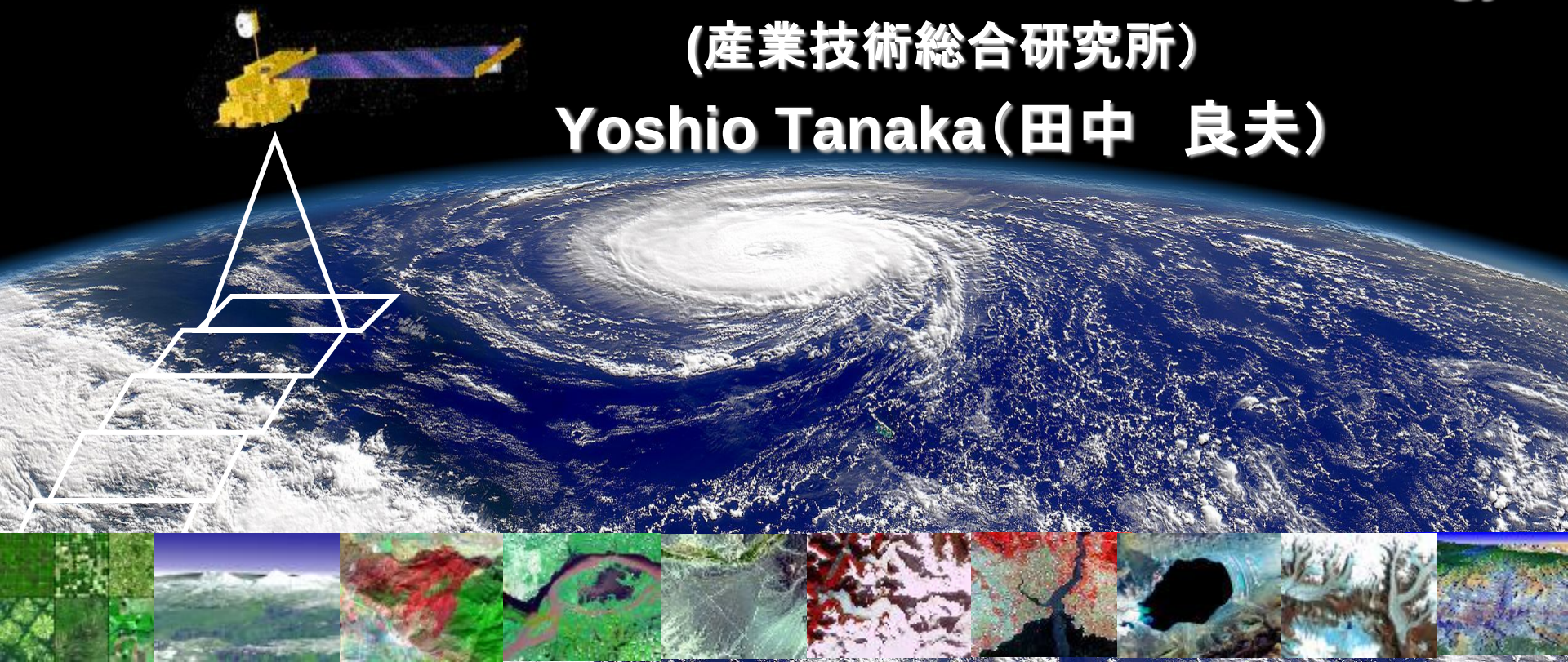
AIST

# Grid and E-Science: discovering and exploiting the future

National Institute of
Advanced Industrial Science and Technology
(産業技術総合研究所)
Yoshio Tanaka(田中　良夫)



# Overview of AIST（産業技術総合研究所）

National Institute of Advanced Industrial Science and Technology

- Headquarters in Tokyo and Tsukuba.
- 9 Centers in Japan
- 8,500 employees including 2,500 full-time researchers.

Chugoku Center

Tsukuba Center

Hokkaido Center

Shikoku Center

Tohoku Center

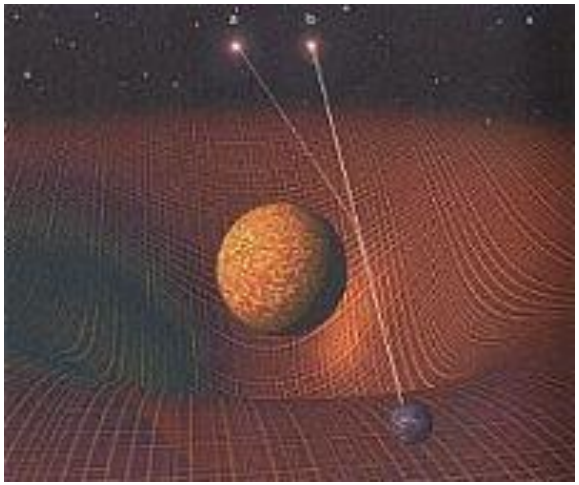
Kyushu Center

Kansai Center

Chubu Center

Rinkai Center

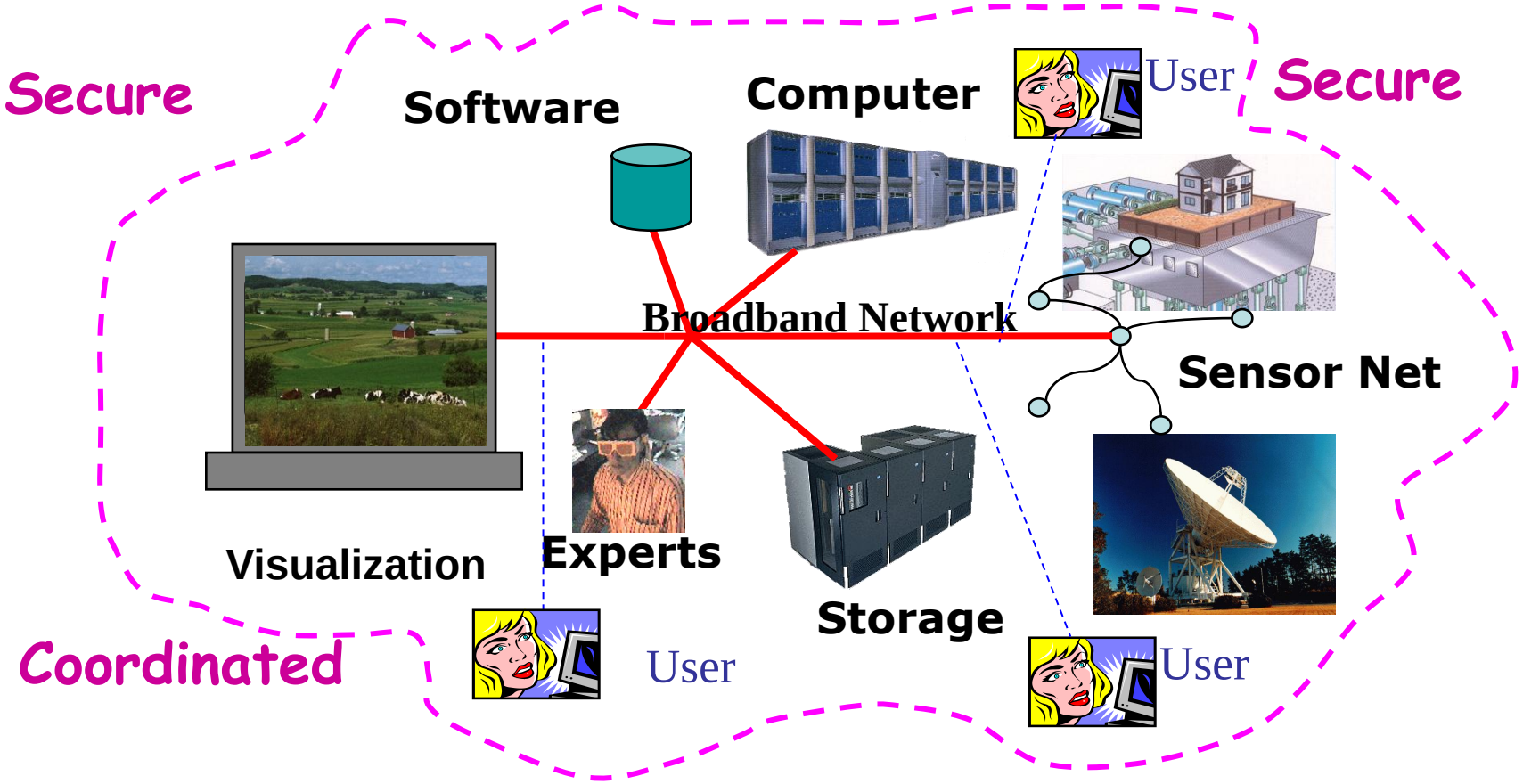
# From Primary science to e-Science



- Primary Science
  - Theoretical Science（理論科学）
- Secondary Science
  - Experimental Science（実験科学）
  - Verify theory by experiments
- Third Science
  - Computational Science（計算科学）
  - Computer simulation and other forms of computation to solve problems in various scientific disciplines.
  - Supercomputers enable accurate large-scale simulations
- Fourth Science
  - E-Science （E-科学）

# Fourth Science == E-Science

- A new methodology of science by federating distributed resources such as computers, databases, and experimental devices.



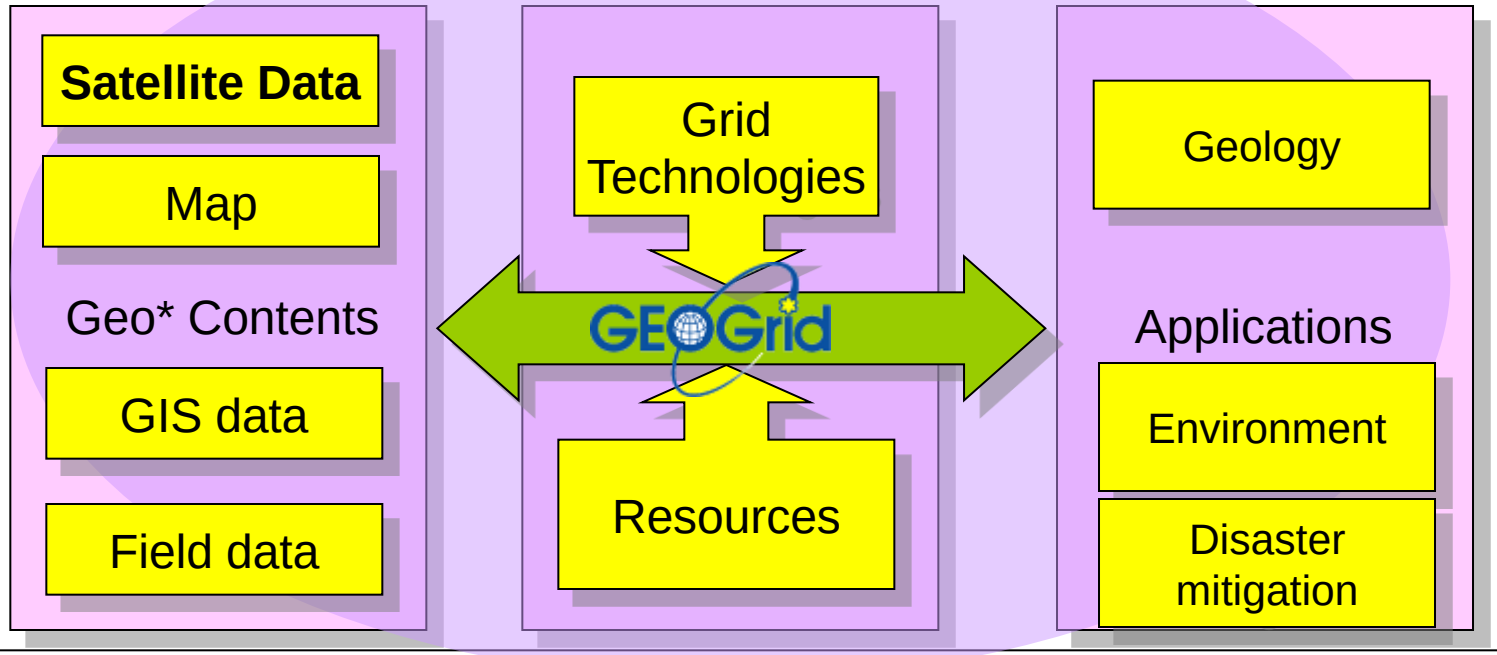

# Global Earth Observation (GEO) Grid

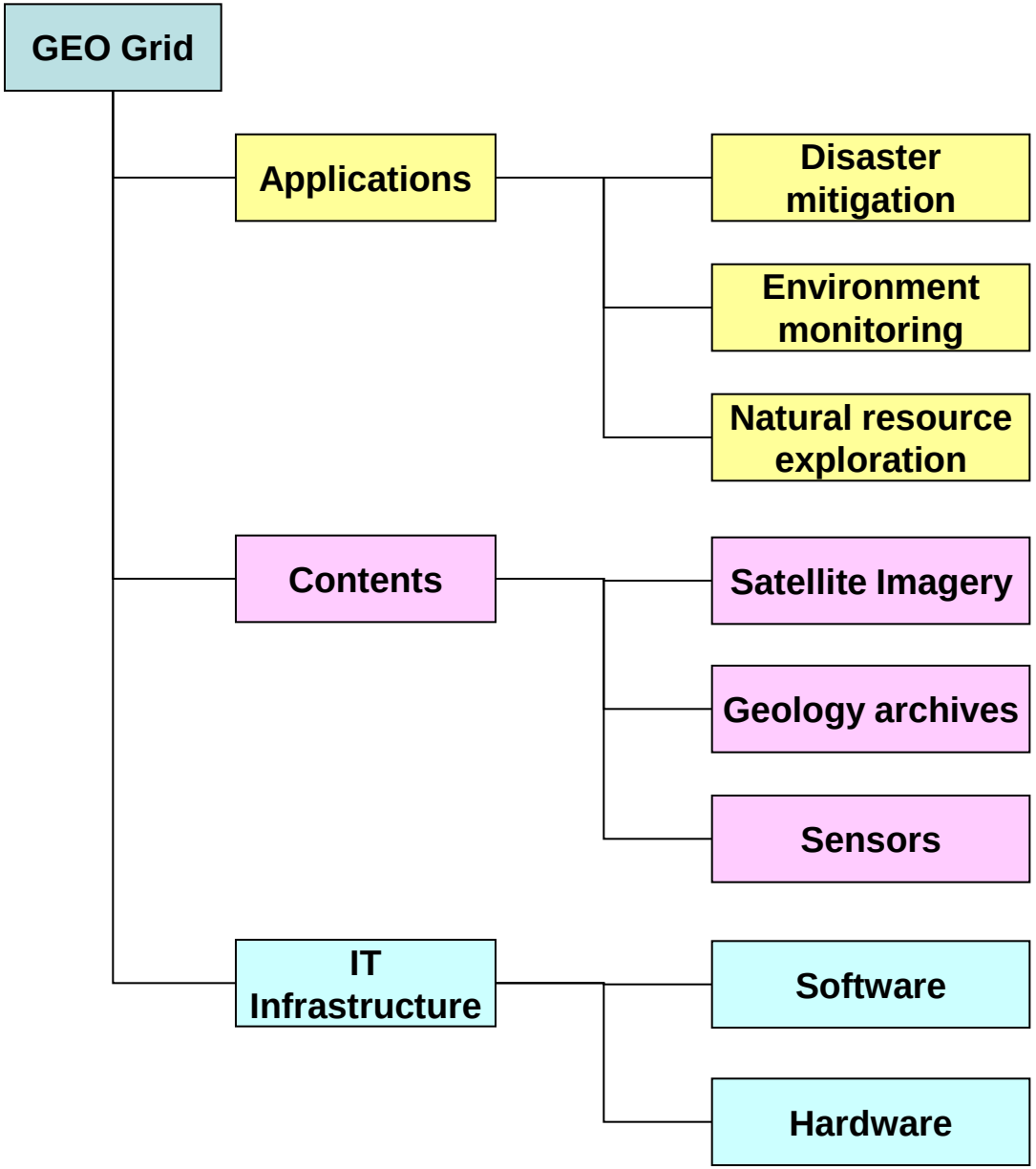
# What is the GEO Grid ?

- 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)**






Land slides, flood

Global warming, $CO_2$ flux estimation

Oil, Gas

Full L0 ASTER on disk

MODIS on disk (East Asia)

Japan, SE Asia

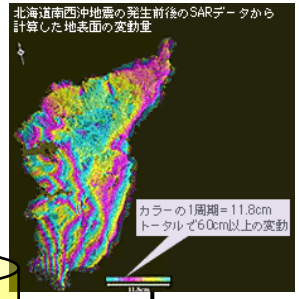
AsiaFlux, Field server

Security, data access, service registry, resource mgmt., Weg GIS, Workflow, U/I Portal, etc.

Storage, Servers
Cluster computers

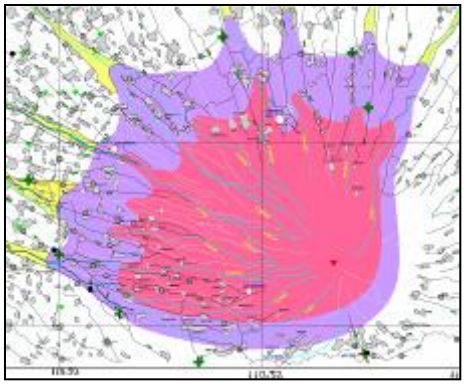
# A Workflow example “Disaster prevention and mitigation”

Monitoring of crustal deformation by PALSAR



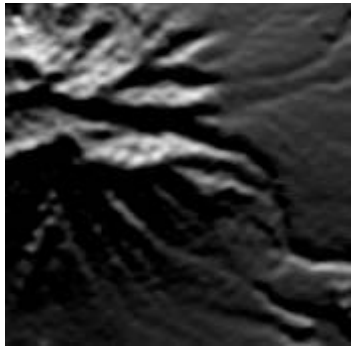

In-situ observations e.g. growth of a lava dome

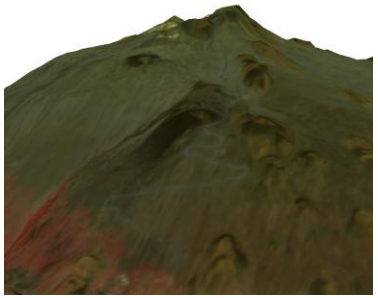
Hazard Map for Evacuation planning



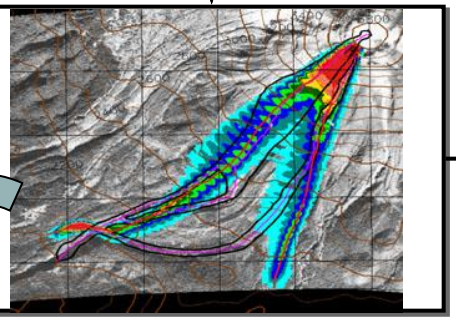
PALSAR

ASTER





High resolution DEM provided from ASTER



Simulation of lava and/or pyroclastic flow on GEO Grid

# Functional 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

# Functional requirements for the IT infrastructure (cont'd)

- Respecting data owner's publication policies
  - Some data are not freely accessible.
    - E.g. commercial data.
- Smooth interaction and loose coupling between data services and computing services
  - 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 environments for application programs and resources, depending on the user's requirements.

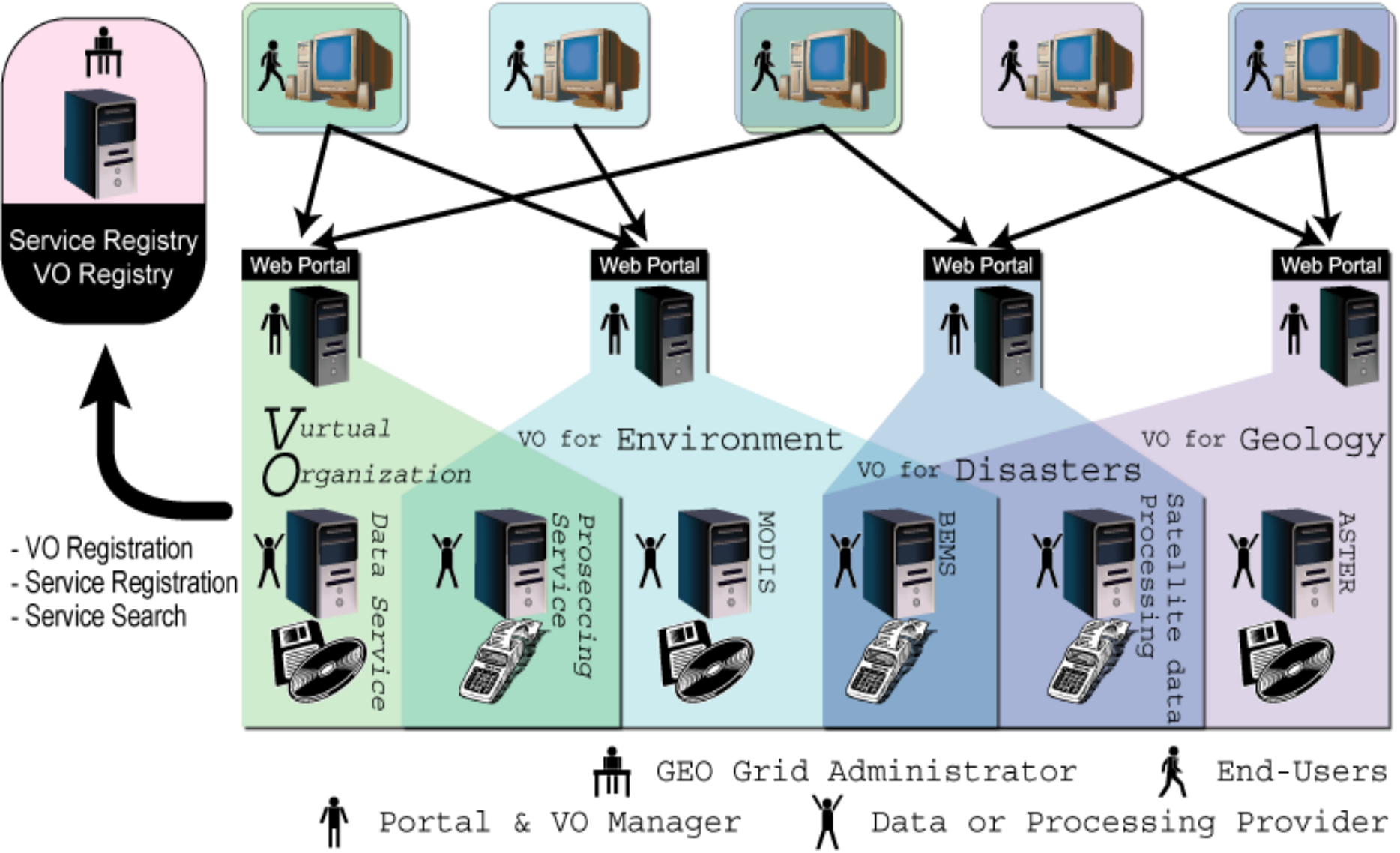
# Functional requirements for the IT infrastructure (cont'd)

- Ease of use
  - End users should be able to access data and computing resources without the burden of installing special software and taking care of security issues (e.g. certificate mgmt.).
  - Data and service providers should be able to easily make their resources available as services with desired access control.
  - Administrators and leaders of communities should be able to create virtual communities easily by configuring appropriate access control.
  - We must provide an ease-of-use framework for publishing services and user interfaces.

# Design Policy

- Introduces a concept of 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



# Key Technologies: GSI and VOMS

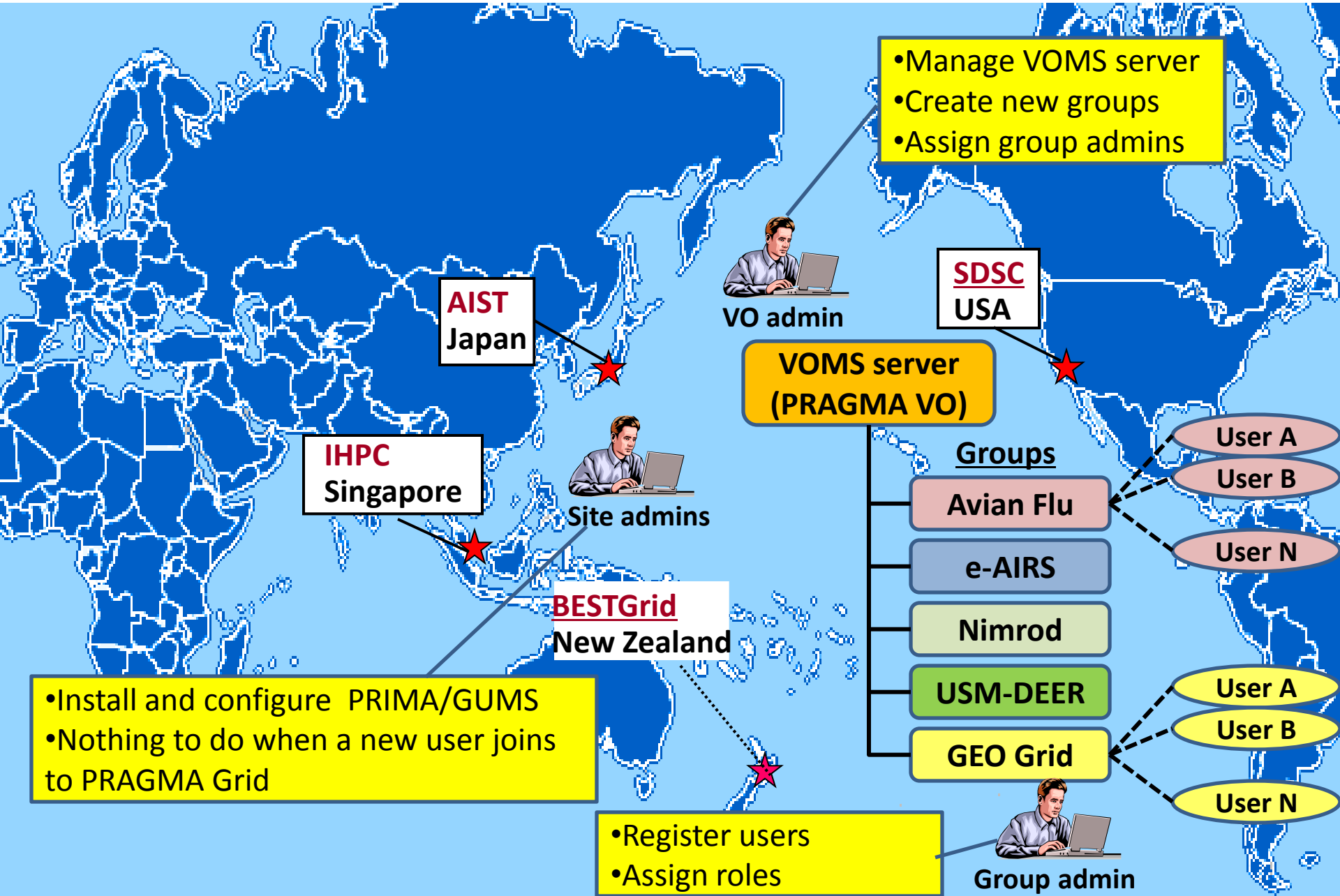
- Grid Security Infrastructure (GSI) is standard security technology used in the current Grid communities.
  - Based on Public Key Infrastructure (PKI) and X.509 Certificates.
- Virtual Organization Membership Services (VOMS) is a software for creating/managing VOs.
  - Developed by European Communities
  - Based on GSI

End users of GEO Grid may not be required to understand GSI, VOMS, etc, but project (VO) admin should understand these technologies correctly.

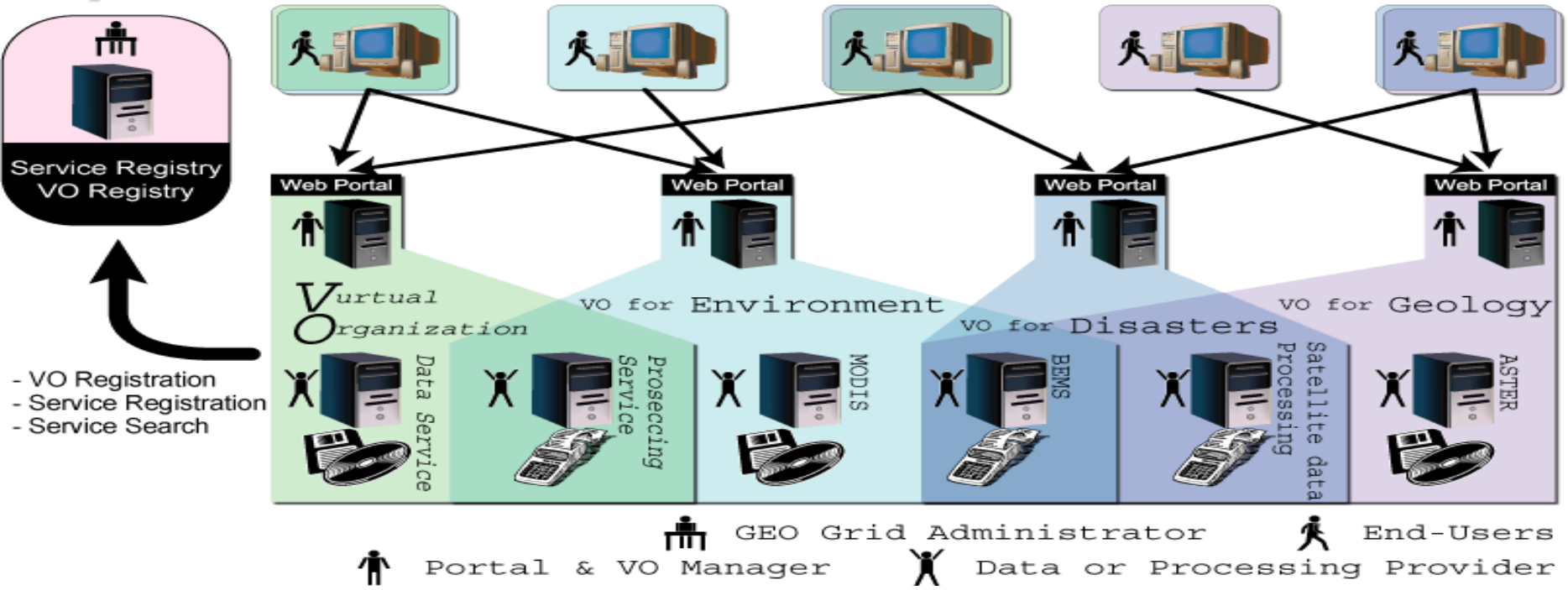
# What's the role of VOMS? (cont'd)

- VOMS provides a mechanism for VO-based authorization.
  - Users are registered to VO(s)
  - Users can belong to Group(s) in the VO
  - Users can be assigned role(s)
  - Service providers can configure the system to control access based on
    - VO-base
      - All users in a VO can access to the service
    - Group-base
      - Users in a specific group can access to the services
    - Group&Role-base
      - Users in a specific group with specific role can access to the services
- It is implemented by embedding "VOMS attributes" in user's proxy certificate.

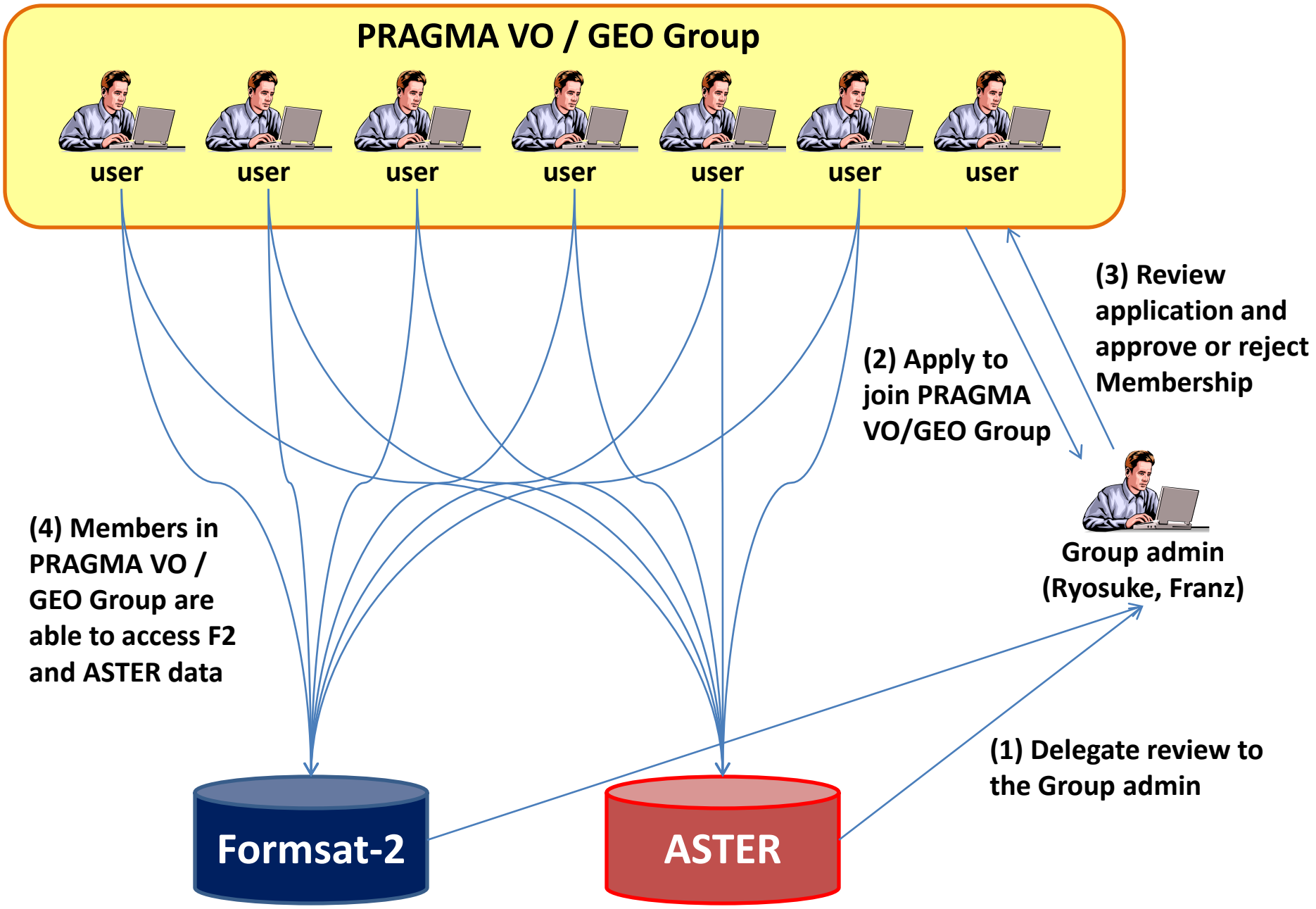
# VOMS on PRAGMA Grid (test phase)



# Overview and usage model of the GEO Grid system

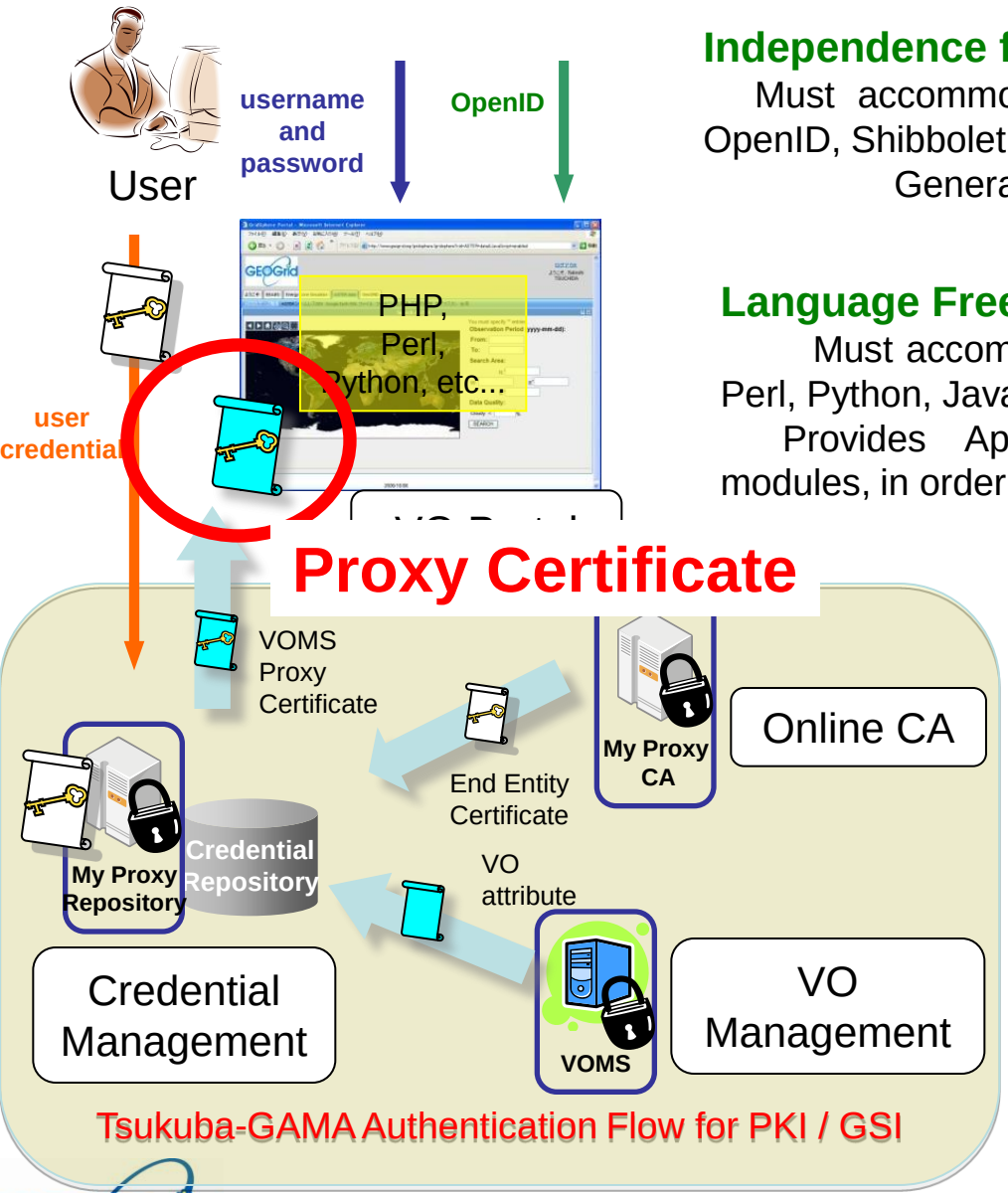



- User-level Authentication and VO-level Authorization
  - User's right is managed (assigned) by an administrator of his belonging VO.
  - Access control to a service is configured by the service provider according to the publication policy. There are some options of the access control
    - VO-level, Group/Role-based, User-level, etc.
  - Scalable architecture for the number of users.


PRAGMA VO / GEO Group
user
user
user
user
user
user
user
(3) Review application and approve or reject Membership
(2) Apply to join PRAGMA VO/GEO Group
Group admin (Ryosuke, Franz)
(4) Members in PRAGMA VO / GEO Group are able to access F2 and ASTER data
(1) Delegate review to the Group admin
Formsat-2
ASTER

# Required functionality for Security

- Credential Management:
  - Non-secure users often manage their private keys for PKI / GSI credentials without careful planning.
- Authentication methods:
  - Must accommodate existing, settled authentication methods, OpenID, Shibboleth, username and password, user credential, etc.
- Portal Development:
  - Must accommodate existing application portals written by PHP, Perl, Python, Java Servlet, etc.




## Independence from Authentication methods:

Must accommodate existing, settled authentication methods, OpenID, Shibboleth, username and password, user credential, etc.
Generates Grid credentials from any method.

## Language Free Portal Development:

Must accommodate existing application portals written by PHP, Perl, Python, Java Servlet, etc.

Provides Apache, Servlet, and GridSphere authentication modules, in order to support any language.

## Credential Management:

Non-secure users often manage their private keys for PKI / GSI without careful planning.

Manages user credentials on the server side, instead of leaving it to inexperienced users.

# Sequence - login



USER

1. input username and pass of user cert

PRAGMA VO portal
http://gfm49.apgrid.org/gridsphere/

voms
proxy cert

4. register proxy cert

Credential
Repository

2. generate gloubs
proxy certificate

PRAGMA VOMS

3. add voms attribute

VOMS Credential Portlet

VOMS Credential Portlet

--- Credential ---

subject : /C=JP/O=AIST/OU=GRID/CN=Naotaka YAMAMOTO

time left : 71 hours 59 minutes 55 seconds.

--- VOMS Extension Information ---

attribute : /PRAGMA

attribute : /PRAGMA/Geo

Same Identity

VOMS Credential Portlet

--- Credential ---

subject : /C=JP/O=AIST/OU=GRID/CN=Naotaka YAMAMOTO

time left : 71 hours 59 minutes 53 seconds.

--- VOMS Extension Information ---

attribute : /testvo2.geogrid.org/Role=NULL/Capability=NULL

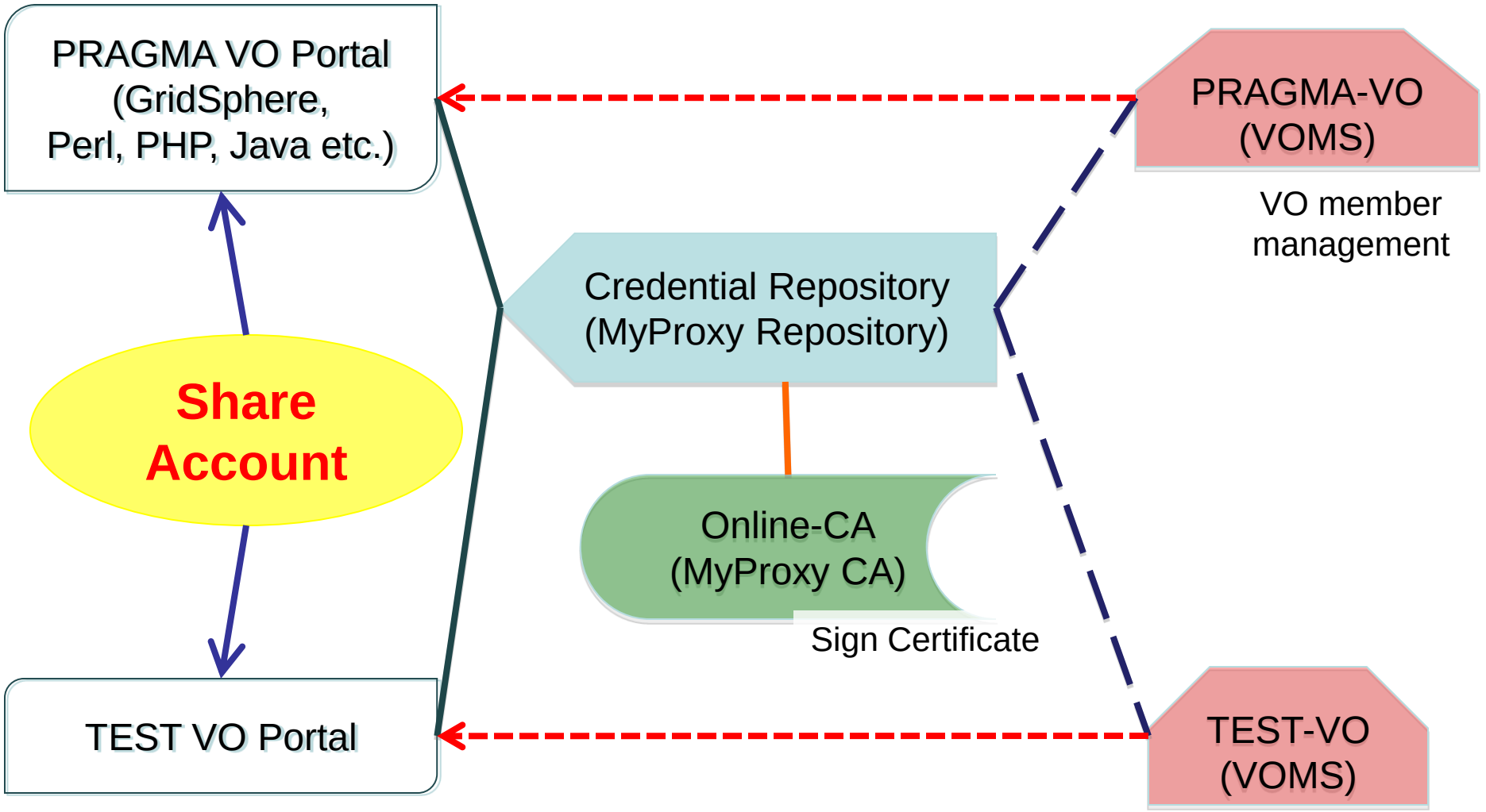
attribute : /testvo2.geogrid.org/ASTER/Role=NULL/Capability=NULL

attribute : /testvo2.geogrid.org/[illegible]/Role=NULL/Capability=NULL

Different Attribute

# Share Account on multiple VOs



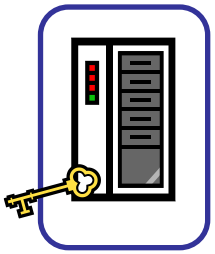


# OpenID authentication module



User

Password for OpenID

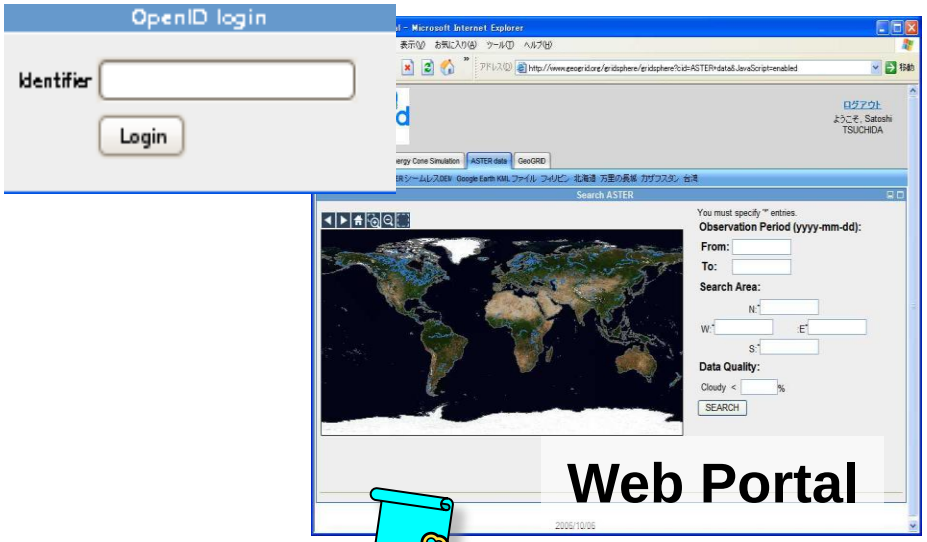


OpenID Server

MyProxy CA

OpenID URL

Request short-lived credential




Web Portal

- Account DB
- Credential Repository

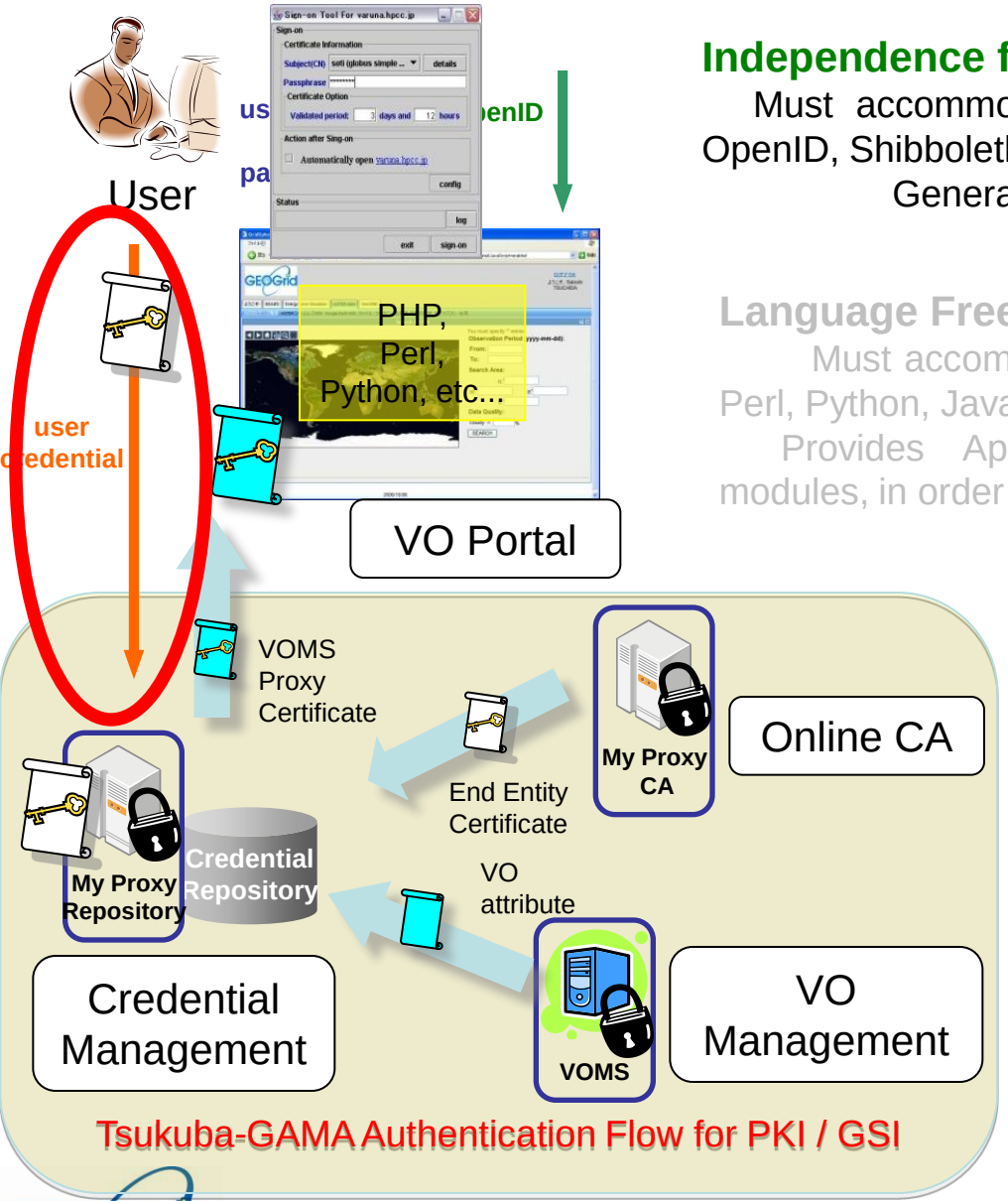
VOMS proxy

VO member DB

VOMS server

# Credential Login






## Independence from Authentication methods:

Must accommodate existing, settled authentication methods, OpenID, Shibboleth, username and password, user credential, etc.
Generates Grid credentials from any method.

## Language Free Portal Development:

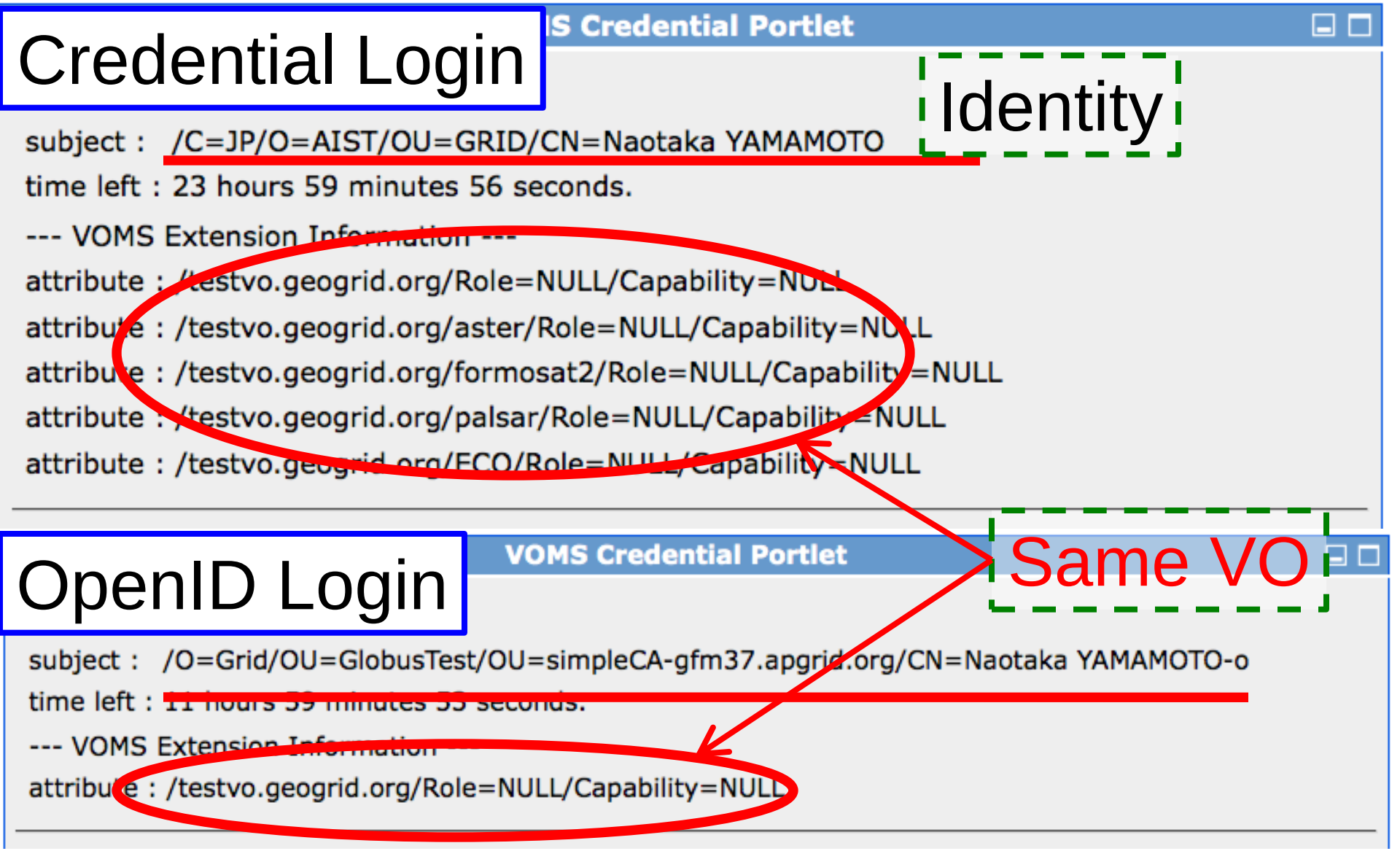
Must accommodate existing application portals written by PHP, Perl, Python, Java Servlet, etc.

Provides Apache, Servlet, and GridSphere authentication modules, in order to support any language.

## Credential Management:

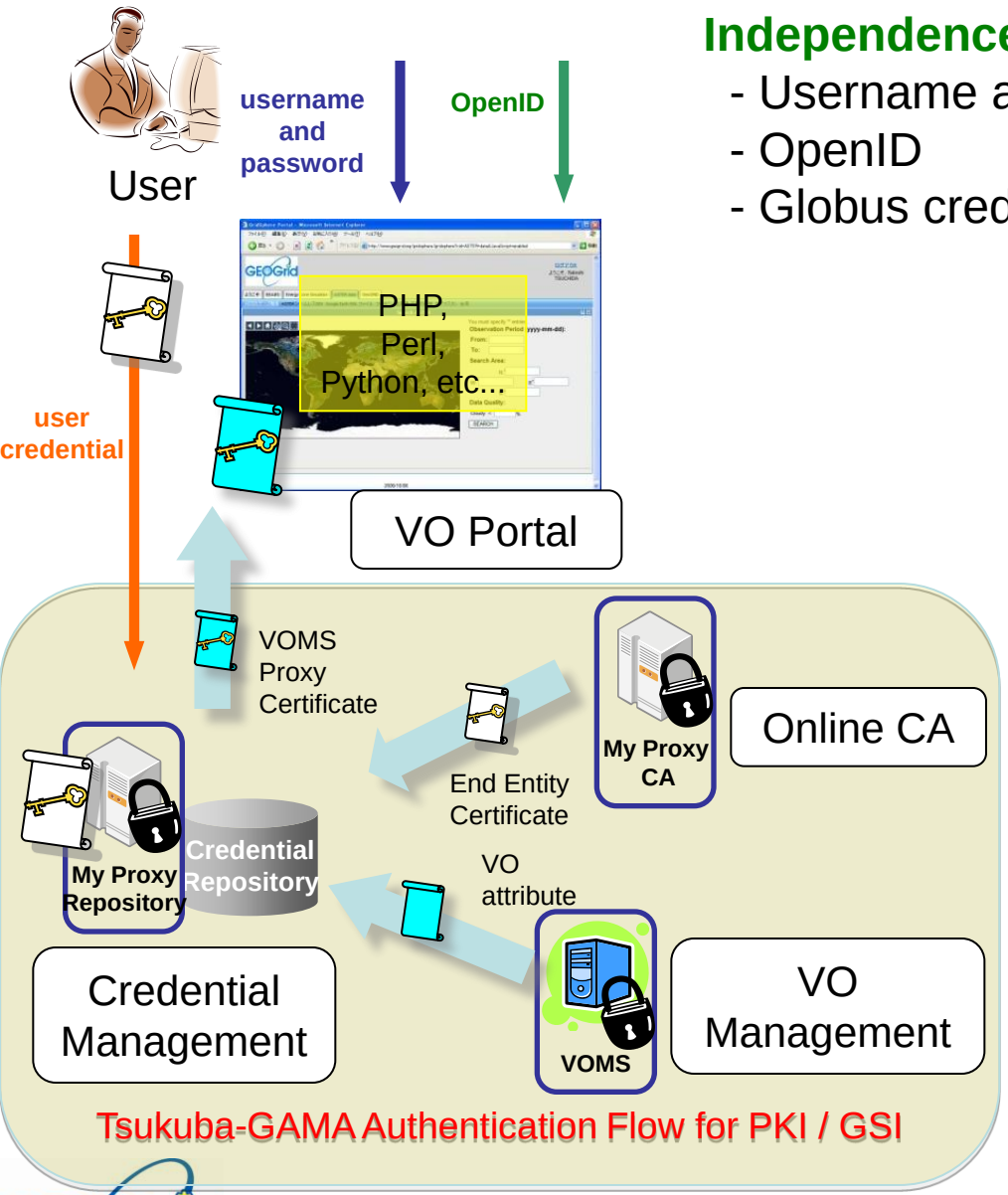
Non-secure users often manage their private keys for PKI / GSI without careful planning.

Manages user credentials on the server side, instead of leaving it to inexperienced users.


Credential Login
S Credential Portlet
Identity
subject : /C=JP/O=AIST/OU=GRID/CN=Naotaka YAMAMOTO
time left : 23 hours 59 minutes 56 seconds.
--- VOMS Extension Information ---
attribute : /testvo.geogrid.org/Role=NULL/Capability=NULL
attribute : /testvo.geogrid.org/aster/Role=NULL/Capability=NULL
attribute : /testvo.geogrid.org/formosat2/Role=NULL/Capability=NULL
attribute : /testvo.geogrid.org/palsar/Role=NULL/Capability=NULL
attribute : /testvo.geogrid.org/ECO/Role=NULL/Capability=NULL
OpenID Login
VOMS Credential Portlet
Same VO
subject : /O=Grid/OU=GlobusTest/OU=simpleCA-gfm37.apgrid.org/CN=Naotaka YAMAMOTO-o
time left : 11 hours 59 minutes 55 seconds.
--- VOMS Extension Information ---
attribute : /testvo.geogrid.org/Role=NULL/Capability=NULL

# Summary of Tsukuba-GAMA






## Independence from Authentication methods:

- Username and Password
- OpenID
- Globus credential

## Language Free Portal Development:

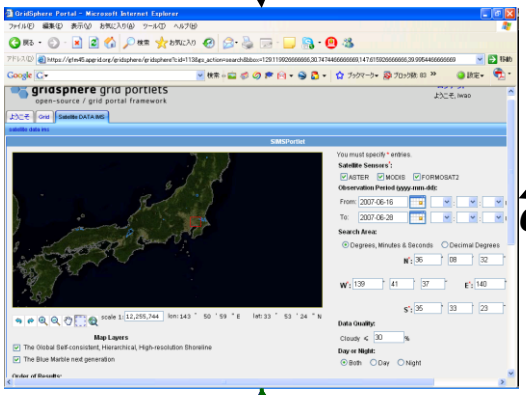
- GridSphere / Satellite database federation
- Geographical portal / OpenLayers
- PHP, Perl

## Credential Management:

- User does not need to manage their credentials
- User can use their certificate

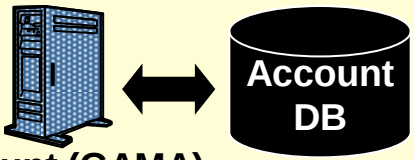

user

login

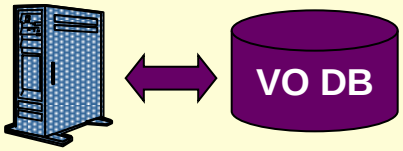


portal server

credential

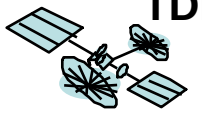



account (GAMA) server



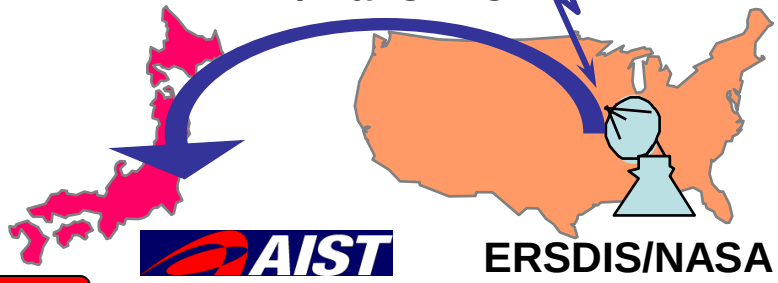

VO (VOMS) server

Terra/ASTER



TDRS

APAN/TransPAC



AIST

ERSDIS/NASA

GET

exec

query

GSI + VOMS

WFS

WCS

WMS

CSW

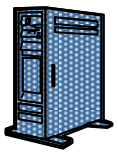
GSI + VOMS

OGSA DAI

GSI + VOMS

GRAM

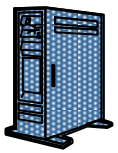
GridFTP



GIS server

map server

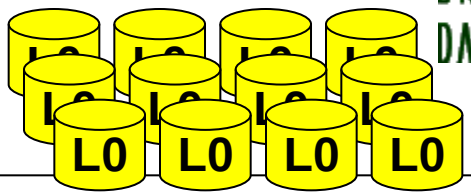
catalogue/ metadata server



gateway server

GEO Grid Cluster

GRID DATA FARM



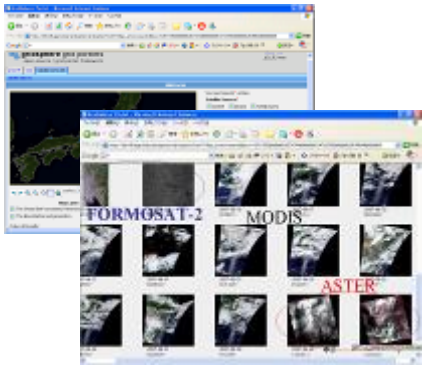

Data

Maps

Meta data

Storage (DEM)

# Demo Environments in 2007 - SIMS (ASTER+MODIS+Formsat2)




Java Program

Integration Framework with OGSA-DAI

OGSA-DAI Client

**SIMS portlet**
- **query data**
- **create web page which shows thumbnail images**

SQL

SQL

SQL

SQL

SQL

Application Server

Globus

OGSA-DAI

VOMS

VOMS

Globus

OGSA-DAI

OGSA-DAI

SQL w/ JDBC

Database Server (Sybase)

FORMOSAT-2

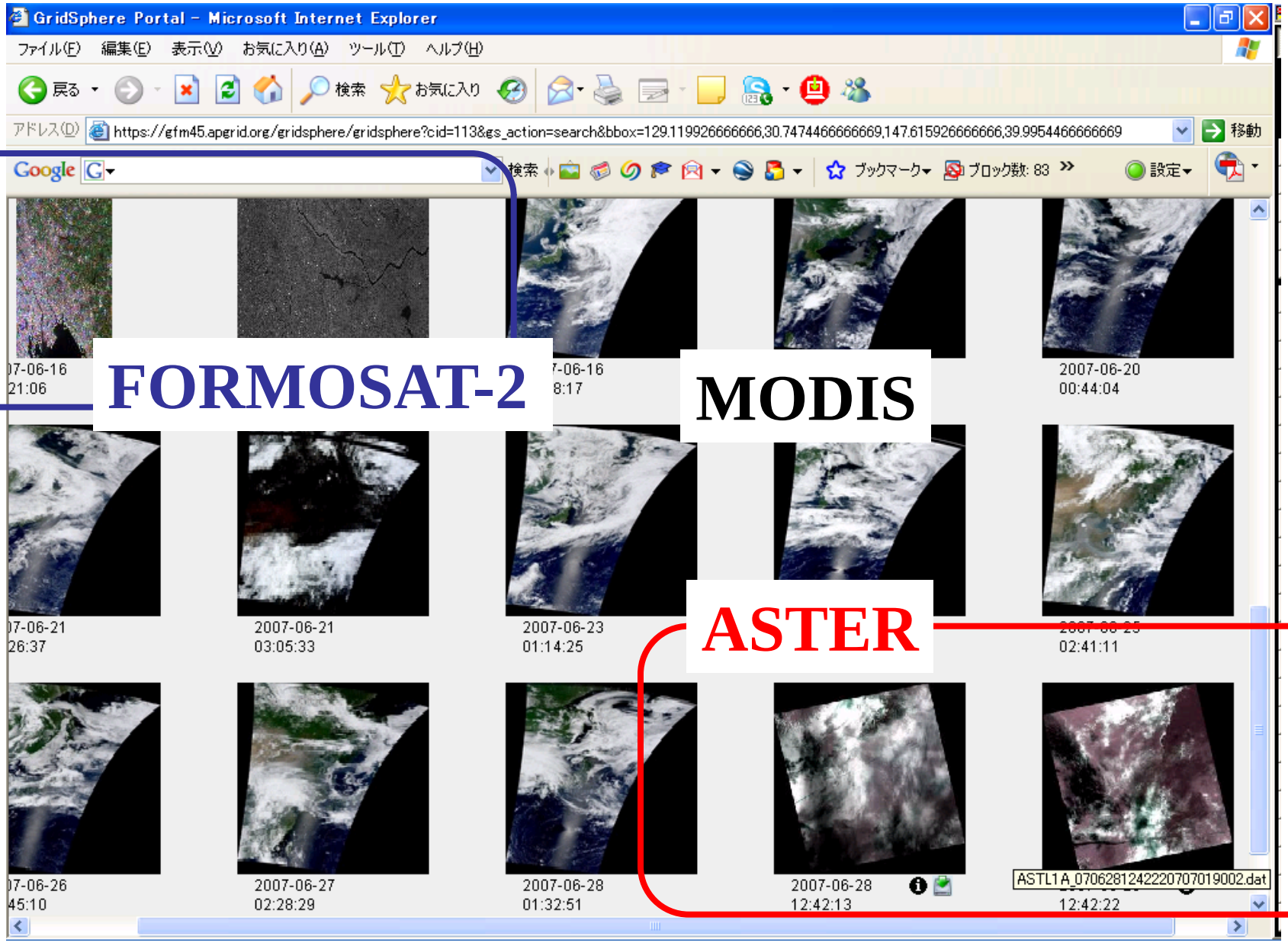
SQL w/ JDBC

Database Server (PostgreSQL)

ASTER

MODIS

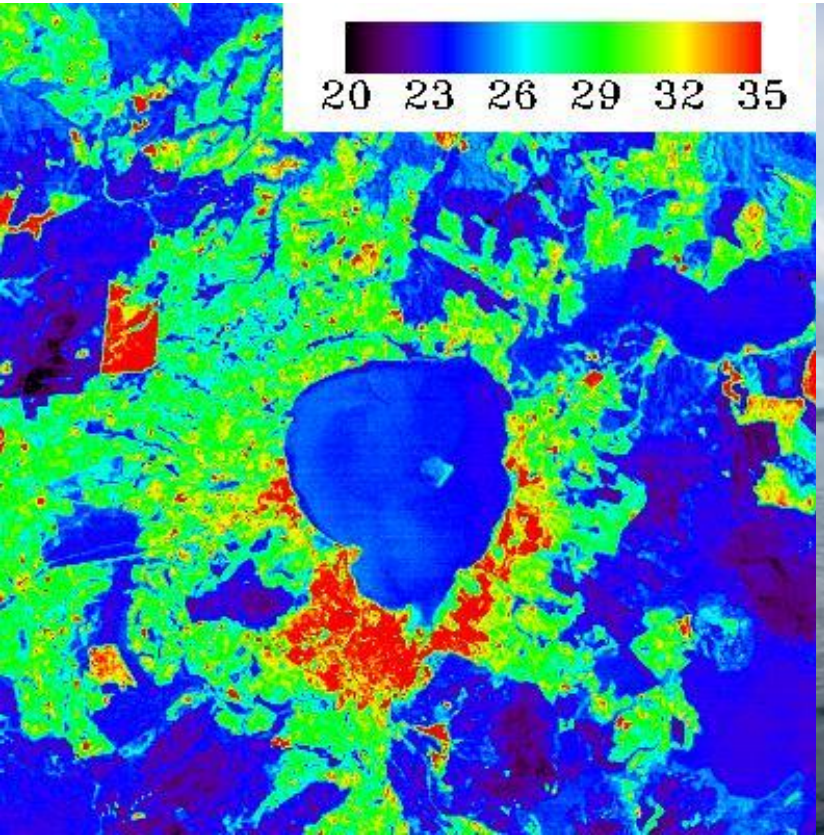
NSPO@TW AIST@JP


GridSphere Portal - Microsoft Internet Explorer
ファイル(F) 編集(E) 表示(V) お気に入り(A) ツール(T) ヘルプ(H)
戻る 検索 お気に入り
アドレス(D) https://gfm45.apgrid.org/gridsphere/gridsphere?cid=113&gs_action=search&bbox=129.119926666666,30.7474466666669,147.615926666666,39.9954466666669
移動
Google 検索 ブックマーク ブロック数: 83 設定
FORMOSAT-2
MODIS
ASTER
2007-06-20 00:44:04
2007-06-21 03:05:33
2007-06-23 01:14:25
2007-06-27 02:28:29
2007-06-28 01:32:51
2007-06-28 12:42:13
ASTL1A_0706281242220707019002.dat

# Collaboration with GLEON

- Cyber infrastructure to support limnology
  - Secure and privileged data access to ASTER/F-2 data (optionally MODIS)
  - Standard data search/access through OGC protocols
  - Common visualization tools to integrate remote sensing and in-situ data
  - Automatic match-up with ~~ground~~ water truth data
- Large-scale computations
  - Atmospheric corrections
  - Full scale simulations of lake metabolism assimilating satellite data

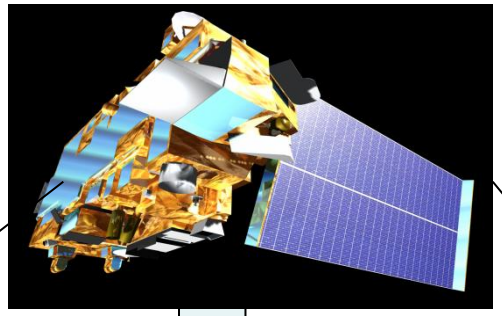
# Synergy between Satellite imagery and in-situ data

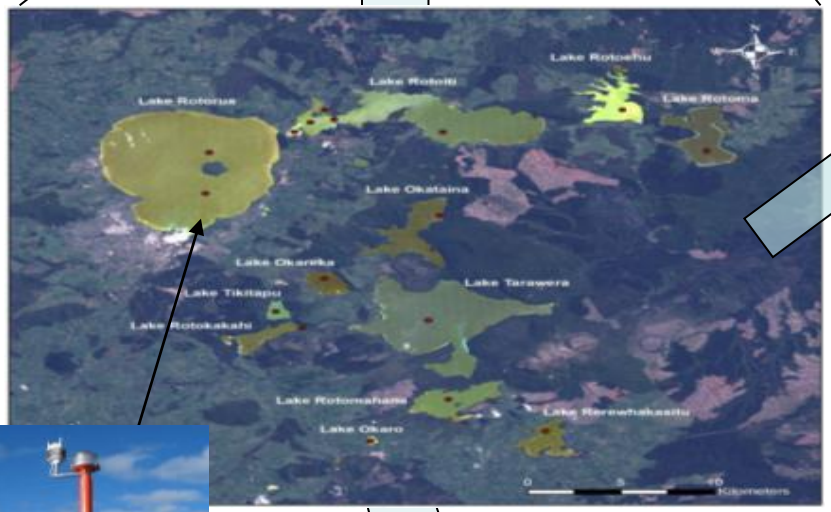



ASTER temperature map

In-situ temperature measurements

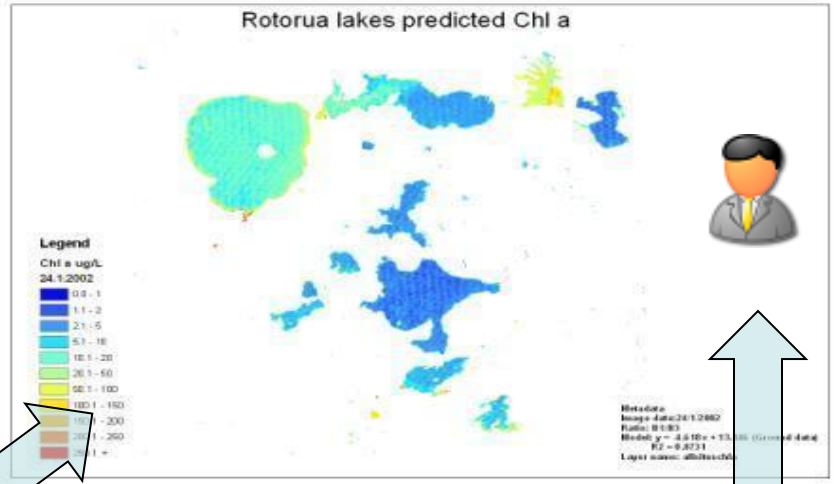
# Synergy of satellite and field data

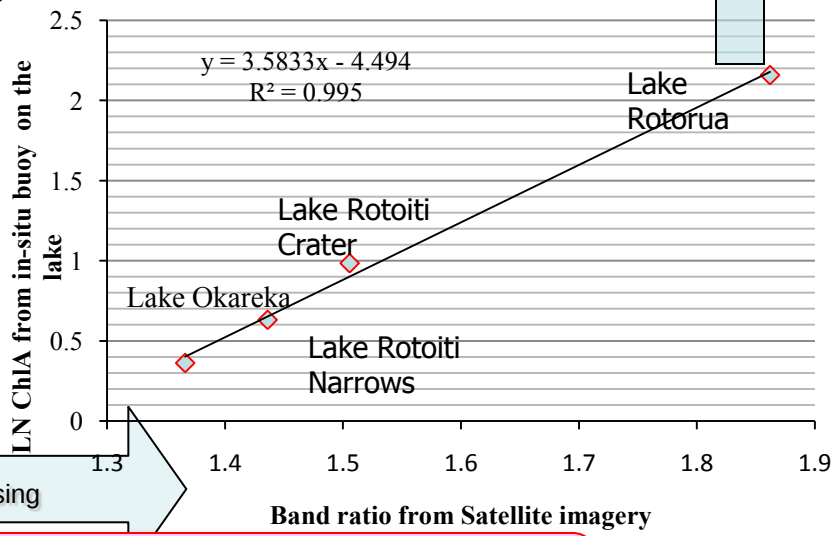





SOS=Sensor Net

WPS= Satellite data processing




Accurate water quality map production with GLEON

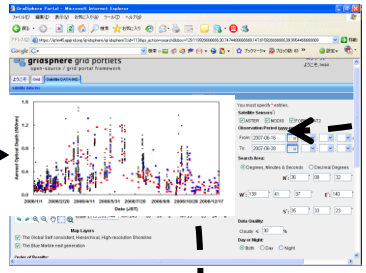
# Calibration / validation of Satellite Data using Field Sensor Data

Calibration and Validation processing

Web Processing Service (WPS)

Access Control by GSI + VOMS

1. login



2. VOMS proxy

AAA Service

Resource WG

Credential Repository

VO Admin

VOMS

VO member DB

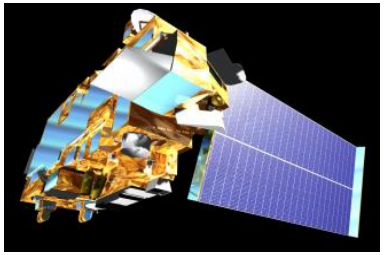
4. Inputs

5. Results

3. Search

Access Control by GSI + VOMS

Web Coverage Service (WCS)

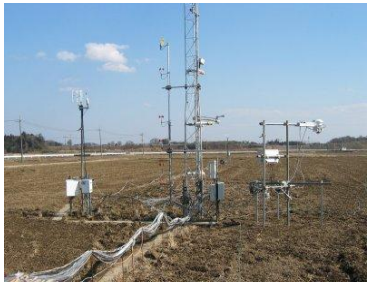


Formsat-2

ASTER / MODIS

Satellite Sensors

Sensor Observation Service (SOS)



GLEON

PEN / Live E!

Field Sensors

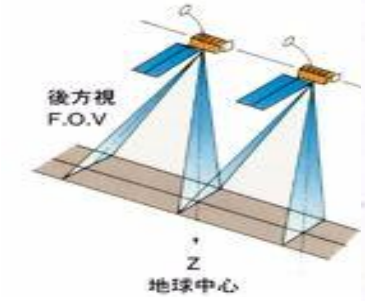
# High Performance GIS
## – Optimization of Stereo-Matching Software –

### Background and Motivation

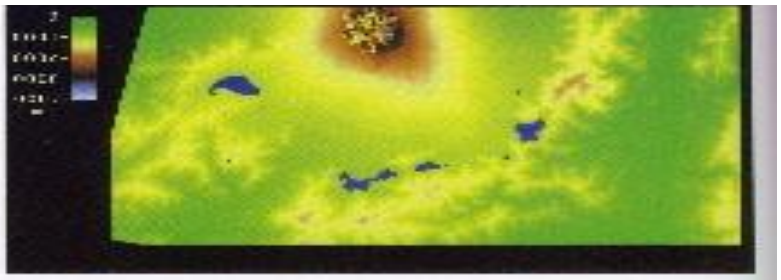
Strong demand for Earth science applications
High-performance Geospatial information System (GIS) is expected to provide an innovative infrastructure for Earth Sciences, enabling near-real-time data and processing services.

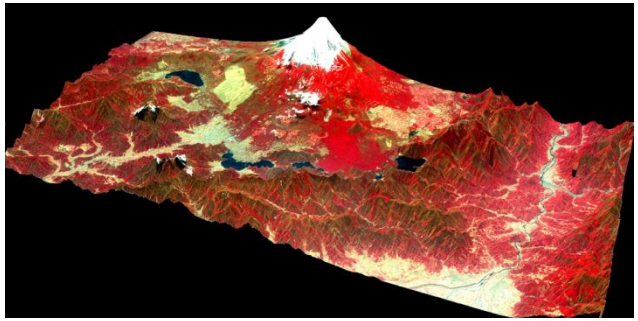
### DEM creation and Stereo-Matching

Stereo-matching software has often been used in generating a Digital Elevation Model (DEM) from a pair of satellite imagery data sets to compute height from parallax views using two photographic images.




DEM ( Digital Elevation Model )



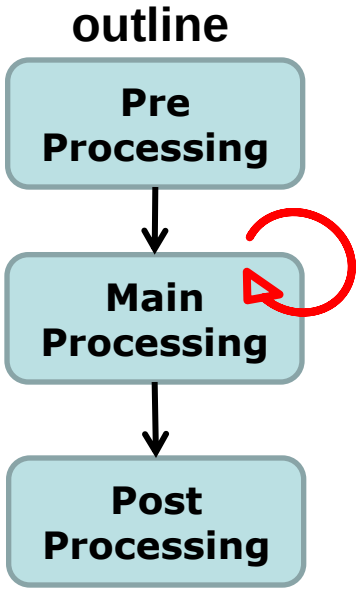


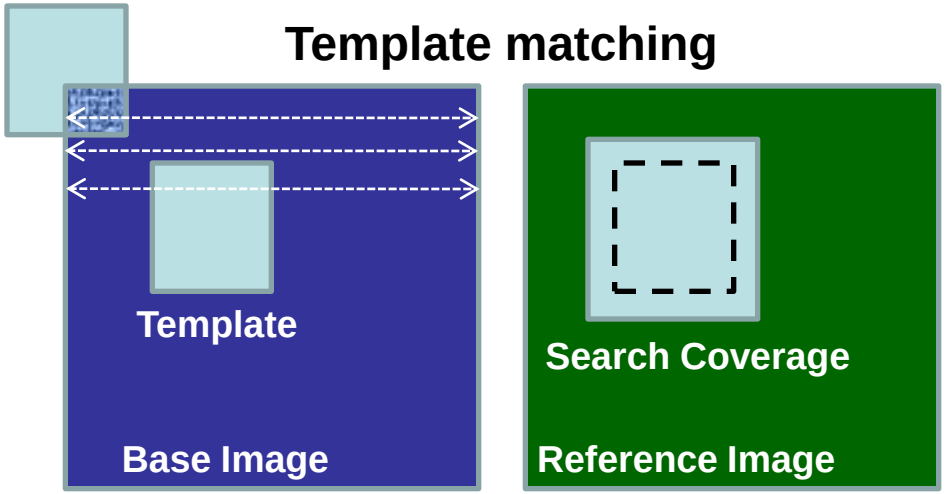
### Optimization of Stereo-Matching Software

- Target architecture
  - Heterogeneous (Cell)
  - Homogeneous (Nehalem)
- How to optimize
  - Manual optimization (Cell)
  - OpenMP + Manual opt. (Nehalem)

# Analysis of the program（outline）

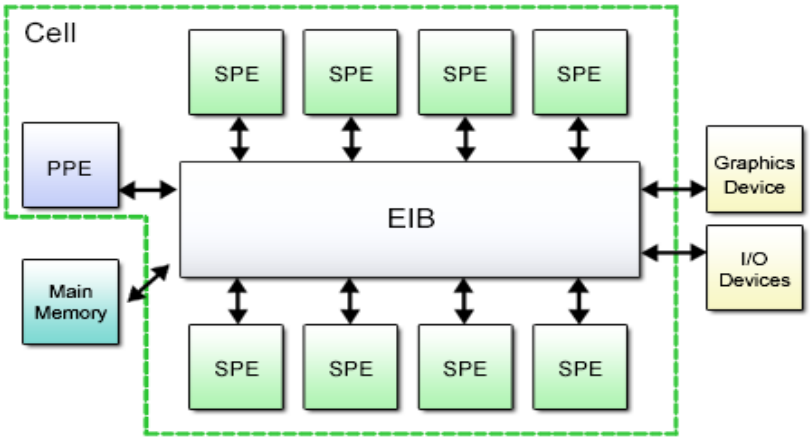
- Outline
  - Compare image data from different sensors
    - Calculate correlative coefficient and identify spots.
    - Complement missing data and generate altitude.
- Pre processing
  - Input data
  - Initialize structures
- Main processing
  - Template Matching
    - Compare two images and identify spots.
  - Interpolation
    - Complement missing data
  - Median Filtering
    - Remove noise
  - Other filtering
  - Output data
- Post processing
  - Free buffers



# What we did

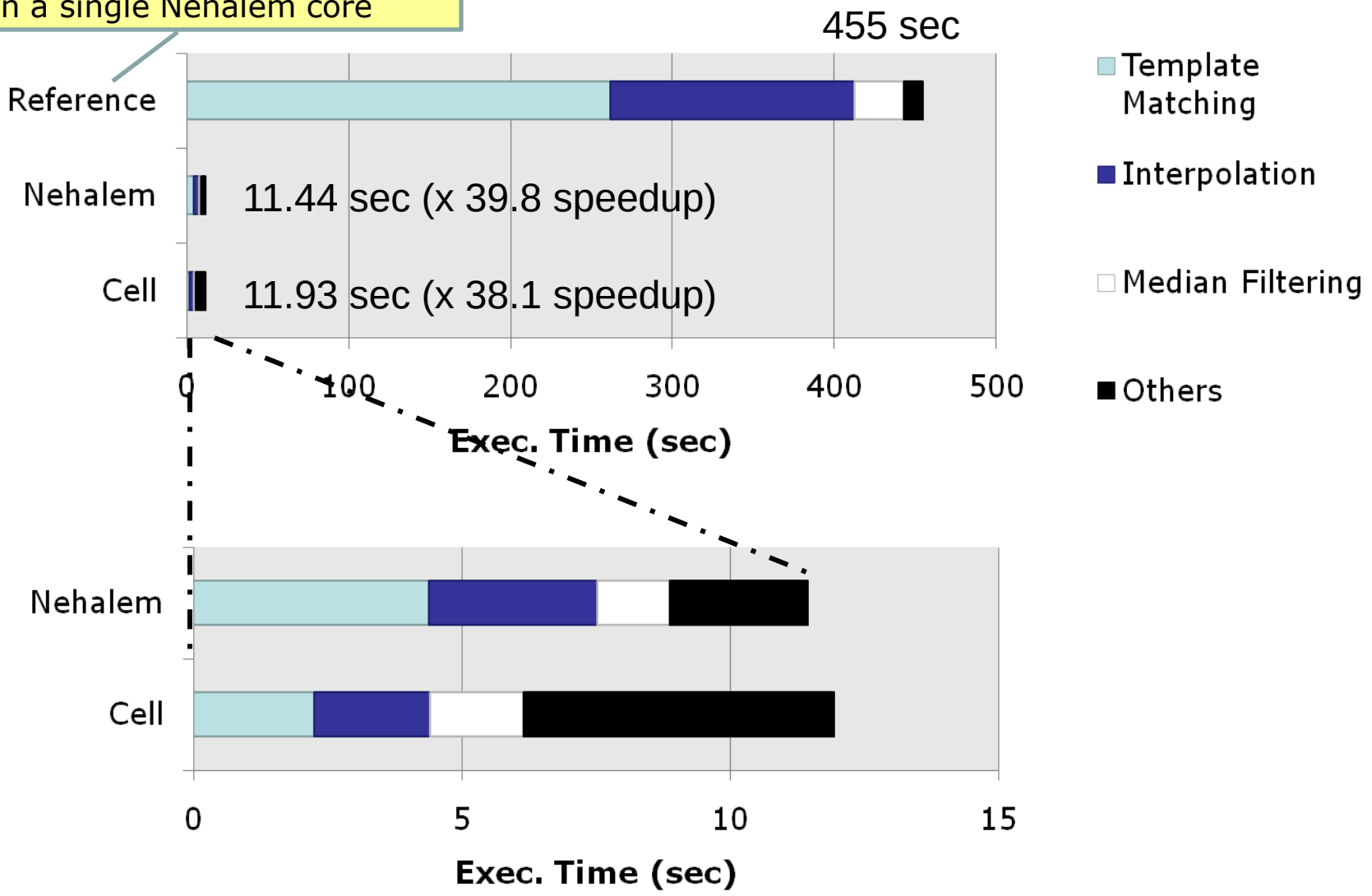
- Optimized Stereo-matching software on multicore processors
- Target architecture
  - Heterogeneous (Cell)
  - Homogeneous (Nehalem)
- How to optimize
  - Manual optimization (Cell)
  - OpenMP + Manual optimization (Nehalem)
- Platforms




| Platform | Processor | #CPU | memory | #core |
|---|---|---|---|---|
| IBM BladeCenter QS22 | IBM PowerXCell 8i (3.2 GHz) | 2 | 8GB | 2 PPE + 16 SPE |
| HP Z800 Workstation | Intel Xeon X5500 (2.66 GHz) | 2 | 4GB | 8 |

# Results of the optimization

Exec. time of the original program on a single Nehalem core



# Summary of GEO Grid

- GEO Grid enables federation of heterogeneous distributed databases and high-performance computation with appropriate access control.
- The system is already used by real GEO science applications.
  - Environmental monitoring
  - Recovery from disasters
  - Natural resource discovery
  - …
- Will continue R&D for new features and improvements of the system.
- The system is useful not only for GEO Science but the other sciences which need federation of distributed databases and computations.
  - Bio Science, Medical Science, Astronomy, …

# Summary

- Key of e-Science is to include not only computing resources but also other resources such as storage, networks, equipments, etc.
- Data and computation are provided as “services” via standard protocols and APIs.
- Such services are integrated and provided to users with appropriate access control but as if they are in a single organization.
- We expect to see new discovery by e-Science.