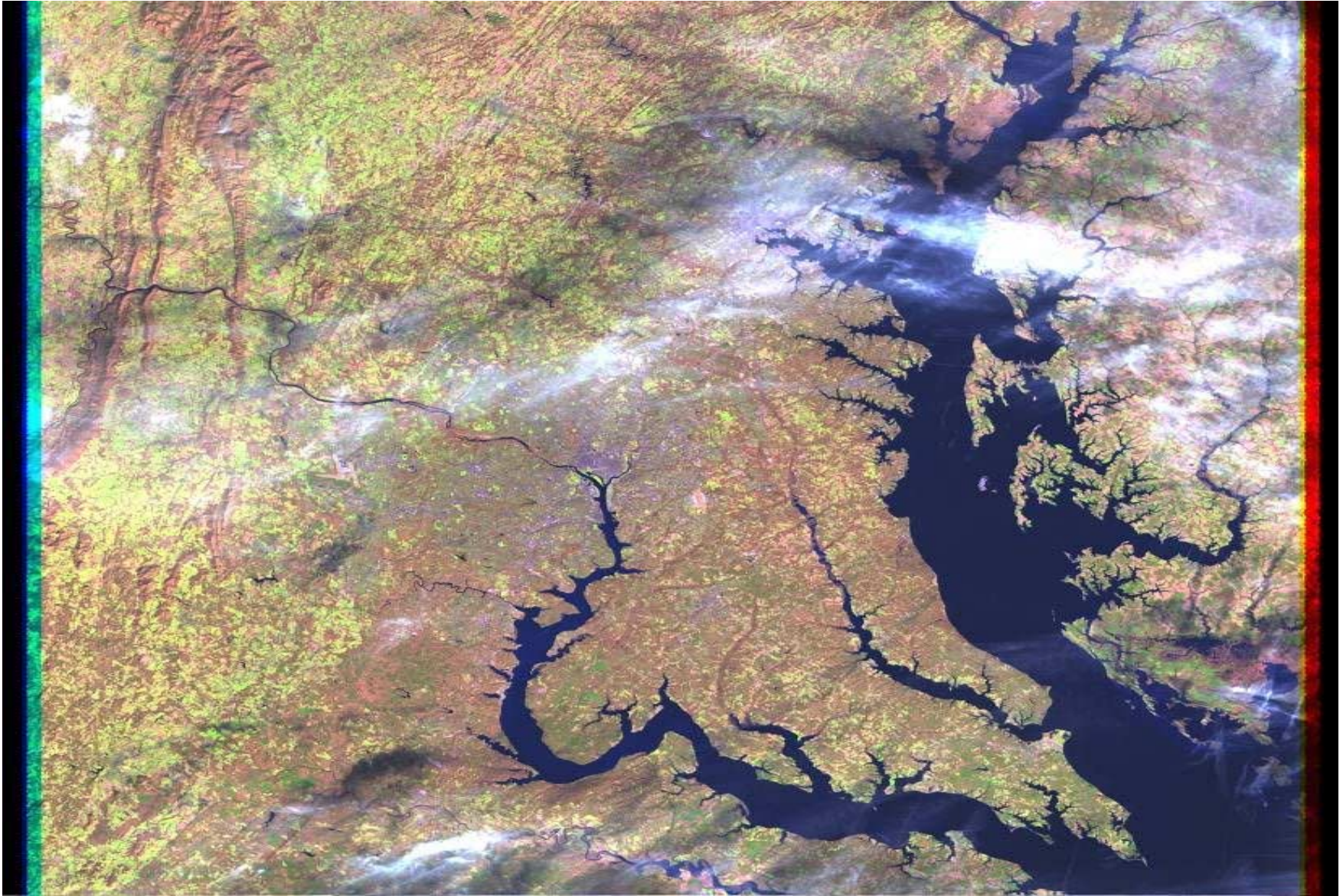


# TERRA Image - MISR





# Landsat Browse Image







## EOSDIS Concept

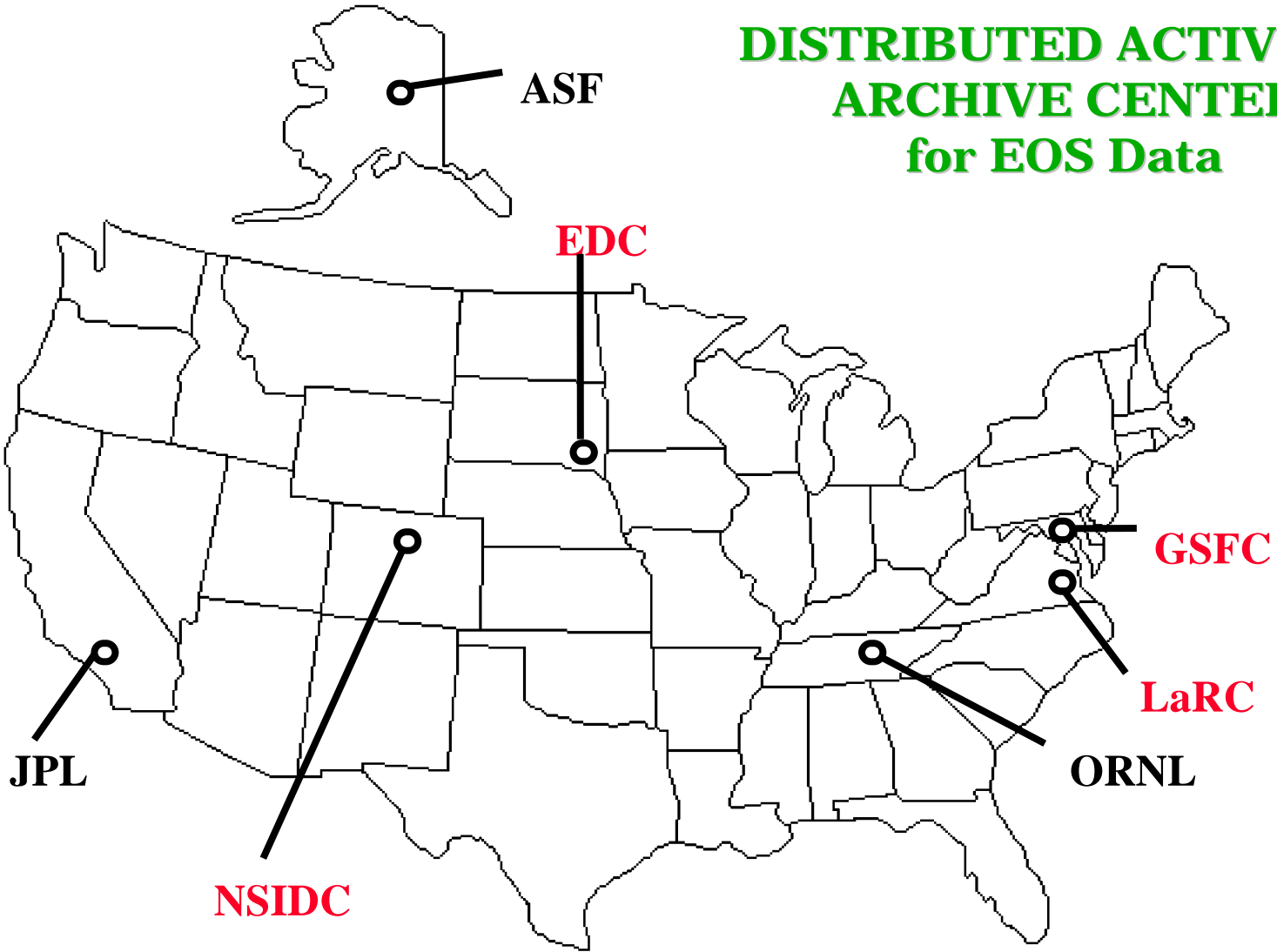
EOSDIS has 3 segments:

- Networks
- Flight Ops
- Science Data Processing System

EOSDIS is composed of several geographically distributed elements that will appear as a single, integrated, logical entity

EOSDIS is working with NOAA and other agencies to ensure long term availability of Earth science data

**DISTRIBUTED ACTIVE  
ARCHIVE CENTERS  
for EOS Data**





## Predicted Data Volumes for 2000

- Expect launch of EOS-Aqua and ADEOS satellites this year
- 260 different data products and sets of raw instrument data
- 1.6 TB of processed data stored daily by end of 2000

Data Center	Archive Volumes GB/Day	# of granules per day	Archive Volumes In TB per year	# of Granules cumulative per year	Distribution via Network GB/day	Distribution via tape GB/day
<b>EDC</b>	522	6886	190	2,513,390	194	159
<b>GSFC</b>	688	5545	251	2,023,925	226	226
<b>LaRC</b>	312	2945	114	1,074,925	102	102
<b>NSIDC</b>	22	1083	8	395,295	6	6
<b>Total</b>	<b>1544</b>	<b>16459</b>	<b>563</b>	<b>6,007,535</b>	<b>528</b>	<b>493</b>

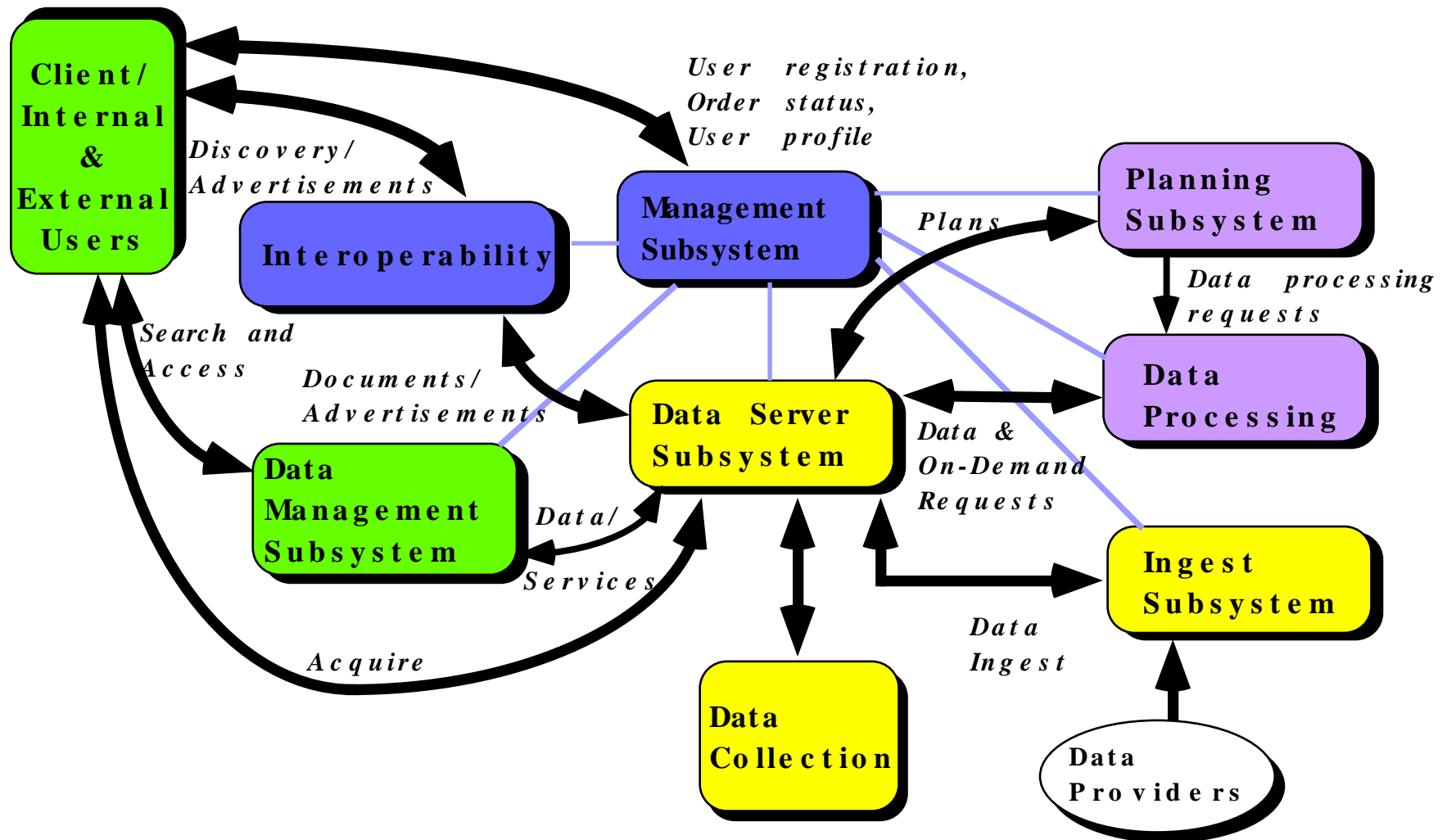


## SDPS System Goals

- Flexible, Scalable, Reliable
- Use Open System standards
- Support standard interface to Earth science to enable coordinated data analysis
- Maximize the use of COTS packages and respond to technological advances and techniques
- Inevitable change and new additions
- *Architecture to support these goals:*
  - *EOSDIS Core System (ECS)*



# ECS Context Diagram





# Data Server Subsystem

“Heart of the ECS system”

- Object-oriented C++ on a multiplatform environment of SUNs and SGIs
- Three Software Configuration Items (CI)
  - Science DataServer CI
    - DBMS, geospatial search, inventory
  - Storage Management CI
    - Manages all peripherals including robotic silos
  - Data Distribution CI
    - Places data in distribution location
- Ingest Subsystem CI is also significant to data archiving

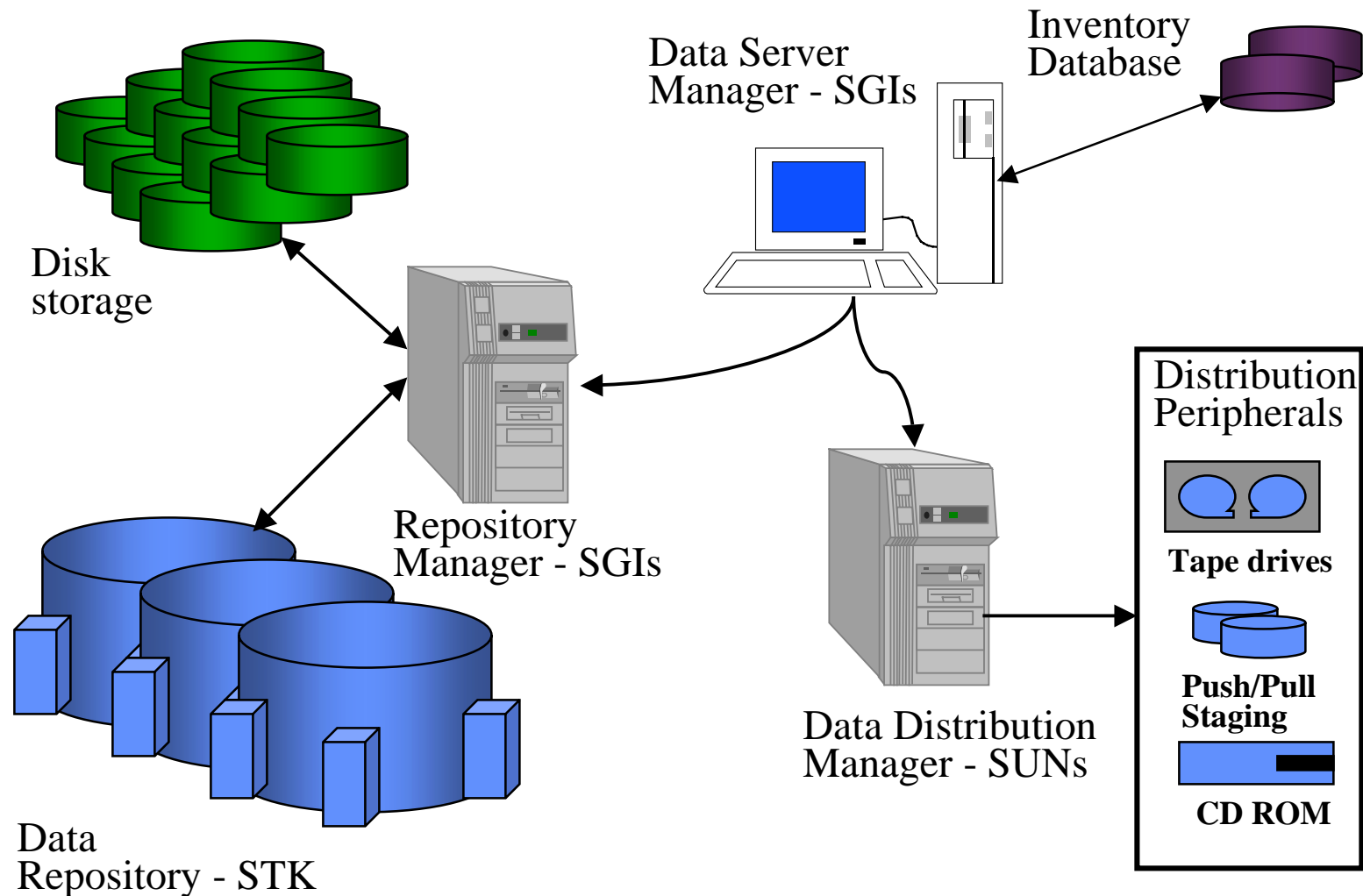


# COTS Packages

- System uses ~ 75 Off-The-Shelf packages from commercial and government sources
- Principal COTS that impact design:
  - Sybase Relational DBMS/SQS - dbms and spatial query system
  - AMASS - file storage management system for robotic storage devices
  - Autosys - scheduling software for the processing system
  - Tivoli - system management tools
  - HP Openview- graphical tool for system management
  - RogueWave - libraries used to map components to objects
  - DCE - distributed computing environment
  - ClearCase - CM tool to manage completion of different builds
  - Remedy - trouble-ticketing software used across project



# Data Server Subsystem Hardware Context Diagram



*\*configuration is similar at all sites*





# Data Repository

## Archive Robotic Storage

DAAC	Make/Model	Qty	Drive Type	Media Capacity	Total # of Media in silos
GSFC	StorageTek STK Powderhorn	4	13 D3 drives	50 GB	12,000
			10 9840 drives	40 GB	
NSIDC	StorageTek STK Powderhorn	1	3 D3 drives	50 GB	600
EDC	StorageTek STK Powderhorn	3	14 D3 drives	50 GB	7,700
			8 9840 drives	40 GB	
LaRC	StorageTek STK Powderhorn	2	8 D3 drives	50 GB	3,100
			8 9840 drives	40 GB	

DAAC	Make/Model	Qty	Drive Type	Media Capacity
GSFC	Exabyte tape drives CD Rom Writers	8	8mm tape	50 GB
		2		600 MB
NSIDC	Exabyte tape drives CD Rom Writers	2	8mm tape	50 GB
		2		600 MB
EDC	Exabyte tape drives CD Rom Writers D3 tape drive	8	8mm tape	50 GB
		2	D3 tape	600 MB
		1		50 GB
LaRC	Exabyte tape drives CD Rom Writers	2	8mm tape	50 GB
		2		600 MB

## Distribution Systems



# STK Silo







# Drive Cabinet





## Typical Cartridge Media for Silo





# Mass Storage I/O System

- Consists of silo, RAID disk, server hosts
  - Capable of 40 MB/s throughput sustained at all times (3.5 TB of data per day)
- Able to push 3 to 4 times as much data because of double buffer mechanism in the storage management system design
  - Minimizes stress on robotics
  - Creates our own persistent cache



# Mass Storage I/O System

- Utilize Volume Groups (groups of tapes together)
  - Group tapes by ‘science data type’ (for example, all Landsat data in a silo is grouped together on a specified collection of tapes)
  - Enables load balancing
  - Assures minimum performance levels
  - Allows logical management of the archive
- Additional information in two poster papers at this conference
  - *Fault Tolerant Design*
  - *Scalable Architecture for Maximizing Concurrency*



# Archive Operations

- Strive for an automated archive system
  - Continuous connection to the archive systems by operations personnel
  - At least two operations personnel at each site:
    - Principle activities include error notification; backup; monitoring; problem resolution
  - Strive for lights out administration
- Support several modes at each site for system upgrade
  - One operational mode; 2 test modes
- Scheduled maintenance includes hardware monitoring, media monitoring, format and cleanup





# Conclusion

*So How Are We Doing?*

Archive  
Size to  
Date

<b>DAAC</b>	<b>Type</b>	<b>Archive Size</b>
EDC	Landsat7/ ASTER/ MODIS Land	20 TB 274 GB/day
NSIDC	MODIS Land (snow products)	9 TB 25 GB/ day
GSFC	MODIS L1/ Atmos/Ocean	50 TB 300 GB/day
LaRC	MISR	18 TB 88 GB/day