EOSDIS: Archive and Distribution Systems in the Year 2000

Jeanne Behnke

Earth Science Data Information Systems

&

Alla Lake

Lockheed Martin

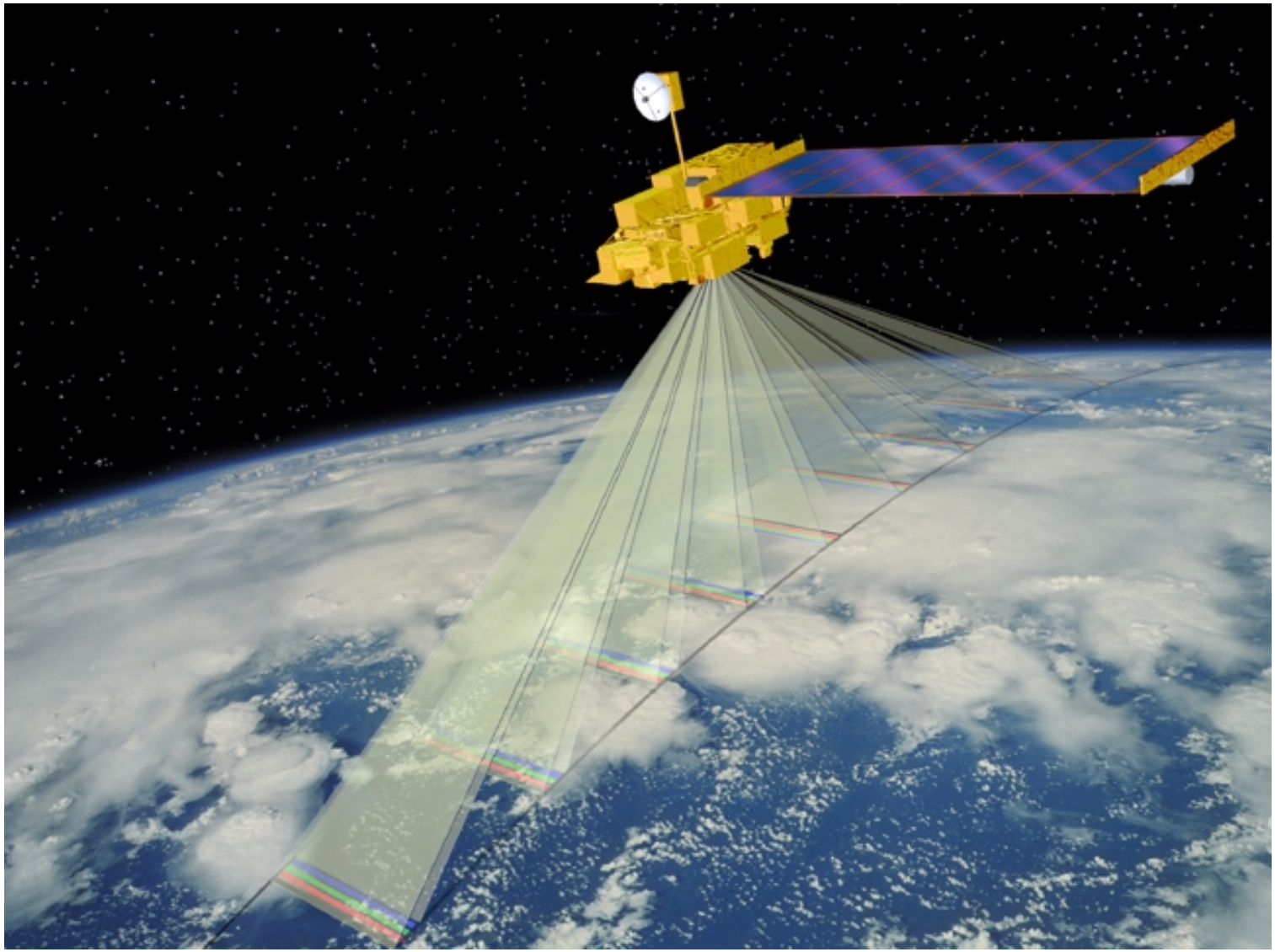
GSFC/IEEE Mass Storage & Technology Conference

March 29, 2000

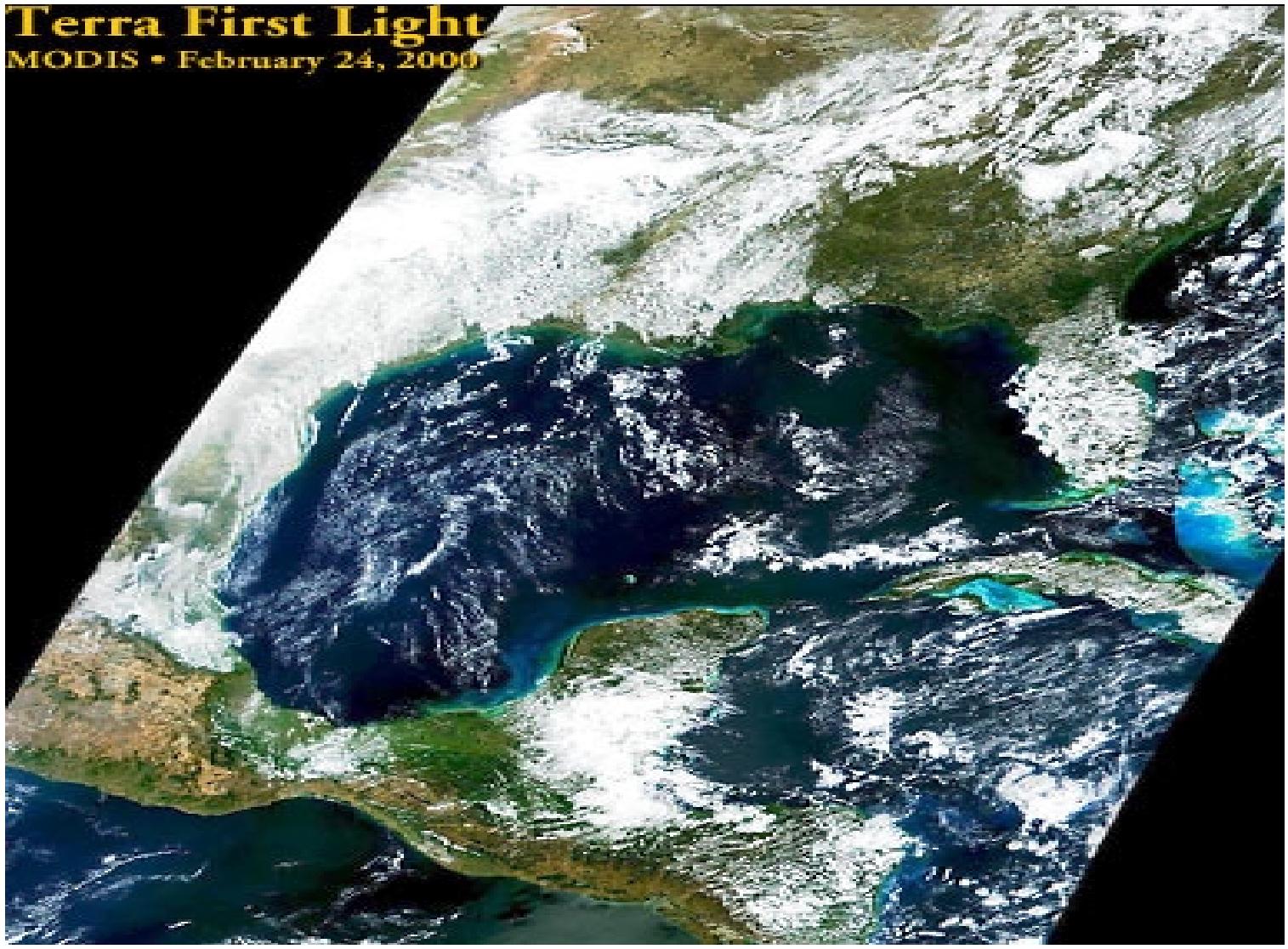
A Big Mission!

- 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)

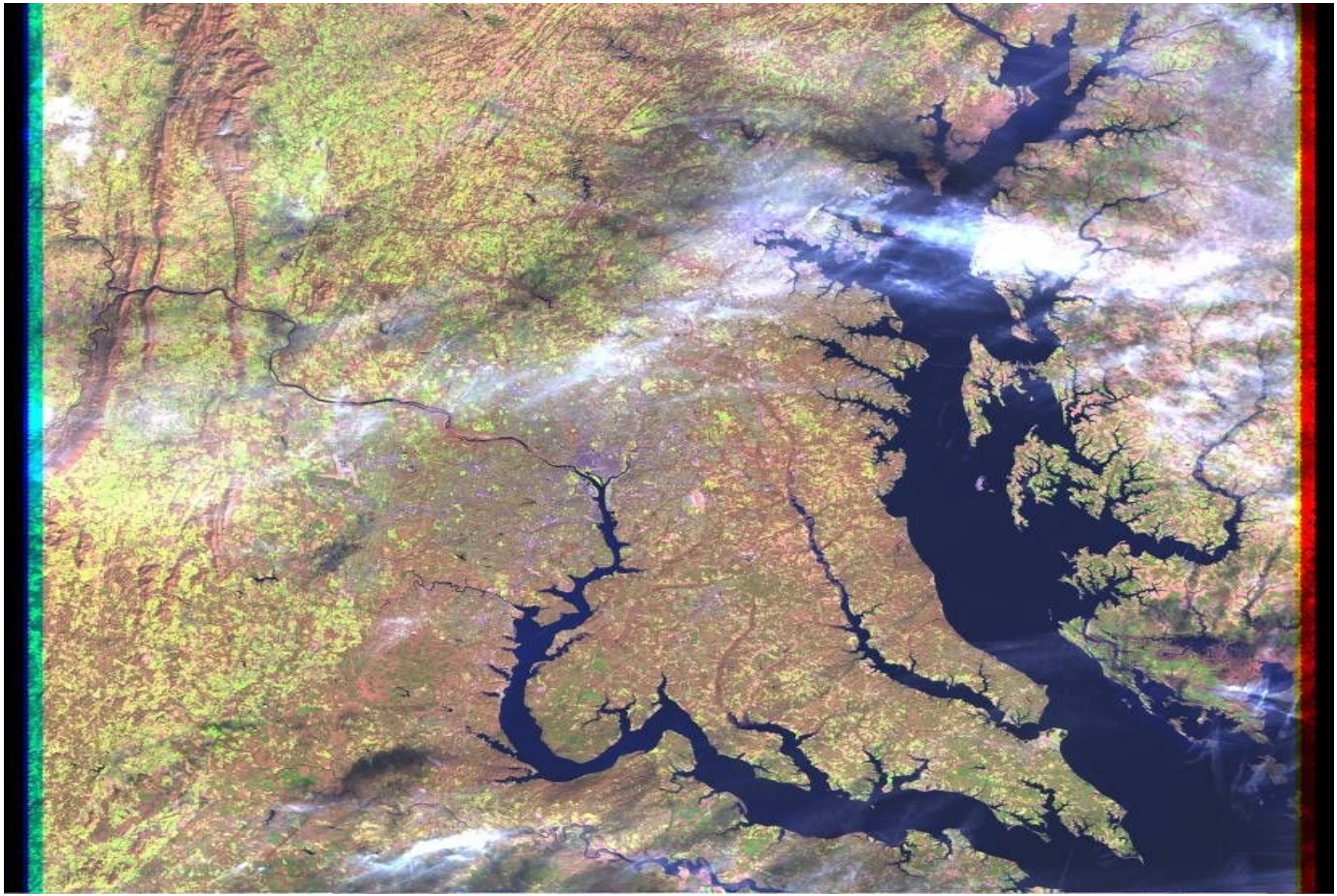
EOS Terra in orbit



Terra Image



Landsat Browse Image



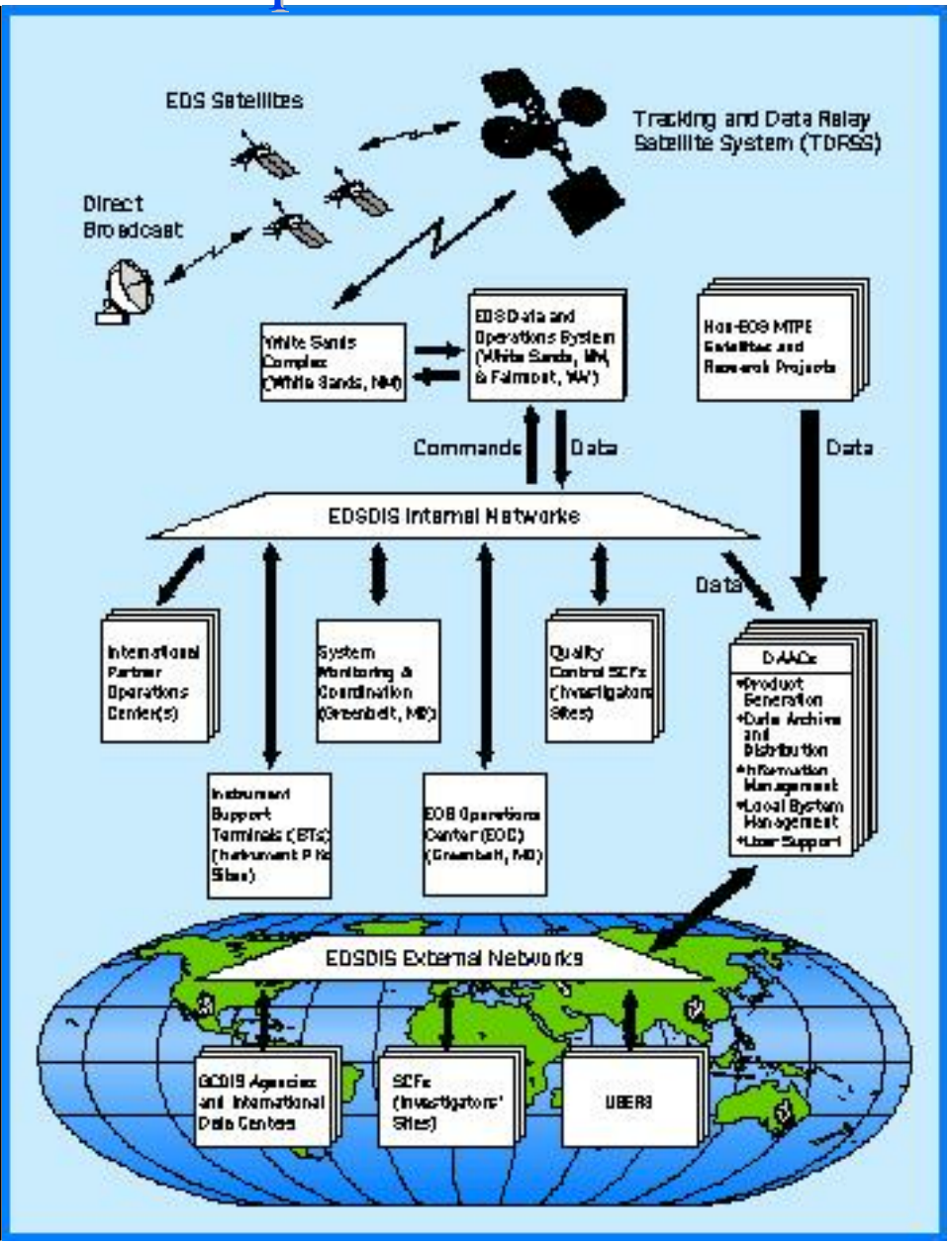
EOSDIS Concept

EOSDIS has 3 segments:

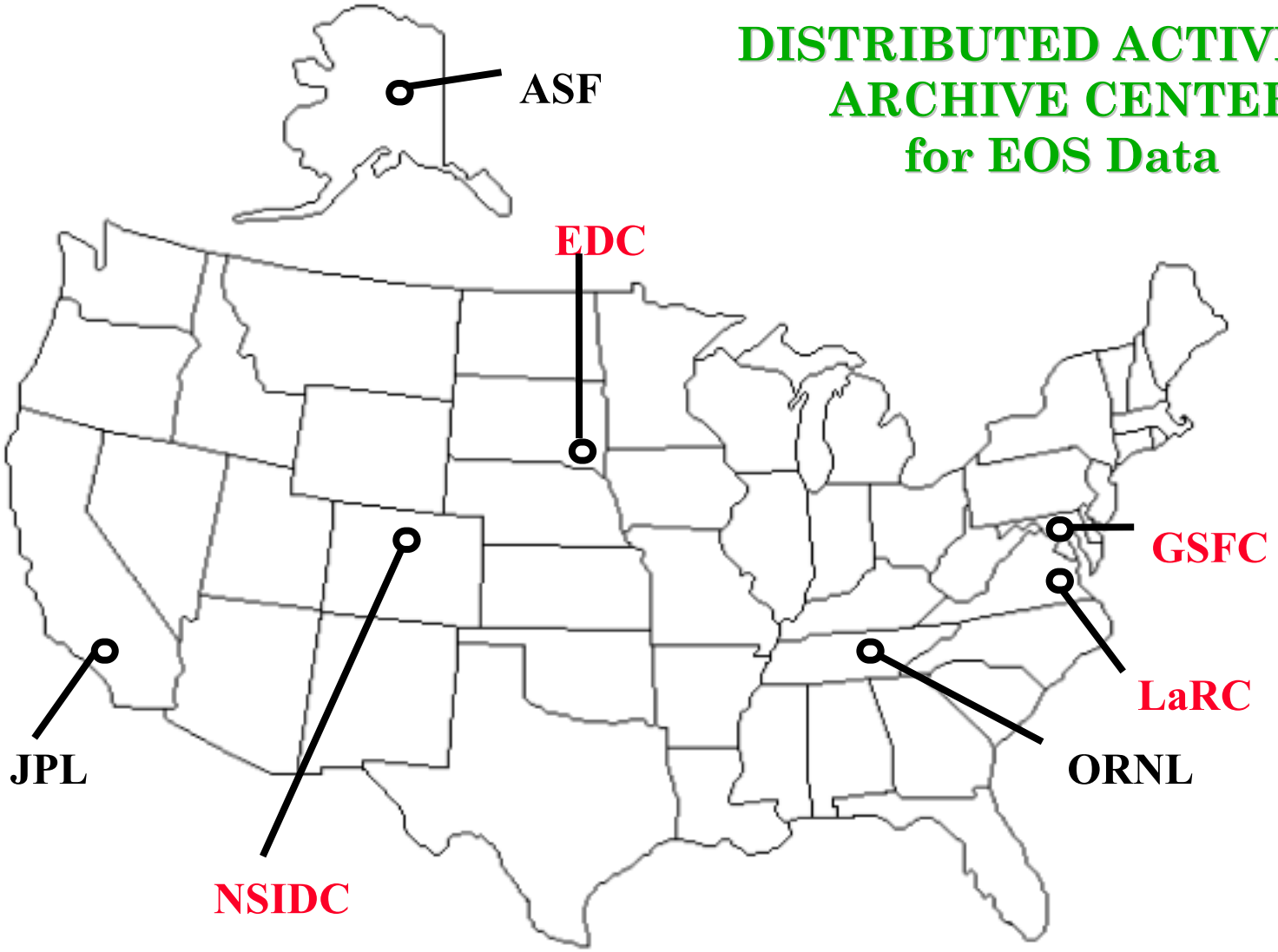
- Networks
- Flight Ops
- Science Data Processing System

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

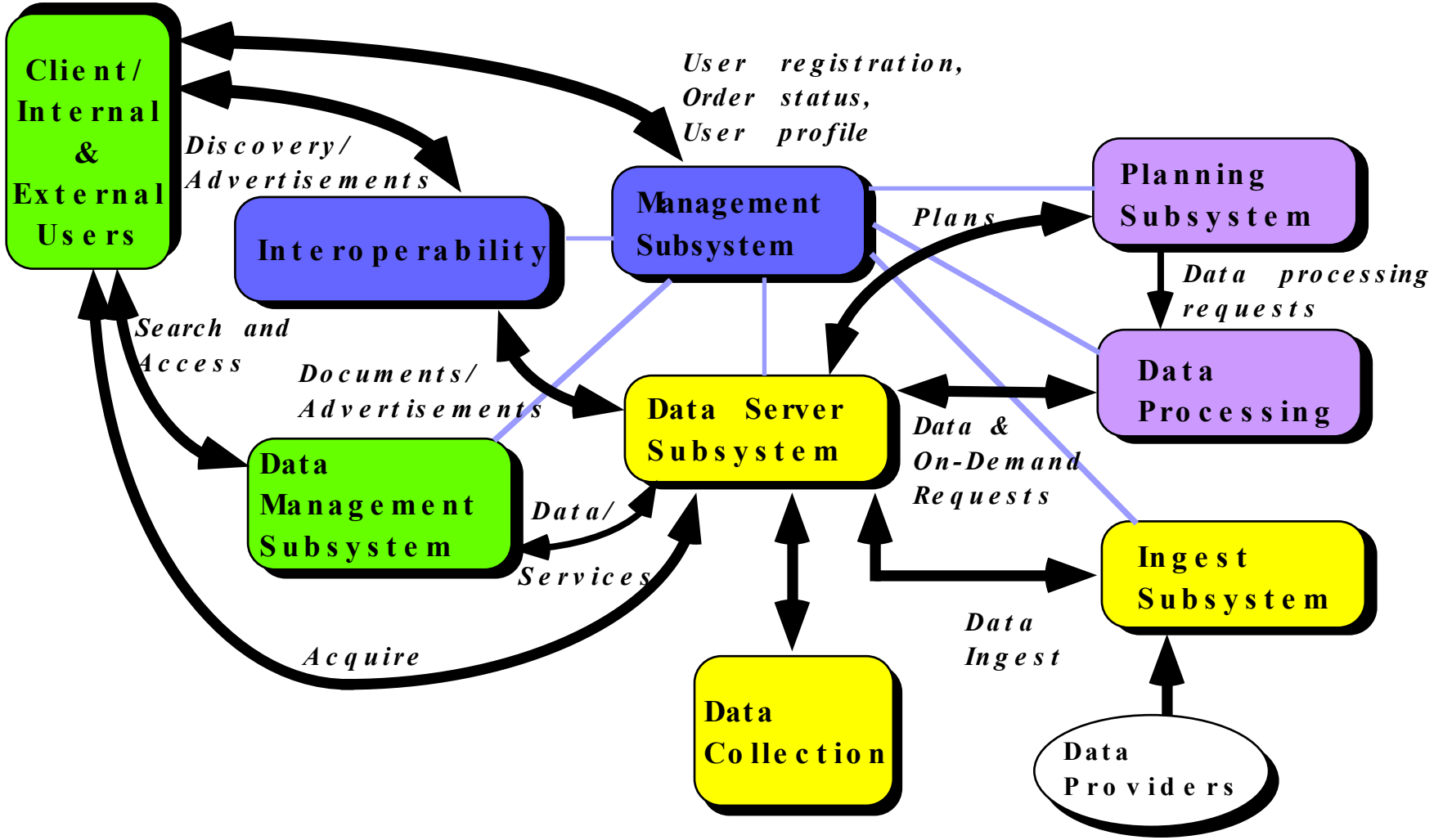
- Expect launch of EOS-Aqua and ADEOS satellites this year
- 260 different data products and sets of raw instrument data
- 1.6 TB of processed data stored daily by end of 2000

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
EDC	522	6886	190	2,513,390	194	159
GSFC	688	5545	251	2,023,925	226	226
LaRC	312	2945	114	1,074,925	102	102
NSIDC	22	1083	8	395,295	6	6
Total	**1544**	**16459**	**563**	**6,007,535**	**528**	**493**

SDPS System Goals

- Flexible, Scaleable, 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

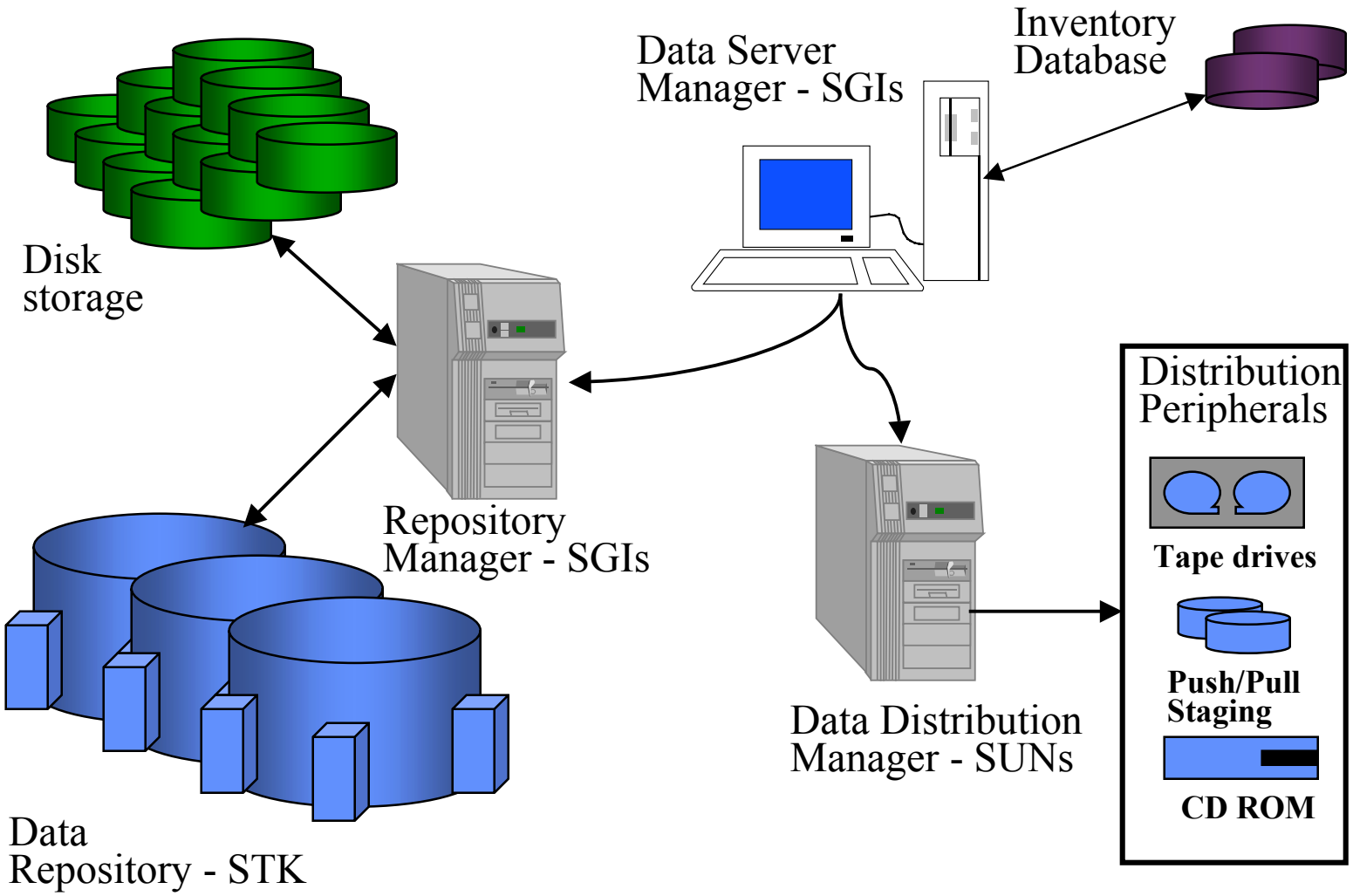
"Heart of the ECS system"

- 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 silo
GSFC	StorageTek STK Powderhorn	4	13 D3 drives 10 9840 drives	50 GB 40 GB	5200
NSIDC	StorageTek STK Powderhorn	1	3 D3 drives	50 GB	5200
EDC	StorageTek STK Powderhorn	3	14 D3 drives 8 9840 drives	50 GB 40 GB	5200
LaRC	StorageTek STK Powderhorn	2	8 D3 drives 8 9840 drives	50 GB 40 GB	5200

Distribution Systems

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

STK Silo


11P

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

Archive Size to Date

DAAC	Date	Archive Size in GB
EDC	26 days 2/23-3/20	4463.3
NSIDC	22 days 2/28-3/21	113
GSFC	27 days 2/23-3/21	7747
LaRC	20 days 2/24-3/15	1375

- For further discussion, contact:
  - jeanne.behnke@gsfc.nasa.gov; 301-614-5326
  - alake@eos.hitc.com; 301-925-0626