CHEP2006 Mumbai, India

Grid Activities in Japan

- Overview of NAREGI Project -

February 16, 2006

Project Leader, NAREGI Project
Professor, National Institute of Informatics

Kenichi Miura, Ph.D.

National Research Grid Initiative (NAREGI) Project:Overview

- Started as an R&D project funded by MEXT
  (FY2003-FY2007)
  2 B Yen(~17M$) budget in FY2003

- One of Japanese Government's Grid Computing Projects
  ITBL, Visualization Grid, GTRC, BioGrid etc.

- Collaboration of National Labs. Universities and Industry
  in the R&D activities (IT and Nano-science Apps.)

- NAREGI Testbed Computer Resources (FY2003)

MEXT:Ministry of Education, Culture, Sports,Science and Technology

National Research Grid Initiative (NAREGI) Project:Goals

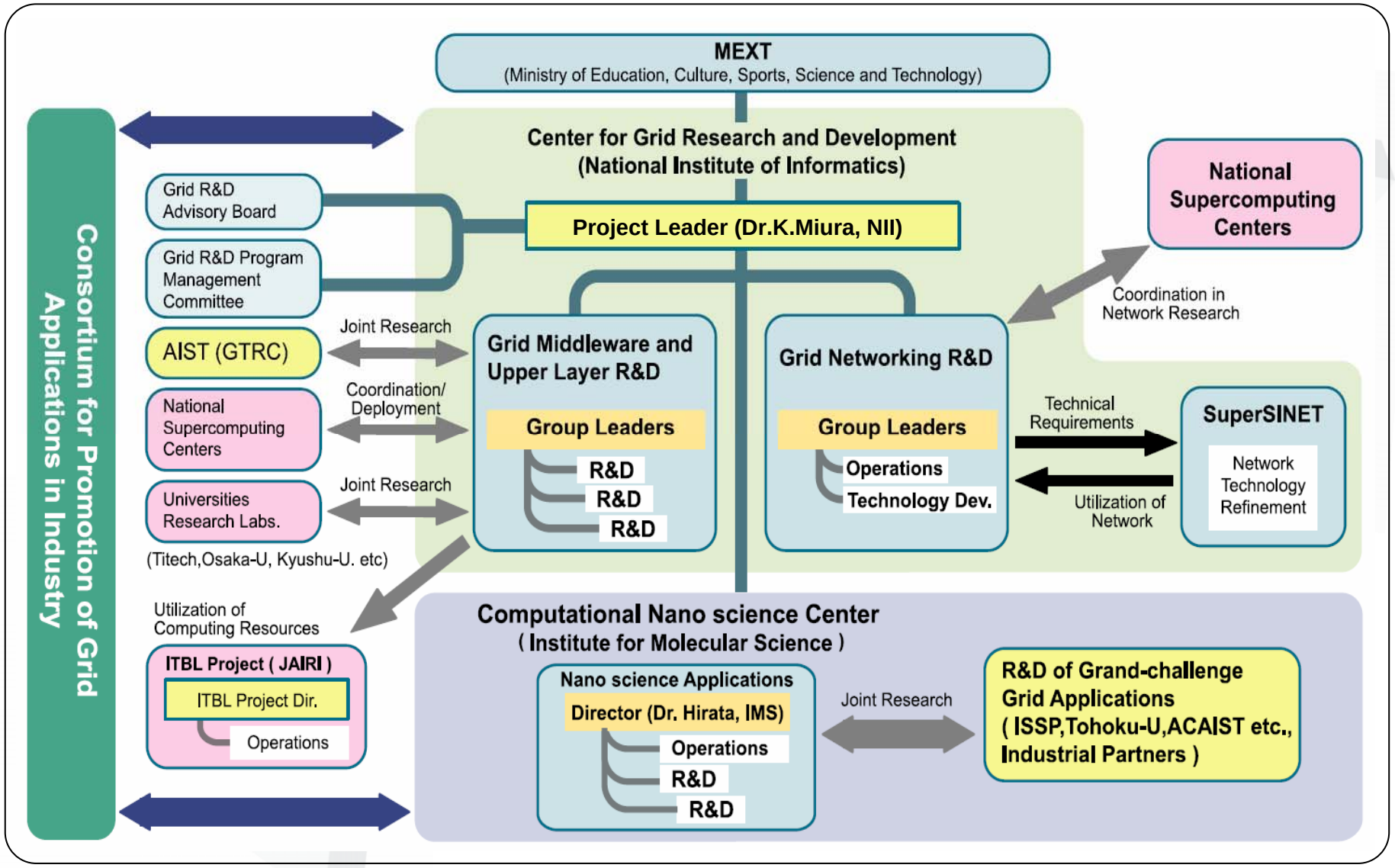
(1) To develop a Grid Software System (R&D in Grid Middleware and Upper Layer) as the prototype of future Grid Infrastructure in scientific research in Japan

(2) To provide a Testbed to prove that the High-end Grid Computing Environment (100+Tflop/s expected by 2007) can be practically utilized in the Nano-science Applications over the Super SINET.

(3) To Participate in International Collaboration (U.S., Europe, Asian Pacific)

(4) To Contribute to Standardization Activities, e.g., GGF

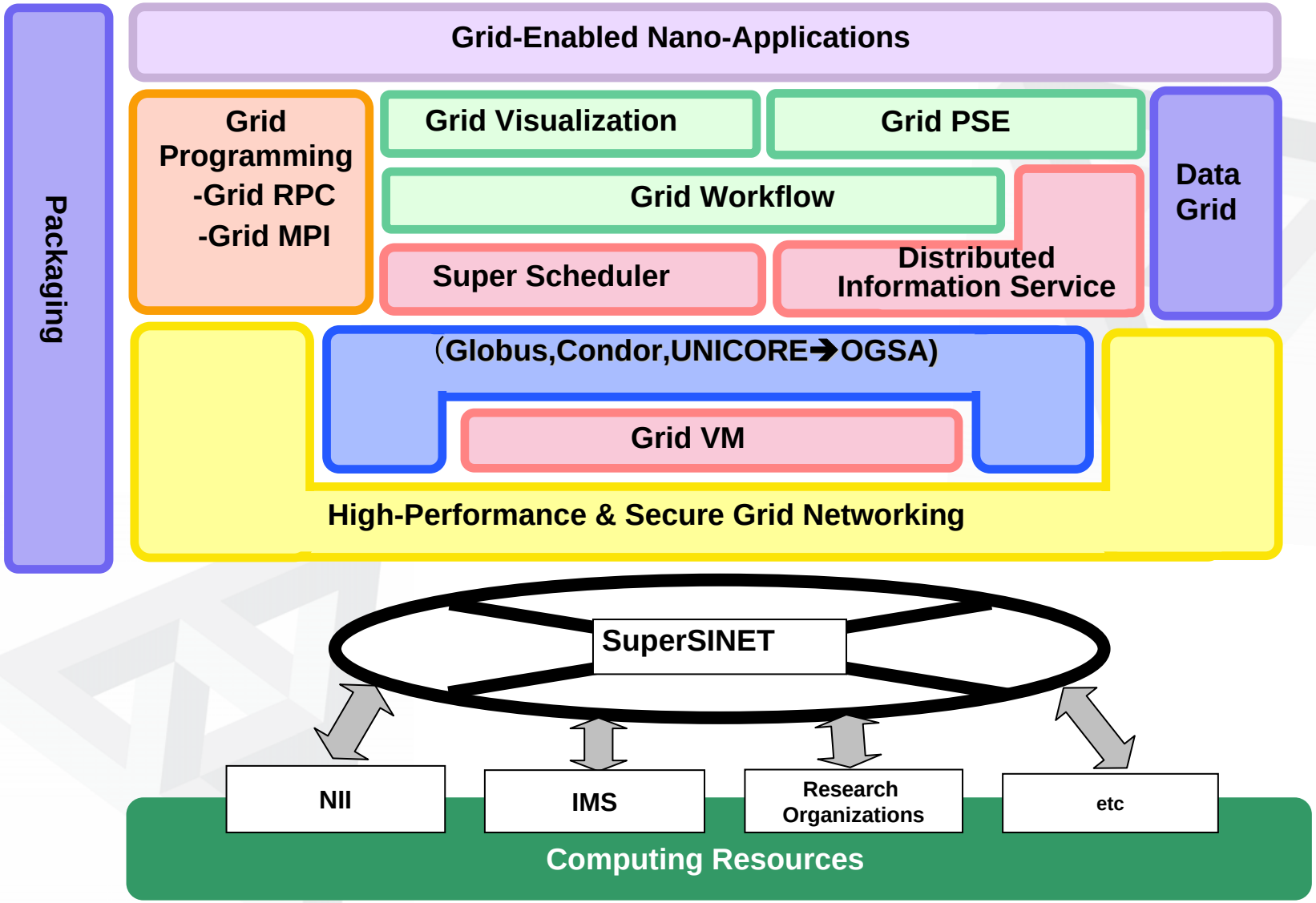
NAREGI Research Organization and Collaboration



Participating Organizations

- National Institute of Informatics (NII)
  (Center for Grid Research & Development)
- Institute for Molecular Science (IMS)
  (Computational Nano‐science Center)
- Universities and National Laboratories(Joint R&D)
  (AIST, Titech, Osaka-u, Kyushu-u, Kyushu Inst. Tech.,
  Utsunomiya-u, etc.)
- Research Collaboration
  (ITBL Project, National Supecomputing Centers, KEK, NAO etc.)
- Participation from Industry (IT and chemical/Material etc.)
  Consortium for Promotion of Grid Applications in Industry

NAREGI Software Stack

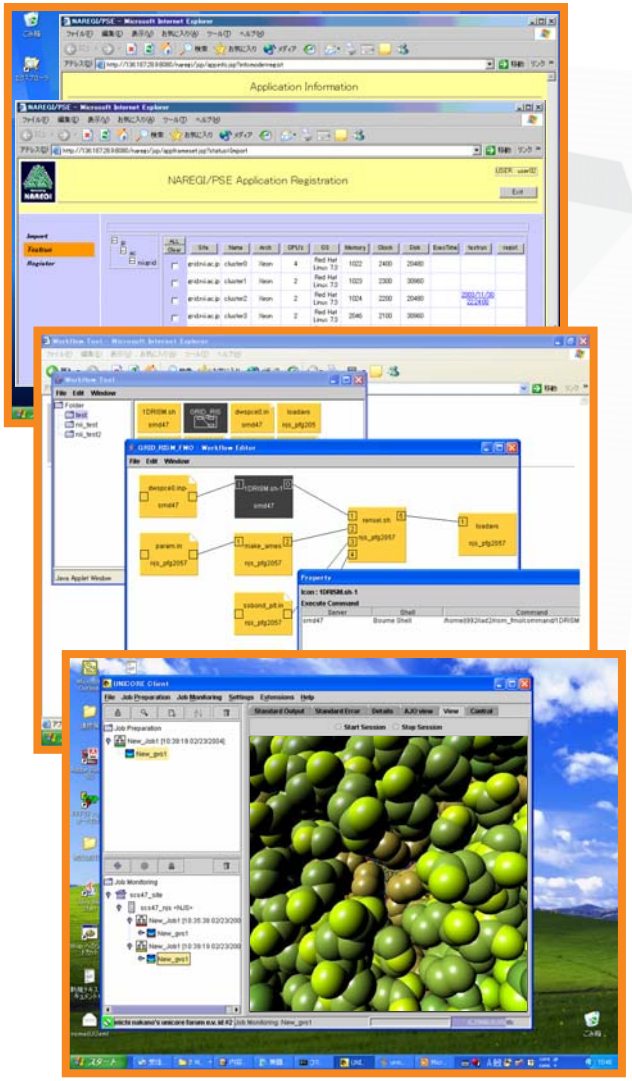


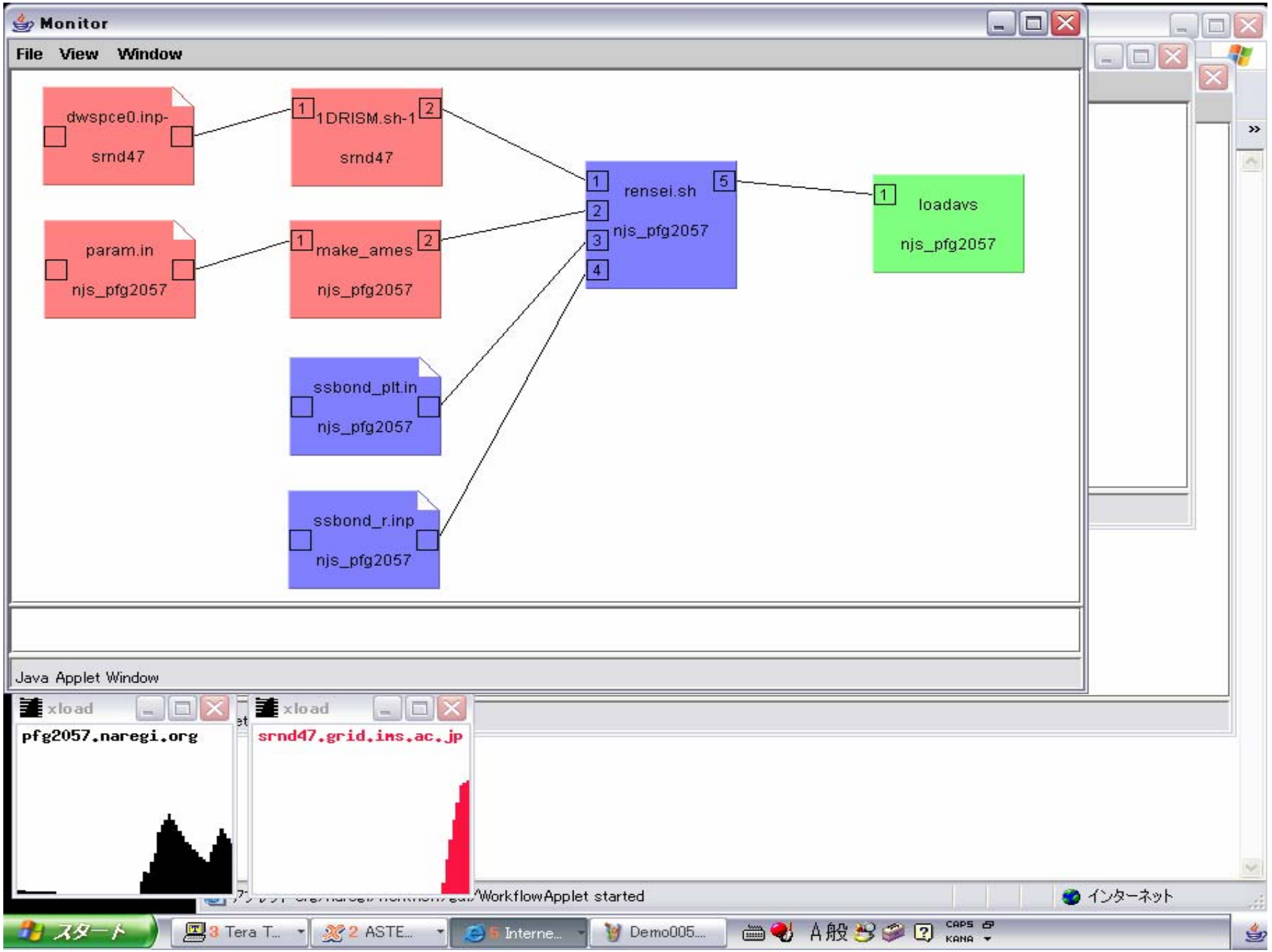
R&D in Grid Software and Networking Area (Work Packages)

- WP-1: Lower and Middle-Tier Middleware for Resource Management:
  Matsuoka (Titech), Kohno(ECU), Aida (Titech)
- WP-2: Grid Programming Middleware:
  Sekiguchi(AIST), Ishikawa(AIST)
- WP-3: User-Level Grid Tools & PSE:
  Usami (NII), Kawata(Utsunomiya-u)
- WP-4: Data Grid Environment
  Matsuda (Osaka-u)
- WP-5: Networking, Security & User Management
  Shimojo (Osaka-u), Oie (Kyushu Tech.),
  Imase（Osaka U.)
- WP-6: Grid-enabling tools for Nanoscience Applications :
  Aoyagi (Kyushu-u)

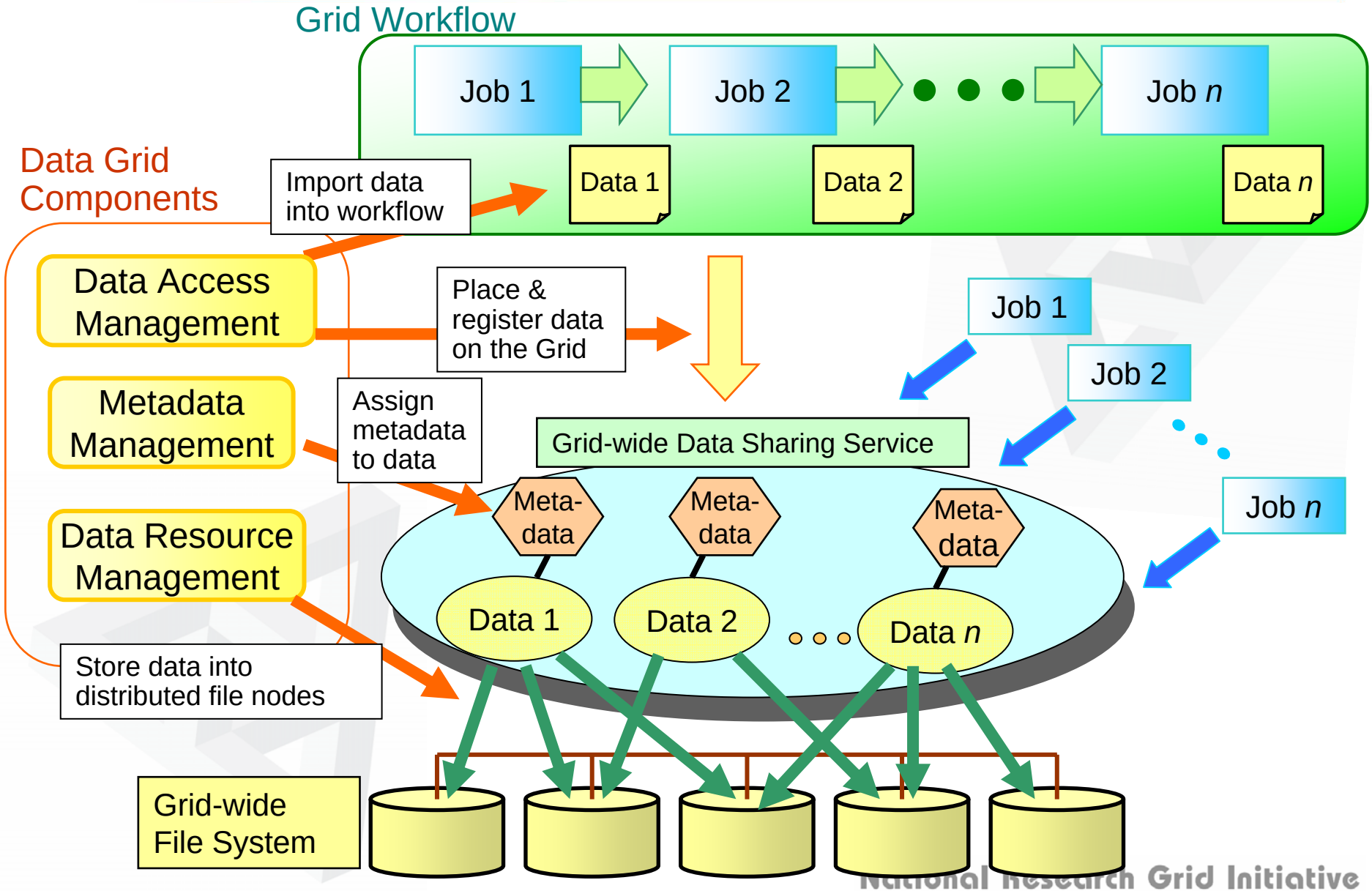
WP-3: User-Level Grid Tools & PSE

- **Grid PSE**
  - Deployment of applications on the Grid
  - Support for execution of deployed applications
- **Grid Workflow**
  - Workflow language independent of specific Grid middleware
  - GUI in task-flow representation
- **Grid Visualization**
  - Remote visualization of massive data distributed over the Grid
  - General Grid services for visualization




Monitor
File View Window
dwspce0.inp-
srnd47
1DRISM.sh-1
srnd47
param.in
njs_pfg2057
make_ames
njs_pfg2057
rensei.sh
njs_pfg2057
loadavs
njs_pfg2057
ssbond_plt.in
njs_pfg2057
ssbond_r.inp
njs_pfg2057
Java Applet Window
xload
pfg2057.naregi.org
xload
srnd47.grid.ims.ac.jp
WorkflowApplet started
インターネット
スタート
3 Tera T...
2 ASTE...
5 Interne...
Demo005...

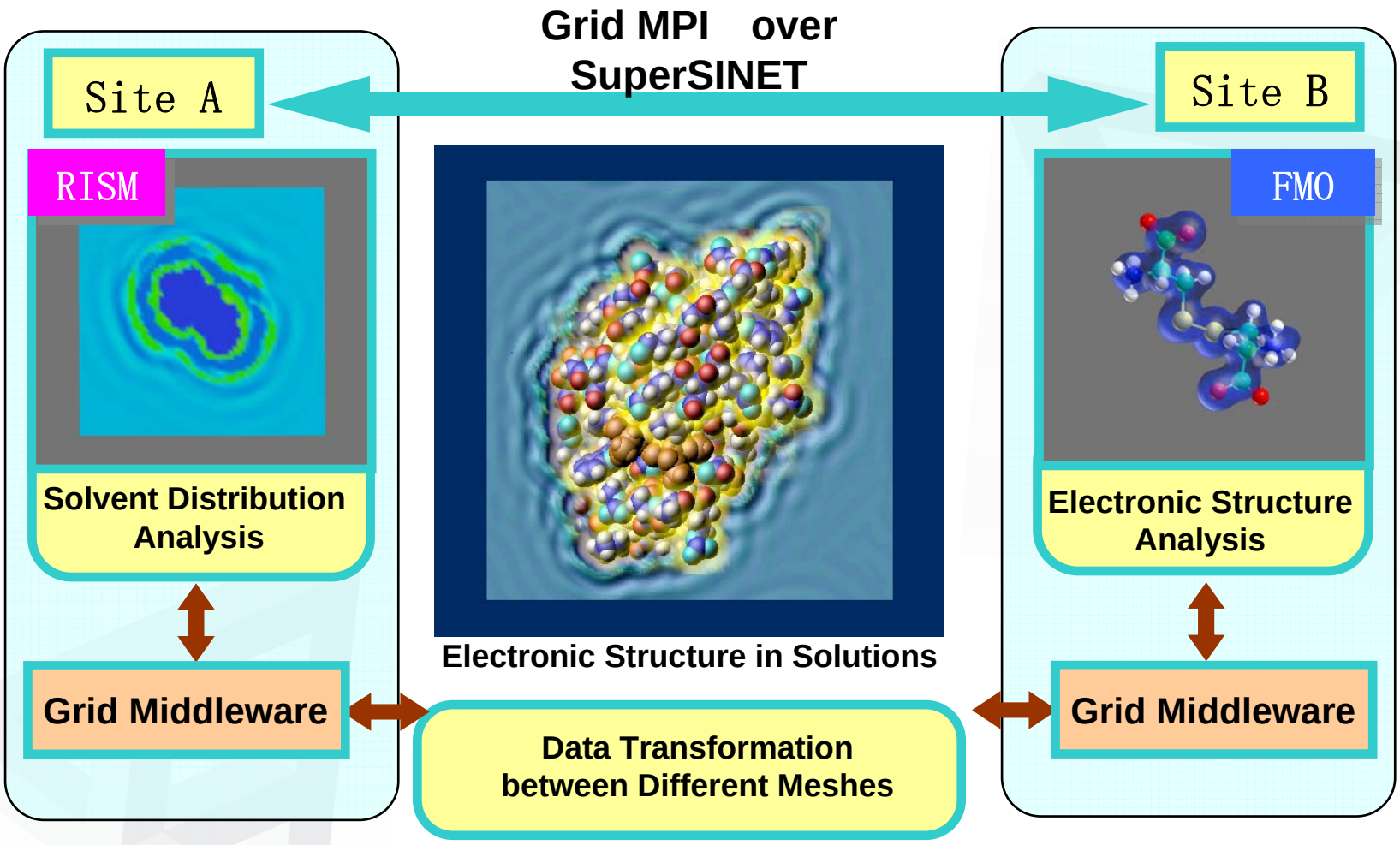
WP-4 : NAREGI Data Grid β-1 Architecture



Collaboration in Data Grid Area

- High Energy Physics
  - KEK
  - EGEE
- Astronomy
  - National Astronomical Observatory (Virtual Observatory)
- Bio-informatics
  - BioGrid Project

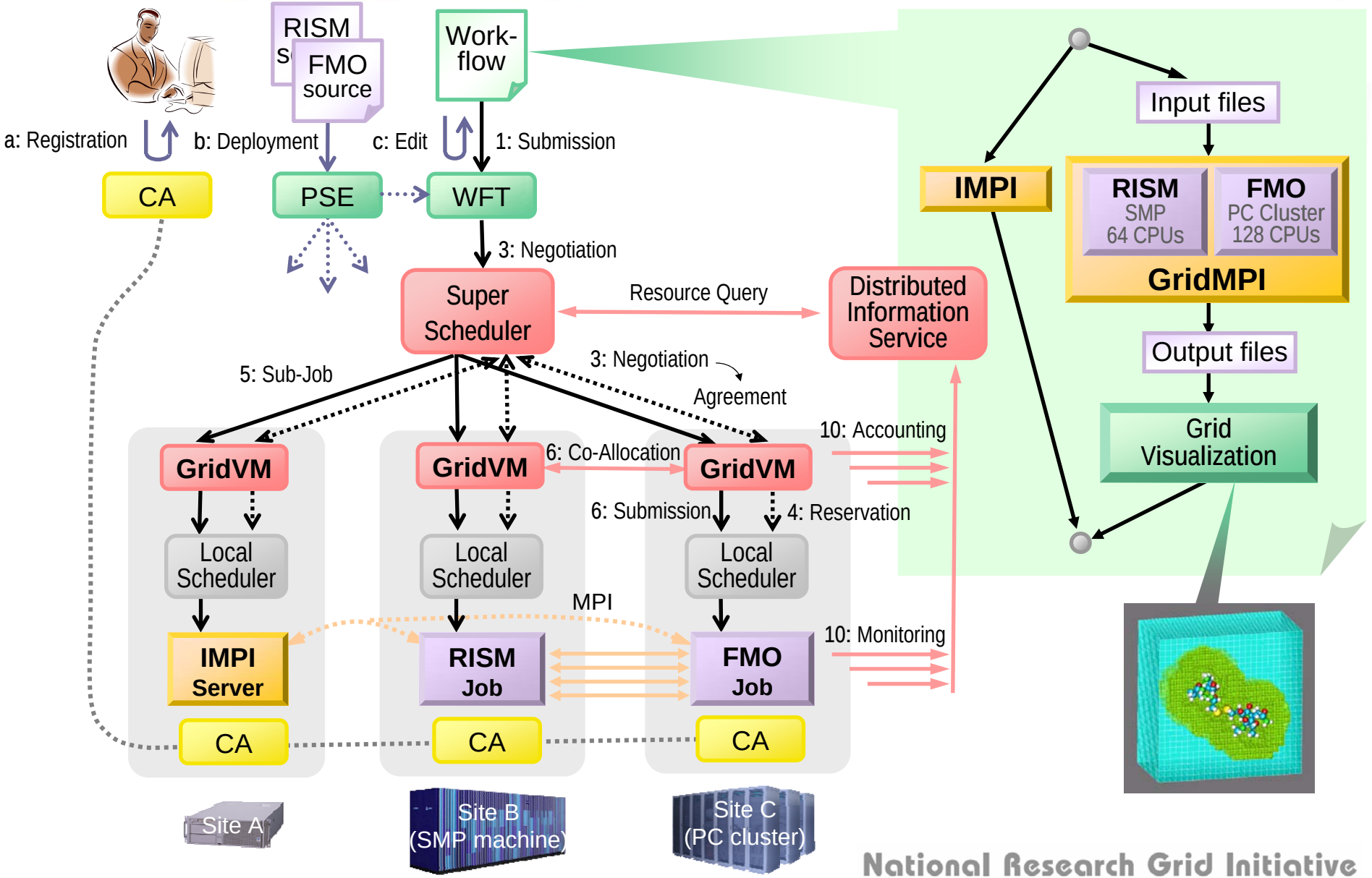
WP-6:Adaptation of Nano-science Applications to Grid Environment



RISM Reference Interaction Site Model

FMO Fragment Molecular Orbital method

Scenario for Multi-sites MPI Job Execution



NAREGI α-Version Middleware (2004-5)

1. Resource Management (WP1)

- **Unicore/Globus/Condor as "Skeletons"**
- **OGSA-EMS Job Management, Brokering, Scheduling**
- Coupled Simulation on Grid (Co-Allocation/Scheduling)
- WS(RF)-based Monitoring/Accounting Federation

First OGSA-EMS based implementation In the world

2. Programming (WP2)

- High-Performance Standards-based GridMPI (MPI-1/2, IMPI)
- Highly tuned for Grid environment (esp. collectives)
- **Ninf-G: Scientific RPC for Grids**

GridRPC GGF Programming Standard

3. Applications Environment (WP3)

- Application Discovery and Installation Management (PSE)
- Workflow Tool for coupled application interactions
- **WSRF-based Large Scale Grid-based Visualization**

WSRF-based Terabyte interactive visualization

4. Networking and Security (WP5)

- **Production-Grade CA**
- **Out-of-band Real-time network monitor/control**

Drop-in Replacement for SimpleCA

5. Grid-enabling Nano Simulations (WP6)

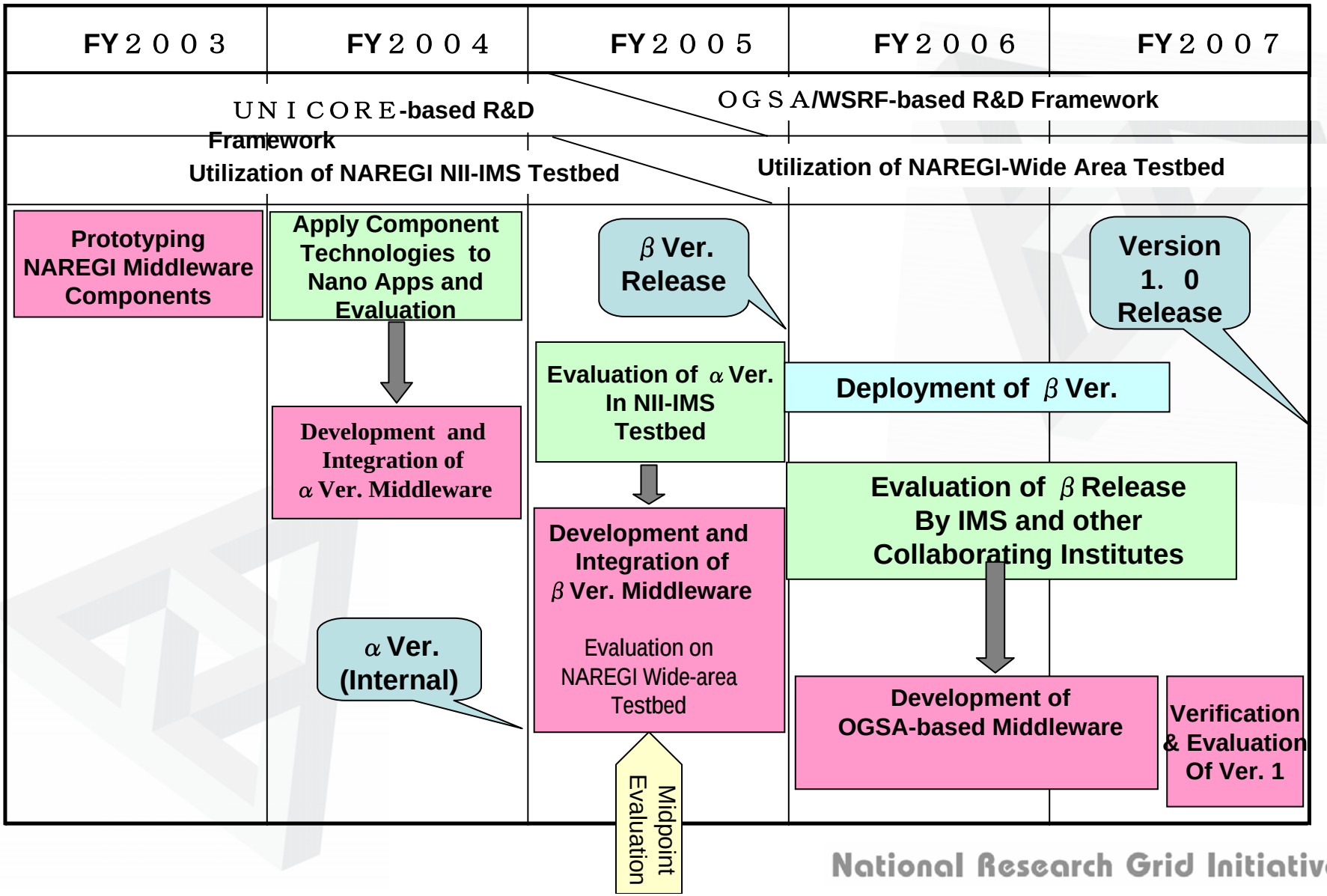
- **Framework for coupled simulations**

Large scale coupled simluation of proteins in a solvent on a Grid

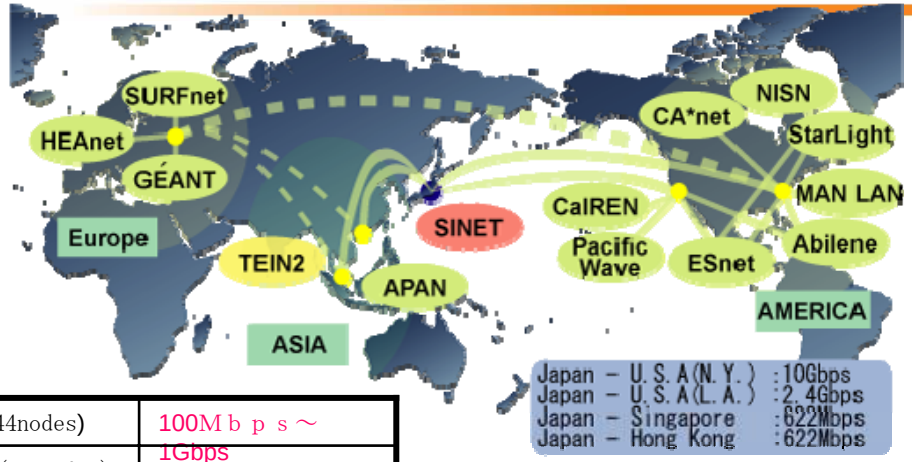
Highlights of NAREGI Beta (2005-2006)

- GT4/WSRF based “full” OGSA-EMS incarnation
  - OGSA-EMS/RSS WSRF components --- no legacy (pre-WS)Unicore/Globus2 dependencies
  - WS-Agreement brokering and co-allocation
  - JSDL-based job submission to WS-GRAM
  - Support for more OSes (AIX, Solaris, etc.) and BQs
- Sophisticated VO support for identity/security/monitoring/accounting (extensions of VOMS/MyProxy, WS-* adoption)
- WS- Application Deployment Support
- Grid-wide file system (GFarm) and other data management tools
- Complex workflow for various coupled simulations
- Overall stability/speed/functional improvements for real deployment

Roadmap of NAREGI Grid Middleware



Network Topology Map of SINET/SuperSINET(Feb. 2006)



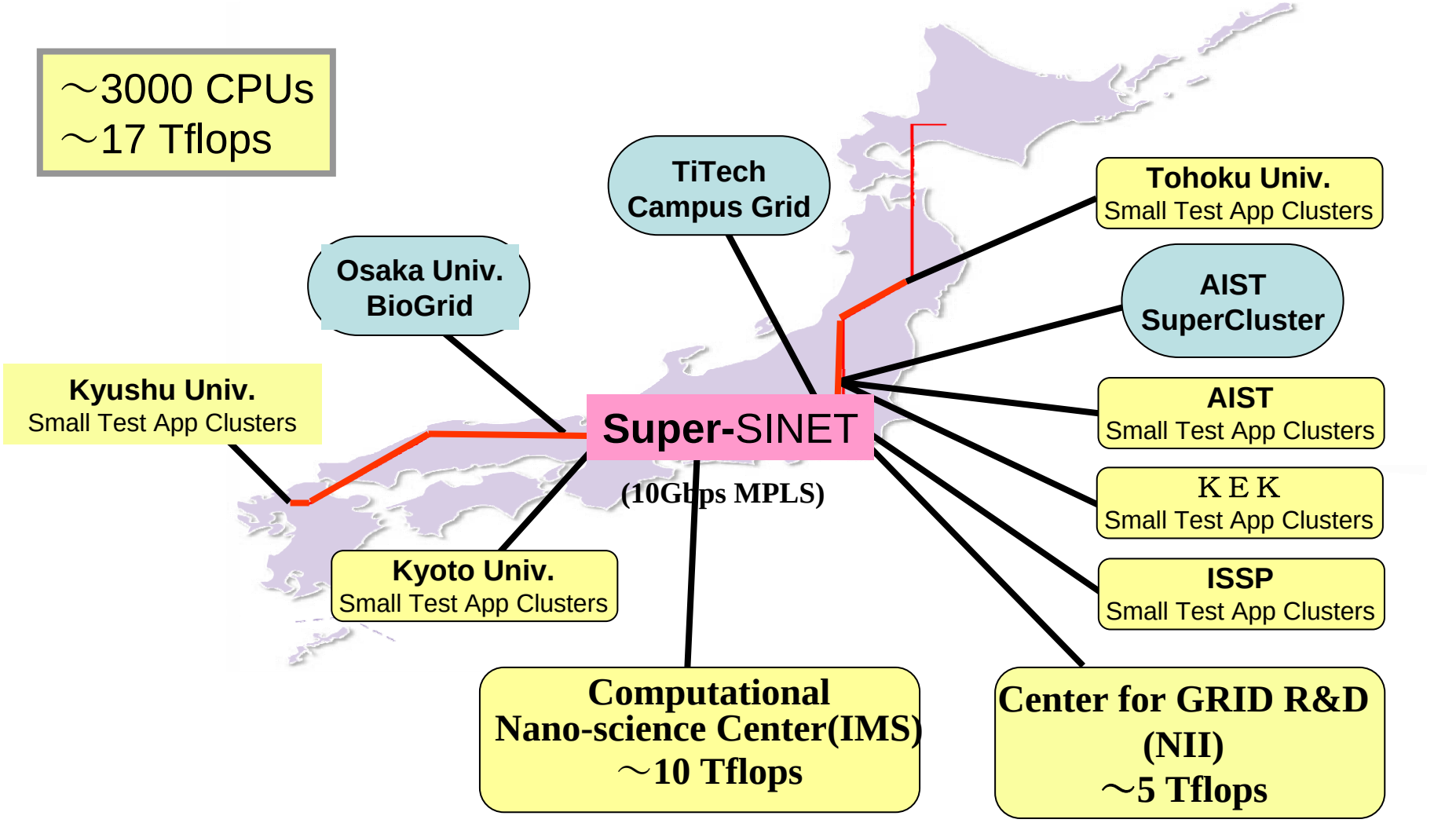
ＳＩＮＥＴ (44nodes)	100Ｍｂｐｓ～1Gbps
Super SINET (32nodes)	10Gbps
International lines	Japan – U.S.A 10Gbps(To:N.Y.) 2.4Gbps(To:L.A) Japan - Singapore 622Mbps Janap – Hong Kong 622Mbps

(Line speeds)

Number of SINET particular Organizations (2006.2.1)

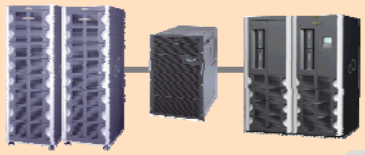
National	Public	Private	Junior Colleges	Specialized Training Colleges	Inter-Univ. Res. Inst. Corp.	Others	Total
81	51	273	68	41	14	182	710

NAREGI Phase 1 Testbed



Computer System for Grid Software Infrastructure R & D

Center for Grid Research and Development （5 Tflops，700GB)



File Server (PRIMEPOWER 900 ＋ETERNUS3000 ＋ ETERNUS)

1node／8CPU (SPARC64V1.3GHz)

Memory 16GB
Storage 10TB
Back-up Max.36.4TB

SMP type Compute Server (PRIMEPOWER HPC2500)

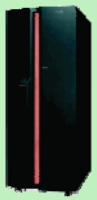
1node (UNIX, SPARC64V1.3GHz/64CPU)

Memory 128GB
Storage 441GB

SMP type Compute Server (SGI Altix3700)

1node
(Itanium2 1.3GHz/32CPU)

Memory 32GB
Storage 180GB



SMP type Compute Server (IBM pSeries690)

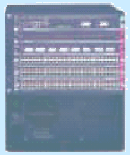
1node
(Power4 1.3GHz/32CPU)

Memory 64GB
Storage 480GB

Intra NW

L3 SW 1Gbps (upgradable To 10Gbps)

Ext. NW



L3 SW 1Gbps (Upgradable to 10Gbps)

SuperSINET

Intra NW-A

High Perf. Distributed-memory type Compute Server (PRIMERGY RX200)

128 CPUs(Xeon, 3.06GHz)+Control Node

InfiniBand 4X(8Gbps)

Memory 130GB
Storage 9.4TB

High Perf. Distributed-memory Type Compute Server (PRIMERGY RX200)

128 CPUs(Xeon, 3.06GHz)+Control Node

InfiniBand 4X (8Gbps)

Memory 65GB
Storage 9.4TB

Intra NW-B

Distributed-memory type Compute Server (Express 5800)

128 CPUs (Xeon, 2.8GHz)+Control Node

GbE (1Gbps)

Memory 65GB
Storage 4.7TB

Distributed-memory type Compute Server(Express 5800)

128 CPUs (Xeon, 2.8GHz)+Control Node

GbE (1Gbps)

Memory 65GB
Storage 4.7TB

Distributed-memory type Compute Server(HPC LinuxNetworx)

128 CPUs (Xeon, 2.8GHz)+Control Node

GbE (1Gbps)

Memory 65GB
Storage 4.7TB

Distributed-memory type Compute Server(HPC LinuxNetworx)

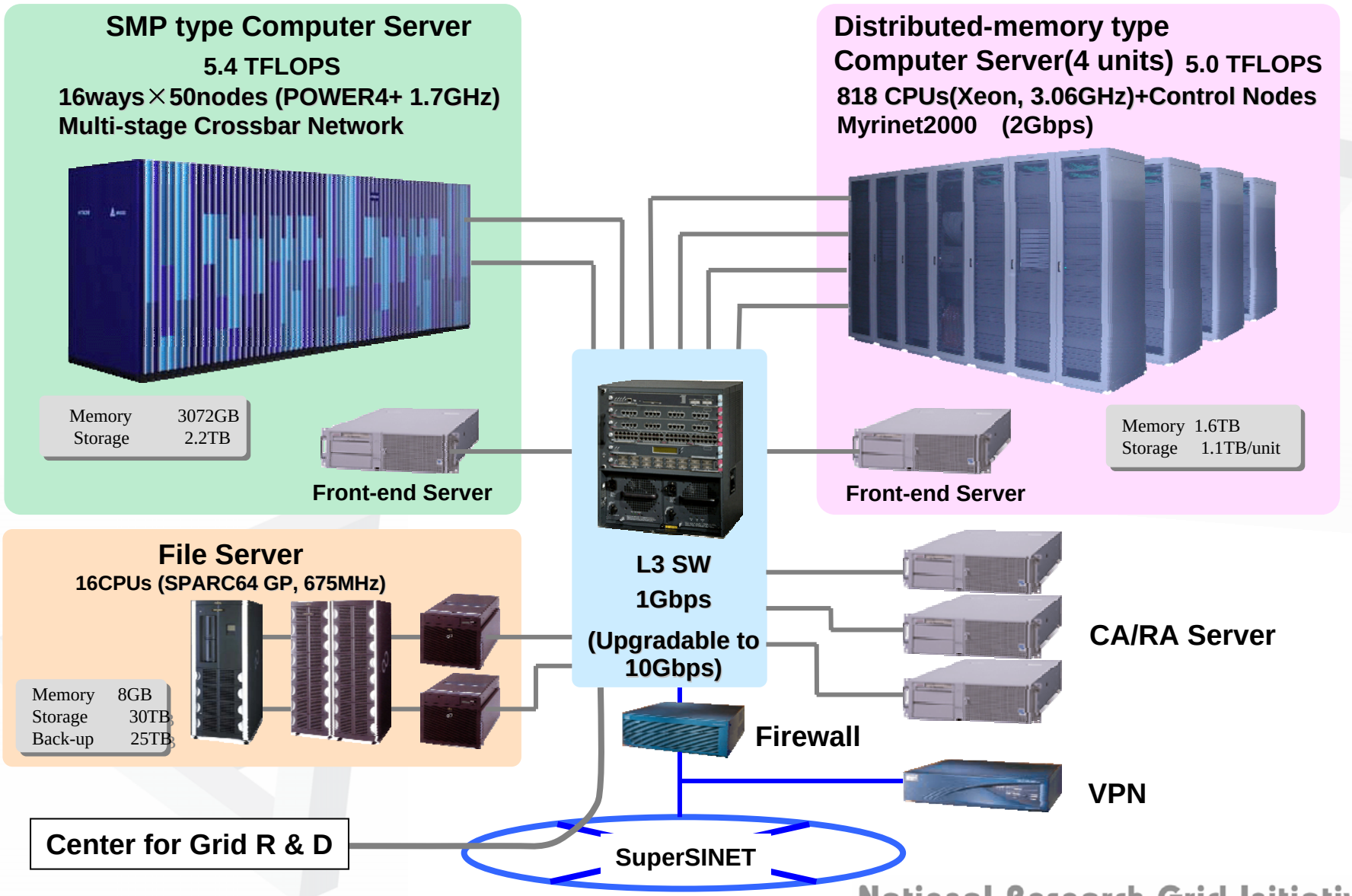
128 CPUs(Xeon, 2.8GHz)+Control Node

GbE (1Gbps)

Memory 65GB
Storage 4.7TB

Computer System for Nano Application R & D

Computational Nano science Center （10Ｔflops，5TB)



Cyber-Science Infrastructure for R & D

Cyber-Science Infrastructure （CSI)

NAREGI Outputs

NII-REO (Repository of Electronic Journals and Online Publications

GeNii (Global Environment for Networked Intellectual Information)

Virtual Labs Live Collaborations

Deployment of NAREGI Middleware

UPKI: National Research PKI Infrastructure

Management Body / Education & Training

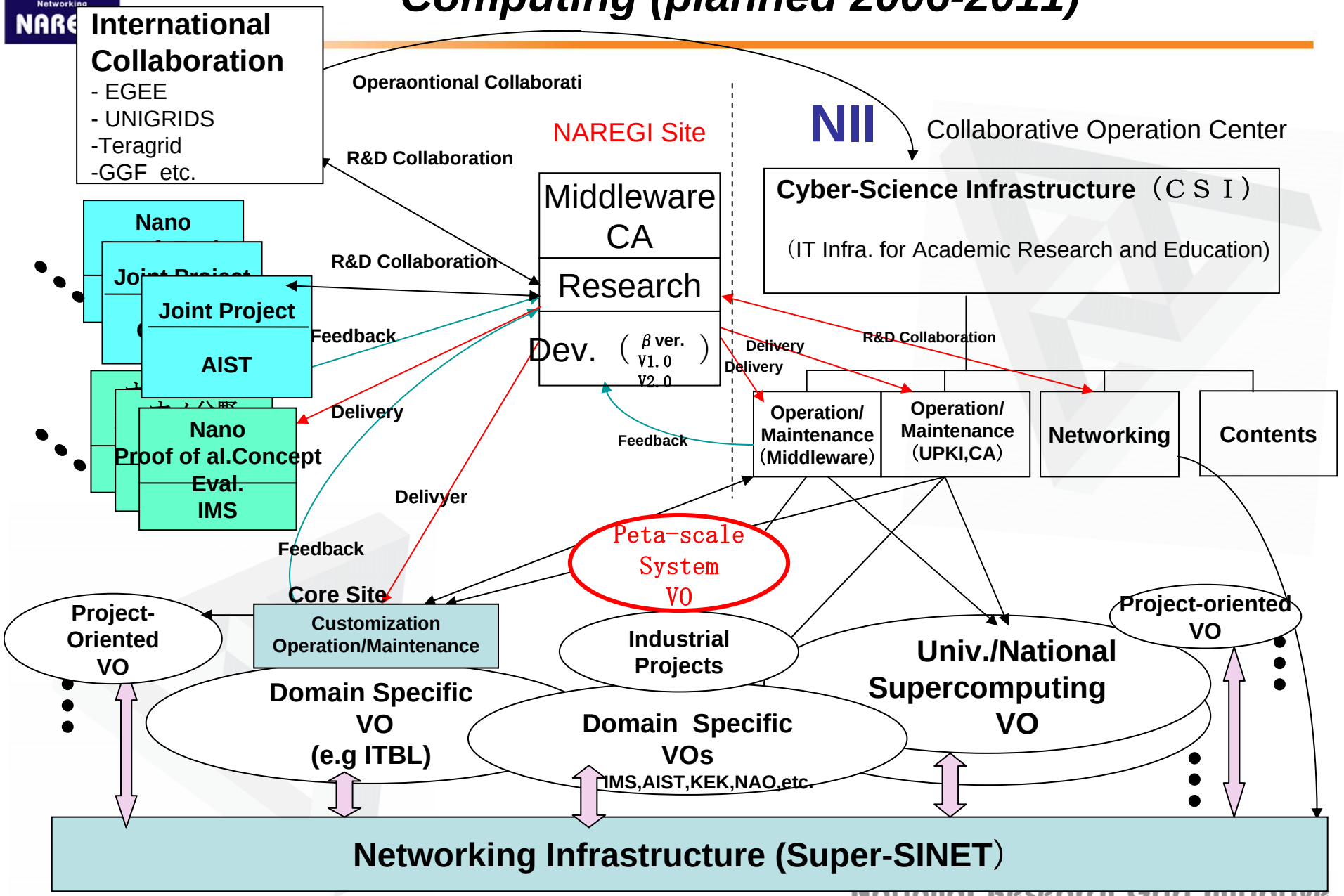
SuperSINET and Beyond: Lambda-based Academic Networking Backbone

Restructuring Univ. IT Research Resources
Extensive On-Line Publications of Results

Industry/Societal Feedback

International Infrastructural Collaboration

Cyber Science Infrastructure toward Petascale Computing (planned 2006-2011)



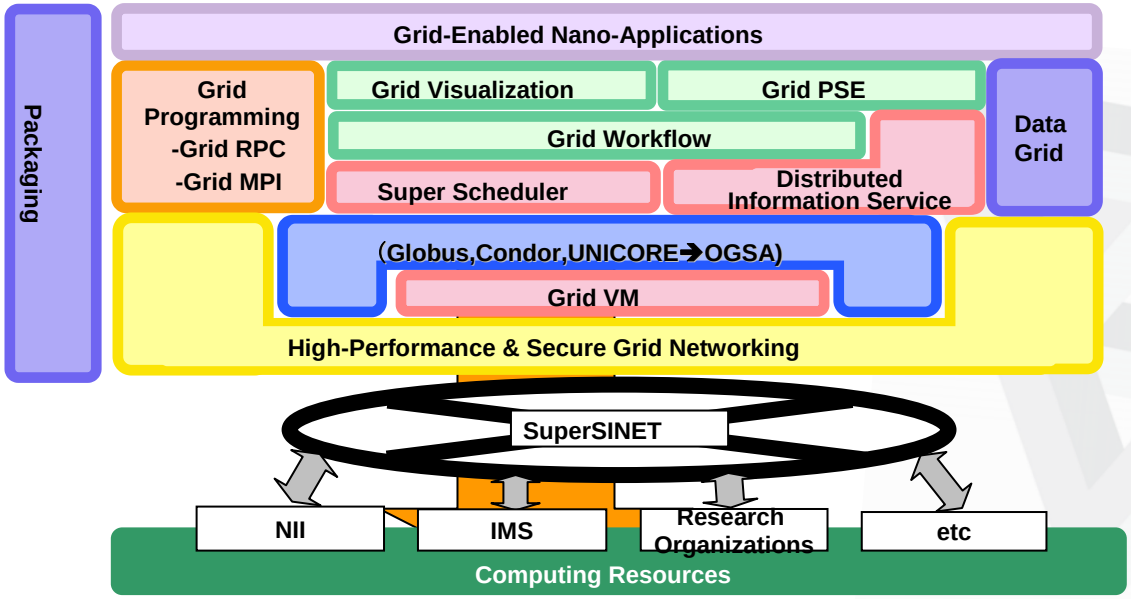
Note: names of VO are tentative）

Summary

- In the NAREGI project, seamless federation of heterogeneous resources is the primary objective
- Computations in Nano-science/technology applications over Grid is to be promoted, including participation from industry.
- Data Grid features has been added to NAREGI since FY'05.
- NAREGI Grid Middleware is to be adopted as one of the important components in the new Japanese Cyber Science Infrastructure Framework.
- NAREGI Project will provide the VO capabilities for the National Peta-scale Computing Project
- International Co-operation is essential.

Additional Slides

Summary of NAREGI Project



Middleware for Science Grid Environment

NAREGI V1.0

Cyber Science Infrastructure

- Grid Middleware for Large Computer Centers
- Productization of General-purpose Grid Middleware for Scientific Computing
- Personnel Training (IT and Application Engineers)
- Contribution to International Scientific Community and Standardization

Vitalization of Industry

- Progress in the Latest Research and Development (Nano, Biotechnology)
- Use in Industry (New Intellectual Product Development)

• •

MEXT: Ministry of Education, Culture, Sports, Science and Technology

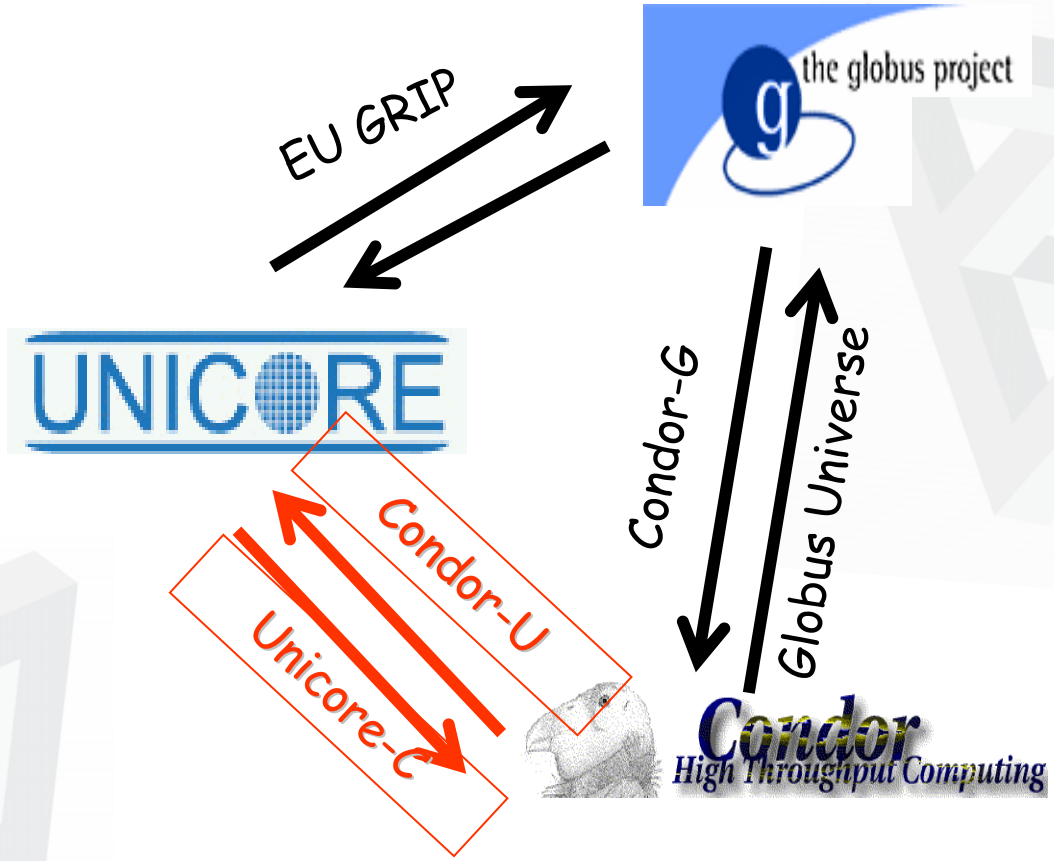
NAREGI Middleware Objectives

- **Mid-range Project Target(β Version)**
  - R&D on scalable middleware infrastructure for server grids and metacomputing resource management, deployable at large computing centers, based on Unicore, Globus and Condor
- **Final Project Target (Version 1)**
  - R&D on building resource management framework and middleware for VO hosting by the centers, based on the OGSA standard
  - Distribute result as high-quality open source software

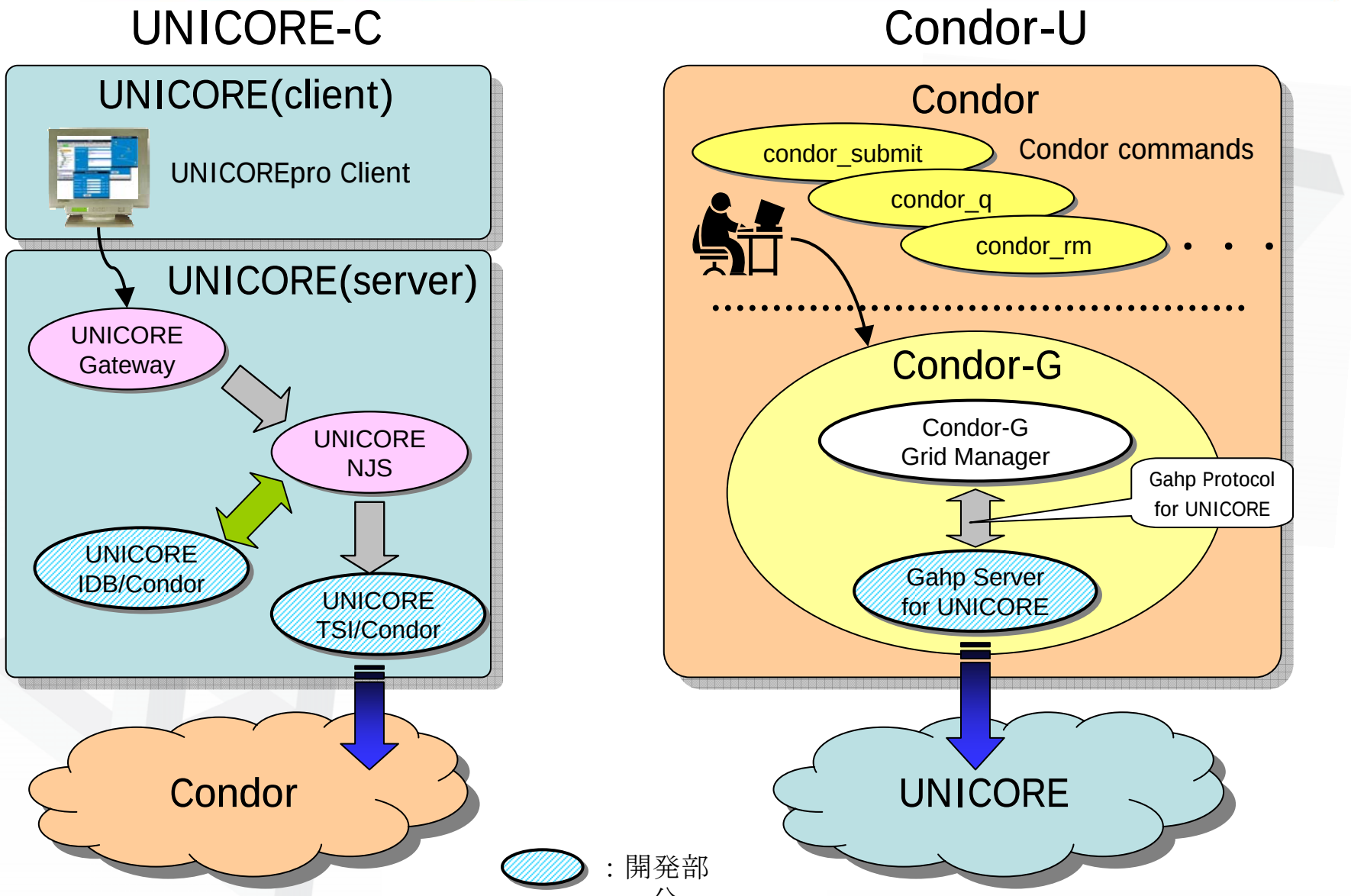
Features of Grid Middleware(WP-1)

- **(Scalable) Super/Meta Scheduler**
  - Schedule large metacomputing jobs
  - “Scalable”, Agreement-based scheduling
  - Assume preemptive metascheduled jobs
- **(Scalable) Grid Information Service**
  - Support multiple monitoring systems
  - User and job auditing, accounting
  - CIM-based node information schema
- **GridVM (Lightweight Grid Virtual Machine)**
  - Metacomputing Support
  - Enforcing Authorization Policies, Sandbox
  - Checkpointing/FT/Preemption
- **Authentication Service**
  - Along with GGF defined assurance level
  - Authentication mechanism across policy domains

UNICORE-CONDOR Linkage



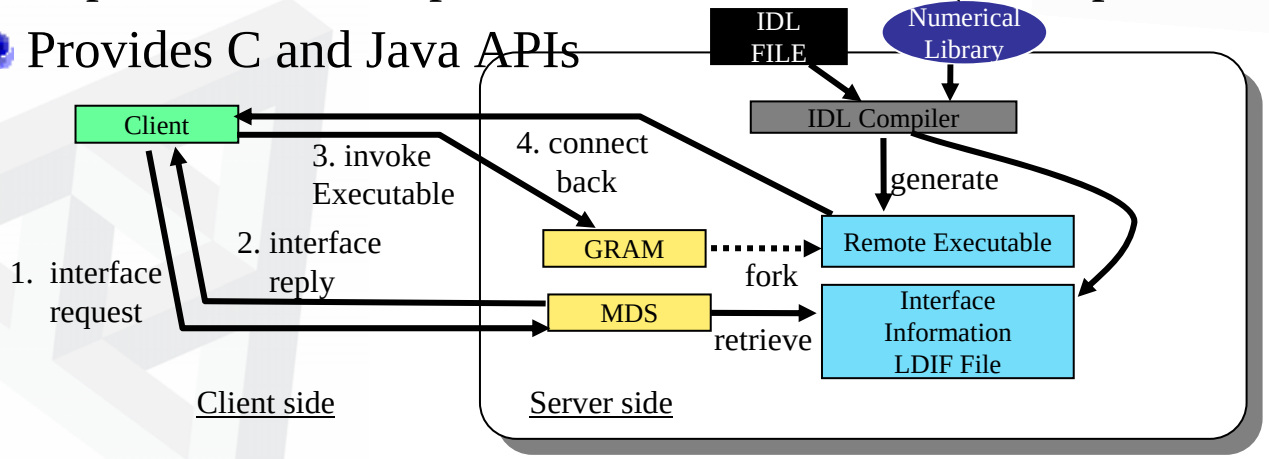
UNICONDORE Architecture



WP-2:Grid Programming – GridRPC/Ninf-G2 (AIST/GTRC)

GridRPC

- Programming Model using RPC on the Grid
- High-level, taylored for Scientific Computing (c.f. SOAP-RPC)
- GridRPC API standardization by GGF GridRPC WG
- Ninf-G Version 2
  - A reference implementation of GridRPC API
  - Implemented on top of Globus Toolkit 2.0 (3.0 experimental)
  - Provides C and Java APIs



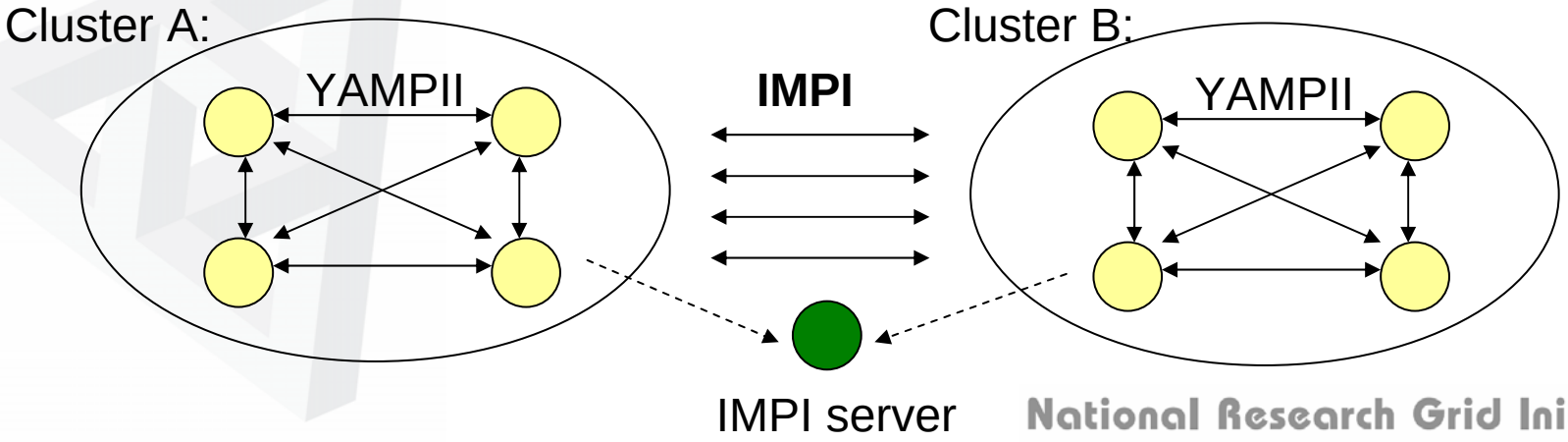
WP-2:Grid Programming -GridMPI (AIST and U-Tokyo)

■ GridMPI is a library which enables MPI communication between parallel systems in the grid environment. This realizes;

- Huge data size jobs which cannot be executed in a single cluster system
- Multi-Physics jobs in the heterogeneous CPU architecture environment

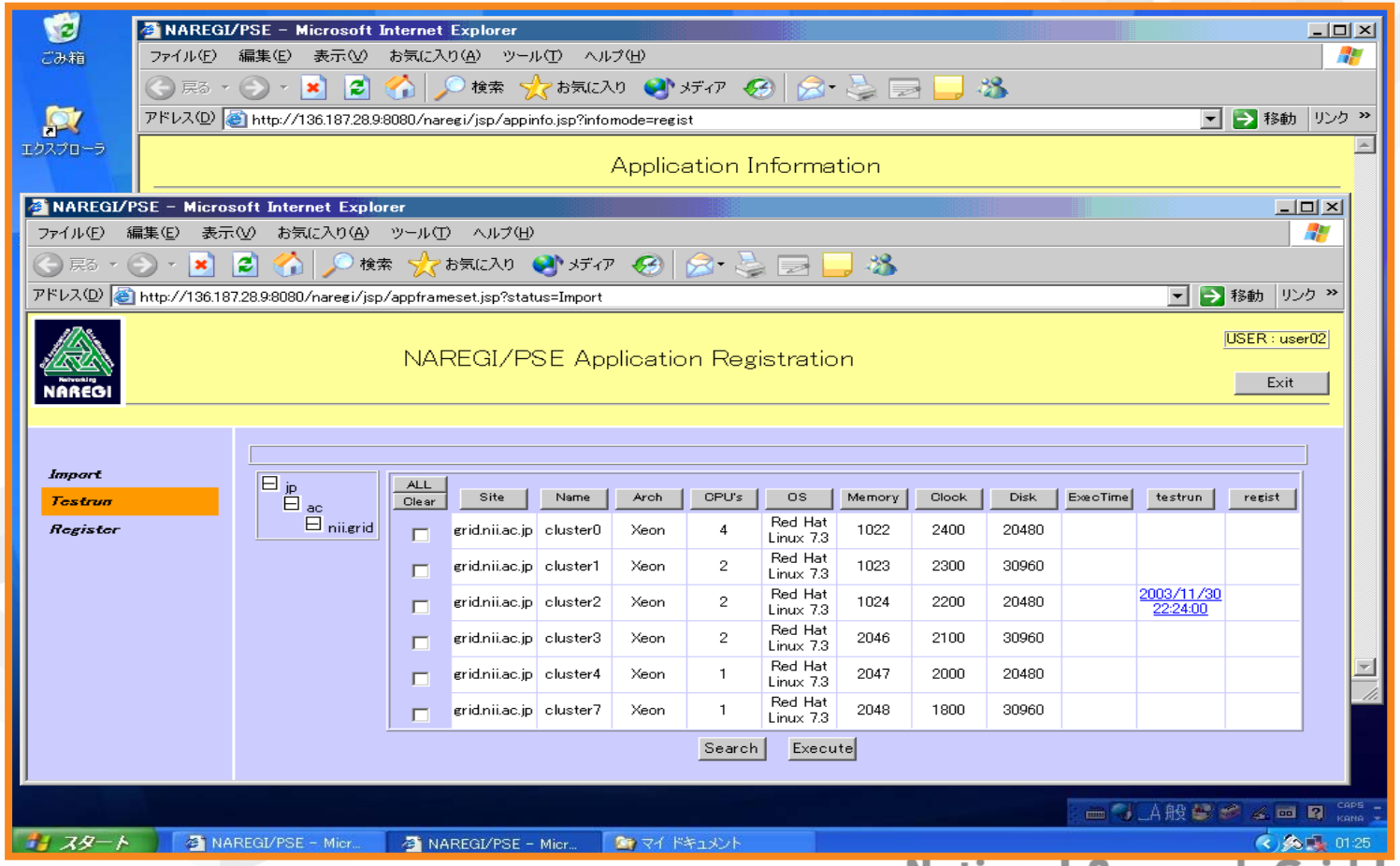
① Interoperability:
- IMPI（Interoperable MPI） compliance communication protocol
- Strict adherence to MPI standard in implementation

② High performance:
- Simple implementation
- Buit-in wrapper to vendor-provided MPI library



Grid PSE: Problem Solving Environment (PSE)

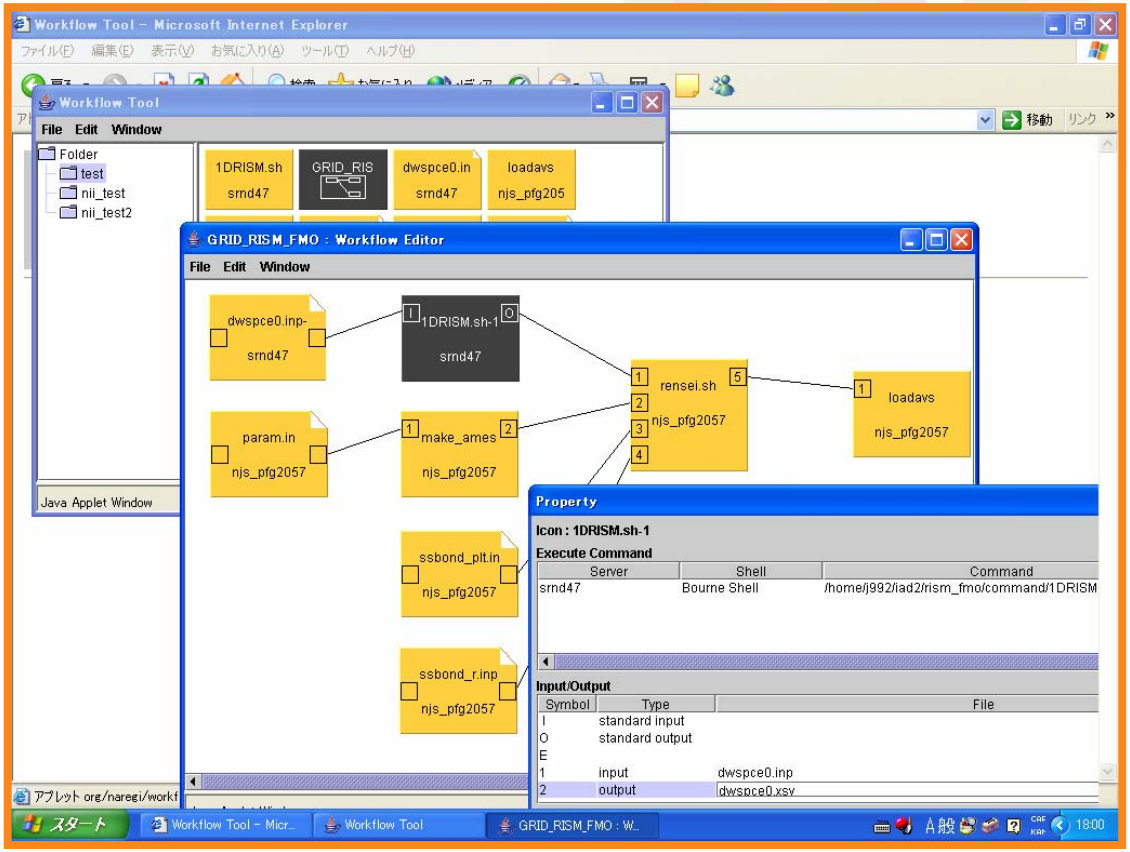
PSE automatically selects appropriate machines on the grid to execute application program before executing program.



Grid Workflow:

Web-based GUI

- Icon: program or data
- Line: execution order
- Middleware independent
  (The GUI does not use UNICORE features.)



Development of Meta-applications

1) Registering components

Registering resources of components needed in meta-applications

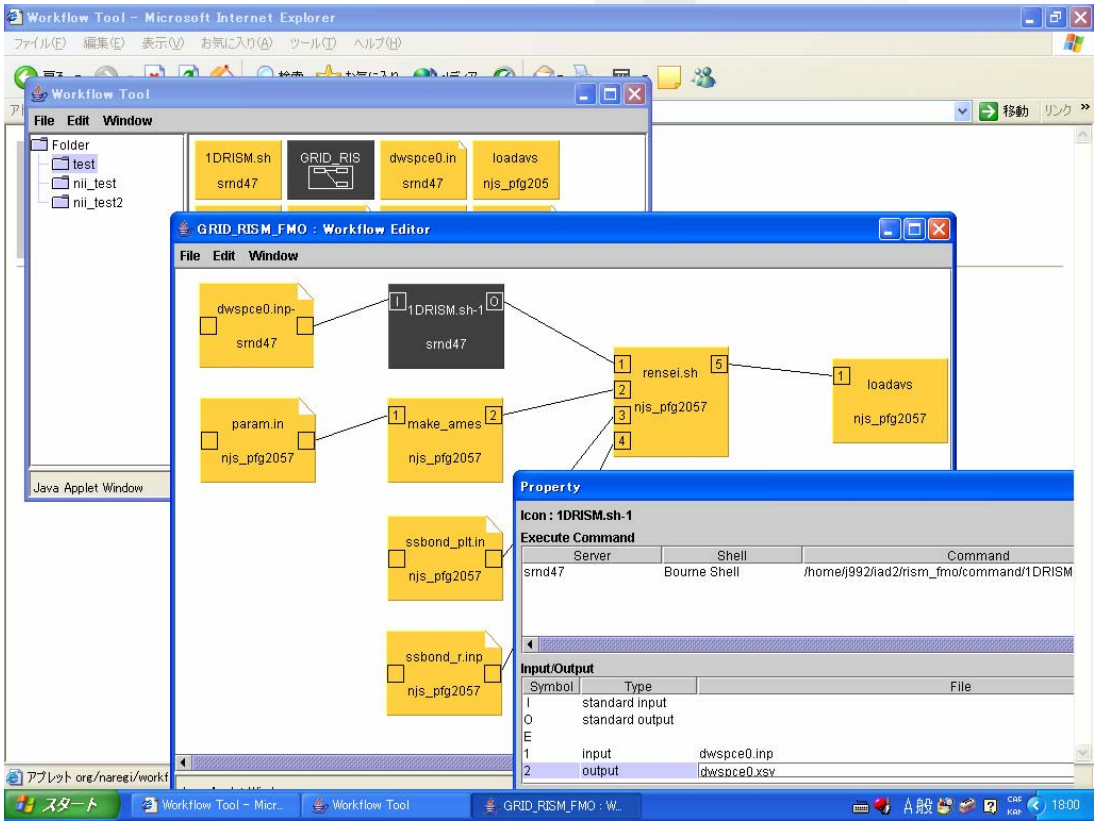
eg. Program, data, target computer,..

2) Connecting components

Specifying dependency between components

3)Monitoring execution

Monitoring execution of each component



Grid Visualization:
Visualization of Distributed Massive Data

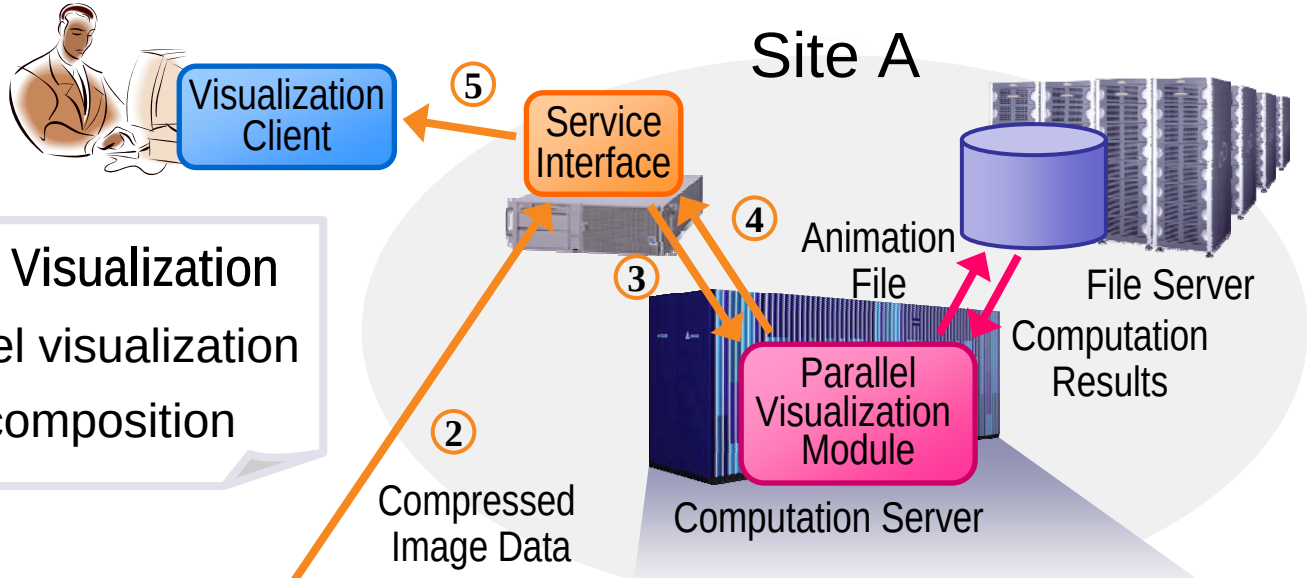


Image-Based Remote Visualization

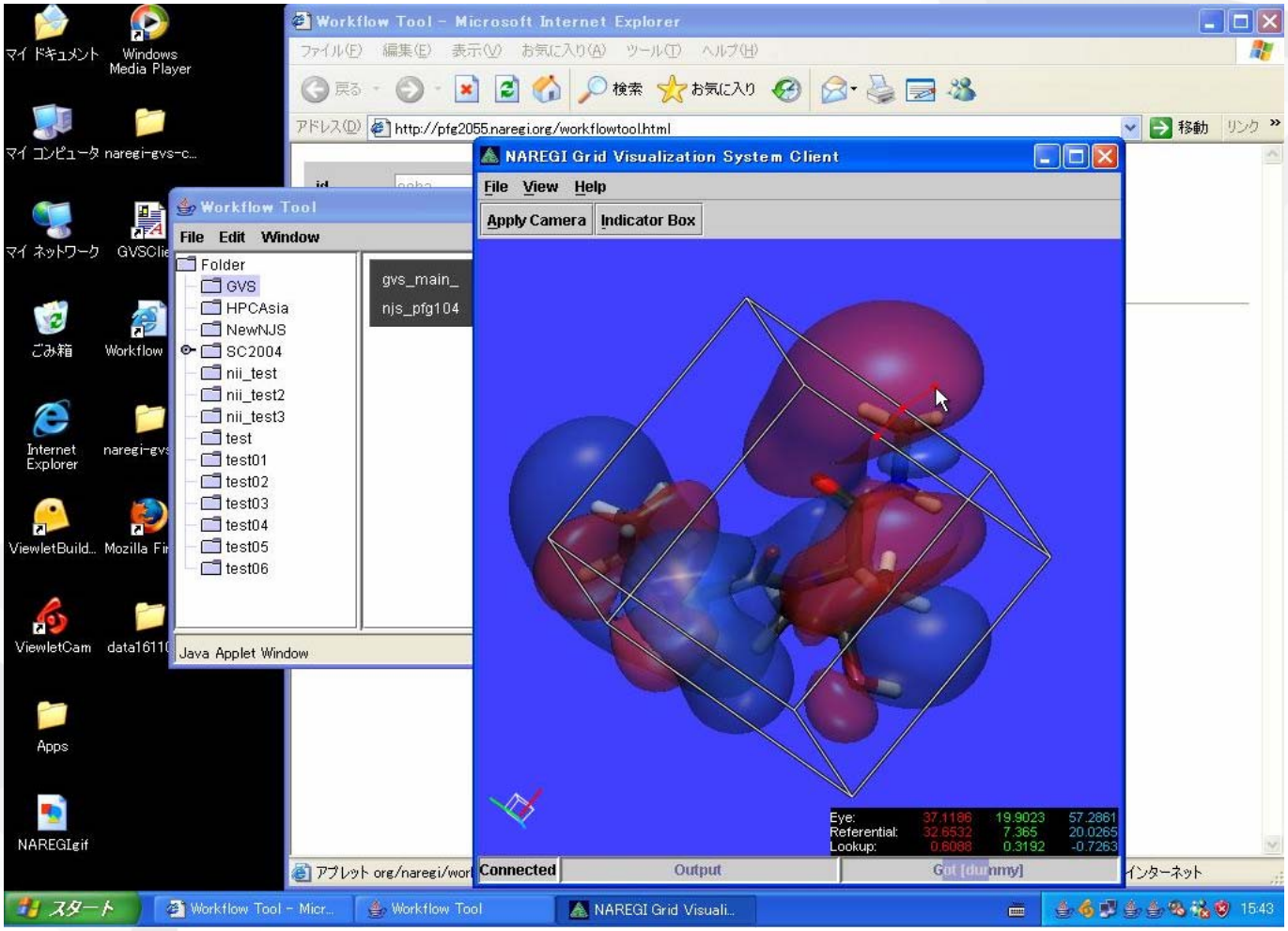
- Server-side parallel visualization
- Image transfer & composition

Advantages

- No need of rich CPUs, memories, or storages for clients
- No need to transfer large amounts of simulation data
- No need to merge distributed massive simulation data
- Light network load
- Visualization views on PCs anywhere
- Visualization in high resolution without constraint

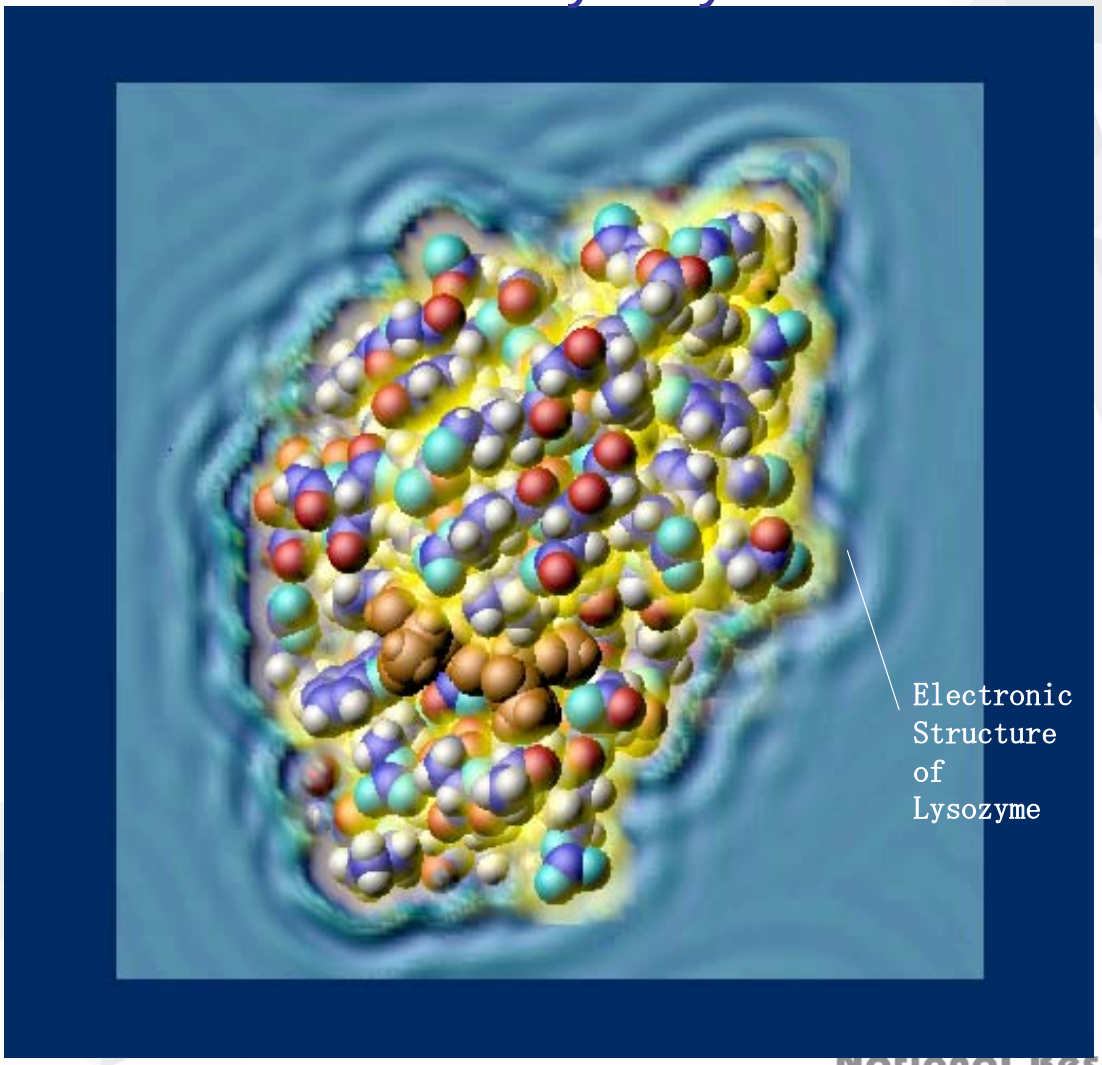
Grid Visualization:

Intermediate and final computed results on the distributed machines can be displayed by Grid Visualization System.

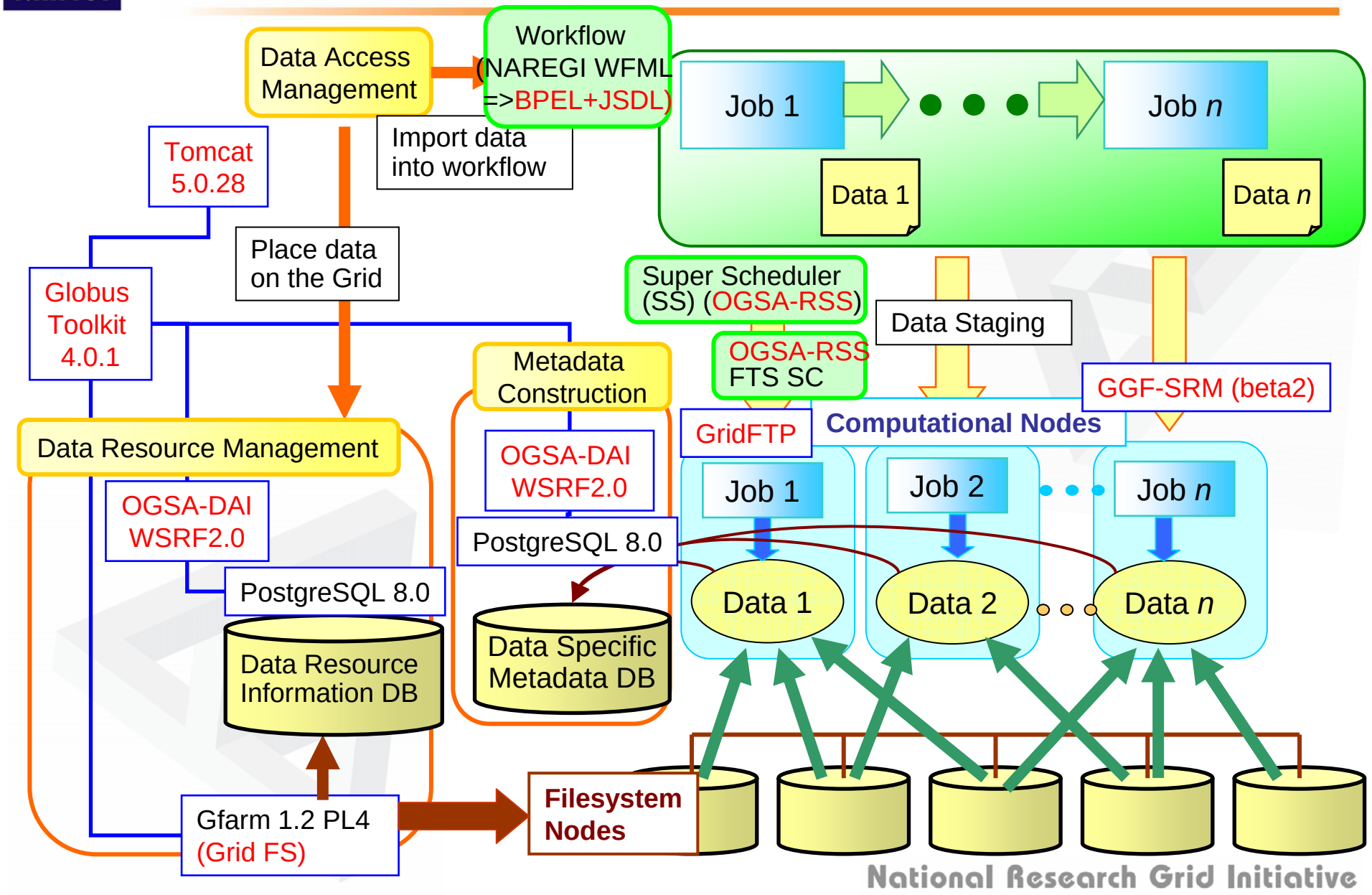


Another Example of Visualization

- Electronic Structure of Lysozyme in Solutions



NAREGI WP4: Standards Employed in the Architecture



NAREGI WP4 Standards

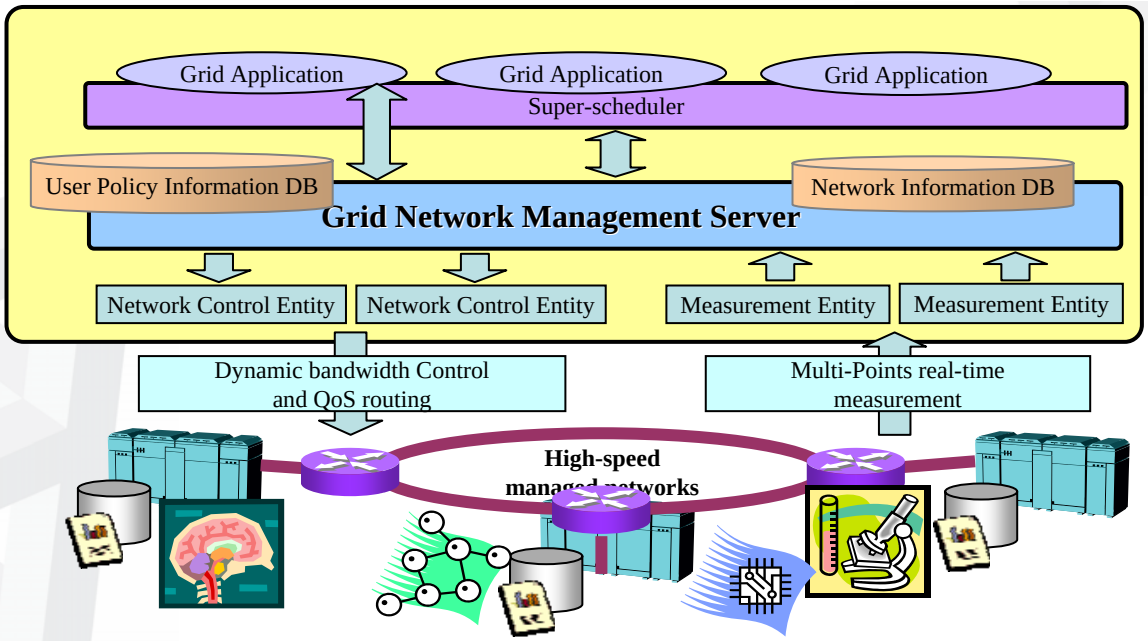
- GGF Standards we help set within a WG
  - Grid FileSystems WG (discussion about functionality and usecase scenario)
- GGF and related Standards we employ
  - OGSA-DAI
  - OGSA-RSS
  - GridFTP
  - WSRF 2.0
  - JSDL
  - SRM (planned for beta 2)
- Other industry standards we employ
  - BPEL
- Other de-facto “standards” we employ
  - Globus 4
  - Tomcat (and associated WS/XML standards)

Roadmaps, Future Plans

- Extension of Data Sharing Service based on Grid FileSystem
  - VO support based on VOMS and/or XACML VO group permissions
  - Shared StorageResource Reservation for Work-Resource Mapping (OGSA-RSS)
- Data Transfer Service
  - Stream-like Data Transfer for huge amount of data (OGSA ByteIO?)
- Virtualization and/or Integration of Metadata
  - For Data Exchange among different storage/file-systems (SRM, SRB, …)
  - Logical Namespace Service (OGSA RNS, WS-Name)

WP-5: Network Measurement, Management & Control for Grid Environment

- **Traffic measurement on SuperSINET**
- **Optimal QoS Routing based on user policies and network measurements**
- **Robust TCP/IP Control for Grids**
- **Grid CA/User Grid Account Management and Deployment**



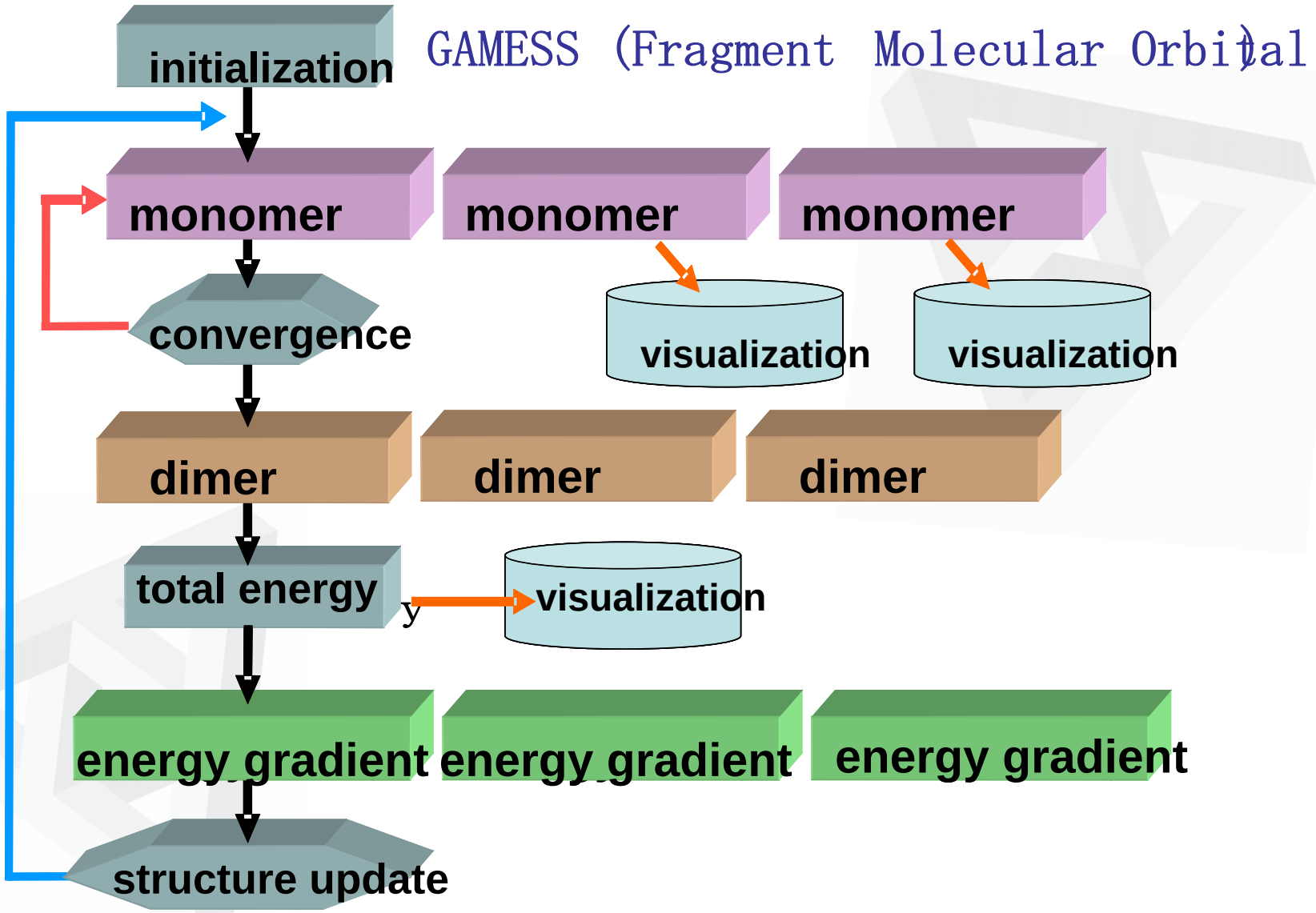
WP-5:NAREGI-CA Features

- Compliance with the basic security level of GGF
  - Independent Registration Authority (RA) Server
  - Practical CP/CPS Template
- License ID management
  - Transfer authentication responsibility to Local RA
- Dual interfaces for certificate request
  - Web & command line enrollment
- Grid operation extensions
  - Batch issue of certificates by command lines
  - Assistance of Grid-mapfile & UUDB creation
- Extension in Version 2.0
  - XKMS interface to realize Web services
  - Management scheme for NAREGI PMA

NAREGI is operating NAREGI CA, which was approved as Production CA by the Asia Pacific Grid Policy Management Authority.

Grid Workflow:

Computational procedure can be defined by Grid Workflow.



Nano-science and Technology Applications Targeted

Participating Organizations:

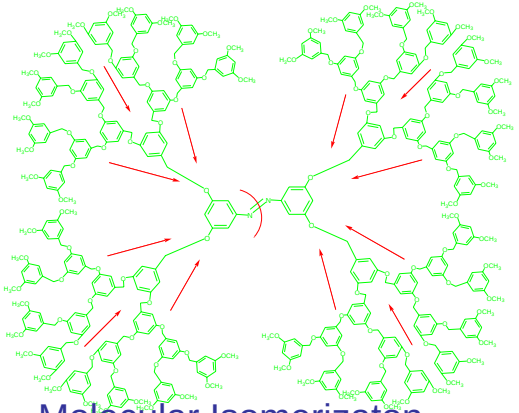
-Institute for Molecular Science
-Institute for Solid State Physics
-AIST
-Tohoku University
-Kyoto University
-Industry (Materials, Nano-scale Devices)
-Consortium for Promotion of Grid Applications in Industry

Research Topics and Groups:

-Functional nano-molecules(CNT,Fullerene etc.)
-Nano-molecule Assembly(Bio-molecules etc.)
-Magnetic Properties
-Electronic Structure
-Molecular System Design
-Nano-simulation Software Integration System
-Etc.

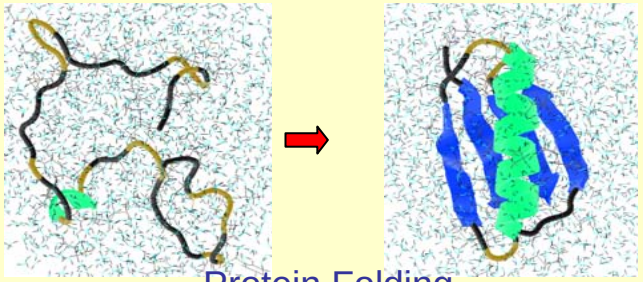
Examples of Nano-Applications Research (1)

Functional Nano-Molecules



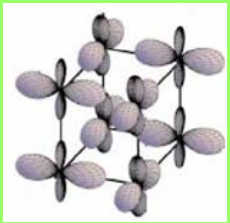
Molecular Isomerizaton
by Low Energy Integration

Nano-Molecular Assembly

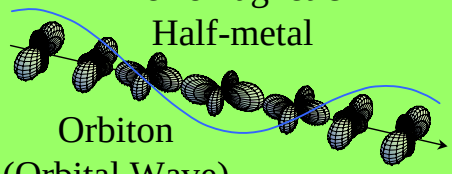


Protein Folding

Nano-Electronic System

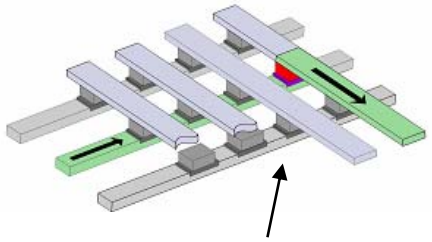


Manganese-Oxide
Ferromagnetic
Half-metal



Orbiton
(Orbital Wave)

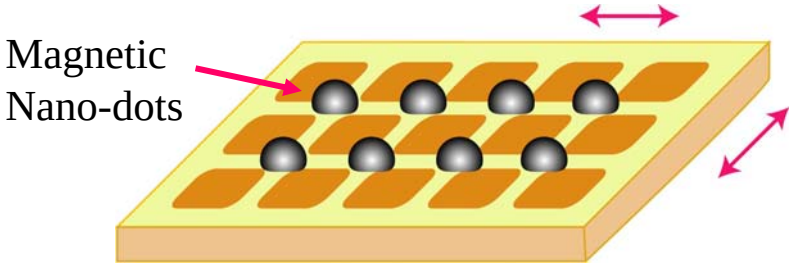
Apps:Half-metal Magnetic
Tunneling Device



Memory Device

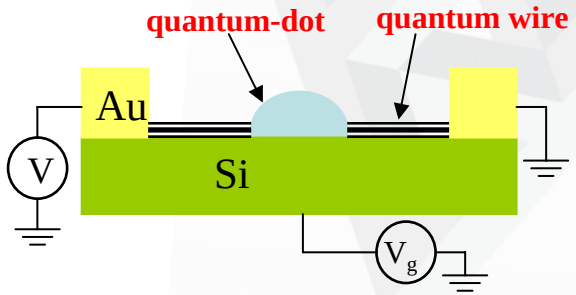
Examples of Nano-Applications Research (2)

Nano-Magnetism



Controling Arrangement of
Nano-dots by Self Organization

Nano-system Design

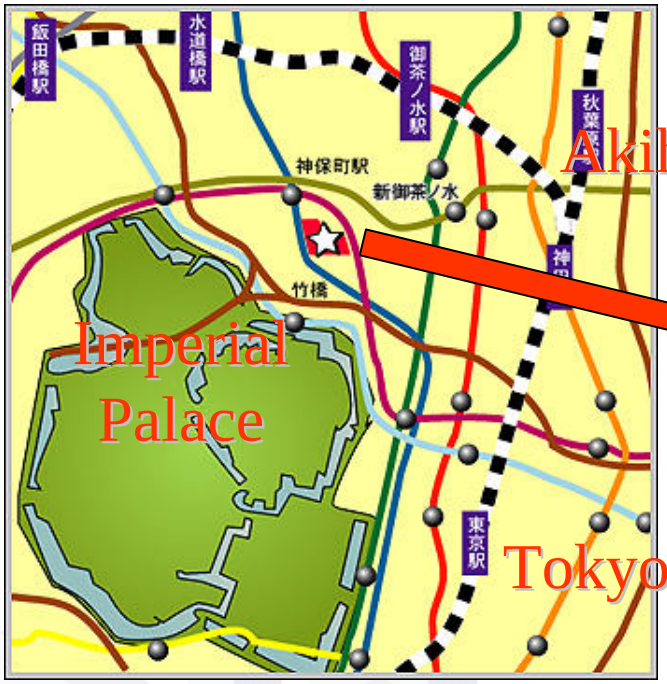


（Metal,organic Molecules）

Nano-device, Quantum Transport

NII Center for Grid R&D (Jimbo-cho, Tokyo)

Mitsui Office
Bldg. 14th Floor



700m^2 office space

(100m^2 machine room)