

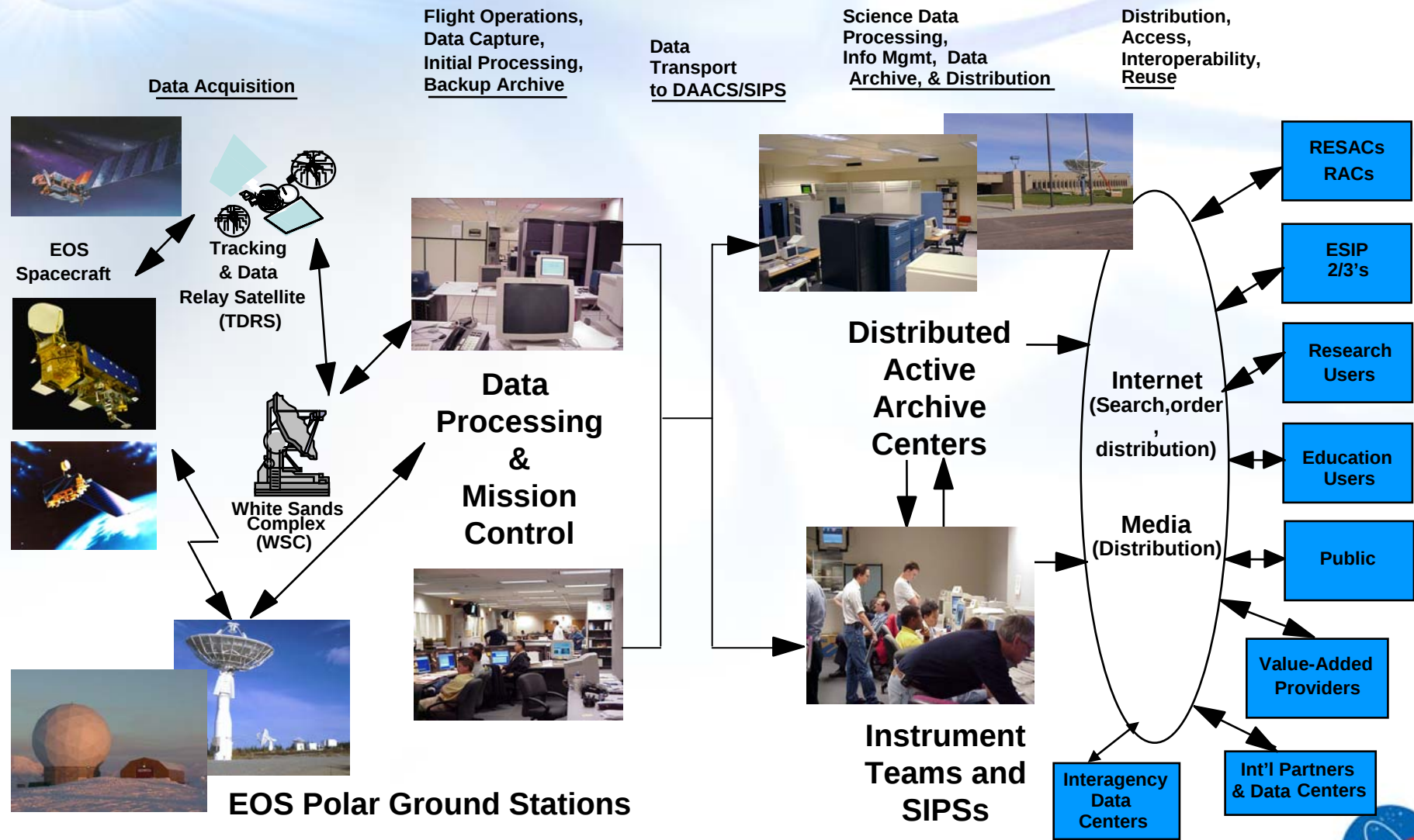
ESE Grid Prototypes at the Goddard Space Flight Center



Ken McDonald
NASA/GSFC
April 7, 2004

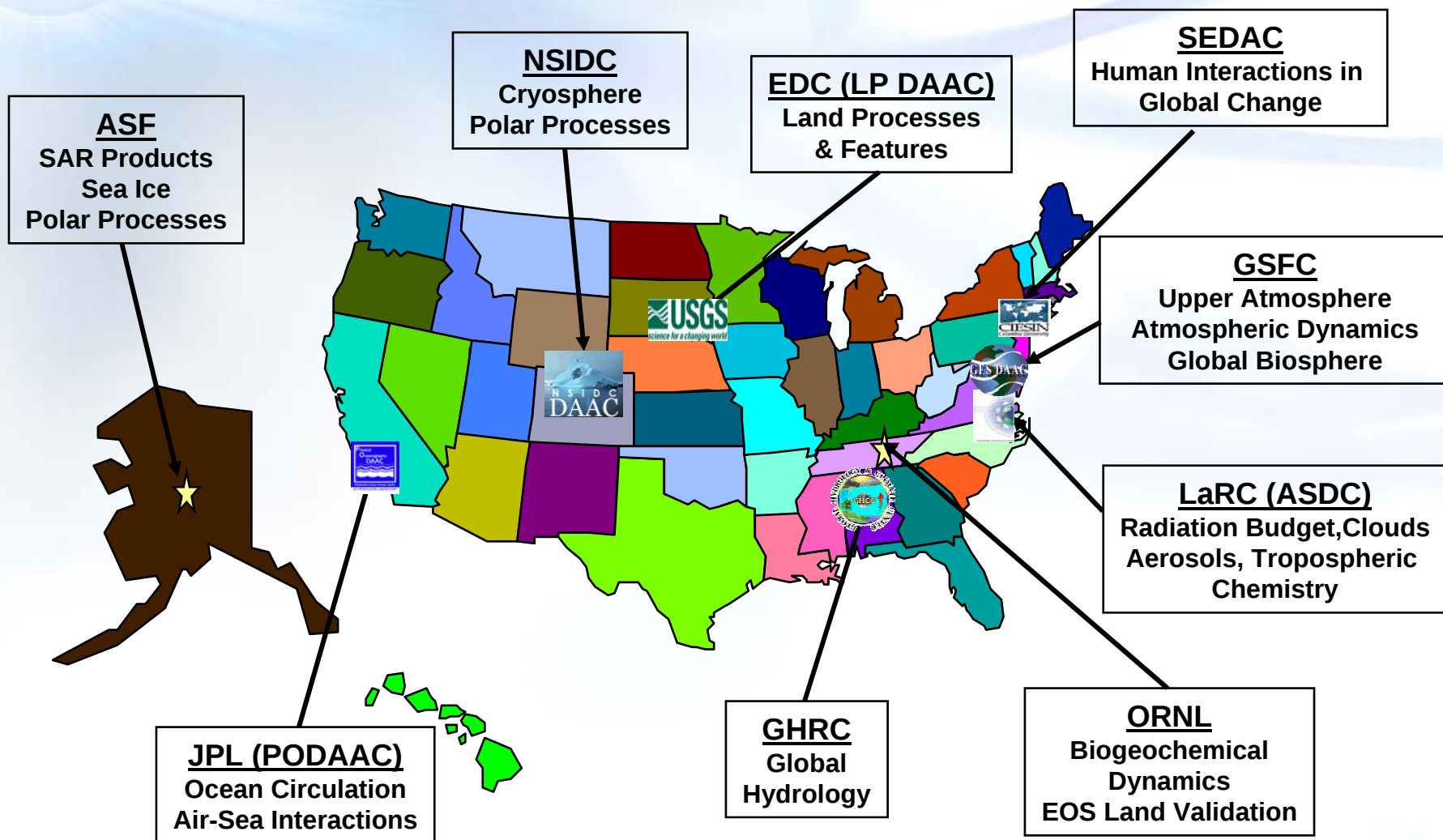


EOSDIS Context



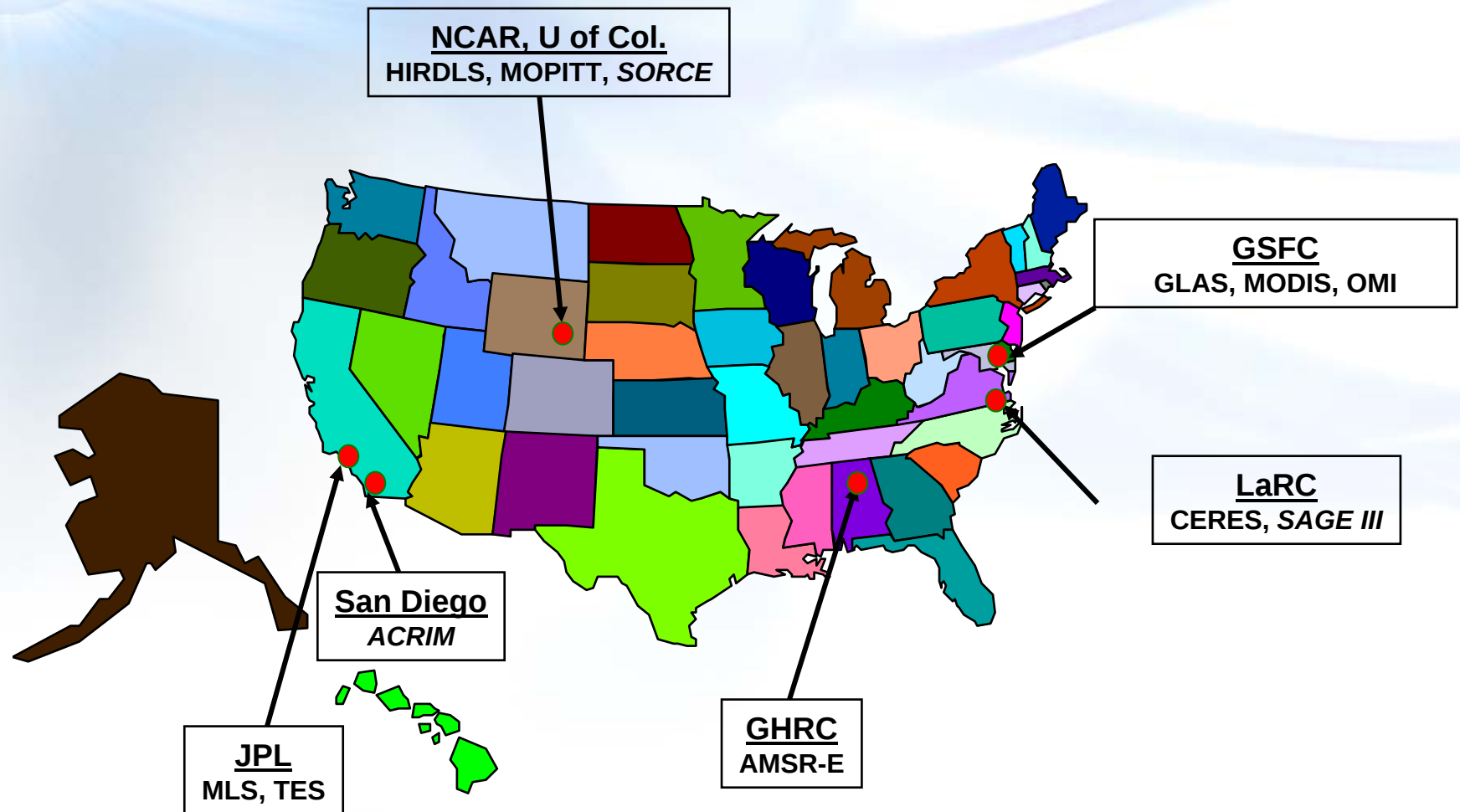


DAAC Alliance Data Centers





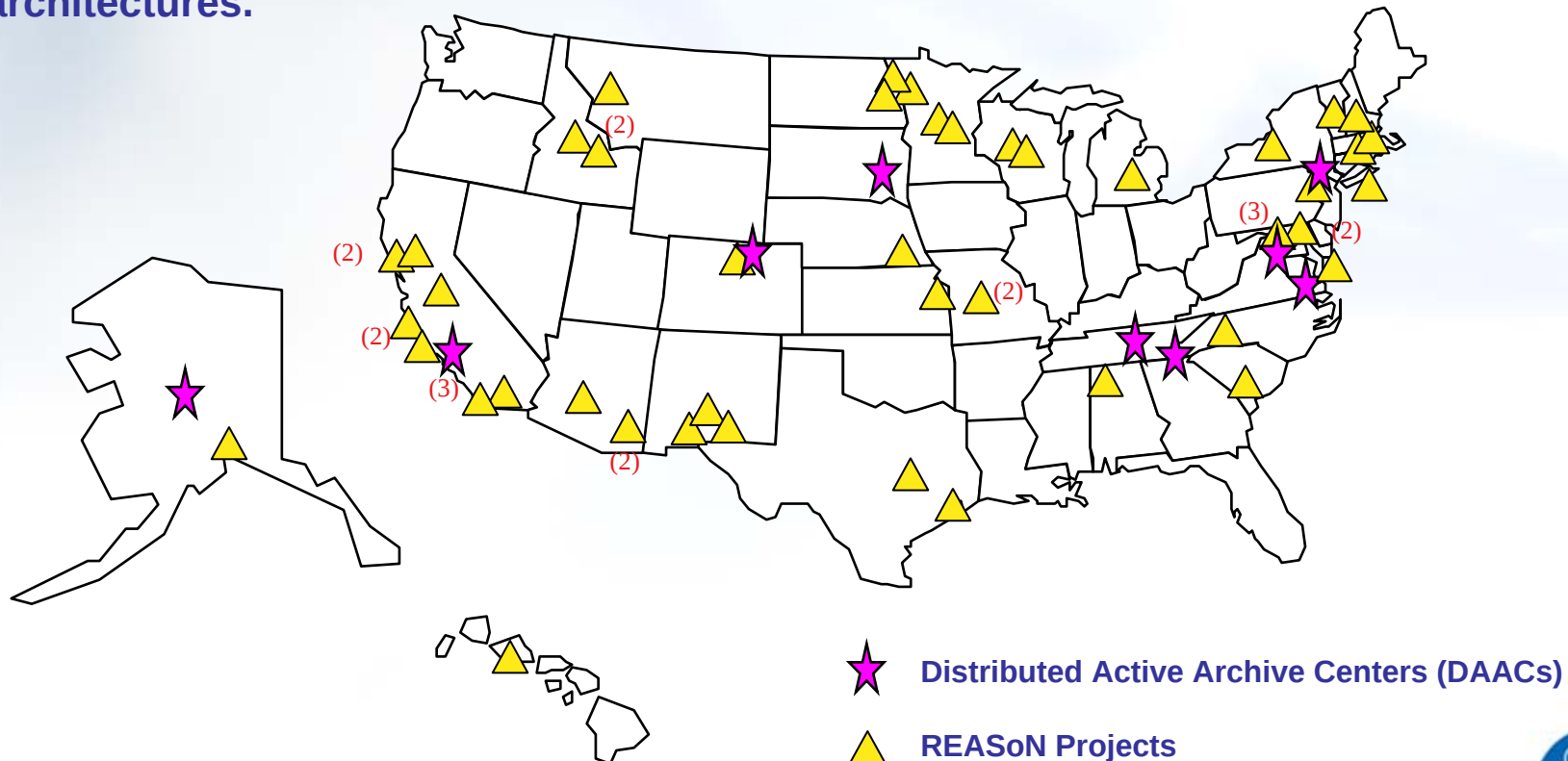
Science Investigator-led Processing Systems (SIPSs)





ESE Data Center Locations

- A total of 68 widely distributed data centers (some of which are at the same location).
- ESE has recently updated our peer reviewed data and information producing centers through the Research, Education and Applications, Solutions Network Cooperative Agreement Notice (REASoN CAN) for development of next-generation architectures.





Grid: Selected Activities Within NASA/ESE

■ Grid Prototypes

- Several projects of the Earth Science Enterprise (ESE) are sponsoring or managing prototypes that use Grid technologies
- Examples:
 - AIST - Integration of OGC and Grid Technologies for Earth Science Modeling and Applications - George Mason U
 - LDCM/GSFC - Advanced Data Grid - GSFC Science Data Systems Branch
 - ESDIS - Remote Data Storage - EDS/Raytheon

■ Grid Collaboration and Information Exchange

- NASA is leading a Grid Task Team of the CEOS Working Group on Information Systems and Services (WGISS)



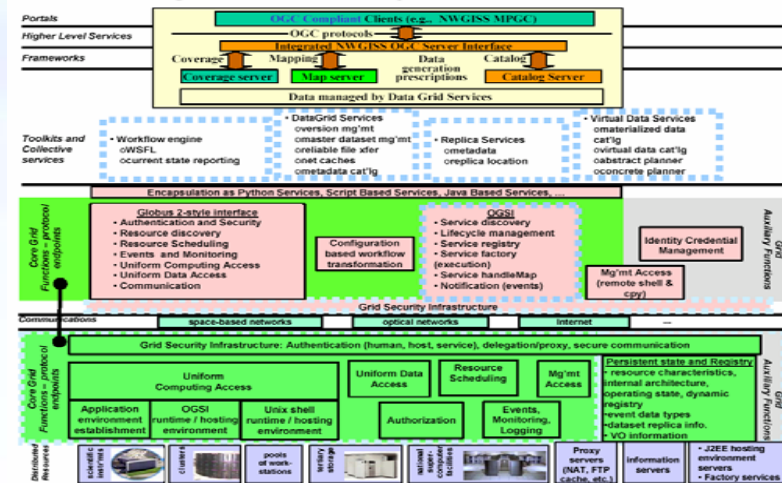
GMU Grid Prototype

OGC and Grid Technology Integration for Earth Science Modeling and Applications

Liping Di, George Mason University

Description and Objectives

- Integrate Grid and OGC technologies to make Grid-managed data accessible through NWGISS OGC servers and allow users to focus on science rather than issues with data receipt, format, and data manipulation
- Leverages the OGC-compliant NASA HDF-EOS Web GIS Software Suite (NWGISS), CEOS Grid testbed, Globus and NASA information Power Grid (IPG) and DOE's Earth System Grid (ESG)



Approach

- Phase 1: Testbed and initial integration (set up development environment, preliminary integration design, implementation of WCS access to Grid-managed data)
- Phase 2: Data naming and location transparency (investigate use of Data Grid & Replica Services)
- Phase 3: Virtual dataset research & development

Co-I's/Partners

William Johnston, ARC/Lawrence Berkley Lab
Dean Williams, Lawrence Livermore National Lab

Science Themes

Atmospheric Composition
Climate
Water & Energy Cycle

Carbon cycle
Weather

Schedule and Deliverables

- Testbed running (8/03)
- Demonstrate Grid-secured NWGISS WCS (4/04)
- Demonstrate WCS access to data pool grid (11/04)
- Demonstrate Grid-enabled WRS, WMS, and WCS for accessing non-virtual data (4/05)
- Demonstrate Grid-enabled NWGISS access to data pools & ESG (11/05)
- Demonstrate NWGISS WCS virtual data access (4/06)

Application/Mission

Facilitate access to EOSDIS data by Earth science modeling and applications communities





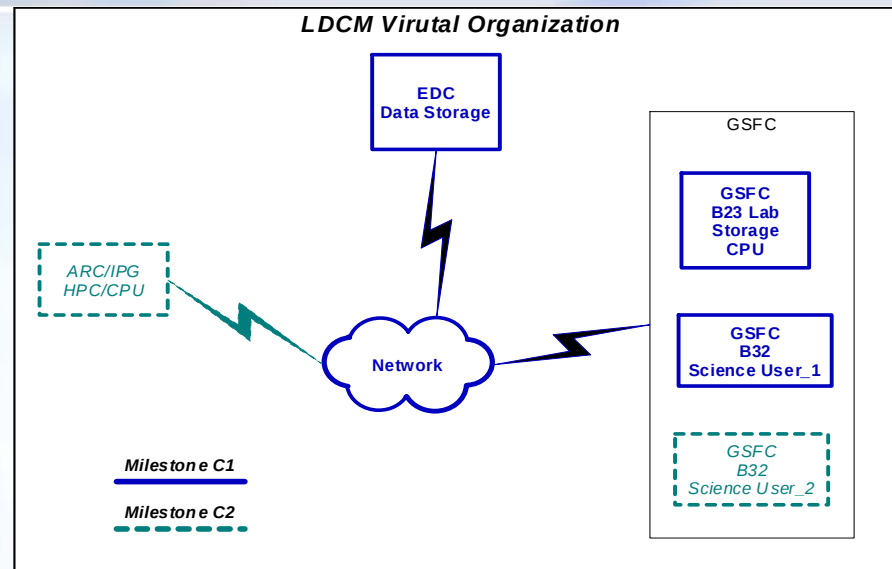
The Advanced Data Grid (LDCM) Prototype

Sponsored by the Landsat Data Continuity Mission and the ISD (580)

POC: Jeff Lubelczyk, 586
Gail McConaughy, 586
Beth Weinstein, 586

Description and Objectives

- The objective of the LDCM ADG prototype is to assess the applicability and effectiveness of a data grid to serve as the infrastructure for research scientists to generate virtual Landsat-like data products.
- Grid technology serves as a key enabler in the creation of scientific Virtual Organizations, promotes a flexible and scalability infrastructure, facilitates the exchange of data, and maximizes the use of available resources



Approach

Phase 1: Provide and demonstrate a basic grid infrastructure that enables a simple data fusion algorithm to access remote heterogeneous instrument data at multiple GSFC labs and EDC.

Phase 2: Enable the data fusion algorithm to obtain datasets, execute, and store the results on any resource within the Virtual Organization (GSFC labs, EDC, ARC IPG).

Co-I's/Partners

EDC, NASA ARC/IPG, GSFC 920 Scientists

Science Themes

Virtual scientific data products
Remote instrument data access
Collaborative computing for the science community
Resource sharing and data discovery

Schedule and Deliverables

- **Prototype start (12/03)**
- **Demo of Phase 1 grid infrastructure (6/04)**
- **Demo of Phase 1 capability (12/04)**
- **Demo of Phase 2 grid infrastructure (3/05)**
- **Demo of Phase 2 capability (6/05)**

Application/Mission

Allow scientists at resource-poor sites access to remote resource-rich sites, enabling greater scientific research. Serve as a key enabler in the creation of scientific Virtual Organizations and by extension, facilities. Maximize utility of existing resources, limiting the expense of building new facilities.



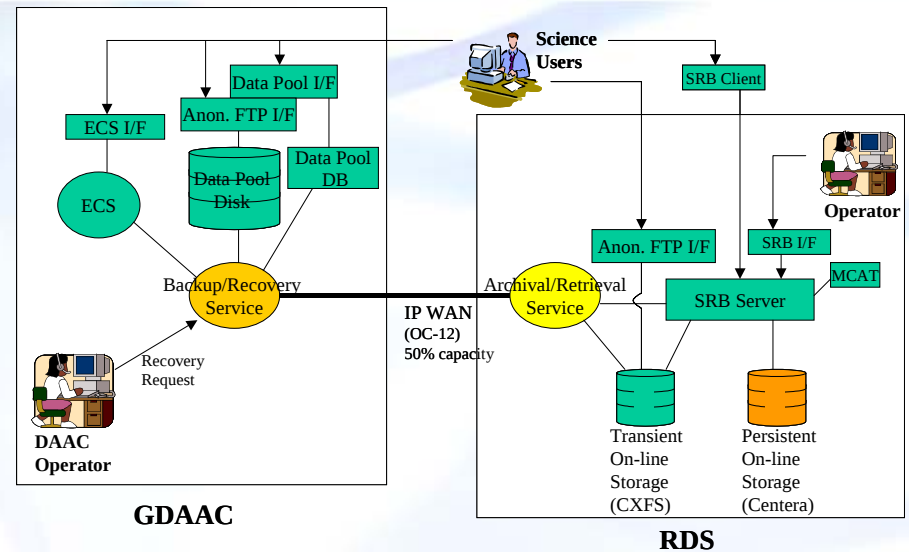


Remote Data Storage (RDS) Prototype

POC: Chris Bock, ESDIS

Description and Objectives

- Demonstrate high volume remote data backup and recovery capability over an IP WAN for portions of the Goddard Earth Sciences DAAC (GDAAC) data holdings
- Provide heterogeneous storage management functionality capable of managing a storage hierarchy including transient and persistent on-line storage resources via a uniform interface
- Implement a preliminary data grid infrastructure making the data holdings accessible to external "data grid" users
- Assess enabling technologies in the context of NASA Earth Science mission needs



Co-I's/Partners

Goddard DAAC
NASA IV&V Facility, Fairmont, WV

Approach

Develop a testbed to prototype the use of advanced technology to provide a back up and restore capability for non-reproducible EOSDIS data.

RDS prototypes 1 through 3 will culminate in an operational prototype system to be deployed at the NASA IV&V Facility and the GDAAC in early 2004

Schedule and Deliverables

- Release 1 prototype (SGI Server & RAID) installed at NASA IV&V Facility in Fairmont, WV in 12/02, receiving MODIS direct broadcast data daily
- Release 2 prototype (integrated SGI server, RAID, CXFS software, EMC Centera storage, Nirvana SRB software) successfully demonstrated at EMC laboratory in Columbia, MD in June 2003.
- High speed OC-12 link between NASA IV&V Facility and GDAAC operational in October 2003
- Delivery of Release 3 prototype to NASA IV&V Facility and GDAAC in January 2004

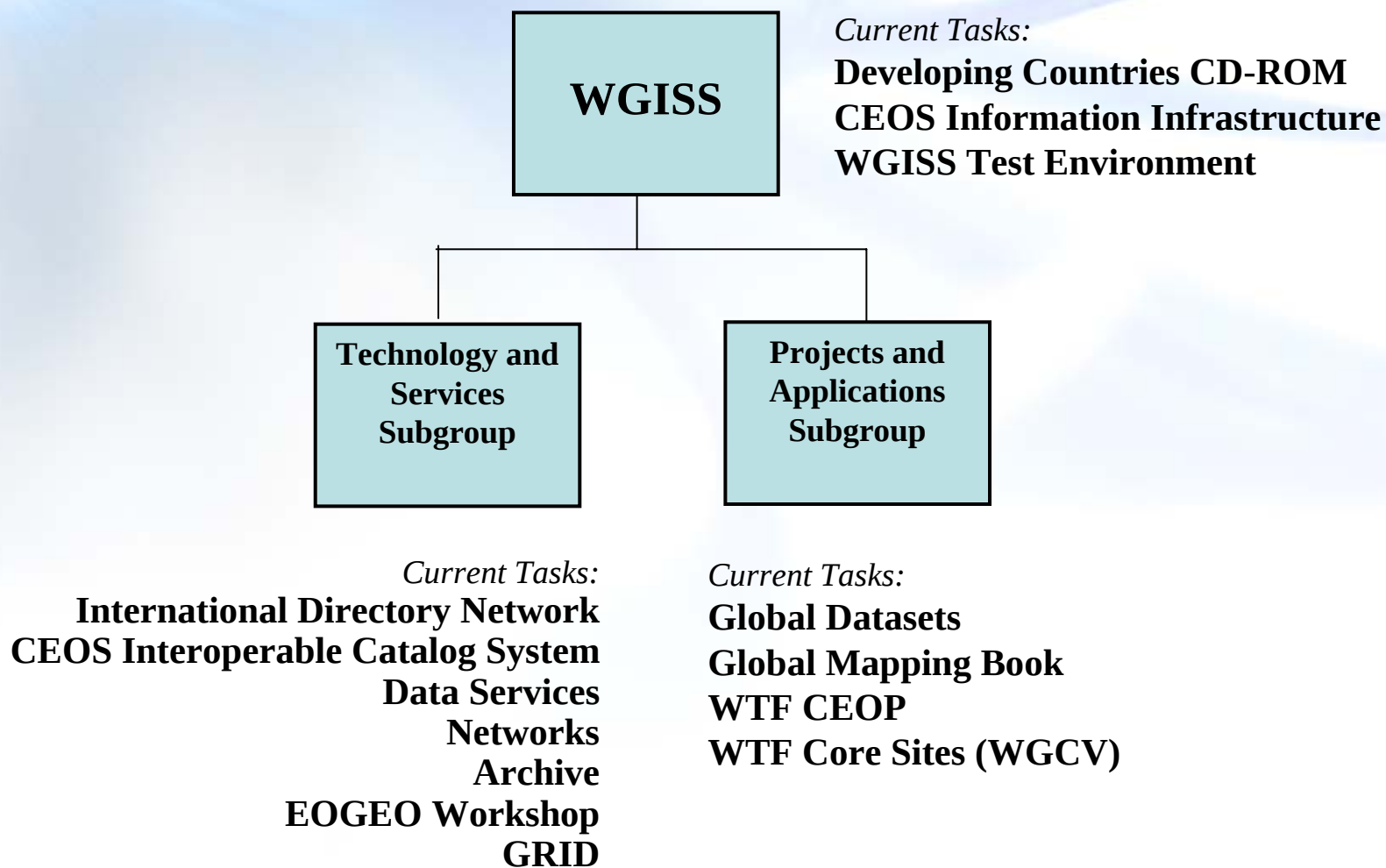


CEOS WGISS

- Committee on Earth Observation Satellites (CEOS)
 - Members are international agencies that operate Earth observing satellites and affiliated science organizations.
 - Purpose is to promote cooperation the acquisition, exchange and utilization of Earth observation data.
- CEOS Working Group on Information Systems and Services (WGISS).
 - One of several working groups established to focus on particular CEOS areas of interest.
 - Provides an effective forum CEOS partners cooperate in applying advanced, data system technology to meet CEOS goals and objectives.



WGISS - Working Group Structure





CEOS Grid Task Team

■ Background

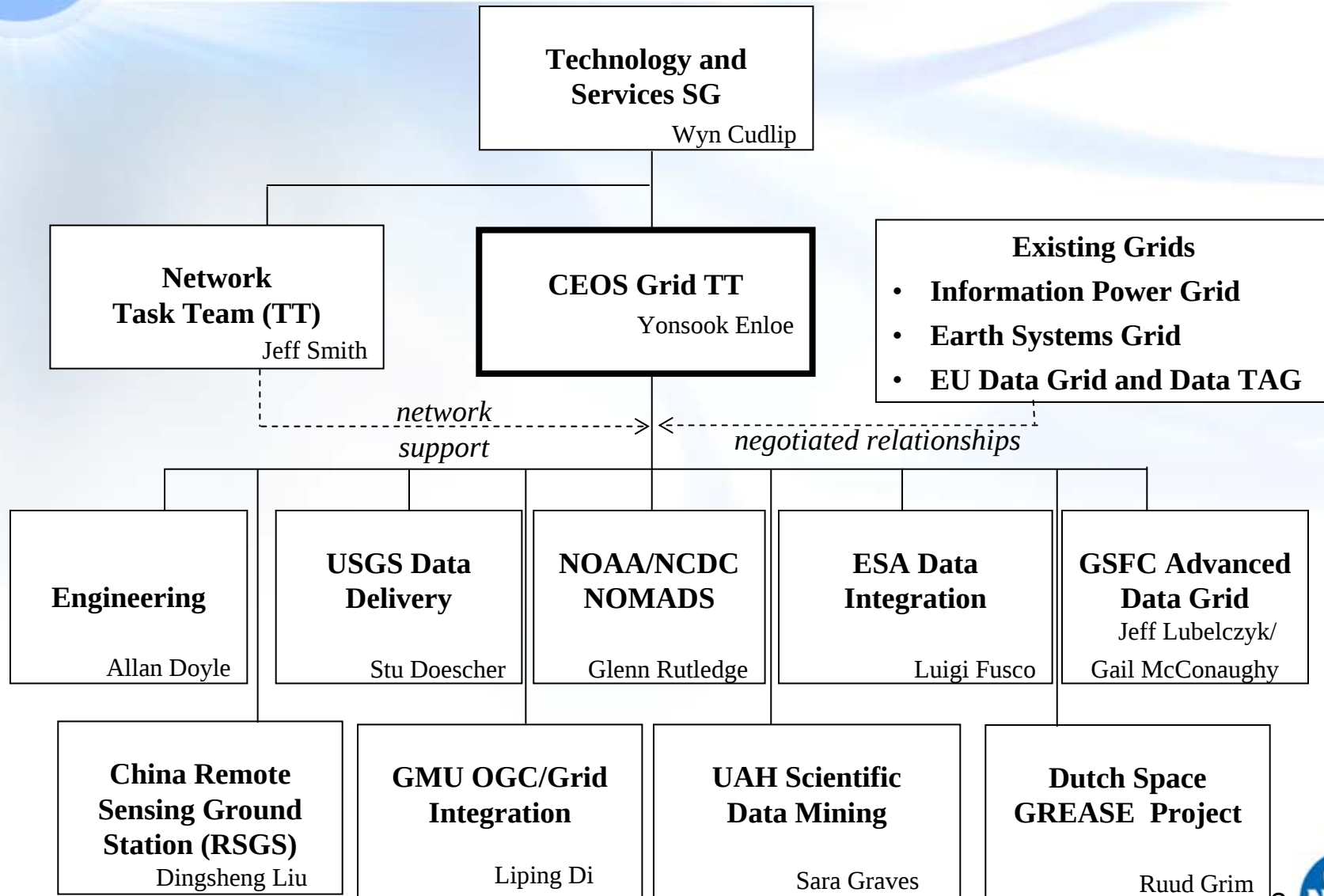
- Initiated about two years ago.
- Grid workshop held in conjunction with regular WGISS meeting.
 - Invited Grid experts interacted with WGISS members.
 - Multiple WGISS agencies initiated the development of Grid technology prototypes.
 - WGISS also interested in potential for incorporating Grid technology into its initiatives.

■ Approach

- Form a task team of interested WGISS members.
- Share lessons among prototype teams.
- Provide engineering support to teams.
- Set up a CEOS Grid Testbed
- Interact with broader Grid technology community.
- Report back to full WGISS membership.



CEOS Grid Task Team Coordination





CEOS Grid Task Team

- Accomplishments - Focus on implementing infrastructure
 - Teams installed Globus Toolkit, CEOS Grid common baselined components
 - Prototyped monitoring tool (to all computers on CEOS Grid)
 - Prototyping resource directory (CEOS Grid computers, applications and processes, CPU processing, data collections)
 - Published user documentation for participants
 - CEOS Grid Implementers Guide
 - Firewall Document
 - Using CEOS version of IPG certificates for security
 - Two way interaction with Open Source community
 - Globus Metadata Catalog Service implementation of spatial and temporal search.
- Plans
 - Facilitate the development of agency applications.
 - Plan multi-agency demonstration(s).
- Additional Information
 - Public Task Team Web site <http://lennier.gsfc.nasa.gov/grid>





NASA Goddard IT Pathfinder Working Group

- Purpose: revitalize Goddard's Information Technology infrastructure and prepare for new paradigms in computing and information systems
 - Center-wide participation from IT professionals
 - Atmospheric Modeling & Geosciences
 - High Performance Computing / Mass Storage
 - Earth Science Data System WGs & EOSDIS
- Activity to Date
 - Kickoff June 03
 - Goddard new Geoscience Network (GEON) collaborator
 - Goddard IRAD: establish GSFC-Scripps Lambda Network (via NLR) & demonstrate (e.g. ESMF/atmospheric modeling application)
 - OptIPuter workshop Jan 04
- On-going Analysis Activities
 - Goddard Semantic Web Interest Group (science applications)
 - "Data web" for Earth Science data analysis
 - NSF cyber-infrastructure paradigm
 - Grid computing (Goddard, Ames, JPL, and Langley)