

EOSDIS Program Update

***THIC Meeting
Turf Valley Hotel
Oct 15, 1996***

**Ben Kobler
NASA GSFC**

- **EOSDIS Overview**
- **EOSDIS Architecture**
- **Data Server Design and Configuration**
- **Risks and Issues**

Mission to Planet Earth

- **Long term NASA research mission to study processes leading to global climate change**
- **US portion of International effort aimed at understanding the Earth System**
- **has 3 major components**
 - **Space based Earth Observing System (EOS)**
 - **Ground based EOS Data and Information System (EOSDIS)**
 - **Interdisciplinary scientific research program**

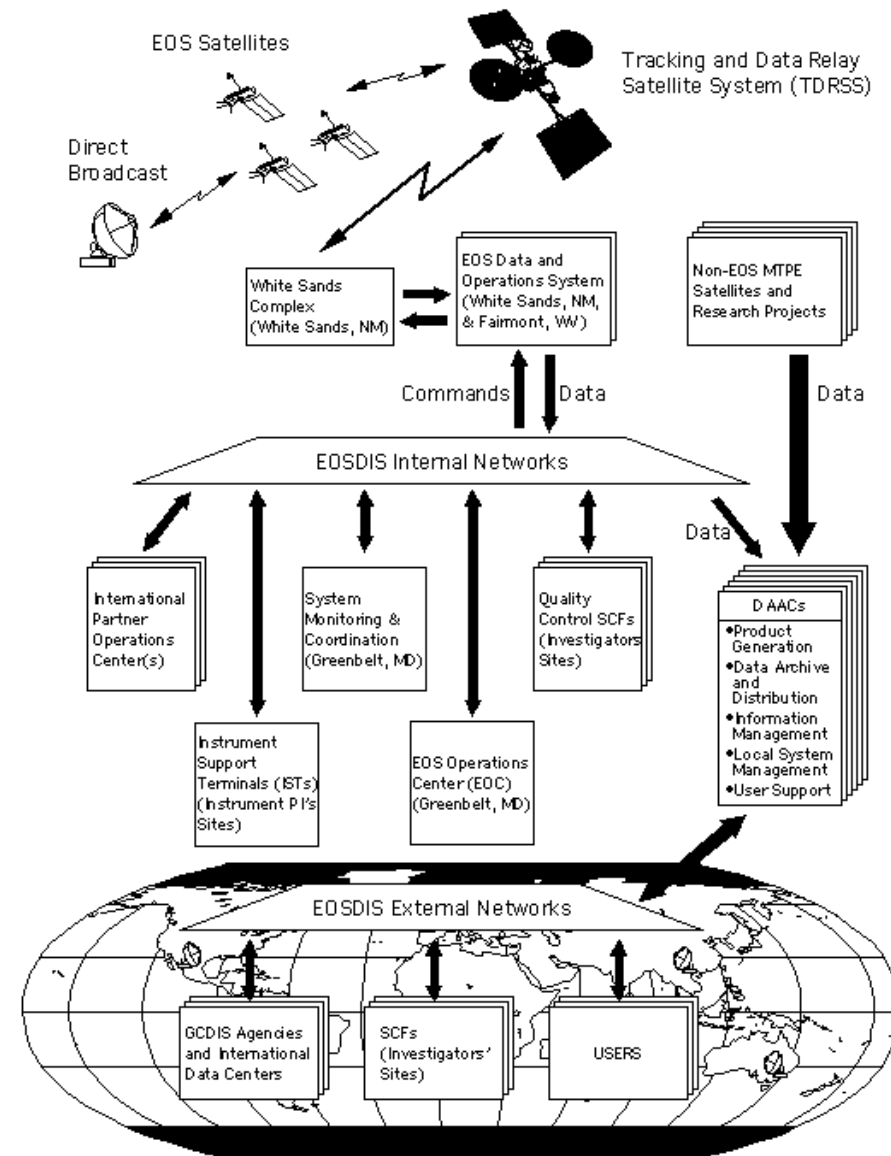
EOS Missions

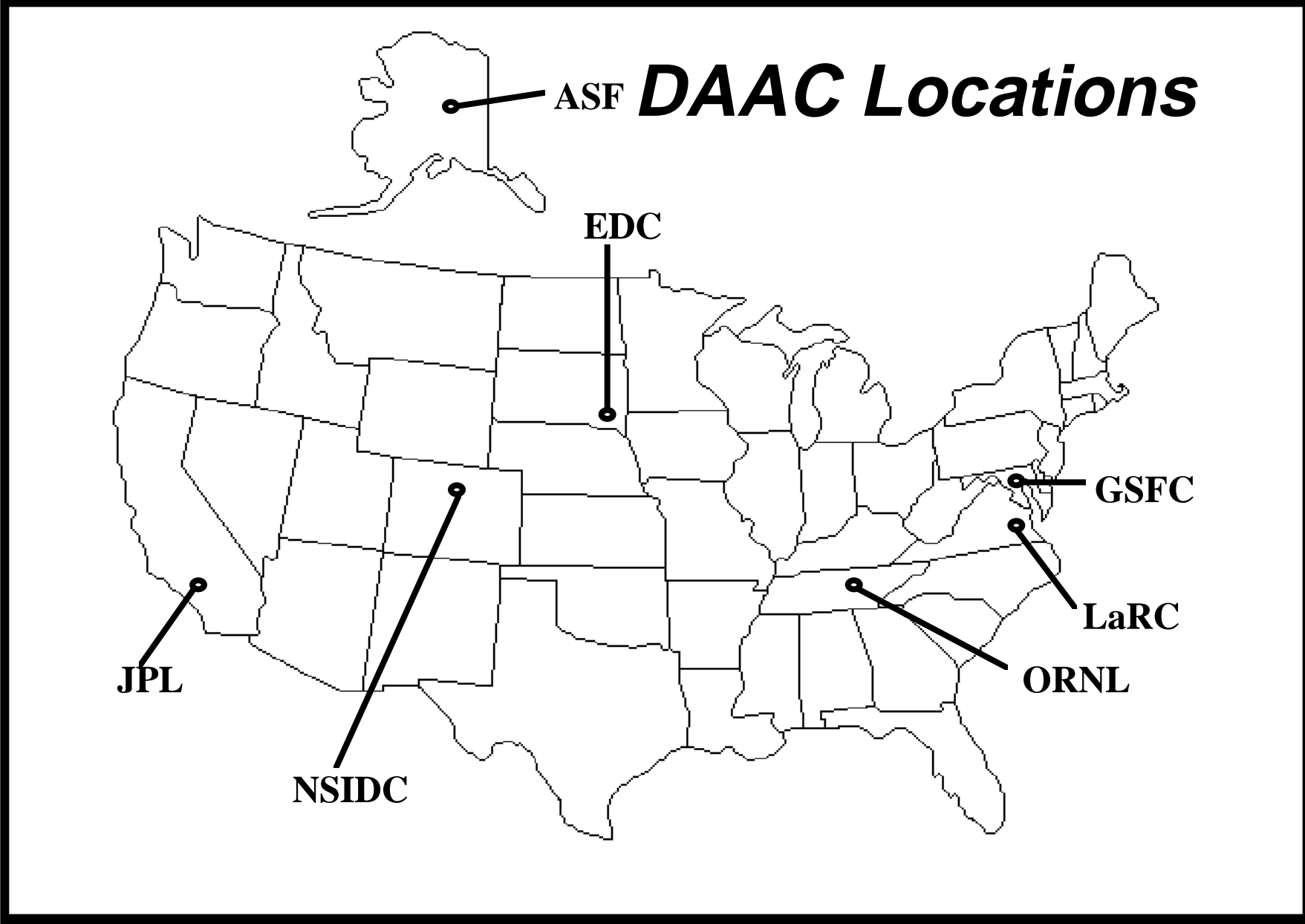
- TRMM Aug 1997**
- Landsat 7 May 1998**
- AM-1 June 1998**
- METEOR Aug 1998**
- ADEOS II Feb 1999**
- RADAR ALT March 1999**
- FOO June 1999**
- PM-1 Dec 2000**
- Space Station June 2001**
- CHEM-1 Dec 2002**
- LASER ALT Jul 2003**
- AM/PM-2 2004, 2006**
- AM/PM-3 2009, 2011**

EOSDIS is implemented using a distributed, open systems architecture

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





Volume Requirements

- **Total storage: approximately 1 TB/day**
- **Distribution rates**
 - **ASF: 1.1 MB/sec (95 GB/day)**
 - **EDC: 2.4 MB/sec (207 GB/day)**
 - **GSFC: 13.6 MB/sec (1175 GB/day)**
 - **JPL: 0.8 MB/sec (69 GB/day)**
 - **LaRC: 5.3 MB/sec (458 GB/day)**
 - **NSIDC: 0.4 MB/sec (34 GB/day)**
- **Average file size: 51 MB (14.3 MB @ NSIDC, 95 MB at EDC)**
- **Average request rate: 17 files/minute @ GSFC**

3 EOSDIS components

- **FOS - Flight Operations Segment**
 - Manages and controls the EOS spacecraft and instruments
- **CSMS - Communications and System Management Segment**
 - Provides an underpinning of communications and management functions
- **SDPS - Science Data Processing Segment**
 - Provides distributed capabilities for processing, storing, and distributing data

7 SDPS subsystems

- **INGEST** - Initial reception of data
- **DATA SERVER** - Physical storage and data distribution
- **PLANNING** - Pre-planning and scheduling
- **DATA PROCESSING** - Data production
- **CLIENT** - User interface which contains tools to search, retrieve, and display data
- **INTEROPERABILITY** - Search and locate data services
- **DATA MANAGEMENT** - Locate and access data

SDPS Subsystem Context

Data Server subsystem

- **Software components**
 - Science Data Server
 - Document Data Server
 - Data Distribution
 - Storage Management
- **Hardware components**
 - Access Control and Management
 - Distribution and Ingest Peripherals
 - Working Storage
 - Data Repository

Science Data Server software

- **Hides underlying metadata and indexing structures from higher level services**
- **Provides a consistent set of interfaces to higher level services**
- **There may be one or several Science Data Servers at any given site**
- **Makes data and their services known to the Interoperability subsystem via advertisements**
- **Developed through a combination of custom code and COTS (Illustra DBMS) software**

Document Data Server software

- **Provides World-Wide-Web style browsing and searching**
- **Future implementations might integrate the Science Data Server and the Document Data Server**
- **Developed through a combination of custom code and COTS (Netscape, Topic, Illustra) software**

Data Distribution software

- **Prepares requested data objects for distribution on specified media or via the network**
- **Data preparation can include data reformatting, data compression, data translation**
- **Generates necessary packing labels and metadata**

Storage Management software

- **Manages allocations of Data Server resources such as working storage, staging disks, and persistent storage**
- **Provides isolation between the search and access views of data and physical storage mechanisms**
- **Combination of custom code and COTS (AMASS FSMS) software**

Access Control and Management hardware

- **Manages Data Server access, maintains client sessions, routes service requests**
- **Provides computing resources for execution of Data Server software components**

Distribution and Ingest hardware

- **Ingest of electronic and physical media where there is not a high RMA requirement**
- **Buffer area for media generation and for electronic data transfers to users**
- **User access to data is isolated**
- **May support a hierarchy of data storage components**

Working Storage hardware

- **Pool of storage used for temporary file buffer space**
- **Shared network attached disks (later releases)**
- **Can include second tier of robotic based storage components**

Working Storage Configuration (AM-1 Release, 1st procurement)

- **EDC**
 - SGI C XL, 51 GB RAID
 - EMASS AML/2, 6 3590 drives, 2 arms
- **NSIDC**
 - (handled by Data Repository)
- **GSFC**
 - SGI C XL, 496 GB RAID
- **LaRC**
 - SGI Challenge XL, 34 GB RAID
- **JPL**
 - SGI C L, 51 GB RAID

Data Repository hardware

- **Provides permanent data storage and maintenance functions**
- **Supports services to monitor and maintain data integrity**
- **Multiple technologies can be instantiated**
- **May support a hierarchy of data storage components**

Data Repository Configurations (AM-1 Release, 1st procurement)

- **EDC**
 - SGI PC XL(x 2), 324 GB RAID
 - STK Powderhorn, 7 D3 drives, SUN Sparc5
- **NSIDC**
 - SGI L (x 2) , 119 GB RAID
 - STK Powderhorn, 3 D3 drives, SUN Sparc5
- **GSFC**
 - SGI Challenge XL (x 2), 432 GB RAID
 - STK Powderhorn, 8 D3 drives, SUN Sparc5

Data Repository Configurations(cont)

- **LaRC**
 - SGI Challenge XL (x 2), 540 GB RAID
 - STK Powderhorn, 8 D3 drives, SUN Sparc5
 - EMASS AML/2, 3 3590 drives, 1 arm
- **JPL**
 - SGI C L (x 2), 324 GB RAID
 - EMASS AML/E, 3 3590 drives

Browse Data Repository

- **Sizing (2002)**
 - ASF: 0.121 TB
 - EDC: 21.3 TB
 - GSFC: 42.6 TB
 - JPL: 1.54 TB
 - LaRC: 10.7 TB
 - NSIDC: 14.2 TB
- **Media: optical platters**
- **Robotics: EMASS GRAU**

Current Status

- **TRMM Release**
 - **Completing Code and Unit Test**
 - **Beginning Integration and Test**
- **AM-1 Release**
 - **Completing Critical Design Reviews**

Risks and Issues

- **Insufficient standards to enable interoperability**
- **Insufficient reliability and performance**
- **Unproven scalability**

Risk: Insufficient Standards to Enable Interoperability

- **Prototype: Standards for Petabyte Archives**
 - File Level Metadata for Portability of Sequential Storage Media (FMP) Working Group
 - IEEE SSSWG (Mass Storage Reference Model)
- **ECS Design: Allows multi-File Storage Management System implementation**
 - Needs to be verified

Risk: Insufficient Reliability and Performance

- **Prototype: Mass Storage Testing Laboratory**
 - Develop standard mass storage benchmarks
 - Run benchmarks to test vendor products and configurations
 - File Storage Management Software
 - Archive media and recorders
 - Archive robotics
- **Monitor technology developments**
 - Optical tape development
 - High performance magnetic tape development
 - Holographic storage development

Risk: Unproven Scalability

- **Prototype: Modeling (Ted Johnson at U. of Florida)**
- **ECS Design: Distributed scalable architecture**
 - Scalability needs to be verified

Cooperative Announcement Notice (CAN-96-MTPE-01)

- **Creative and Innovative Working Prototype Earth Science Information Partnerships in Support of Earth System Science**
- **Includes Data Archive Elements:**
 - High-capacity storage systems, integration of off-the-shelf tape and optical technologies
 - Data organization schemes for optimum file access based on user access patterns
 - Scalable file storage and management systems
 - Concepts for storage of information
 - Concepts for information stewardship
 - System wide techniques for compression and decompression of files, objects, and databases
 - System techniques to optimize data flow

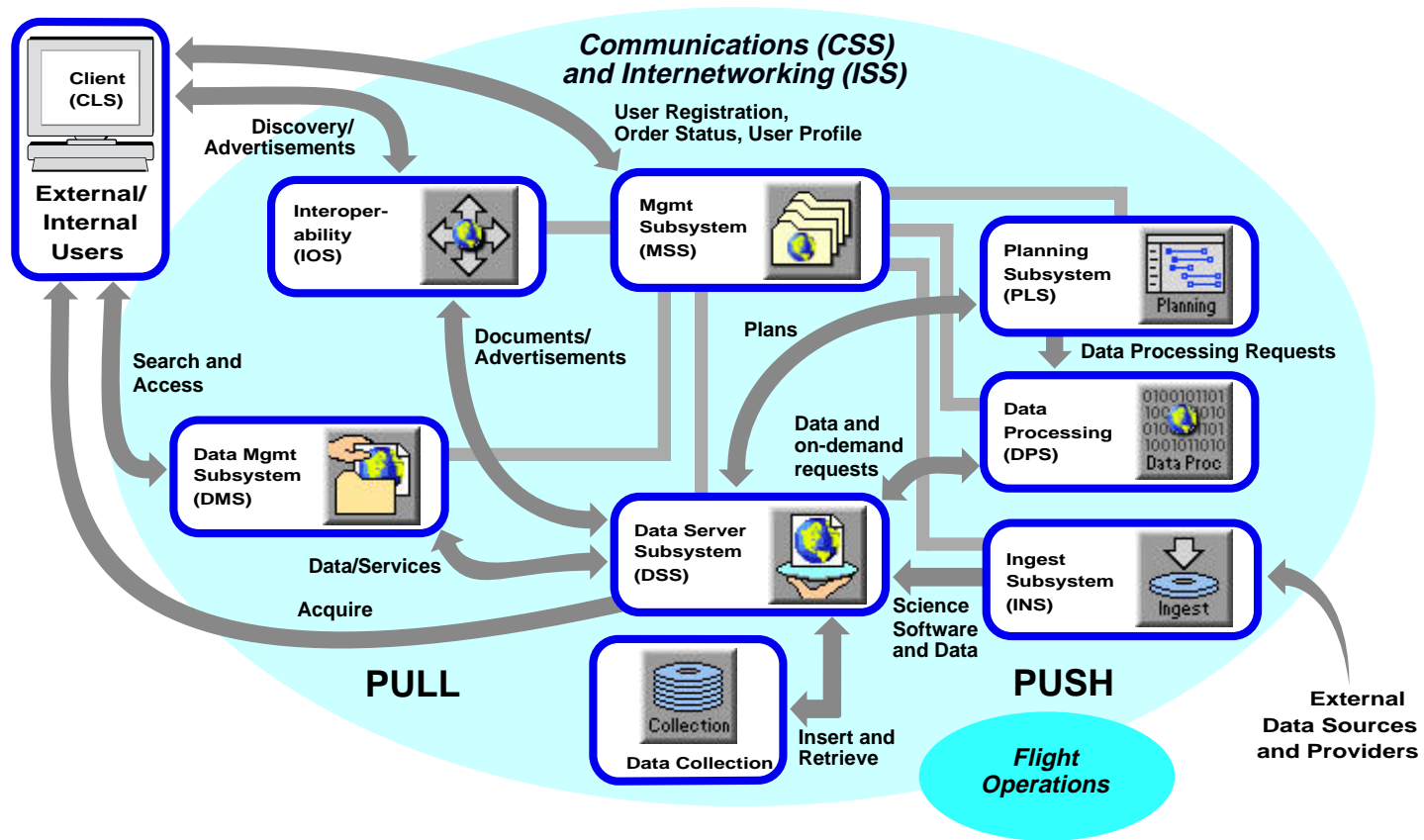


Figure 4.3-1. ECS Context Diagram