GSFC Cyberinfrastructure Initiative

Enabling Transparent Collaborative Science Through Lambda Optical Networks

Status Report of the ITPWG

by

M. Halem*

* Distinguished Information Scientist, Emeritus

Background

- Motivation and Inspiration (L. Smarr)
- Earth System Science Drivers (Data, Modeling)
- Objectives of Goddard Cyberinfrastructure Initiative
- Enabling Optical lambda Networks
- GSFC/SIO Science Applications
- Summary

Larry Smarr: " Why optical networks are emerging as 21st century IT Driver" …5/21/03 @ GSFC

- **Computer Chips** (Number of Transistors- 70 nm.- Intel) (Doubling time 18 Months)
- **Data Storage** (bits per sq. in. - 100's of Gb/sq. in.- IBM) (Doubling time 12 Months)
- **Optical Fiber** (bits per second- 10Gb/s/w-Telecoms) (Doubling time 9 Months)

NASA High End Networks

Motivation

- NASA has fallen significantly behind the state of the art in advanced networks as indicated in figure 1

- GSFC and Cal(IT2) initiated a collaborative prototype to explore feasibility of multi wave optical networks

- With the introduction of NASA's newest supercomputer system(Project Columbia at Ames) the lack of bandwidth is a significant barrier to collaboration and data sharing-(e.g. a 2TB per day data set cannot be effectively transferred between research teams

- The agency is addressing this problem for all the centers partialy through Project Columbia and Agency/ Center initiatives
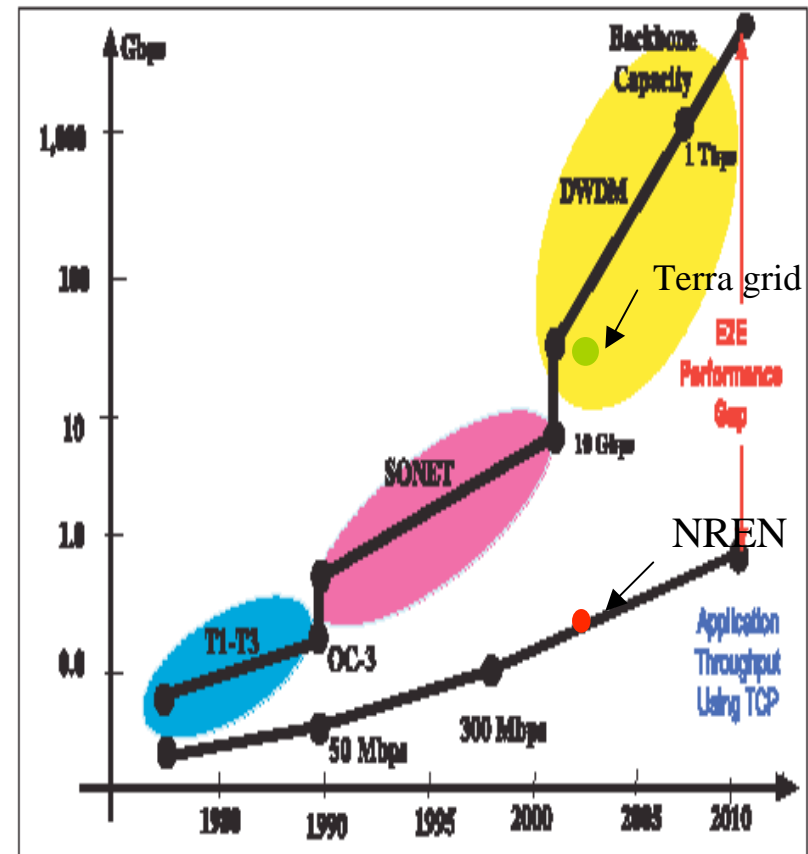
NLR – Proposed Southern Route
Wiltel Fiber in the West -- AT&T Fiber in the Southeast
Qwest Fiber North-South
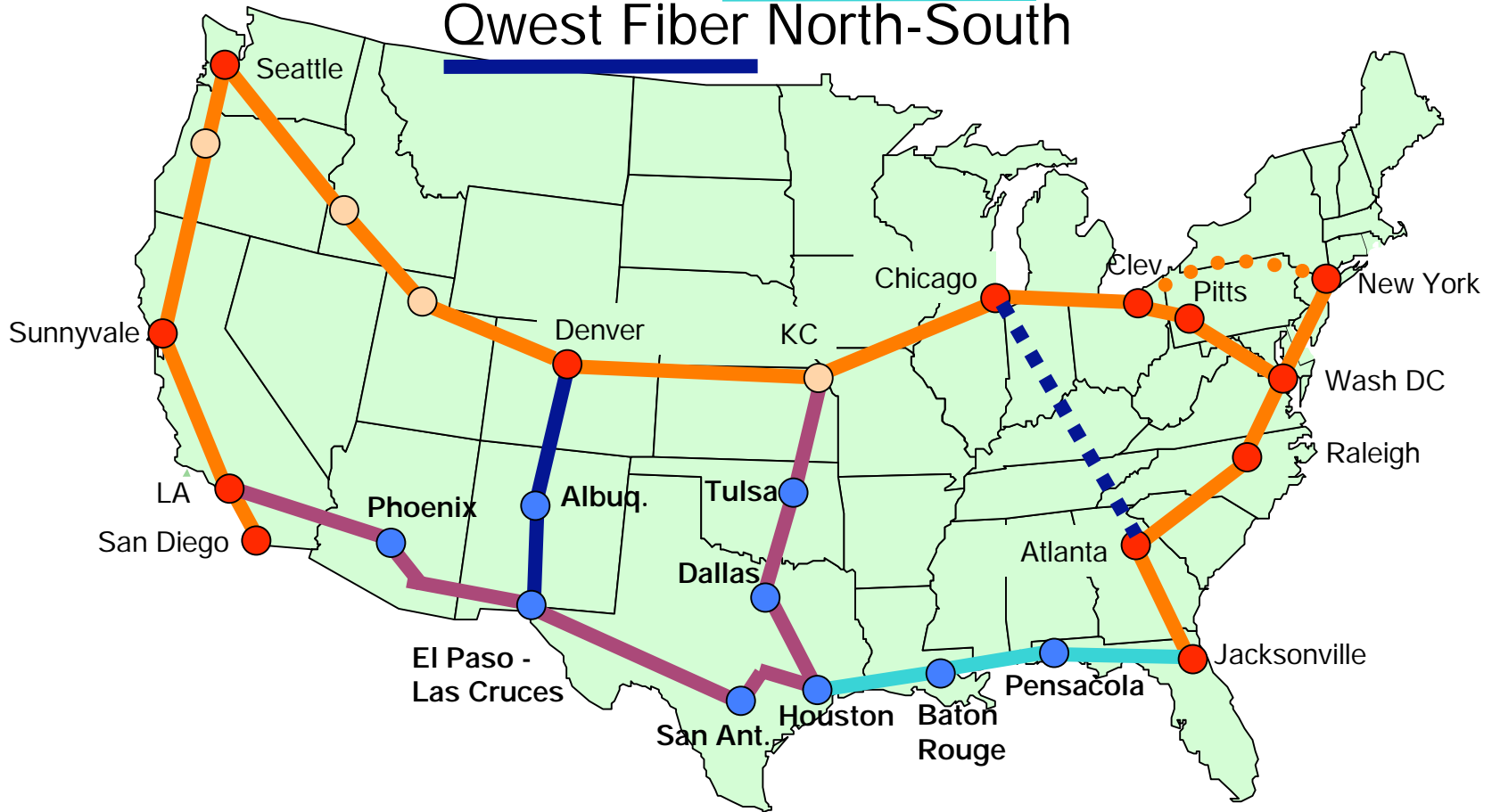
Using Cisco 15454 DWDM gear

5

NLR Members as of 19Jul04

- CENIC
- Pacific Northwest GigaPOP
- Pittsburgh Supercomp. Center
- Duke (coalition of NC univers.)
- Mid-Atlantic Terascale Partnership
- Cisco Systems
- Internet2
- Florida LambdaRail
- Georgia Institute of Technology
- Committee on Institutional Cooperation (CIC)

- Texas / LEARN
- Cornell
- Louisiana Board of Regents
- University of New Mexico
- Oklahoma State Regents
- UCAR/FRGP

Plus Agreements with:

- SURA (AT&T fiber donation)
- Oak Ridge National Lab (ORNL)

NASA/ARC Project Columbia 'Space Exploration Simulator'

Addressing NASA shortfalls in supercomputing and bandwidth

- 20 SGI (512 CPU) Altix systems
- (2) 2048 CPU single image processors
- Serves the needs of the Agency
- Establish 10 Ge connections between centers to ARC 'SES' via NLR
- Outlining framework to support NSF & DOE users on 2048 system
- Available on floor today.

GSFC IRAD Proposal "Preparing Goddard for Large Scale Team Science in the 21st Century: Enabling an All Optical Goddard Network Cyberinfrastructure"

Objectives Summary

- "…establish a 10 Gbps Lambda Network from GSFC's Earth Science Greenbelt facility in MD to the Scripps Institute of Oceanography (SIO) and ARC over the National Lambda Rail (NLR)"

- "…make data residing on Goddard's high speed computer disks available to SIO with access speeds as if the data were on their own desktop servers or PC's."

- "…enable scientists at both institutions to share and use compute intensive community models, complex data base mining and multi-dimensional streaming visualization over this highly distributed, virtual working environment."

GSFC IRAD Proposal "Preparing Goddard for Large Scale Team Science in the 21st Century: Enabling an All Optical Goddard Network Cyberinfrastructure"

Benefits

- Creates a virtual laboratory and an SIO wing within GSFC's Building 33
- Develop real-time interactive collaborations with other leading Earth and space science academic institutions
- Enable scientists at both institutions to share and use compute intensive community models, complex data base mining and multi-dimensional streaming visualization
- Supports several important NASA missions and programs that can benefit significantly from the high bandwidth connection
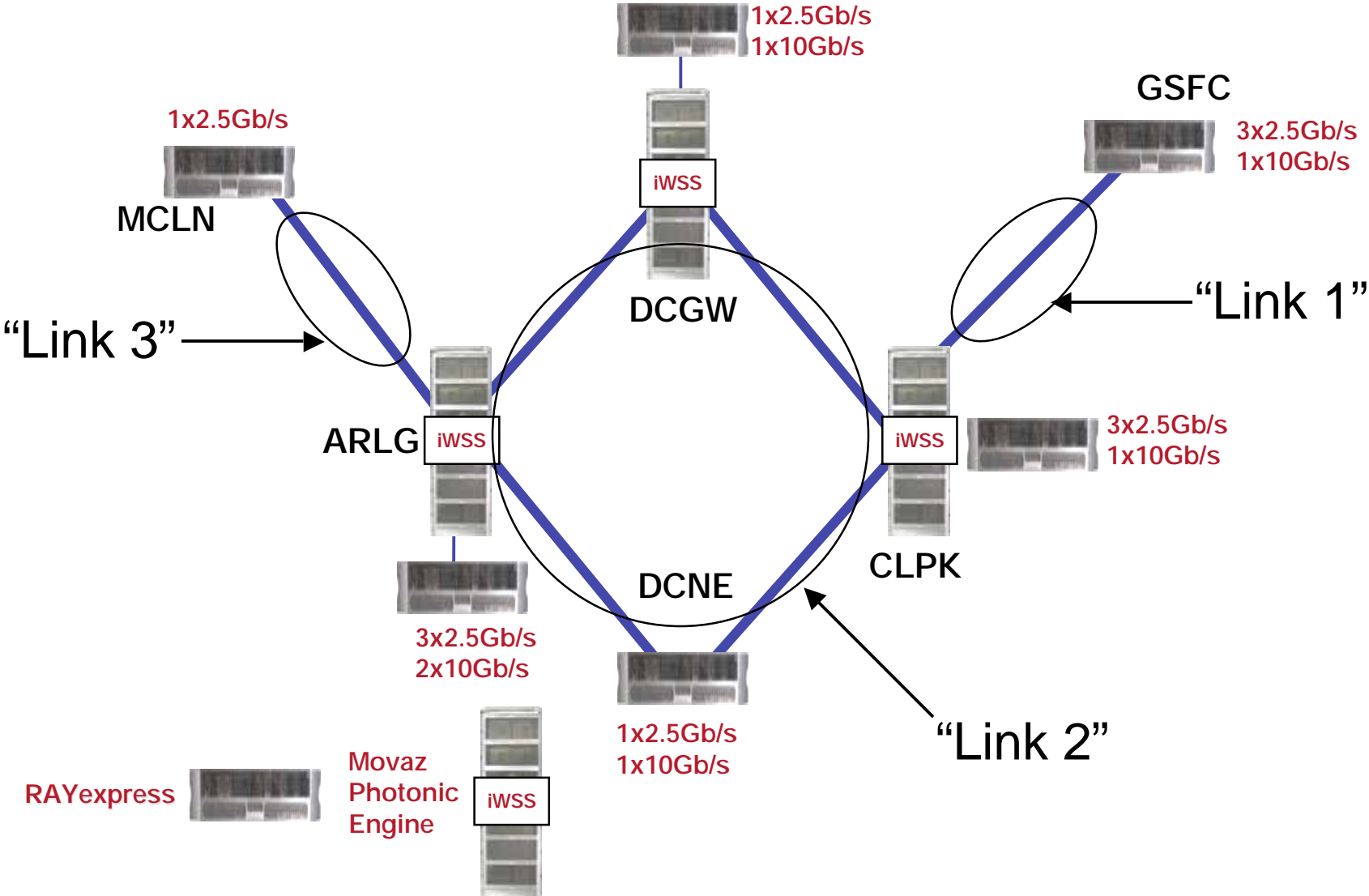
Synergies

- Takes advantage of next generation networking technologies (Lambda-Nets)
- Makes use of NSF funded compute, storage and visualization resource being implemented at UCSD and Cal-IT2 (OptiPuter and GEON)
- Falls in line with reinvigorating the nation's cyber infrastructure

GSFC-SIO Cyberinfrastructure Project Status

- Will be connected with 10Ge by end of Sept or early Oct. to SIO and ARC over the NRL from Goddard
- Introducing new IT capabilities between GSFC, SIO and ARC (e.g. optical network protocols, SAN's, hyperwall visualizations, compute/data grid software)
- Identified and supporting several joint research applications as prototypes to stress these resources
- Using the NLR coastal hubs (N.Y. and Seattle) to provide access to international science data archives
- Implementing an internal GSFC Optical Network to enable 10 Ge inter building science and engineering on the Goddard campus

Dynamic Resource Allocation with GMPLS on Optical Networks (DRAGON) Configuration

1x2.5Gb/s
1x10Gb/s

GSFC

1x2.5Gb/s

3x2.5Gb/s
1x10Gb/s

MCLN

iWSS

DCGW

"Link 1"

"Link 3"

ARLG iWSS

iWSS

3x2.5Gb/s
1x10Gb/s

CLPK

DCNE

3x2.5Gb/s
2x10Gb/s

RAYexpress Movaz
Photonic
Engine iWSS

1x2.5Gb/s
1x10Gb/s

"Link 2"

GSFC IRAD Proposal "Preparing Goddard for LargeScale Team Science in the 21st Century: Enabling an All Optical Goddard Network Cyberinfrastructure"

Key Features of GSFC L-Net Design (continued)

- Regional Network Part

    – Two 10-GE connections with DRAGON at GSFC in Greenbelt

    – Two 10-GE connections with DRAGON at Level3 POP in McLean

    – Two 10-GE and multiple 1-GE connections for network test stations at Level3 POP in McLean

- Transcontinental Network Part

    – 10-GE connection with NLR/MATP's IP Backbone and Switched Ethernet lambdas

    – 10-GE connection with NLR/Internet2's HOPI lambda

High Performance Networking and Remote Data Access
GSFC L-Net for NCCS and Science Buildings

Legend
- Dark Fiber
- 10 Gbps GE
- 2 Gbps FC
- 1 Gbps GE

GSFC at Greenbelt

ARC/ Project Columbia

ORNL

GISS

Level3 POP at McLean

DRAGON

NLR

CMU
JPL
SIO
UCSD
UIC

GWU
ISI-E
UMBC
UMCP
ATDnet
BosSNet
MIT/ Haystack

High Performance Remote Data Cache Facility (creating Inter-Facility virtual SANs using SAN-over-IP technologies)

FC

iFCP Gateway
FCIP Gateway
iSCSI Gateway

Op. Sw/ OADM

B G P

F W

10-GE Sw/Rtr w/OSPF

NCCS "Classic"

10-GE SW/RTR

CPU(s)
CPU(s)
CPU(s)
CPU(s)

F C

VIS

Other GSFC Science Data Facilities

10-GE SW/RTR

VIS VIS
GE Sw/Rtr

VIS VIS
GE Sw/Rtr

VIS VIS
GE Sw/Rtr

VIS
GE Sw/Rtr

GE Sw/Rtr VIS

C P U (s)
C P U (s)
C P U (s)
C P U (s)
C P U (s)
C P U (s)
C P U (s)
C P U (s)
C P U (s)
C P U (s)

FC
FC
FC
FC
FC

JPG 8/05/04

GSFC IRAD Prototype Applications

Examples of Initial Primary Users/Applications (1 of 2)

- Drs. Paul Houser and Mike Bosilovich of Code 970 are collaborating with Dr. John Roads of SIO on the **Coordinated Earth Observing Program (CEOP)**
- Dr. Y. Kaufman and Dr W. Lau .of Code 910 are collaborating with Dr. V. Ramanathan of SIO on the **Project ABC** (Atmospheric Brown Clouds)

- SIO's Prof. Richard Sommerville is introducing his **Cloud-Radiation** parameterization as part Goddard development of the Earth System Modeling Framework for the GMAO Seasonal Climate Model
- The **EOSDIS Clearing House (ECHO)** metadata gateway is developing a web brokering and chaining service for providers and clients
- Dr. H. Mitchell (SVS Code 930) is developing **hyperwall streaming visualizations** to enable scientists to analyze model output fields in their offices collaboration with SIO/UIC and Ames
- Developing SETI like **Century climate ensemble simulations** using the GISS Model E of J. Hansen running on grid Apple G5's and other platforms at GSFC and partnering universities
- Dr. J.Centrella et. al., will conduct large scale numerical relativity **simulations of gravitational wave sources for LISA** on supercomputers at Project Columbia and other national sites over the NLR

GSFC IRAD Prototype Applications

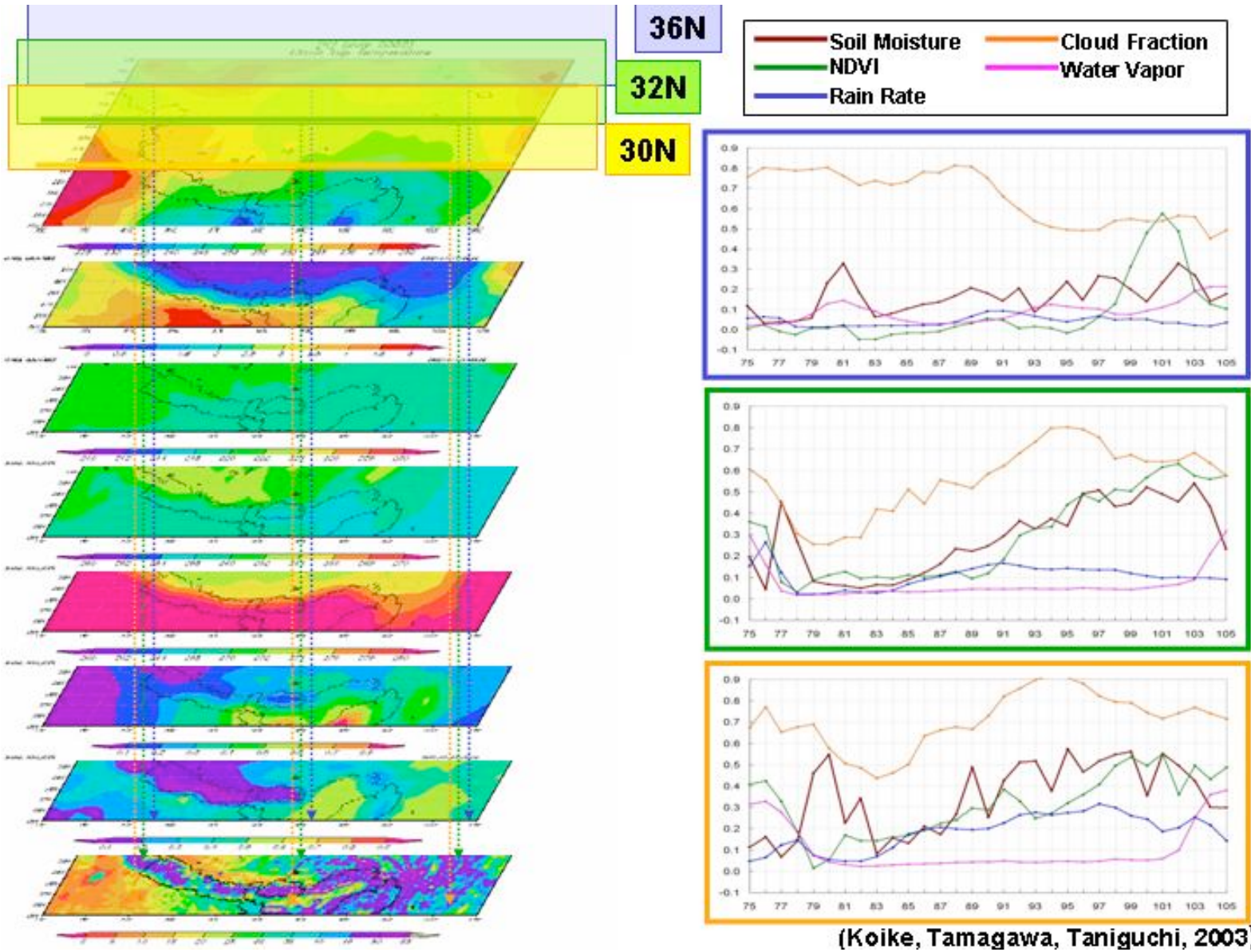*Examples of Initial Primary Users/Applications (1 of 2)*

- Dr. Michele Rienecker of Code 900.3 is collaborating with Dr. Tim Barnett of SIO on the **assimilation of global sea height data** from TOPEX and GRACE

- UCSD's Geosciences Network PI Dr. Dogan Seber has identified some of **GSFC's solid earth research data sets and models** for developing collaborative research efforts with Dr. Weijia Kuang and others from Code 920

- Dr. J. Roads (SIO) with Dr. M. Suarez of Code 900.3, Mike Seablom of Code 560, are planning to run **interactive distributed regional model forecasts** using boundary forcing conditions from the Global Modeling and Assimilation Office (GMAO)

- Dr. J. Herman of Code 910 is the Co–I with Dr. Francisco Valero of SIO who is the PI on the **Triana** mission

An Example of Application Requiring L-NET

- Drs. Paul Houser and Mike Bosilovich of Code 970 are collaborating with Dr. John Roads of SIO on the **Coordinated Earth Observing Program (CEOP)** accessing 300TB's of Observational Data in Tokyo and 100TB's of Model Assimilation Data in MPI in Hamburg Germany and analyzing remote data using GRaD-DODS at these sites over the NLR and Starlight.

36N

32N

30N

Soil Moisture   Cloud Fraction
NDVI   Water Vapor
Rain Rate

(Koike, Tamagawa, Taniguchi, 2003)

17

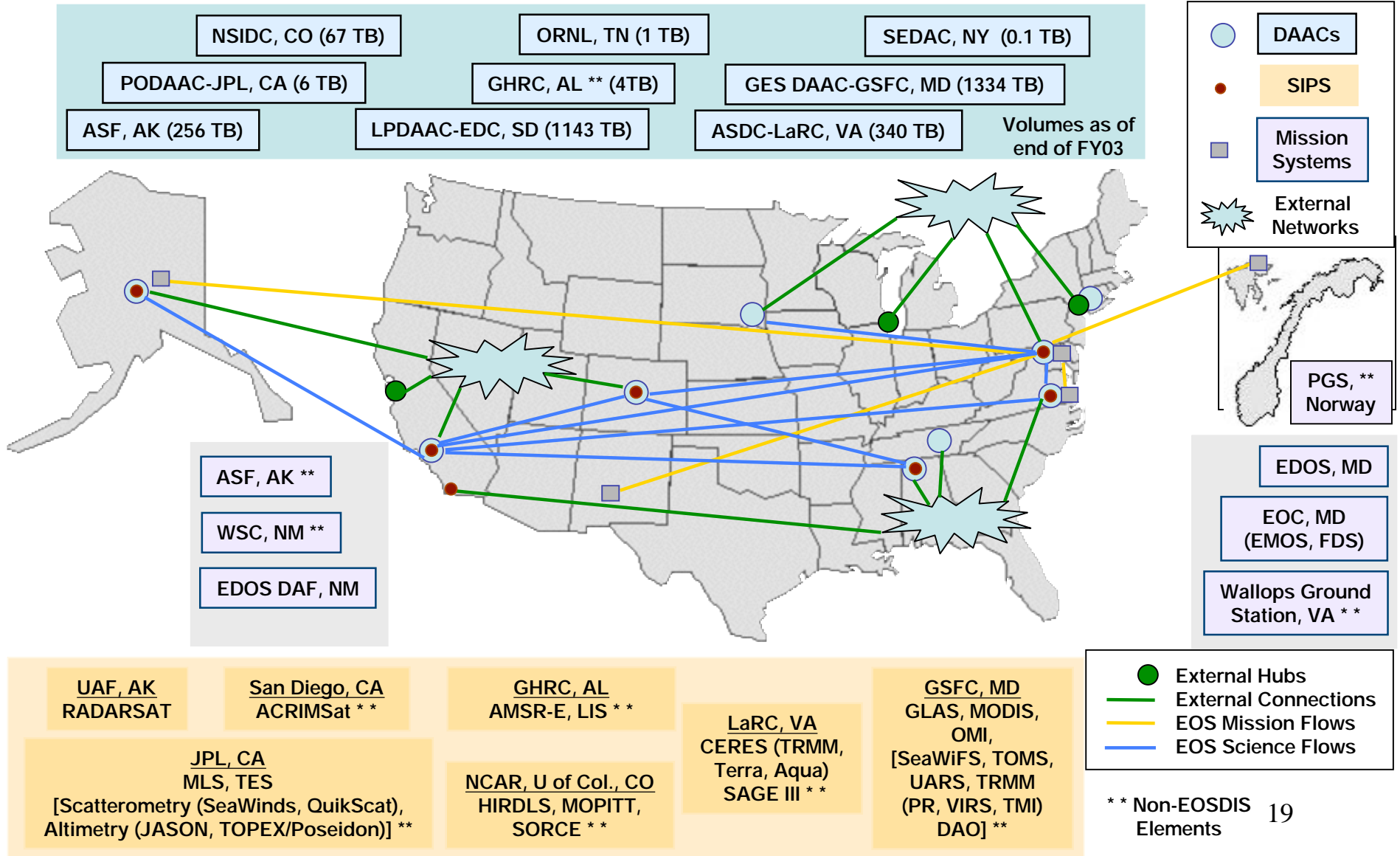# GEOS3 CEOP In-situ Validation

(Tamagawa, Koike, Williams, 2003)
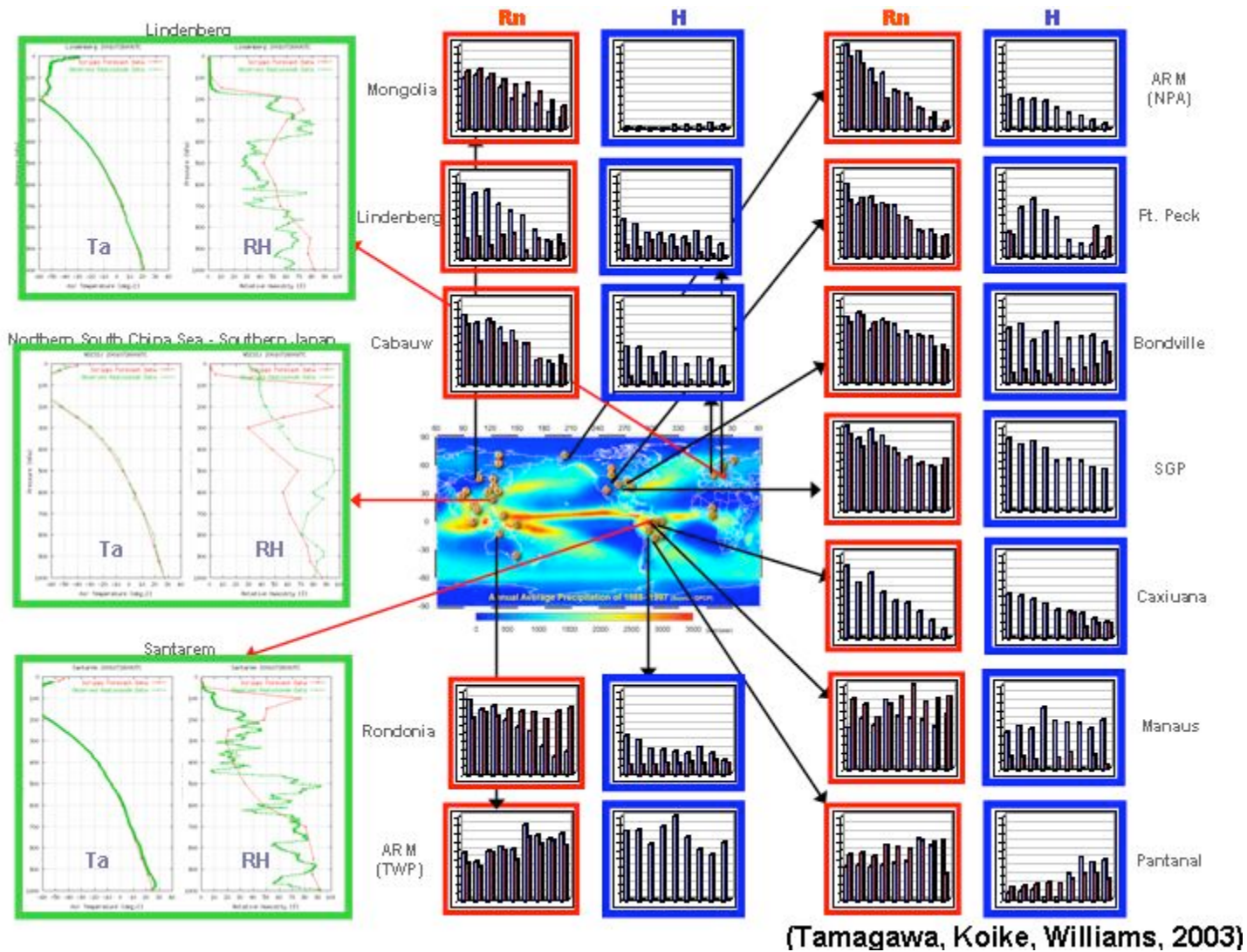
- CEOP is collecting and coordinating global insitu, remotely sensed and model analysis data

Background and Context

Geographic Distribution of EOSDIS and Interfacing ESE Elements

NSIDC, CO (67 TB)

ORNL, TN (1 TB)

SEDAC, NY (0.1 TB)

PODAAC-JPL, CA (6 TB)

GHRC, AL ** (4TB)

GES DAAC-GSFC, MD (1334 TB)

ASF, AK (256 TB)

LPDAAC-EDC, SD (1143 TB)

ASDC-LaRC, VA (340 TB)

Volumes as of end of FY03

DAACs

SIPS

Mission Systems

External Networks

PGS, ** Norway

ASF, AK **

WSC, NM **

EDOS DAF, NM

EDOS, MD

EOC, MD (EMOS, FDS)

Wallops Ground Station, VA * *

External Hubs
External Connections
EOS Mission Flows
EOS Science Flows

UAF, AK
RADARSAT

San Diego, CA
ACRIMSat * *

GHRC, AL
AMSR-E, LIS * *

LaRC, VA
CERES (TRMM, Terra, Aqua)
SAGE III * *

GSFC, MD
GLAS, MODIS, OMI,
[SeaWiFS, TOMS,
UARS, TRMM
(PR, VIRS, TMI)
DAO] **

JPL, CA
MLS, TES
[Scatterometry (SeaWinds, QuikScat),
Altimetry (JASON, TOPEX/Poseidon)] **

NCAR, U of Col., CO
HIRDLS, MOPITT,
SORCE * *

* * Non-EOSDIS Elements

19

System Overview
About ECHO
http://eos.nasa.gov/echo
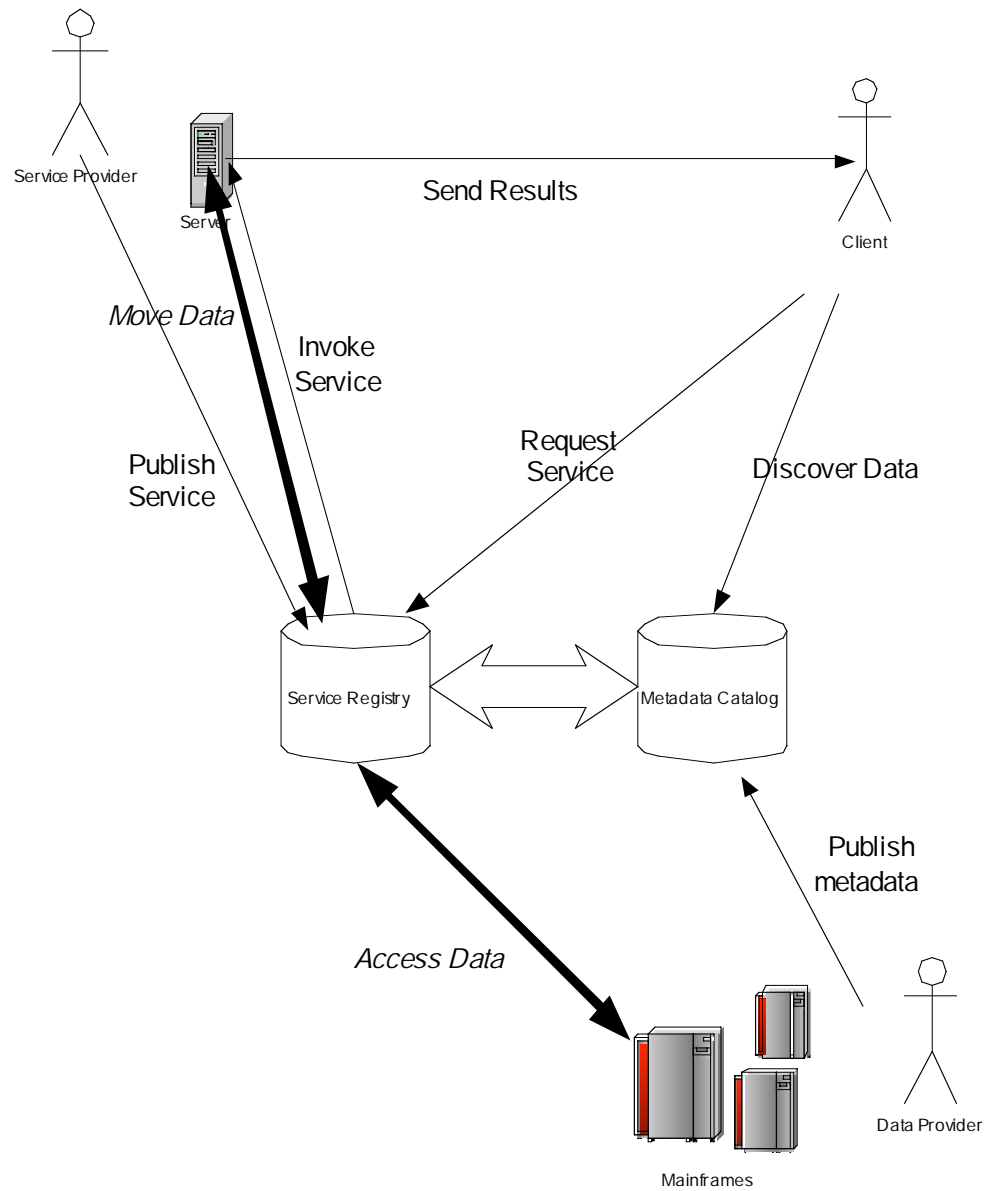
- Metadata Clearinghouse (Registry)
  - ECHO makes the data holdings of the Earth Science Enterprise available to clients as if they had a copy of all the metadata in their system

- Order Broker
  - ECHO provides a single interface for clients to place orders through and deals with the complexity of ordering from its partner systems on behalf of the client

- User Account Management
  - While not a primary feature of the system, ECHO provides flexible user account management for both registered users and providers in support of its registry and brokering roles

- Service Registry (6.0 feature – in testing now)
  - ECHO provides a mechanism for clients to find out what Earth Science services are available dynamically. ECHO provides the mechanism for services to express their relationship to Earth Science data

- **Service Broker** (proposed as Cyberinfrastructure prototype)
  - ECHO will allow a client to order data from a data provider and then have a series of services applied to it by relying on ECHO to do the coordination
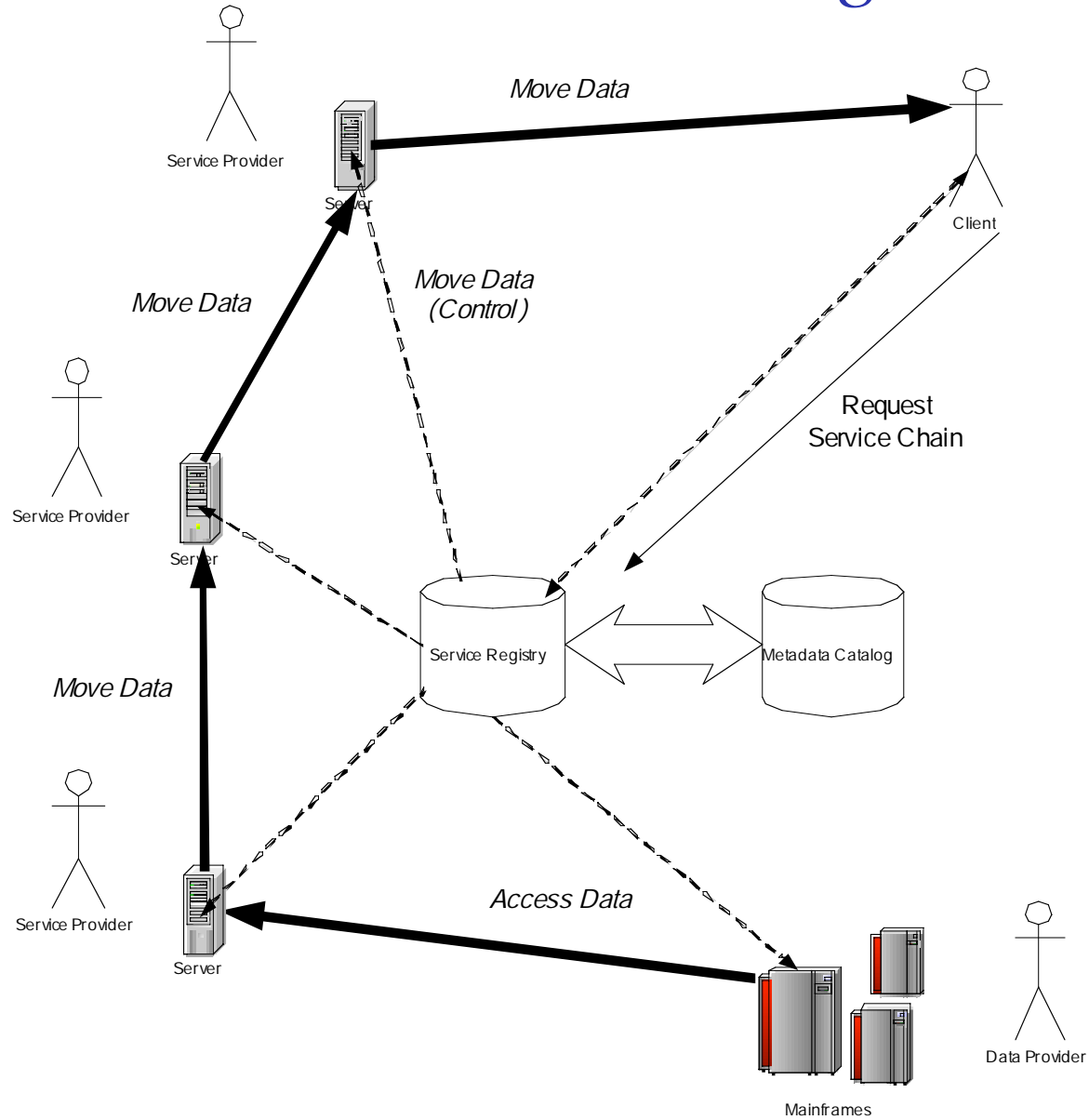
Services in ECHO
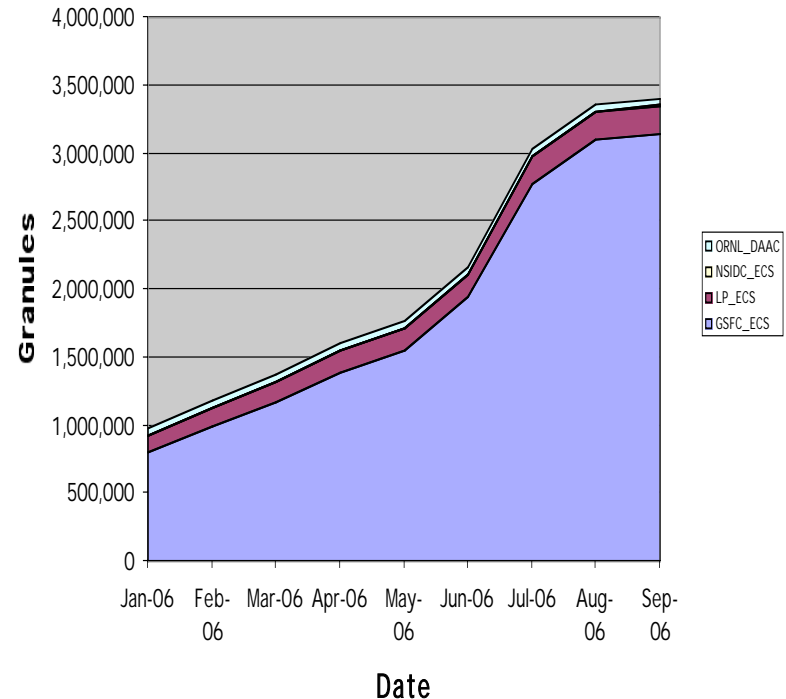## *Service Brokering*

Service Provider

Server

Send Results

Client

Move Data

Invoke
Service

Request
Service

Discover Data

Publish
Service

Service Registry

Metadata Catalog

Publish
metadata

Access Data

Mainframes

Data Provider

Services in ECHO
Service Chaining

Service Provider

Server

Move Data

Client

Move Data

Move Data (Control)

Move Data

Request
Service Chain

Service Provider

Server

Service Registry

Metadata Catalog

Move Data

Service Provider

Server

Access Data

Mainframes

Data Provider

22

ECHO Data Providers

- Current ECHO Data Providers
  - EOSDIS Core System DAACs (17% thus far)
    - EDC Land Processes DAAC
    - Goddard DAAC
    - NSIDC DAAC
  - ORNL DAAC (100%)
- ECHO holds the metadata for over 3 million granules, and growing

ECHO Metadata Holdings

Legend:
ORNL_DAAC
NSIDC_ECS
LP_ECS
GSFC_ECS

Y-axis: Granules (0 to 4,000,000)
X-axis: Date (Jan-06 through Sep-06)

23

Benefits from ECHO

Overall

- Interoperable functions allows everyone to benefit from distributed development of functions
- Decentralizes development of end user functionality. Externally developed functions can interoperate through ECHO.
- Can help move away from ECS's complex infrastructure and its related high maintenance costs.

For Providers

- Reduces machine workload from ESDIS system by removing search loads.
- Allows for community provided services and increased functionality.

For Users

- Supports more flexible data access paradigms.
- Can search and find data regardless of provider down time. If the provider is down when the user submits a requests a service, ECHO will continue to attempt work arounds on behalf of a user.
- Allows users to build their own search UIs.
- Users benefit from a broader range of interoperable functions than can be provided by ESDIS development team alone.

Summary

- NASA putting major IT assets (i.e., Supercomputers, massive data sets, visualizations, models, web services) on the NLR to improve accessibility

- Setting up 'One NASA' through internal collaborations among NASA centers based on the use of optical networks and external collaborations through joint virtual centers (i.e., SIO/UCSD.CEAS,CIRA/ Colo. St.,, JCES/UMD, etc.)

- Creating opportunities for internal and external scientific community for discovery and knowledge exploration of Sun-Earth Systems, Solar Systems and the Universe through the leveraging of next generation Optical Network Technologies