An A Design of the GEO Grid: Systems of Systems federating Geospatial Data and Services

Satoshi Sekiguchi
Director, Grid Technology Research Center
National Institute of Advanced Industrial Science and Technology (AIST), Japan

Grid Technology Research Center

<table>
<thead>
<tr>
<th>Grid Technology Research Center  (Sep 1, 07)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Employees</td>
<td>69</td>
</tr>
<tr>
<td>Tenured Researchers</td>
<td>20</td>
</tr>
<tr>
<td>Fix-term Researchers</td>
<td>8</td>
</tr>
<tr>
<td>Administrative Staff</td>
<td>11</td>
</tr>
<tr>
<td>Contractors</td>
<td>30</td>
</tr>
</tbody>
</table>

Tsukuba Campus

Cal-IT2 at UC San Diego

AKIBA Office

Akihabara, Electric Market Street

GridMPI™
What is the AIST?

- One of the largest Nat’l Labs in Japan
  - 3,500 employee+3,000 staff
- Research topics include
  - Environment
  - Material
  - Bio/Life science
  - Standards (JIS/ISO)
  - Geology
  - Semiconductor device
  - IT/Computer Science
  - etc.
- NIST+USGS+DOE labs+NIH

AIST Tsukuba Main Campus
7 other campuses across Japan

It is a little story of the GEO Grid

Geo meets a vehicle
Satellite Data
Map
Geo* Contents
GIS data
Field data
Grid Technologies
IT Infrastructure
Resources
Geology
Applications
Environment
Grid meets needs
ASTER : Main content of AIST GEO Grid

- ASTER was developed by METI with AIST scientific & engineering supports
- Global land coverage of Digital Elevation Model of 15m spatial resolution
- Excellent geo-location accuracy
  - Easy to mosaic (or make a seamless image/DEM)
  - Easy to overlay to GIS data
- Powerful spectral analysis
  - VNIR 3 bands (+backward 1band)/SWIR 6 bands/TIR 5 bands
- 2000~
- ASTER archive 1.4 million images = 140 TB on a HD

Landslides volume estimation using ASTER DEM

Synthetic bird eyes view with ASTER image and DEM

Volume : 0.06 km³, area : 0.5 km²
at DEM difference > 50 m

14NOV2000 11OCT2005
OK, no problem. Grid would help you
What is the GEO Grid?

The GEO (Global Earth Observation) Grid is aiming at providing an E-Science Infrastructure for worldwide Earth Sciences communities to accelerate GEO sciences based on the concept that relevant data and computation are virtually integrated with a certain access control and ease-of-use interface those are enabled by a set of Grid and Web service technologies.

AIST: OGF Gold sponsor (a founding member)
AIST: OGC Associate member (since 2007)

---

AIST GEO Grid Architecture (IT Infrastructure)

...
**Why “GRID”?**

- **Data Grid capability**
  - large (>100TB) satellite imagery data
    - storage design, networking design
  - loosely couple of a wide variety of geographically distributed data
    - meta data (access method, server location,), ontology,

- **Computing Grid capability**
  - on-demand generation of high level data products
    - adopt the most accurate geometric-, radiometric- and atmospheric-correction methods on-the-fly
  - simulation jobs may consume computing resources
    - a “common” requirement of computing grid

- **Grid Basic Service**
  - compliance with owners’ access control policy of data/service
    - Grid Security Infrastructure – AuthN, AuthZ, Accounting
  - complex workflow support in portals incl. data access, simulation execution, visualization, etc.

---

**GEO Grid use case scenario: Volcano monitoring**

[Image of volcano monitoring data with temperature gradient and color coding]
Simulation of Pyroclastic flows on volcanos

Performs numerical simulations of lava and/or pyroclastic flows on volcanos for prediction and mitigation of the hazard.

Position: 436815.547 UTM(x)
9166345.806 UTM(y)

Hc_height: 10 m (>0; e.g. 10)
H/L min: 0.2 (>0; e.g. 0.2)
H/L max: 0.4 (<1; e.g. 0.4)
H/L interval: 0.02 (e.g. 0.02)

Simulate | Show confirm window
A Workflow example

“Disaster prevention and mitigation (Volcano)”

Monitoring of crustal deformation by PALSAR
In-situ observations by CCOP members
Hazard Map for Evacuation planning

High resolution DEM provided from ASTER
Simulation of lava and/or pyroclastic flow on GEO Grid

A Workflow example

“Disaster prevention and mitigation (Landslide)”

Geology map (GSJ)
Rain fall (AMEDAS/Fieldserver)
Early warning system based on Susceptibility map

High resolution DEM provided from ASTER
Large scale Computer simulation using actual landslide DB
Distributed Workflow Design

A scientist is going to try one's new version with better physics.
- Add a new version of the application to the Grid
- Reconfigure the workflow to use this version.

Data Search

Select a region and a time range and desired attributes
But, it was not so simple ...

- Need more sophisticated security architecture
- Need more flexible Data access and integration
Requirements for the IT infrastructure

- **Size scalability** in near-real-time data handling and distribution
  - Need to manage hundreds tera-bytes to peta-byte of data.
  - Such data will be made available with minimum time delay and at minimum cost.
- Handling wide **diversification of data** types, associated metadata, products and services.
  - Research communities wish to integrate various data according to their interests.
  - IT infrastructure must support
    - the creation of user groups which represent various types of virtual research/business communities
    - Federation of distributed and heterogeneous data resources which is shared in such communities
Requirements for the IT infrastructure (cont’d)

- Respecting data owner's publication policies
  - Some data are not freely accessible.
    - E.g. commercial data.
  - IT infrastructure must provide a security infrastructure which supports flexible publication policies for both data and computing service providers.
- Smooth interaction and loose coupling between services (data, computing)
  - A desirable IT architectural style would achieve loose coupling among interacting software agents to allow users both to create services independently, and to produce new application from them.
  - IT infrastructure must support sharing, coordination, and configuration of environment for application programs and resources, depending on the user’s requirements.

---

GEO Grid Players & Tools

<table>
<thead>
<tr>
<th>Player</th>
<th>Tools</th>
<th>Product</th>
</tr>
</thead>
<tbody>
<tr>
<td>End Users</td>
<td>Web browser</td>
<td>GEO Grid Client</td>
</tr>
<tr>
<td>Portal &amp; VO mgr</td>
<td>Project Goal</td>
<td>Customized Web portal VO Management System</td>
</tr>
<tr>
<td>Data Provider</td>
<td>Scientific data</td>
<td>GEO Grid Data Service</td>
</tr>
<tr>
<td>Resource Provider</td>
<td>Application, IT resource</td>
<td>GEO Grid Hosting Service</td>
</tr>
<tr>
<td></td>
<td>GEO Grid Portal Dev. Kit</td>
<td>GEO Grid Tool kit</td>
</tr>
<tr>
<td></td>
<td>VO tools</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GEO Grid Service Dev. Kit</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
FOUR categories of Data Access Attribute

OPEN

Cat. 2
Open to Many
(Black list)

Cat. 1
Open to All
(Public)

CLOSE

Cat. 3
Close to Many
(White list)

Cat. 4
Close to All
(Private)

You might need to have like this

Virtual Organization (VO) enabled by GRID

Cat 3 Project/Community

Cat 2 Domain Common

Cat. 1 Public
Security Architecture Design Policy

- Introduces VO (Virtual Organization)
- Data and computation are provided as “services” via standard protocols and APIs.
- A VO is created dynamically by integrating available services and resources according to the interests and requirements of the VO.
- User-level Authentication and VO-level Authorization
  - User’s right is managed (assigned) by an administrator of his belonging VO.
  - Access control to a service is configured by the service provider according to the publication policy. There are some options of the access control: VO-level, Group/Role-based, User-level, etc.
  - Scalable architecture for the number of users.

Overview and usage model of the GEO Grid system
Data Integration
GEO Grid solution at a glance

OGC Standards

Workflow engines

OGSA-DQP

Service Coordination

OGC SQL Spatial query

Distributed Joins

GWS

WPS

CSW

WFS

WMS

Services

OGSA-DAI@AIST

OGSA-DAI

OGSA-WebDB

Data Services

GTRC

GSJ

Data Resources

ASTER

MODIS

Parallel SQL processing

Data Integration Requirements (1)

Support of Geo-Spatial Standards

OGC standards

- CSW for catalog service
- WMS, WFS, WCS for geospatial data services
- WPS for data processing
- OGC spatial query

Data Format standards

- Data
  - GML and other XML data format
- Metadata(schema)
  - ISO19115, Dublin Core, ebXML etc.
Data Integration Requirements (2)

- Support of Distributed Database Federation
  - Support of multi-level distributed database federation
    - Local = Different Resources under one Service
    - Enterprise
    - Global = Distributed Services over the world

- Support of Heterogeneous Resources
  - Relational databases
  - XML databases
  - Web databases

Data Integration Requirements (3)

- Support of the Grid Technology
  - Support of the VO concept
    - VO based database authorization/security management
  - Support of Grid Standards & Middlewares
    - OGF Standards & Middleware
      - WS-RF/Globus, MDS, GSI (VOMS)
    - OASYS/W3C
      - BPEL, Service Registry (ebXML/UDDI)
**GEO Grid data integration - at a glance**

- **OGC Standards**
  - WPS
  - CSW
  - WFS
  - WMS

- Support of OGC standards with Grid-based federated Databases

- Workflow engines
- Service Coordination

- Distributed Joins

- OGSA-DQP

- OGSA-DAI@AIST
- OGSA-DAI
- OGSA-WebDB

- Data Services
- GTRC
- GSJ

- Data Resources
- ASTER
- MODIS

- Parallel SQL processing

---

**Approach**

- **Multi-level-distributed data integration**
  - Service-based Database Integration Middleware

- **Support**
  - OGSA-DAI with VOMS
  - Grid based Database Middleware
Approach

- Support of Heterogeneous Data Resources esp. Web databases
  - ex. NASA MODIS image service

OGSA-DAI based
Web Access Interface

---

OGSA-WebDB

OGSA-DAI based SQL interface for WebDB

OGSA Environment

- OGSA-DAI Compliant Database Access
- Proxy Databases
  - NASA MODIS
  - AIST ASTER

Mediator

- Wrapper
- join

Web Databases

- NASA MODIS
- AIST ASTER
- OGSA-DAI Databases
- Wrapper
- join

SQL
OGSA-DAI 3.0 & VOMS

OGSA-DAI with VOMS
= One key component to support data integration with VO concept.

1. OGSA-DAI 3.0 = Just released last week!
2. DAI 3.0 with VOMS = Not yet released.

Great Technical Helps & Supports by OMII team
☐ AIST had a chance to be a PAL member of the OMII project
☐ AIST had a chance to be a beta tester of OGSA-DAI 3.0
☐ Information exchange with OMII Europe team on future DAI3.0 with VOMS.
DAI3.0 beta experience

- ASTER Grid System
  Test application which evaluates a new activity SQLBag

- OGSA-DAI-RDF & OGSA-WebDB
  Convert our own activities to fit to 3.0 model

AIST experience of DAI3.0 beta

- Programming Flexibility
  - Resource Definitions and Creations
  - Activity I/Os

- Rich Set of New Activities
  - Many new functionalities
    - We tested SQLBag for our ASTER application
    - Activity Monitoring and Management
  - Seems to be sufficient for one data service

- Focus on Java Programming
  - Via Client API = No notions of Perform Document
    - No standard XML access way
Insights from the preliminary evaluation

- Most software components (e.g. GT4, VOMS, GAMA, OGSA-DAI) worked as we expected.
  - Will evaluate OGSA-DAI w/ VOMS.
- We have been able to confirm that the GEO Grid system will fulfill the functional requirements.
  - 170TB ASTER data in ASTER Grid system.
    - We used Gfarm on a cluster, but it is one of options.
- Open- and standard-protocol based architecture is appropriate for smoothly interacting with and loosely coupling data services and computing services, E.g.
  - It is easy to use Dtmsoft on the CELL computer if it is available.
  - It is easy to integrate the other data sources.
- We can create VOs by deploying and configuring GAMA+VOMS
  - ASTER data and Dtmsoft can also be provided for the other projects (VOs).
- Users are freed from burden with certificate managements.
SEE YOU ALL AGAIN in SC07 in Reno
AIST booth is next to UK e-Science!!

Special thanks to:
Yoshio Tanaka – Security
Isao Kojima – Data Integration