GSAW 2005 Breakout Session 9E :
Components, Frameworks, & Web/Grid Services for
Ground Systems

Shirley Tseng
Infinite Global Infrastructures
www.igillc.com
Stseng@igillc.com
714-832-5373, 949-300-1899 cell
Survey of space projects using Grid & Web Services components

• Who is doing what? Where? Contact information
  – Project implementing Grid component layers
  – Projects implementing Web Services & SOA (WSDL, UDDI, SOAP)
  – XML in Space definition initiatives
  – Layer view to map projects to

• A few representative charts from the Projects

• Observations from Survey
  – Early US initiatives - NASA IPG, DOE
  – Concentrated European initiatives – SpaceGrid study (2001-2003), buyin by European management and companies, projects initializations
  – US Grid Rollout (driven by science data users/organizations)
    ■ driven by users close to the science data – Principle Investigators using satellite data
    Ex: NVO, CEOS, OGC
  – NASA ESTO SEEDS (Strategy for Evolution of ESE Data Systems study 1998)
    ■ ESDSWG (Earth Science Data Systems Working Groups)
  – Mixed pockets of implementations – OODT, MERS CIP
NASA’S Information Power Grid (IPG)

NASA Grid Development

Application and User Oriented Services
- Portal Services
- Workflow Specification
- Query Structuring

Execution Management Services
- Workflow Management
- Resource Broker
- Naturalization Services
- Remote Execution
- Task Graph Execution

Data Management Services
- Replica Management
- Meta-data Catalogues
- Data Movement

Grid Management Services
- Distributed Information Management
- Dynamic User Access Management
- Event Management
- Test & Monitoring Service
- Information Service
- Security Infrastructure

Compute & Data Resources

ARC in-house R & D Efforts
ARC partnership with external organizations
External Technologies
The IPG, funded by the Computing, Information and Communications Technology (CICT) program at NASA Ames Research Center, is a collaborative effort between NASA Ames, NASA Glenn, and NASA Langley Research Centers, and the NSF PACI programs at SDSC and NCSA.

IPG Applications:

- Remote Access to High Data-Rate Instruments
  - The DARWIN system at Ames has users scattered across the country
  - UCSD TeleScience system has a NASA user at Wallops manipulating the instrument at UCSD.

- Management and access to massive data sets

- The IPG Data Mining application:
  - Persistent and uniform access to heterogeneous, multi-organizational archival storage systems
  - The SDSC Storage Resource Broker (SRB - an IPG Grid service) provides a standard data access interface for heterogeneous data archive systems
  - SRB’s MCAT is a catalogue service that provides a standard way to define, manage, and search metadata for all files in a collection, where a collection may span many data archive systems

- Aviation Safety (Glenn on distributed simulation)
# Remote Access to High Data-Rate Instruments

## Users

- **Web user interface**
- **Tele-Science Portal**
- **UC San Diego**
  - authentication
  - compute resource access
  - user data access
- **NASA Ames**
  - authentication
  - user data access
- **Web user interface**
  - DARWIN/DREAM data server / portal
  - instrument data storage

## Grid Services: Uniform access to distributed resources

- **Grid Information Service**
- **Uniform Resource Access**
- **Brokering**
- **Global Queuing**
- **Global Event Services**
- **Collaboration Instrument Services**
- **Uniform Data Access**
- **Cataloguing**
- **Global Event Services**
- **Security Services**
  - Authentication
  - Authorization
  - Auditing
  - Monitoring

## Security Services

- **Fault Management**
- **Communication Services**
- **Network Cache**

## UC San Diego

- National Center for Microscopy and Imaging Research

## SDSC compute and data resources

- **Boeing**
- **ARC**
- **JPL**
- **NGIX**
- **NCSA**
- **GSFC**
- **LaRC**
- **NTON-II/SuperNet**
- **IPG compute and data resources**
- **300 node Condor pool**
- **EDC**
- **CMU**
- **GRC**
- **NREN**

## Ames Wind Tunnels:

- National Full-Scale Aerodynamics Complex
- 9x7 ft Supersonic and 11 ft Transonic
- 12 ft Pressure

---

Remote Access to High Data-Rate Instruments

<table>
<thead>
<tr>
<th>Grid Services: Uniform access to distributed resources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grid Information Service</td>
</tr>
</tbody>
</table>

Security Services

- Authentication
- Authorization
- Auditing
- Monitoring

Fault Management

Communication Services

Network Cache

UC San Diego

National Center for Microscopy and Imaging Research

SDSC compute and data resources

Boeing
ARC
JPL
NGIX
NCSA
GSFC
LaRC
NTON-II/SuperNet
IPG compute and data resources

300 node Condor pool
EDC
CMU
GRC
NREN

Ames Wind Tunnels:

- National Full-Scale Aerodynamics Complex
- 9x7 ft Supersonic and 11 ft Transonic
- 12 ft Pressure
• Data access capabilities of IPG are demonstrated by *parallel data mining*
• 512 node SGI Origin at Ames uses IPG uniform interface data access tools (SRB) to simultaneously mine hydrology data from four sites
  - SDSC
  - CalTech
  - GRC
  - Washington U.

Tom Hinke, NASA Ames
Who is using Grids? ….CEOS Grid Prototypes

- Committee on Earth Observation Satellites (CEOS http://wgiss.ceos.org/ceos.htm)
- Next Generation Prototyping AI: Dick desJardins, NASA Ames
- CEOS Grid Team Objectives:
  - Establish CEOS Grid Testbed with multiple nodes
  - Demonstrate Grid enable applications
  - Infuse grid technologies into CEOS information systems and test facilities
- 4 CEOS Grid Prototypes
  - GSFC Advance Data Grid, Debbie Ladwig, Prototype advanced data management technologies like Storage Resource Broker with nodes at NASA GSFC, NASA Ames, Aerospace Corp., El Segundo, CA
  - USGS Data delivery utilizing GridFTP, Stuart Doescher, Prototype Globus Replica Catalog for keeping track of replicated files and the Globus Replica Management for managing files in a high performance, wide area data storage environment.
  - NOAA Operational Model Archive and Distribution System (NOMADS)”. Glenn K. Rutledge, Prototype remote access to high volume numerical weather prediction and global climate models and data, the National Climatic Data Center (NCDC), along with the National Centers for Environmental Prediction (NCEP) and the Geophysical Fluid Dynamics Laboratory (GFDL )
  - ESA Data Integration via Grid Portal, ES data warehouse, Grid service management for Ozone data application
National Virtual Observatory

- NSF funded - $10M 5 year project starting Nov 01
- Objective: provide observers with ‘transparent’ access to astronomy data (across data centers, instruments, tools)
  - Enabling new science, new user classes, integrated tool, common data access
- WGs: System Architecture, Metadata, data model,
  - image services, registry, user interfaces,
- PI and project director: Alex Szalay (JHU); CoPI: Roy Williams (Caltech/CACR)
- Demo candidates
  - Brown Dwarf science search – distributed query across data archives, searching into raw data for new correlation
  - Galaxy morphology science demo – on the fly image analysis and recognition
  - Gamma Ray burst science demo – event follow-up service for interoperability demonstration
  - SkyQuery – dist query tool using set of reusable services
- International Virtual Observatory Alliance (IVOA) and roadmap
- The National Academy of Sciences and Astronomical Survey Committee recommended in its decadal survey (NAS99) the establishment of a National Virtual Observatory (NVO) to utilize the latest computer and networking technologies to connect the archival and real-time resources of many earthbound and orbital astronomical observatories :http://www.us-vo.org/
GRIDs AND APPLICATIONS (Cont’d)

NVO Conceptual Overview

- More wavelengths
- Expansion across time
- Links to other data sets, related and not

Archival Storage
- (microwave)
- (visible)
- (X-Ray)
- (other)

Space Based Observatories & Instruments

Virtual Ground Station/s

- More reliable access to resources
- More collaboration (RT and non-RT)
- Networked/grid resources efficiencies and flexibilities

Government Research Facilities (*)

University Research Facilities (*)

Public Access

- Easier access to more resources, allowing choices to meet needs and schedules
- Greater continuous access
- Collection of more data for later analysis

Ground Based Observatories

- More timely sharing of techniques and raw data
- Shared computing resources for greater efficiency and effectivity

* Includes computers, staff, local storage, etc.
Astronomy Applications on the TeraGrid

Palomar-Quest Survey

Wide-area Mosaicking – HyperAtlas
O. George Djorgovski, Majid Majid, Ray Williams

2MASS Mosaicking Portal
A TeraGrid Science Gateway

N-point galaxy correlation
Robert Nichol, Jeffery Gardner, Andrew Connolly, Jon Gray

Quasar Science
Con Verekin Bath, Chris Drakhtoff, Robert Nichol

from catalog pipelines

from image pipeline
National Virtual Observatory Data Grid

1. Portals and Workbenches

2. Knowledge & Resource Management

3. Metadata View | Data View | Catalog Analysis | Bulk Data Analysis

4. Grid Security Caching Replication Backup Scheduling

5. Information Discovery | Metadata delivery | Data Discovery | Data Delivery

Standard Metadata format, Data model, Wire format

6. Catalog Mediator | Data mediator

Catalog/Image Specific Access

7. Compute Resources | Derived Collections | Catalogs | Data Archives

Concept space
Teragrid Impediments

- Learn Globus
- Learn MPI
- Learn PBS
- Port code to Itanium
- Get certificate
- Get logged in
- Wait 3 months for account
- Write proposal

and now do some science....
A better way:
Graduated Security
for Science Gateways

- Write proposal
  - own account

- Register
  - logging and reporting

- Authenticate
  - X.509

- big-iron computing

- more

- science....

- Web form
  - anonymous

- power user

- some science....

- own account

- browser or cmd line
GGF tech in VObserve? 4+1 options

1. No Grid, no way!

2. Grid throughout.

3. Grid services as leaf nodes.

4. Leaf nodes + pervasive OGSA-SEC + pervasive GridFTP.

5. Grid only “within” web services

Preferred by GWS-WG of IVOA (options 4 & 5 are not mutually exclusive)
AstroGrid-2 Layer Diagram

CLI | Portal | Workbench

Science Applications

Dataset Access | Registry
VObs Support Services | Community | Resource Discovery

Agent Framework | Data Mining Framework | Visualization Framework
MySpace | Auth/Auth Security

Grid & Web Services Middleware

Astronomical Datasets

Legend
- Existing Component
- AstroGrid-2 Component
- External Component
ES Data System Architecture (current)

Data Acquisition
- Spacecraft
- Tracking & Data Relay Satellite (TDRS)
- Ground Stations
- Polar Ground Stations

Flight Operations, Data Capture, Initial Processing & Backup Archive
- Data Processing & Mission Control

Data Transport to DAACs
- NASA Integrated Services Network (NISN) Mission Services

Science Data Processing, Data Mgmt., Data Archive & Distribution
- EOSDIS Science Data Systems (DAACs)
- REASoNs

Distribution, Access, Interoperability & Reuse
- WWW IP Internet
- Research
- Education
- Value-Added Providers
- Interagency Data Centers
- International Partners
- Use in Earth System Models
- Benchmarking DSS

Science Teams
- Measurement Teams

TECHNOLOGY
See SpaceGrid presentation 2001
... From Final SpaceGrid report

*Figure 7: High-Level SpaceGRID Infrastructure*
More from Final SpaceGRID report

SpaceGRID Applications “at large”
(modelling/simulation, data manipulation, end-to-end services, group-to-group collaboration...)

Virtual Organisation User Access

Services Orchestration: Definition & Execution

Services to Share and Integrate “resources”

Enabling Technology: GRID

Distributed “Hard” Resources

Distributed “Soft” Resources

Figure 6: ESA wide SpaceGRID infrastructure with different flavours
Figure 15: Technological view of Grid layered architecture for EO
DOD NCES (Network Centric Enterprise Services) Operational Concept

NCES Services Will Be Located at 5 DECC Equivalents Plus 15 SDNs (Which Includes 6 Teleport Sites)

Users Will Be Able to Access NCES Services From Any Location Worldwide Using NIPRNet and SIPRNet

Net-Centric Enterprise Services “Increment One”

Locations shown for CONUS DISA DECCs are notional

OpenGroup NCES presentation
NCES Vision

Edge Users

C2I Capabilities

- C2
- Intell
- Logistics
- Personnel
- Finance
- Etc.

Comms

Backbone

COI-Specific Capabilities

Core Enterprise Services

Notional only - does not imply one “box” per service etc.

© 2002, DISA
The PSE layer provides the scientist’s / engineer’s interface to Grid services. It is an application domain-specific collection of tools (e.g. simulations, databases, instruments), and a “workbench” environment that makes it easy to use those tools and to collaborate with others working on the same problem.

Applications, e.g. simulations, sit below the PSE and use middleware services.

The middleware layer provides different styles of service interfaces for application developers to access the basic Grid services.

Grid services are “standard” interfaces for the functions needed to build and manage distributed applications of all sorts.

Most “resources” are “local” and will have their own resource managers and use policies. It is the use mechanisms and interfaces for the local resources that the Grid common services are intended to homogenize.
Grid Architecture – Information Environments

Domain Independent Portals
- Domain Independent Web Services
  - Workflow Management
  - Data Management
  - Monitoring
  - Events
  - Programming Services
  - Experiment Management
  - Collaboration Services
  - Visualization
  - Data Processing & Analysis
- Grid Common Services
- Domain Specific Web Services
  - Computational Simulation
  - Flight Simulation
  - Instrument & Sensor Gateways
  - System Models
  - Archive Gateways
  - Zooming
  - Coupling

Domain Specific Portals
- User Environment Portals
- Composition Frameworks
- STS/SLI Mission Analysis
- ISS Training
- ES Modeling
- MER/CIP
- Aviation Capacity

Piyush Mehrotra, NASA Ames

So: GGF4 Arch WG presentation http://grid.lbl.gov/GPA/GPA.GGF-4.1.ppt

Pulled 3/5/2002

Shirley Tseng 3/2/05
Combined Grid and Web Services Architecture

So: GGF4 Arch WG presentation http://grid.lbl.gov/GPA/GPA.GGF-4.1.ppt

Pulled 3/5/2002
Grid Applications in a Generic Space Ops Architecture

Data Acquisition

Virtual Ground Station (VGS)

Ground Link Optimization

CSA, ESA, NASA, RSA

WAN

Mission Monitoring and Control and IP Facilities around the World

Virtualization for Distributed Operations

Enhanced Collaborative Environments

User Community

Enhanced Data Sharing

Mission & Ops Data Storage

Grid SW

Servers/WSs

Virtual Computer and Storage System/s