



# Adopting and Setting Standards The NAREGI Strategy

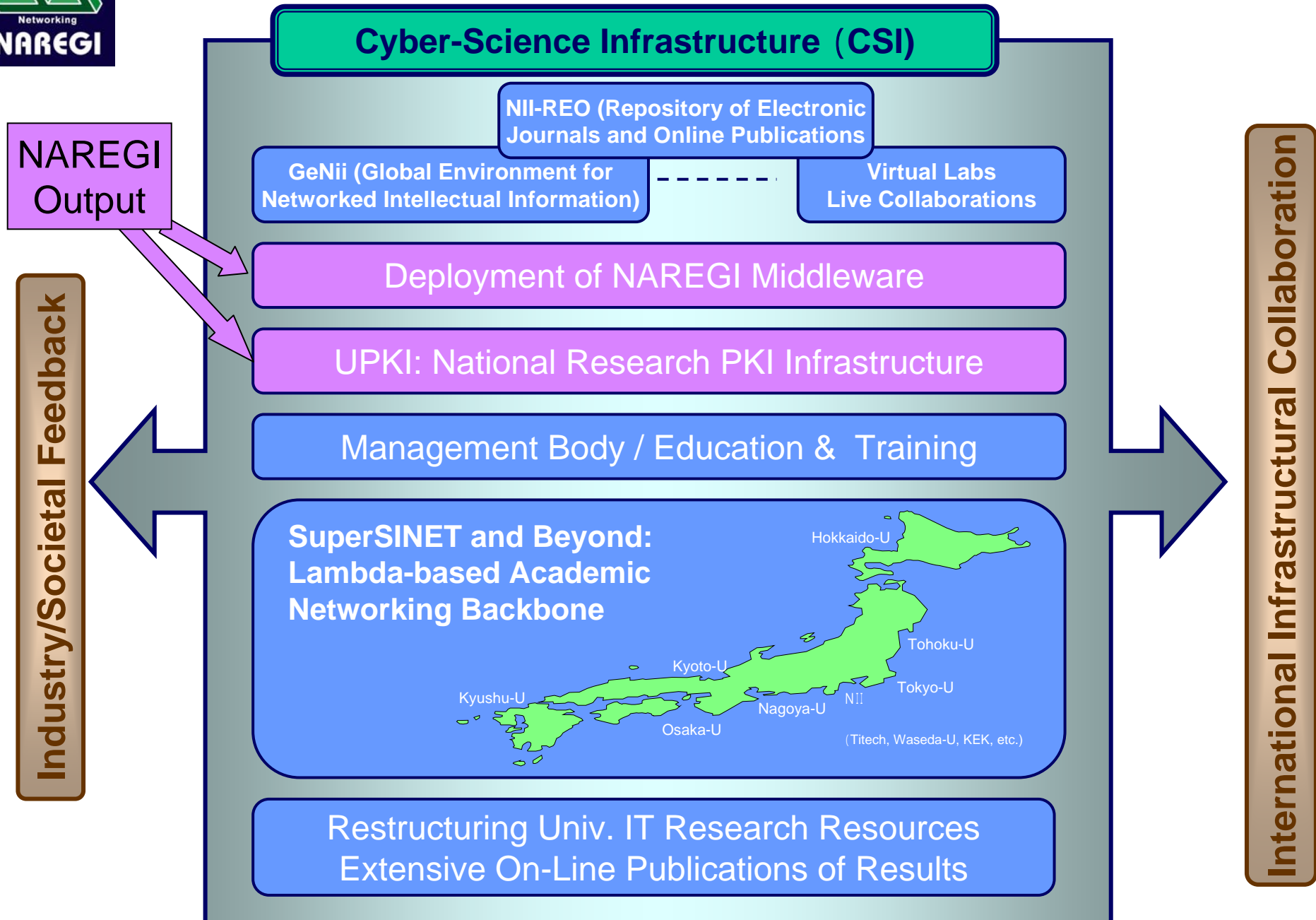
Satoshi Matsuoka

Professor, Global Scientific Information and  
Computing Center,

Deputy Director, NAREGI Project  
Tokyo Institute of Technology / NII



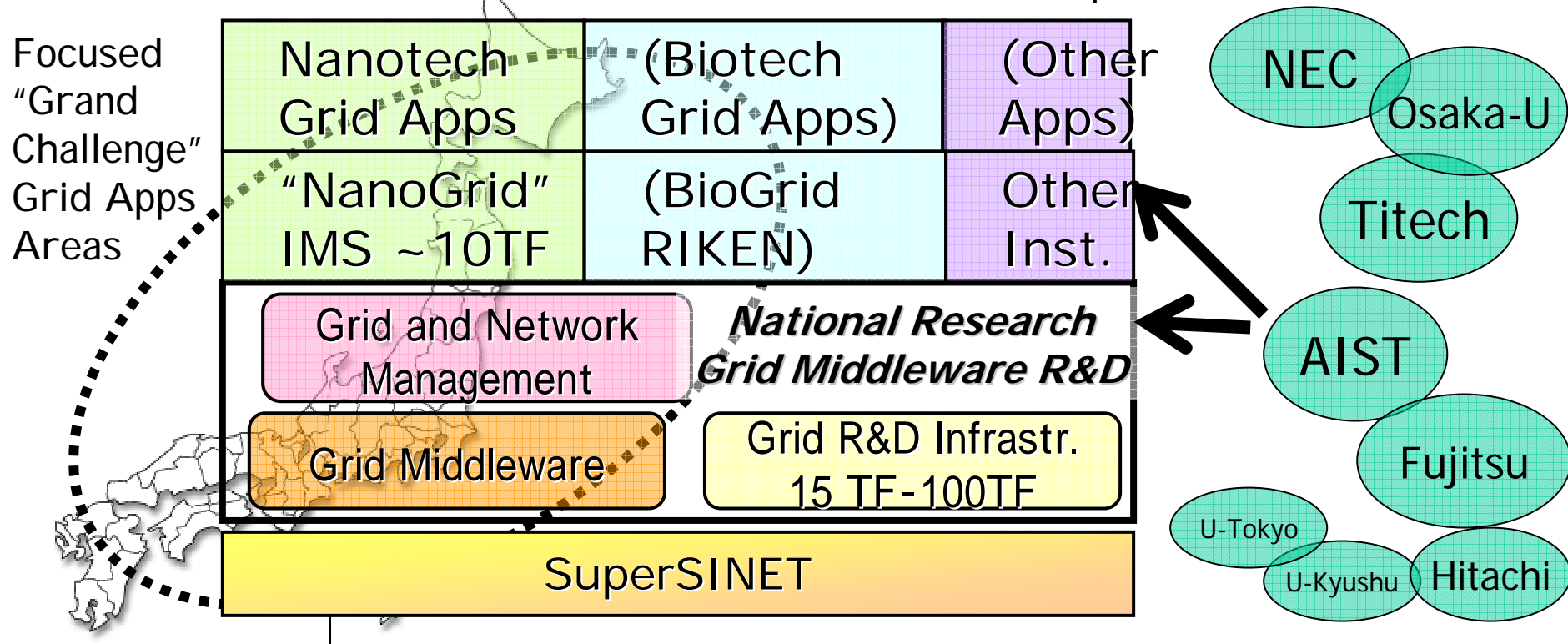
# Towards a Cyber-Science Infrastructure for R & D





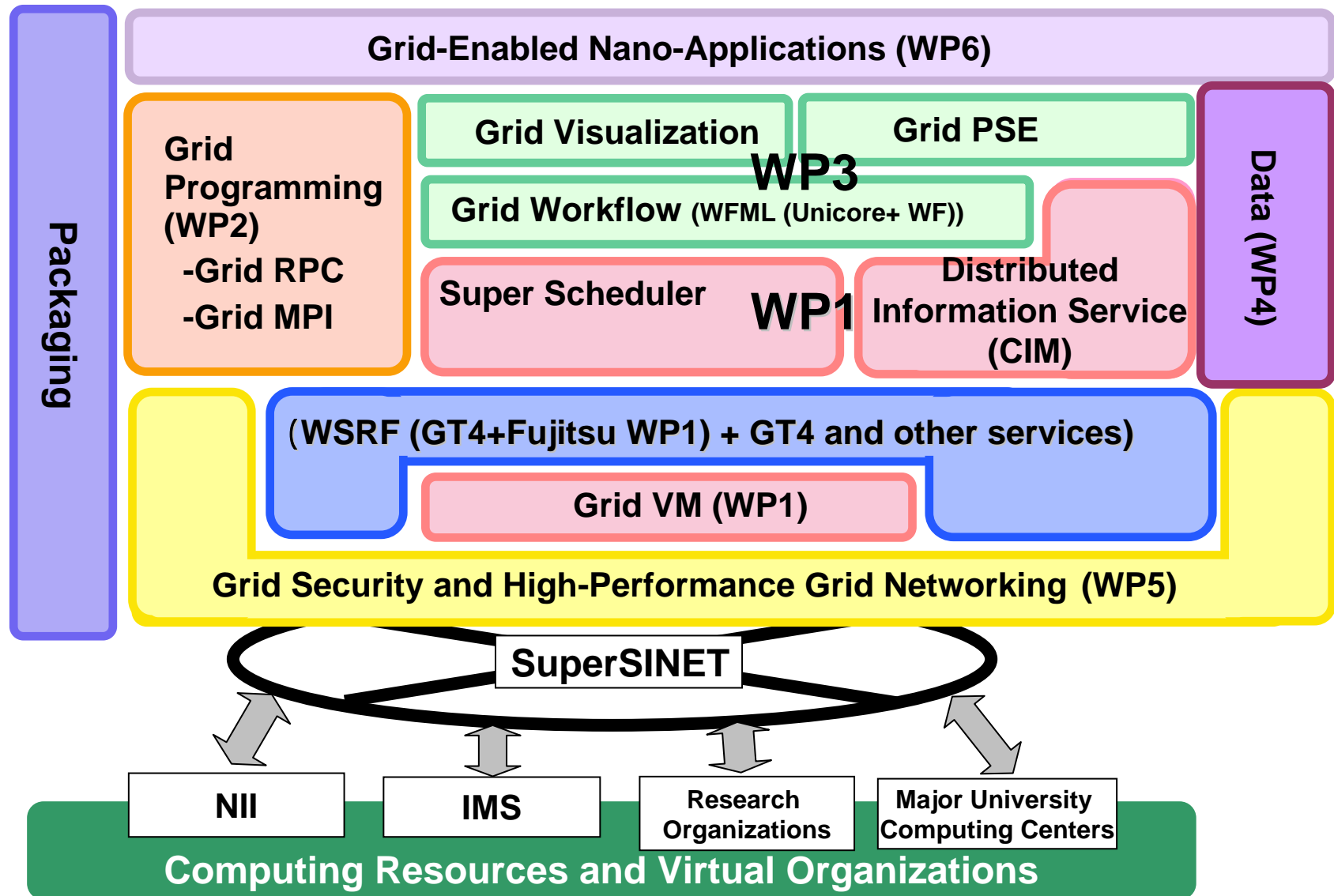
# National Research Grid Infrastructure (NAREGI) 2003-2007

- Petascale Grid Infrastructure R&D for Future Deployment
  - \$45 mil (US) + \$16 mil x 5 (2003-2007) = \$125 mil total
  - Hosted by National Institute of Informatics (NII) and Institute of Molecular Science (IMS)
  - PL: Ken Miura (Fujitsu → NII)
    - Sekiguchi(AIST), Matsuoka(Titech), Shimojo(Osaka-U), Aoyagi (Kyushu-U)...
  - Participation by multiple (>= 3) vendors, Fujitsu, NEC, Hitachi, NTT, etc.
  - Follow and contribute to GGF Standardization, esp. OGSA





# NAREGI Software Stack (Beta Ver. 2006)



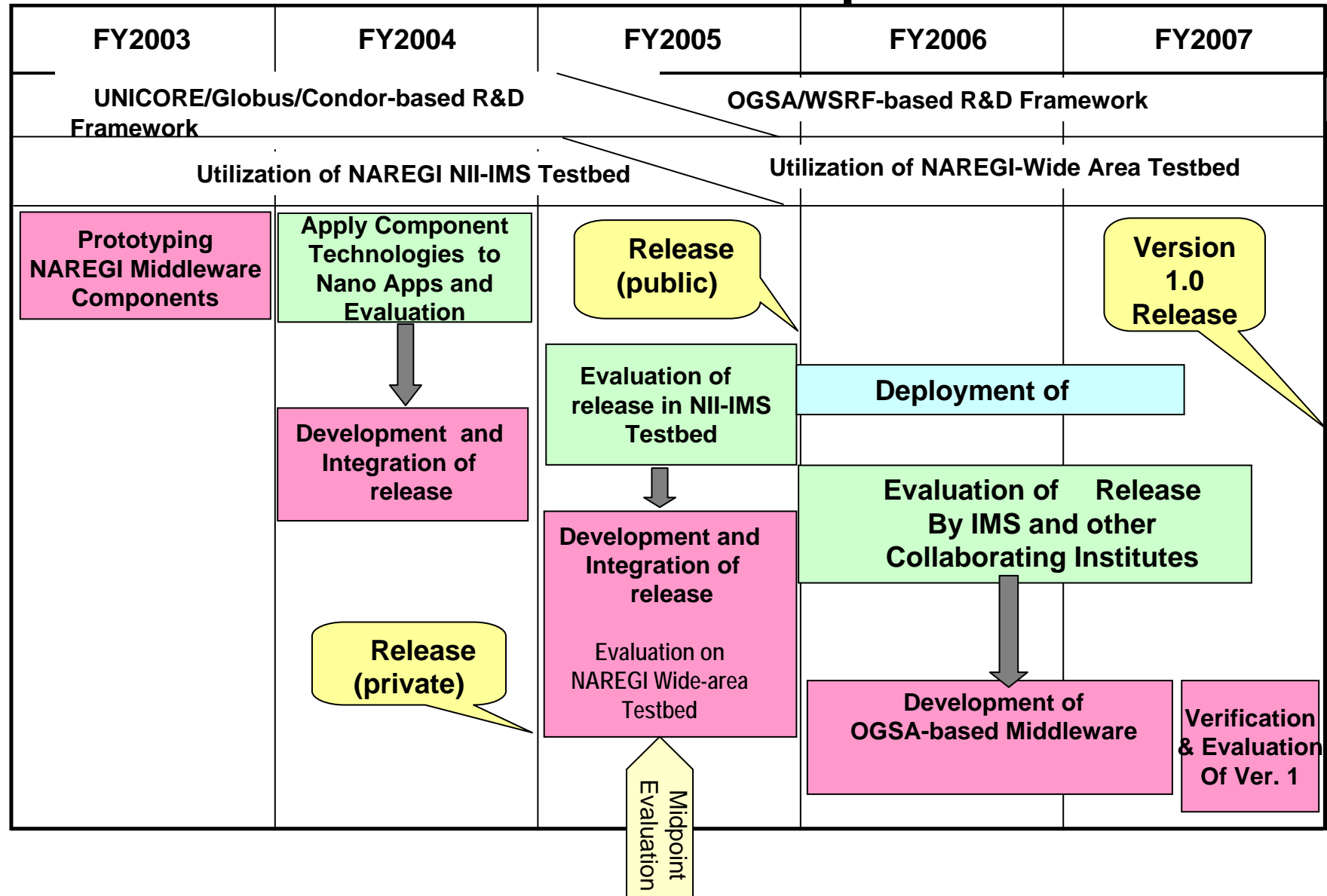


# R&D in Grid Software and Networking Area (Work Packages)

- Work Package Structure (~150 FTEs):
  - Universities and National Labs: technology leadership
  - Vendors (Fujitsu, NEC, Hitachi, etc.): professional development
- WP-1: Resource Management:
  - Matsuoka(Titech), Nakada(AIST/Titech)
- WP-2: Programming Middleware:
  - Sekiguchi(AIST), Ishikawa(U-Tokyo), Tanaka(AIST)
- WP-3: Application Grid Tools:
  - Usami (new FY2005, NII), Kawata(Utsunomiya-u)
- WP-4: Data Management (new FY 2005, Beta):
  - Matsuda (Osaka-U)
- WP-5: Networking & Security
  - Shimojo(Osaka-u), Oie( Kyushu Tech.)
- WP-6: Grid-enabling Nanoscience Appls
  - Aoyagi(Kyushu-u)



# NAREGI Middleware Roadmap





# NAREGI is/has/will...

- Is THE National Research Grid in Japan
  - Part of CSI and future Petascale initiatives
  - METI “Business Grid” counterpart 2003-2005
- Has extensive commitment to WS/GGF-OGSA
  - Entirely WS/Service Oriented Architecture
  - Set industry standards e.g. 1st impl. of OGSA-EMS
- Will work with EU/US/AP counterparts to realize a “global research grid”
  - Various talks have started, incl. SC05 interoperability meeting
- Will deliver first OS public beta in May 2006
  - To be distributed @ GGF17/GridWorld in Tokyo



# NAREGI is not/doesn't/won't...

- Is NOT an academic research project
  - All professional developers from Fujitsu, NEC, Hitachi, NTT, ...
  - No students involved in development
- Will NOT develop all software by itself
  - Will rely on external components in some cases
  - Must be WS and other industry standards compliant
- Will NOT deploy its own production Grid
  - Although there is a 3000-CPU testbed
  - Work with National Centers for CSI deployment
- Will NOT hinder industry adoption at all costs
  - Intricate open source copyright and IP policies
  - We want people to save/make money using NAREGI MW



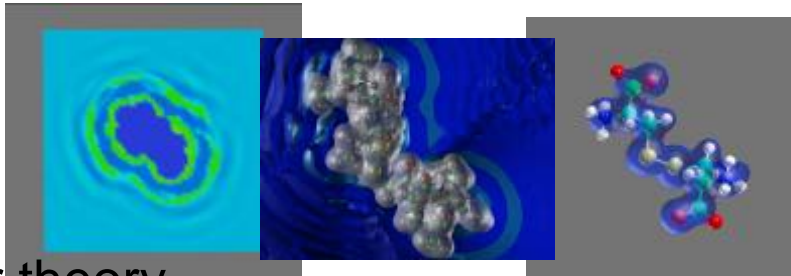


# Nano-Science : coupled simulations on the Grid as the sole future for true scalability

--- Coupled Simulations as “Killer Apps” for Grid ---

Material physics  
(Infinite system)

- Fluid dynamics
- Statistical physics
- Condensed matter theory



Molecular Science

- Quantum chemistry
- Molecular Orbital method
- Molecular Dynamics

...

$10^{-6}$

$10^{-9}$

m

...

Limit of  
Idealization

Multi-Physics  
Multi-Resolution

Limit of  
Computing  
Capability

Old HPC environment:

- decoupled resources,
- isolated users,
- special software, ...
- no coupling!

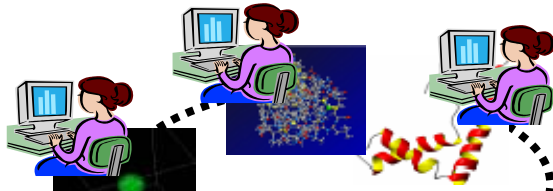
Coordinates decoupled resources on Grid;

Meta-computing,  
High throughput computing,  
Multi-Physics simulation

w/ components and data from different groups  
within VO composed in real-time



The only way to achieve true scalability!



## LifeCycle of Grid Apps and Infrastructure

Application Contents Service

HL Workflow  
NAREGI WFML (Unicore+)

VO Application Developers&Mgrs  
Workflows and Coupled Apps / User

Many I/O Users

SuperScheduler (OGSA-RSS)

Grid Info Service (CIM)

Place & register data on the Grid

Assign metadata to data

MetaComputing

GridRPC/Grid MPI

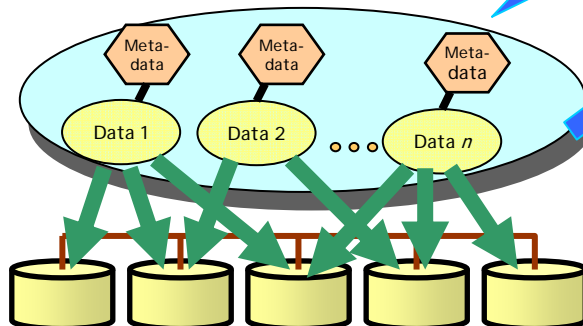
User Apps  
GridVM+GRAM4

Distributed Servers

User Apps  
GridVM+GRAM4

User Apps  
GridVM+GRAM4

Implementation of OGSA-EMS Architecture



Grid-wide Data Management Service (GridFS, Metadata, Staging, etc.)



# Highlights of NAREGI Beta (May 2006, GGF17/GridWorld)

- Professionally developed and tested
- “Full” OGSA-EMS incarnation
  - Full C-based WSRF engine (Java -> Globus 4)
  - OGSA-EMS/RSS WSRF components
  - Full WS-Agreement brokering and co-allocation
  - GGF JSDL1.0-based job submission, authorization, etc.
  - 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 via GGF-ACS
- Comprehensive Data management w/Grid-wide FS
- Complex workflow (NAREGI-WFML) for various coupled simulations
- Overall stability/speed/functional improvements
- To be interoperable with EGEE, TeraGrid, etc. (beta2)



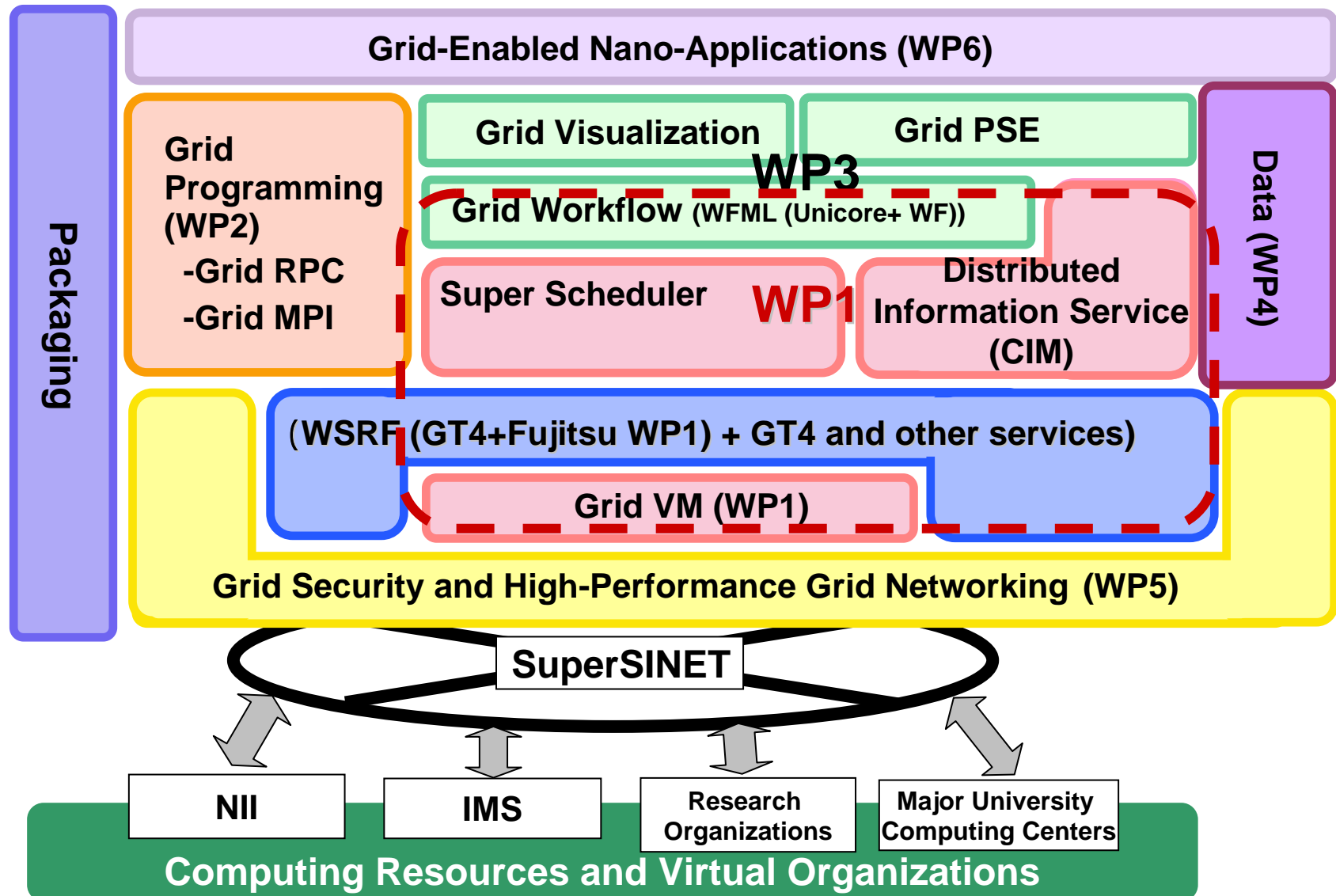
# List of NAREGI “Standards”

(beta 1 and beyond)

- GGF Standards and Pseudo-standard Activities set/employed by NAREGI
  - GGF “OGSA CIM profile”
  - GGF AuthZ
  - GGF DAIS
  - GGF GFS (Grid Filesystems)
  - GGF Grid CP (GGF CAOPs)
  - GGF GridFTP
  - GGF GridRPC API (as Ninf-G2/G4)
  - GGF JSDL
  - GGF OGSA-BES
  - GGF OGSA-Byte-IO
  - GGF OGSA-DAI
  - GGF OGSA-EMS
  - GGF OGSA-RSS
  - GGF RUS
  - GGF SRM (planned for beta 2)
  - GGF UR
  - GGF WS-I RUS
  - GGF ACS
  - GGF CDDL
- Other Industry Standards Employed by NAREGI
  - ANSI/ISO SQL
  - DMTF CIM
  - IETF OCSP/XKMS
  - MPI 2.0
  - OASIS SAML2.0
  - OASIS WS-Agreement
  - OASIS WS-BPEL
  - OASIS WSRF2.0
  - OASIS XACML
- De Facto Standards / Commonly Used Software Platforms Employed by NAREGI
  - Ganglia
  - GFarm 1.1
  - Globus 4 GRAM
  - Globus 4 GSI
  - Globus 4 WSRF (Also Fujitsu WSRF for C binding)
  - IMPI (as GridMPI)
  - Linux (RH8/9 etc.), Solaris (8/9/10), AIX, ...
  - MyProxy
  - OpenMPI
  - Tomcat (and associated WS/XML standards)
  - Unicore WF (as NAREGI WFML)
  - VOMS

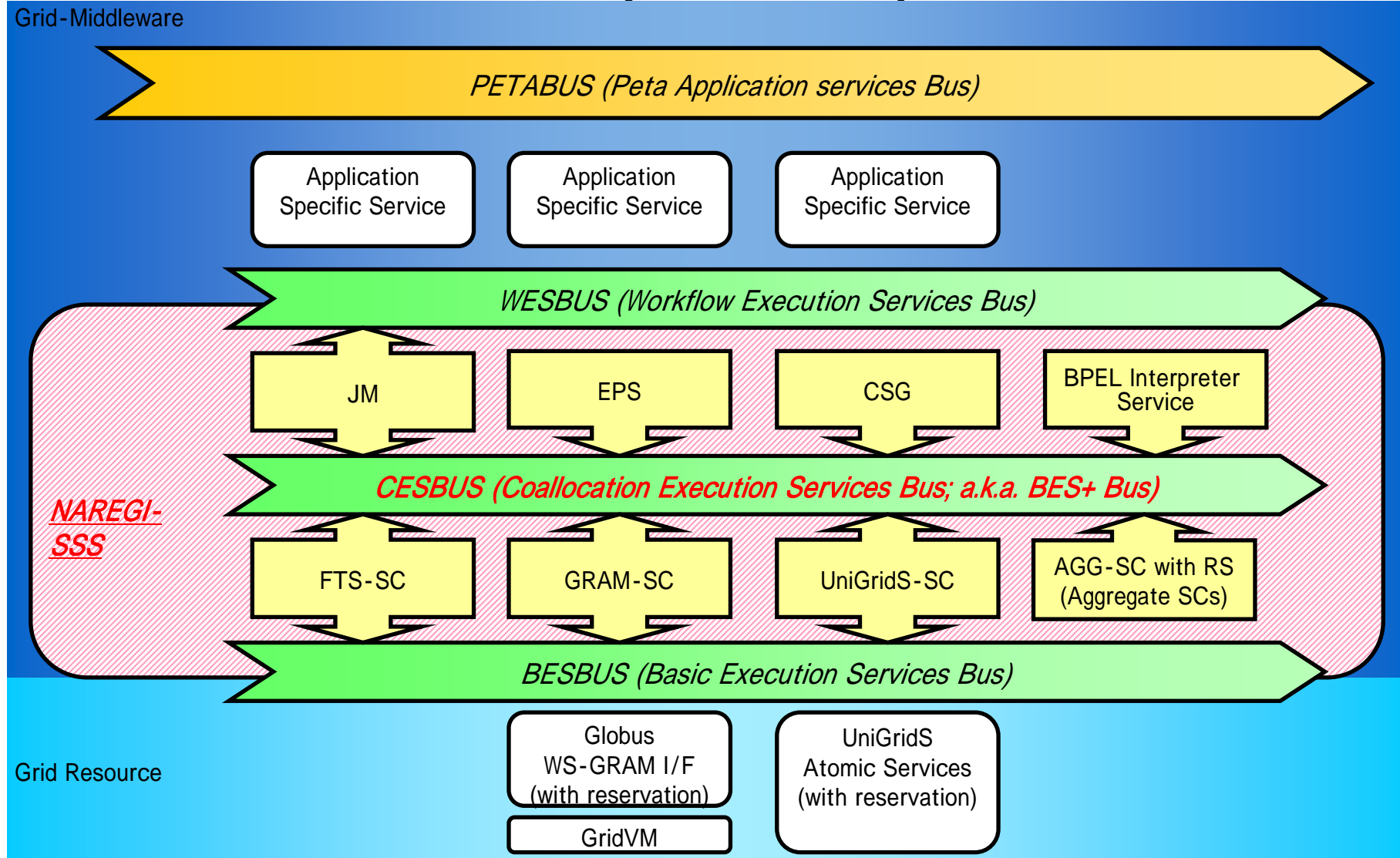


# NAREGI Software Stack (Beta Ver. 2006)





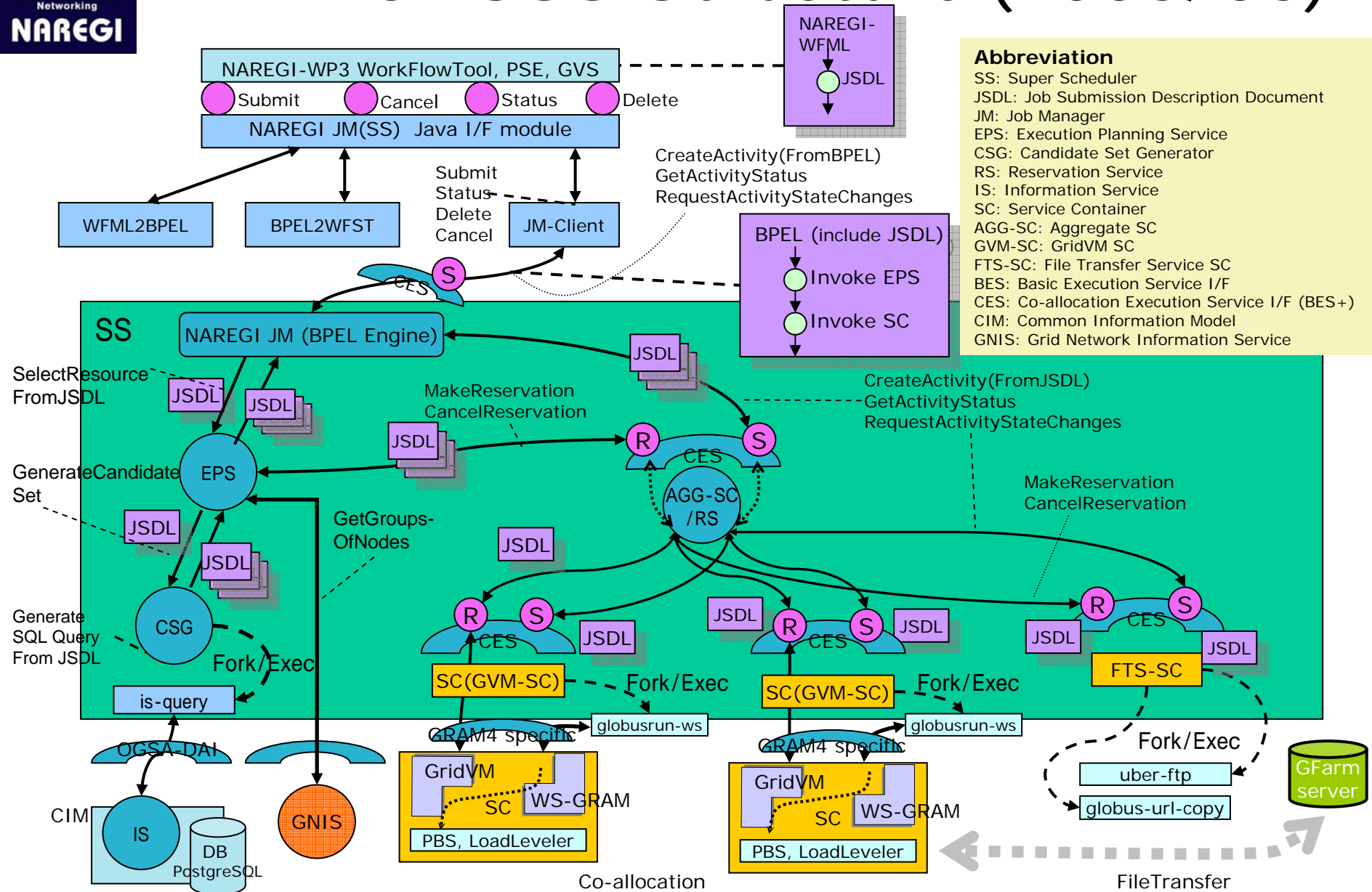
# The NAREGI SSS Architecture (2007/3)







# NAREGI SSS Structure (2006/06)





# NAREGI SSS Standards (1/3)

Service Name	Description	Standards
NAREGI -JM	<ul style="list-style-type: none"> <li>• is an OGSA-EMS Job Manager implementation.</li> <li>• accepts NAREGI-WFML (Work-Flow Markup Language), which is an intermediate workflow language for workflow-to-workflow language conversion</li> <li>• generates a BPEL document for each NAREGI-WFML (workflow-job) as a OGSA-EMS/RSS service orchestration script.</li> </ul>	@ OGSA-WG [OGSA-EMS] \$ OGSA WFM-RG # OASIS WS-BPEL
NAREGI -BPEL	<ul style="list-style-type: none"> <li>• is an BPEL script execution service implementation.</li> <li>• accepts a BPEL document.</li> <li>• interpret and execute a BPEL document.</li> </ul>	# OASIS WS-BPEL
NAREGI -EPS	<ul style="list-style-type: none"> <li>• is an OGSA-RSS Execution Planning Service implementation.</li> <li>• accepts a NAREGI Complex Job Instance document or a (simple) Job Instance document, which express a job execution instance for a coupling job. Each Job Instance document contains a JSDL document as a resource requirement description.</li> <li>• provides a co-allocation-enabled resource selection service as a whole.</li> <li>• is a aggregation point of the following underlying selection services: single CSG (Candidate-Set Generator) service and multiple "Narrowing down candidates" services such as Grid Storage Service, Network Neighborhood Information Service and Reservation Cache Service.</li> </ul>	@ OGSA-RSS-WG @ JSDL-WG \$ GSM-WG (Grid Storage Management WG) \$ GHPN-RG (Grid High Performance Networking RG; network services) \$ GSA-RG (Grid Scheduling Architecture RG) \$ CDDLM-WG

@: GGF Standard we use, \$ GGF Standard we will use, # Industry Standard % De-facto Standard





# NAREGI SSS Standards (2/3)

Service Name	Description	Standards
NAREGI -CSG	<ul style="list-style-type: none"> <li>- is an OGSA-RSS Candidate Set Generator service implementation. - accepts a NAREGI Complex Job Instance document or a (simple) Job Instance document, which express a job execution instance for a coupling job.</li> <li>- extracts JSDL documents from a Job Instance document, generates resource-database query expressions from these JSDL documents, and reflects query results (resource candidates) into the Job Instance document.</li> <li>- Virtual Organization aware resource query by VOMS</li> </ul>	@ OGSA-RSS-WG @ JSDL-WG @ DAIS-WG @ OGSA-WG [Resource Management] % MyProxy, VOMS # DMTF CIM, ANSI/ISO SQL \$ RUS-WG
NAREGI -AGG with RS	<ul style="list-style-type: none"> <li>- is a virtual Service Container for a coupling job at coallocation execution services level, and aggregates "real" Service Containers at basic exeuction services level.</li> <li>- accepts a NAREGI Complex Job Instance document or a (simple) Job Instance document, including final candidates.</li> <li>- provides logical service operations: MakeReservations() and CommitReservations() to co-allocate and co-schedule a coupling job, and Cancel() and Destroy() to control the coupling job execution.</li> <li>- contains Reservation (Cache) Service to efficiently select available resources over "time-space" during final candidates. A Reservation Cache Service tracks each resource usage (including reservation) on Aggregation Service Container as possible.</li> </ul>	@ OGSA-RSS-WG \$ GRAAP-WG [WS-Agreement] \$ GSA-RG

@: GGF Standard we use, \$ GGF Standard we will use, # Industry Standard % De-facto Standard



# NAREGI SSS Standards (3/3)

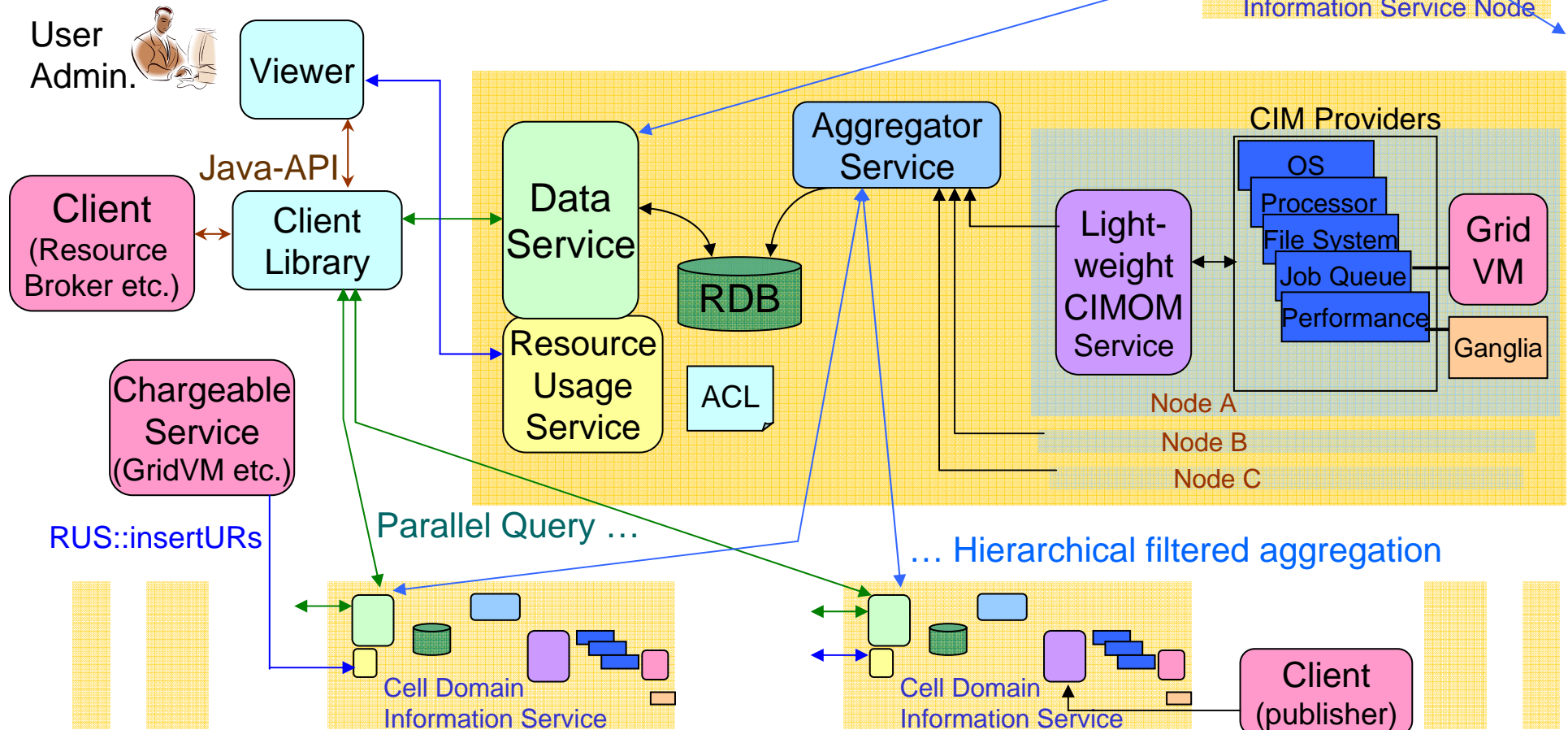
Service Name	Description	Standards
NAREGI -FTS	<ul style="list-style-type: none"> <li>- is a file operation service, support basic file operations such as 3rd-party-file-transfer and check-file-existence</li> <li>- accepts single NAREGI Job Instance document, including DataStaging job without Application in JSDL.</li> </ul>	@ GridFTP-WG % GFarm (a Grid File System) \$ WS-Naming-WG, \$ GFS-WG (Grid File System WG)
NAREGI -SC (GRAM-SC) (UniGridS-SC)	<ul style="list-style-type: none"> <li>- is a tentative wrapper service to hide grid-middleware-specific service interfaces (in particular for OGSA-BES to become widely used and standardize reservation operation in OGSA-BES).</li> <li>- accepts single NAREGI Job Instance document.</li> </ul>	@ OGSA-BES @ JSDL-WG % Globus [GRAM4] % UniGridS [Atomic Services] % Proxy Certificate, MyProxy \$ GRAAP-WG [WS-Agreement]
(infrastructure )	<ul style="list-style-type: none"> <li>- native https/soap and wsrf processor [WS/Rf 1.2 r/rp/r/bf, WS/N 1.3 bn/topic] in C</li> <li>- native WSRF-aware WSDL compiler (WSDL to C compiler)</li> <li>- native BPEL interpreter (BPEL4WS) in C</li> </ul>	@ OASIS WS/Rf @ OASIS WS/N # OASIS WS-BPEL

@: GGF Standard we use, \$ GGF Standard we will use, # Industry Standard % De-facto Standard



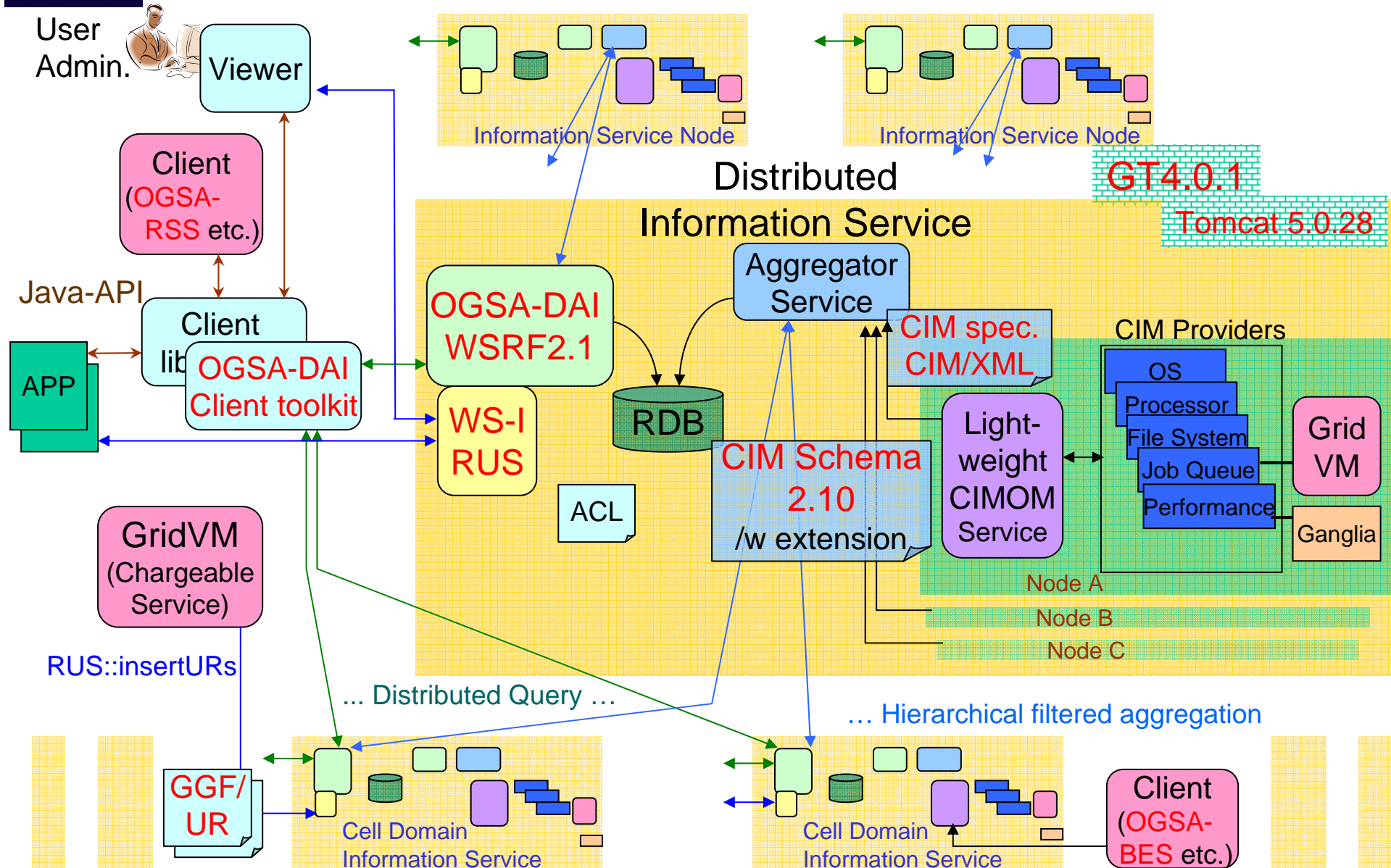
# NAREGI Info Service (beta) Architecture

- CIMOM Service classifies info according to CIM based schema.
- The info is aggregated and accumulated in RDBs hierarchically.
- Client library utilizes OGSA-DAI client toolkit.
- Accounting info is accessed through RUS.





# NAREGI IS: Standards Employed in the Architecture





# NAREGI Dist. Info Standards

- GGF Standards we help set within a WG
  - “OGSA CIM profile”
- GGF and related Standards we employ
  - OGSA-DAI
  - WS-I RUS
  - GGF UR
  - OGSA-EMS
- Other industry standards we employ
  - DMTF-CIM
- Other de-facto “standards” we employ
  - Globus 4
  - Tomcat (and associated WS/XML standards)



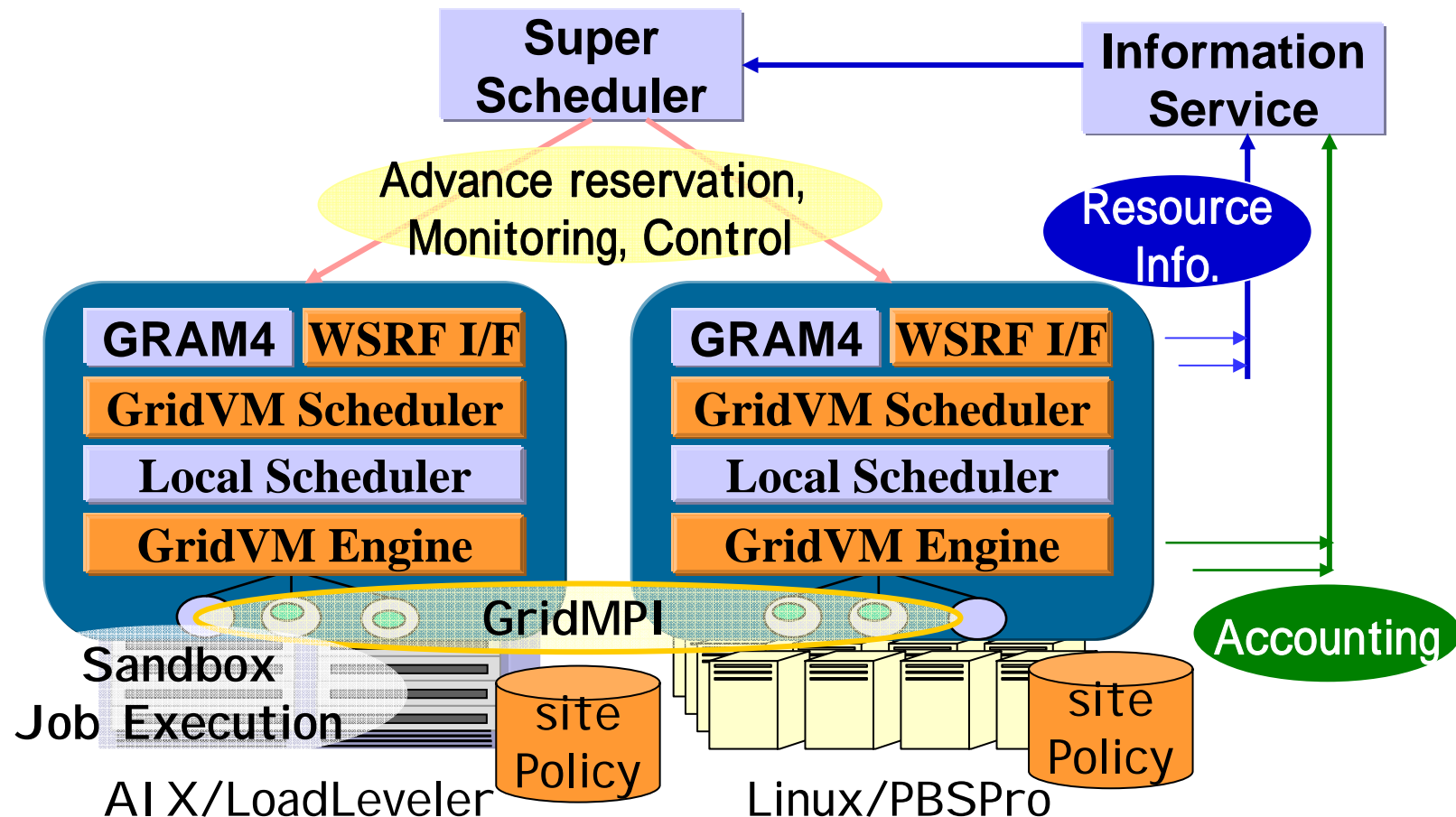
# IS Future Plans

- GLUE CIM translation
  - having a set of sites all communicate some minimal set of data (a simple subset of GLUE schema)
  - getting the information from GLite/R-GMA consumer i/f, Globus/WS-MDS Aggregator
  - translating it to CIM/XML format, archiving it in RDBs and retrieving it through OGSA-DAI Data Service.
- OGSA Information producer / consumer i/f
- Access control with XACML policy
- Service separability



# NAREGI GridVM (beta) Architecture

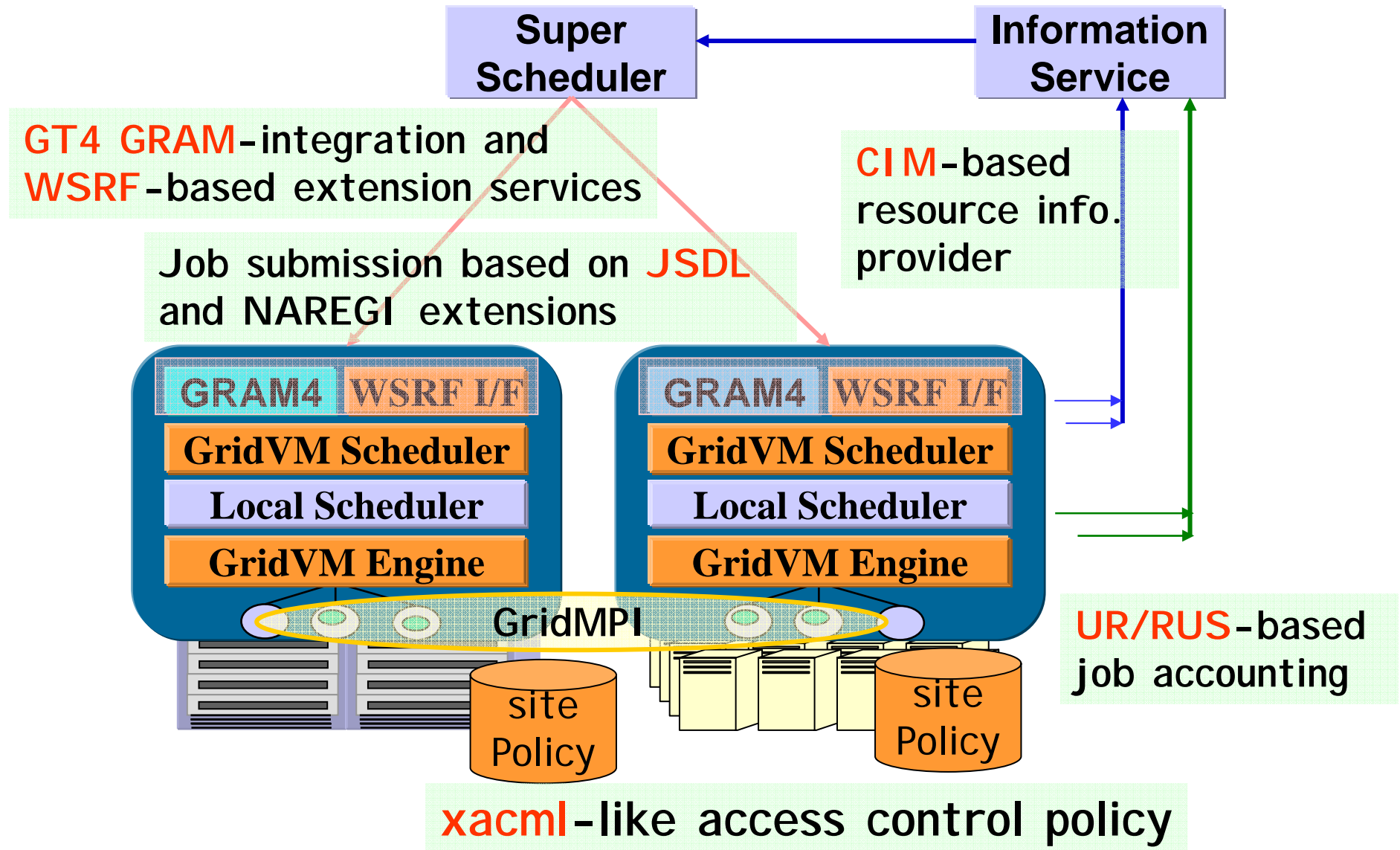
- ✓ Virtual execution environment on each site
  - Virtualization of heterogeneous resources
  - Resource and job management services with unified I / F







# NAREGI GridVM: Standards Employed in the Architecture

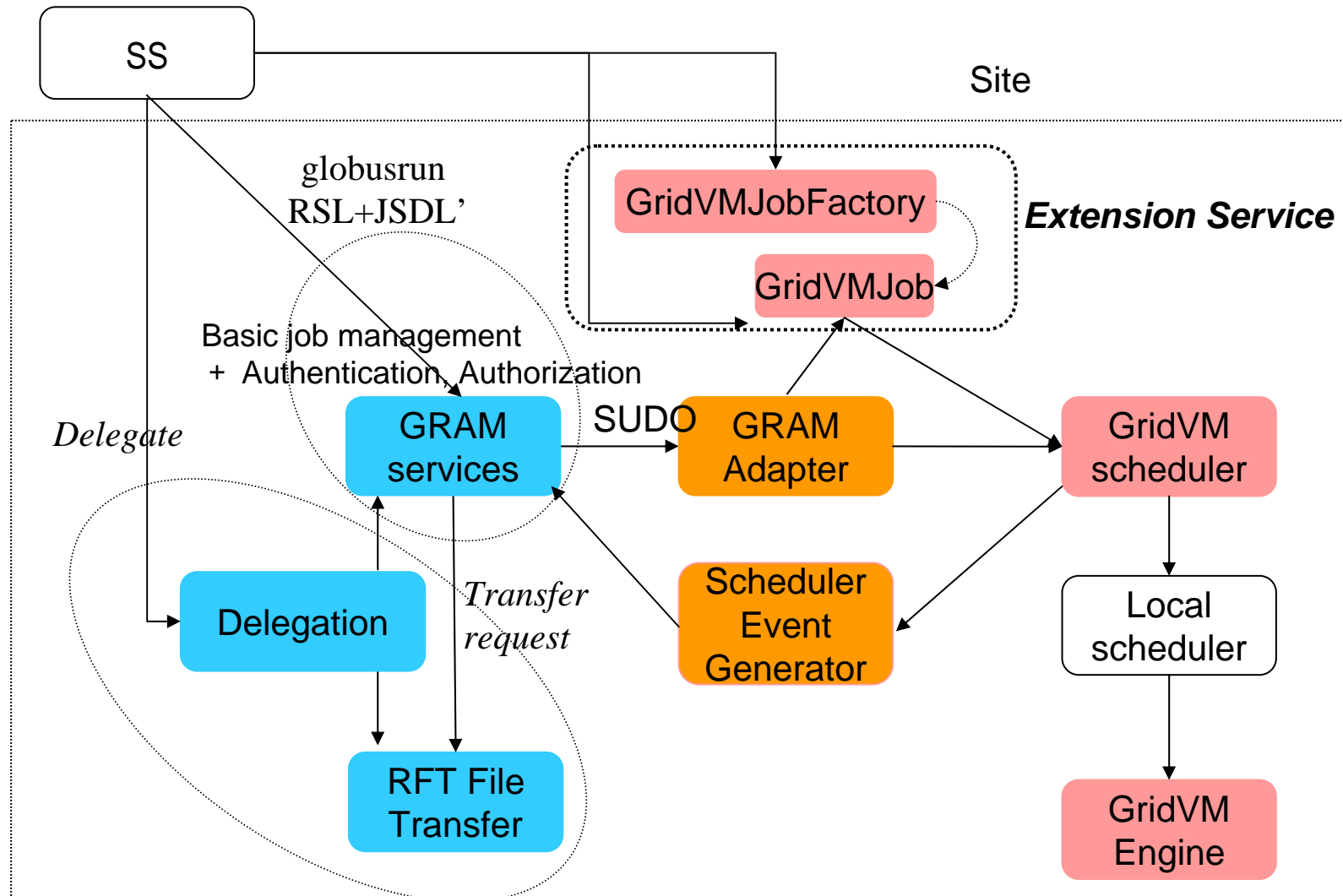






# GT4 GRAM-GridVM Integration

- ✓ Integrated as an extension module to GT4 GRAM
- ✓ Aim to make the both functionalities available



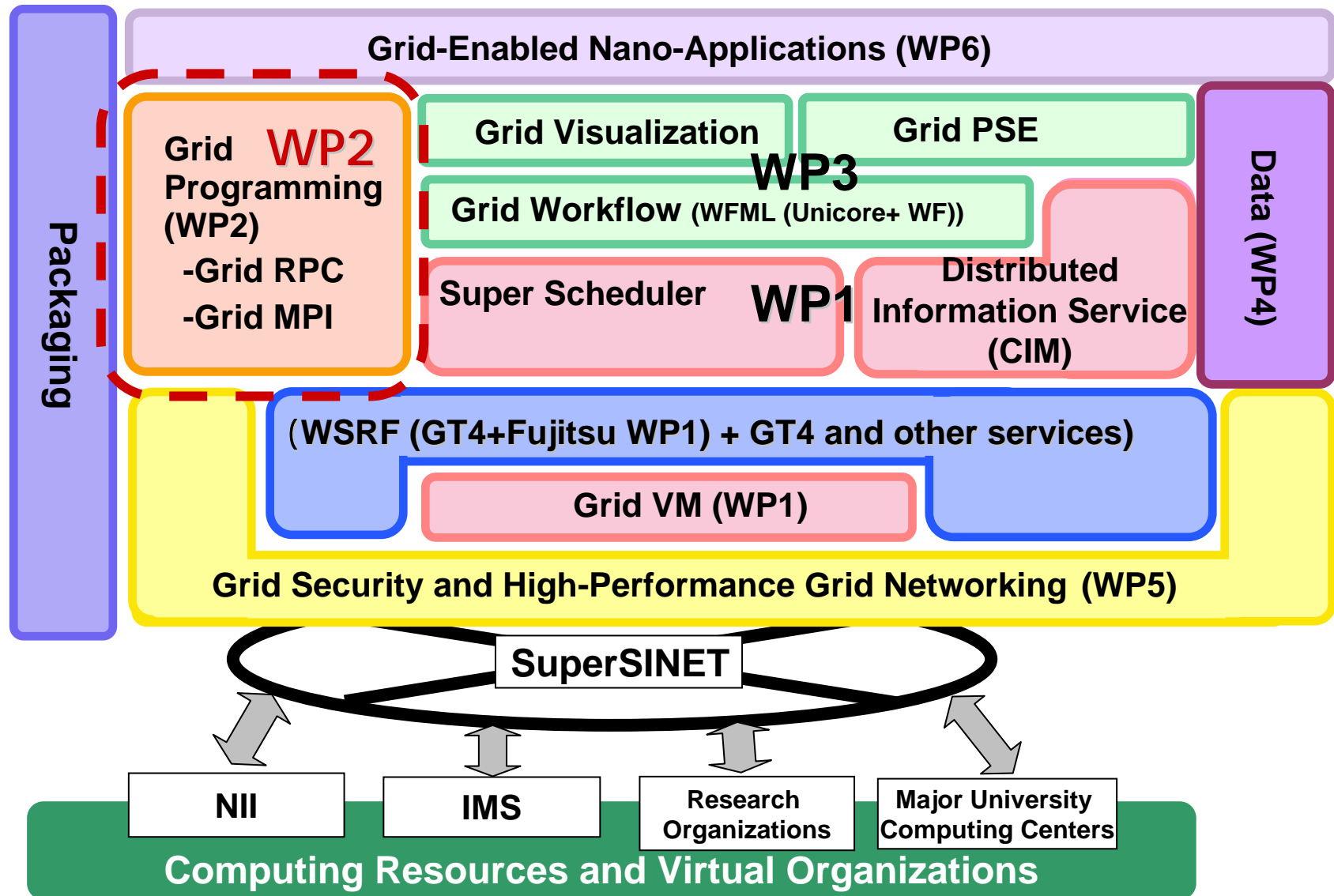


# NAREGI GridVM Standards

- GGF Standards we help set within a WG
  - OGSA BES (made a presentation at BOF of GGF13)
- GGF and related Standards we employ
  - JSDL
  - WS-I RUS
  - GGF UR
  - WSRF2.0
- Other industry standards we employ
  - DMTF-CIM
  - XACML
- Other de-facto “standards” we employ
  - Globus Toolkit version 4



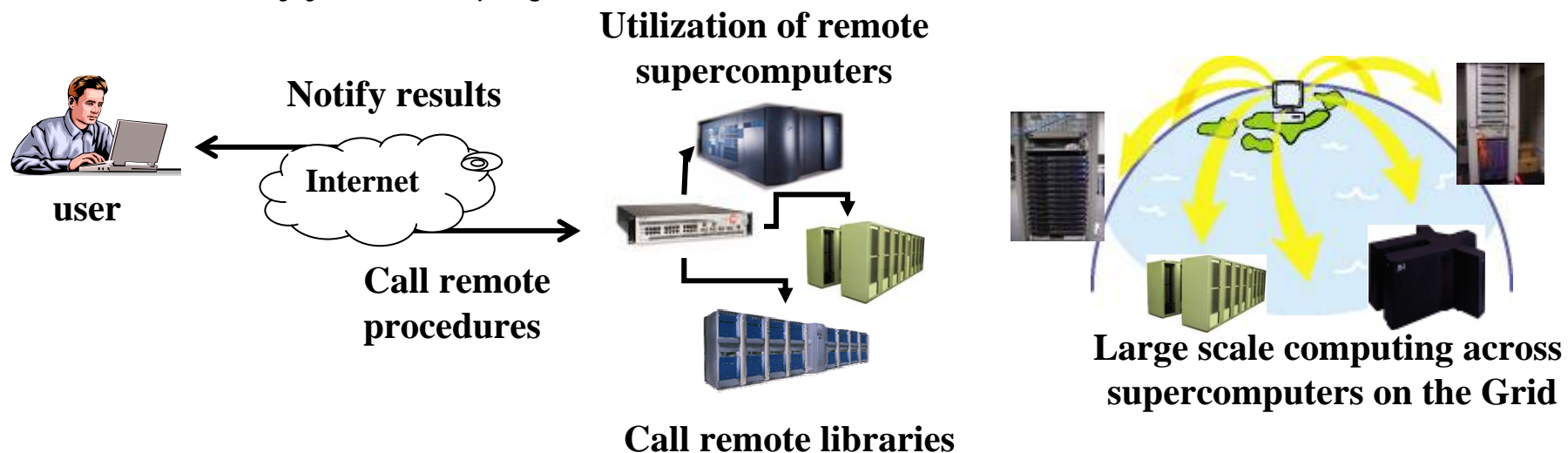
# NAREGI Software Stack (Beta Ver. 2006)





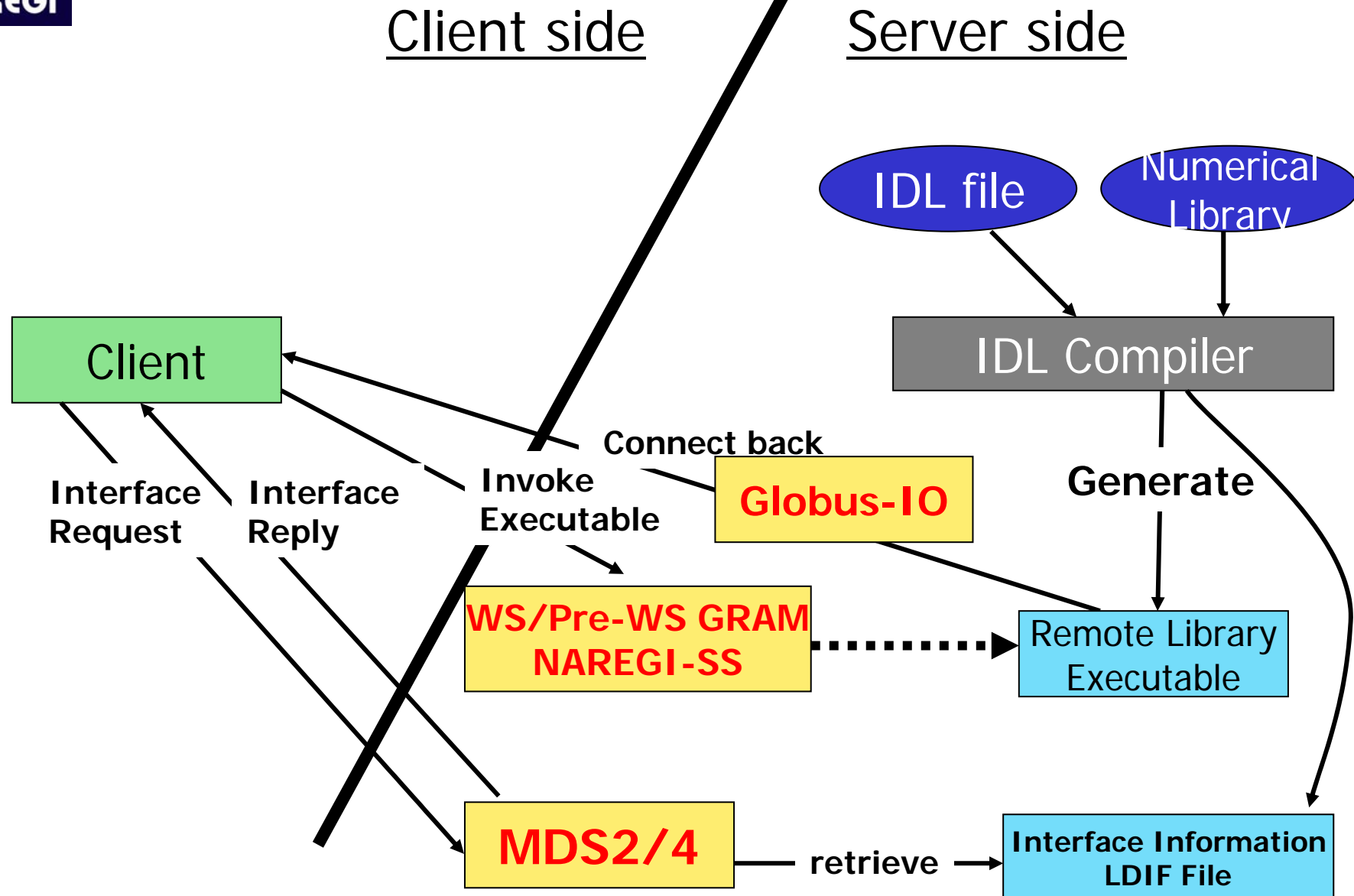
# Ninf-G: A Reference Implementation of the GGF GridRPC API

- What is GridRPC?
  - Programming model using RPCs on a Grid
  - Provide easy and simple programming interface
  - The GridRPC API is published as a proposed recommendation (GFD-R.P 52)
- What is Ninf-G?
  - A reference implementation of the standard GridRPC API
  - Built on the Globus Toolkit
  - Now in NMI Release 8 (first non-US software in NMI)
- Easy three steps to make your program Grid aware
  - Write IDL file that specifies interface of your library
  - Compile it with an IDL compiler called ng\_gen
  - Modify your client program to use GridRPC API





# Architecture of Ninf-G





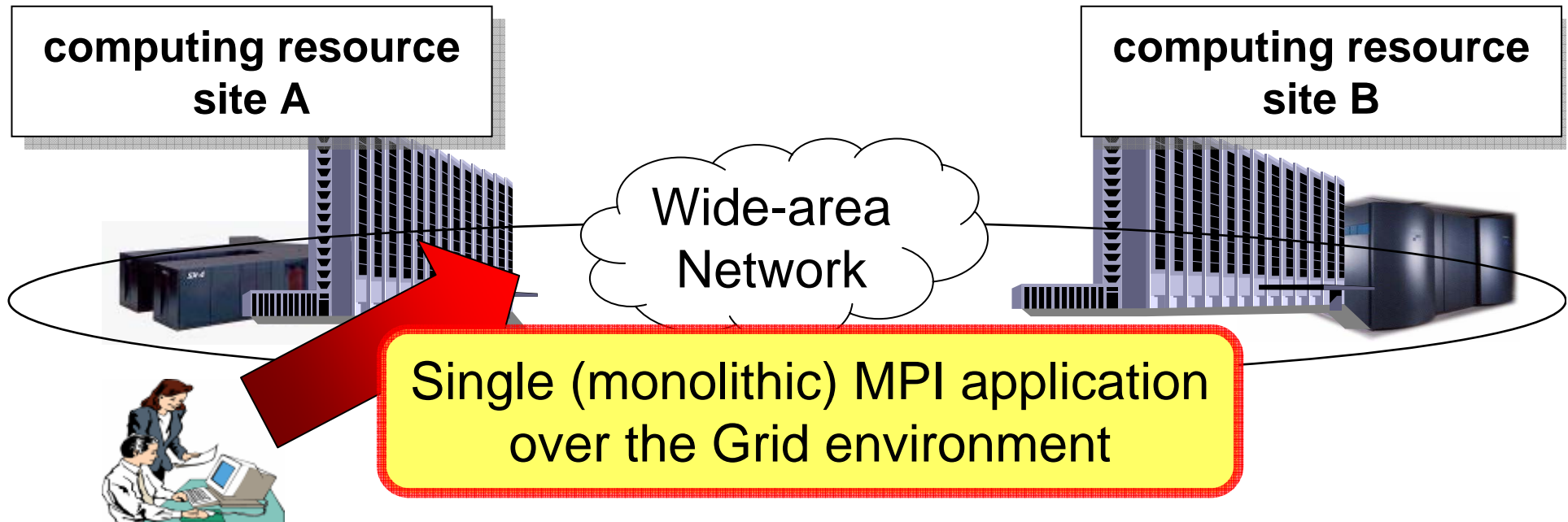
# GridRPC Standards and contributions to GGF

- GGF Standards we help set within a WG
  - Grid RPC WG (WGではなく、仕様自身へのrefが必要)
- GGF and related Standards we employ
  - GridRPC API
  - GridFTP
  - WSRF 2.0
  - JSDL
- Other de-facto “standards” we employ
  - Globus Toolkit Versions 4
  - UNI CORE
- Ninf-G is a reference implementation of the GridRPC API
- Standardization of the GridRPC API
  - GGF GridRPC API WG (spinned off from APM RG)
    - Co-chair: Hidemoto Nakada (AI ST)
    - Secretary: Yoshio Tanaka (AI ST)
  - GridRPC API is published as a proposed recommendation (GFD-R.P 52)
    - A GridRPC Model and API for End-User Applications
      - Co-Authors: Hidemoto Nakada (AI ST), Satoshi Matsuoka (TI TECH)



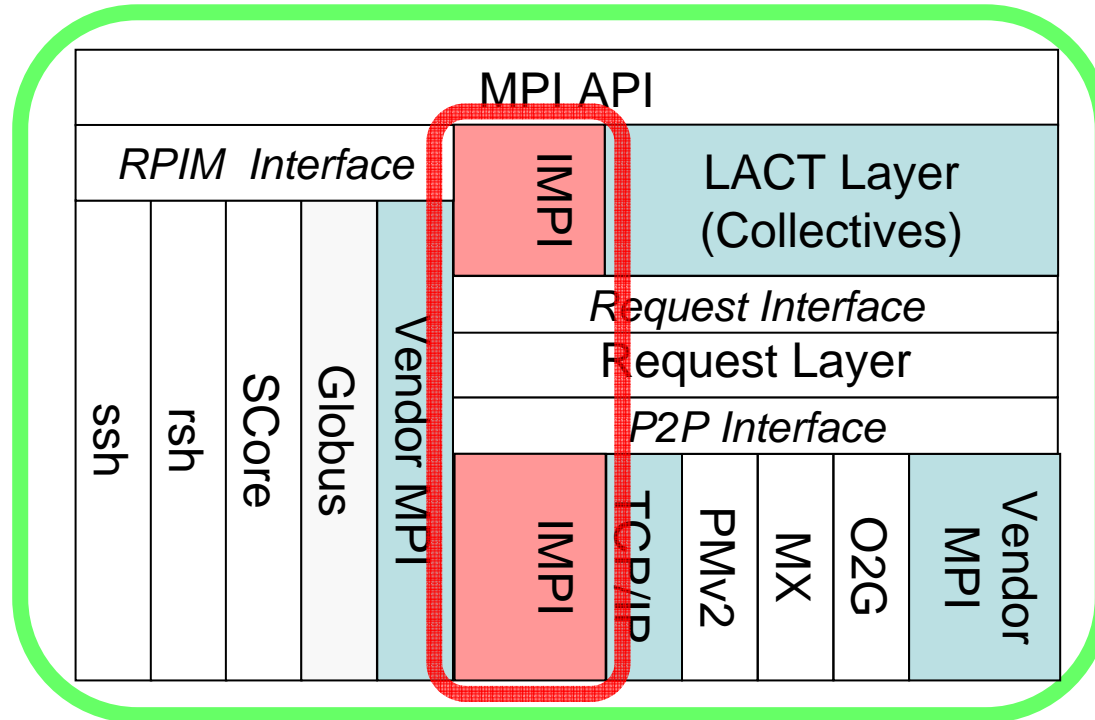
# GridMPI

- MPI applications run on the Grid environment
  - Metropolitan area, high-bandwidth environment:  $\geq 10$  Gpbs,  $\leq 500$  miles (smaller than 10ms one-way latency)
    - Parallel Computation
  - Larger than metropolitan area
    - MPI-I/O





# GridMPI Software Architecture and Standards

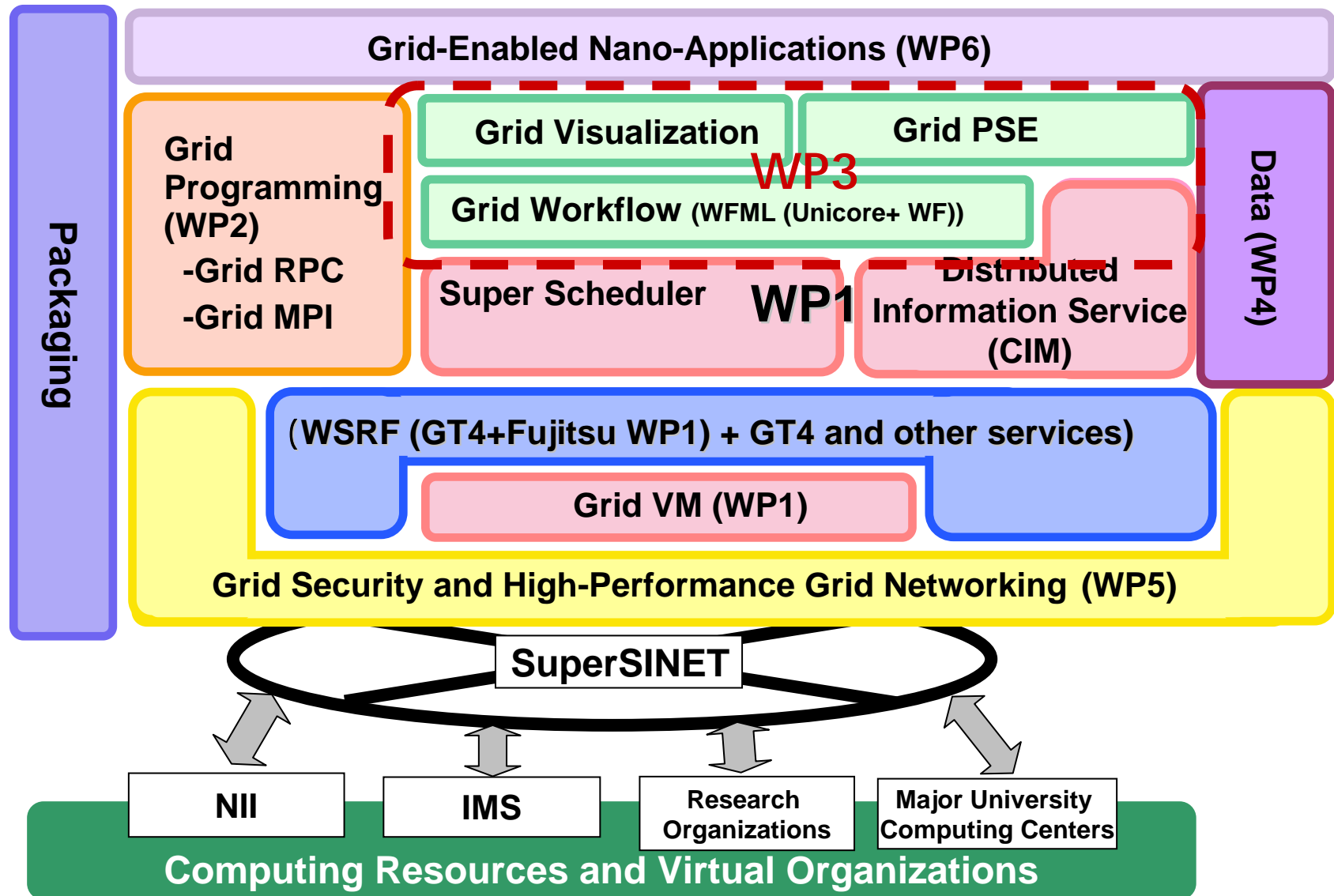


- MPI 2.0 (test suite compliant)
- IMPI (Interoperable MPI)
  - The original IMPI is defined only for the MPI-1.2 feature
  - Extension for MPI-2
- Porting the extended IMPI protocol to Open MPI
- Planning to submit the protocol to NIST

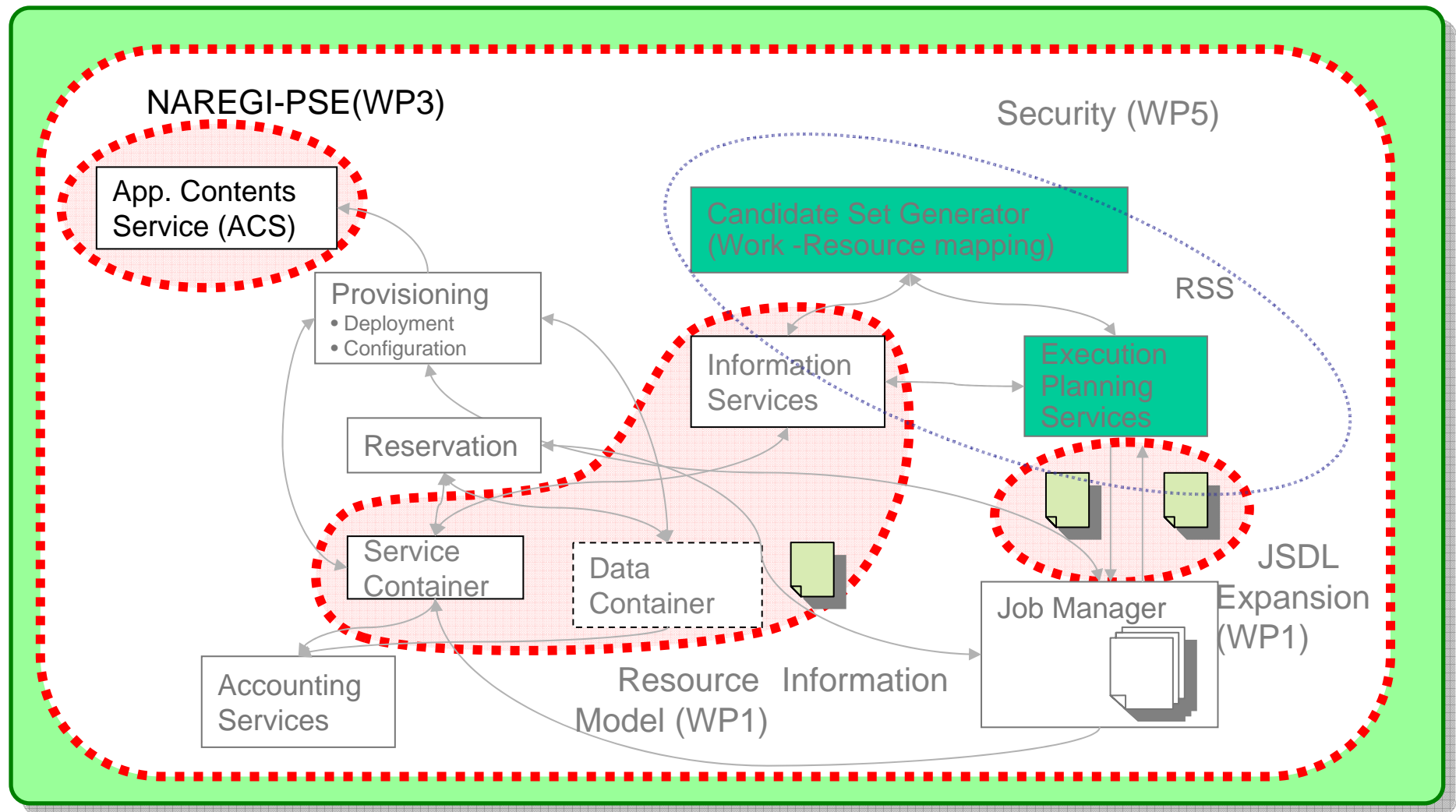




# NAREGI Software Stack (Beta Ver. 2006)

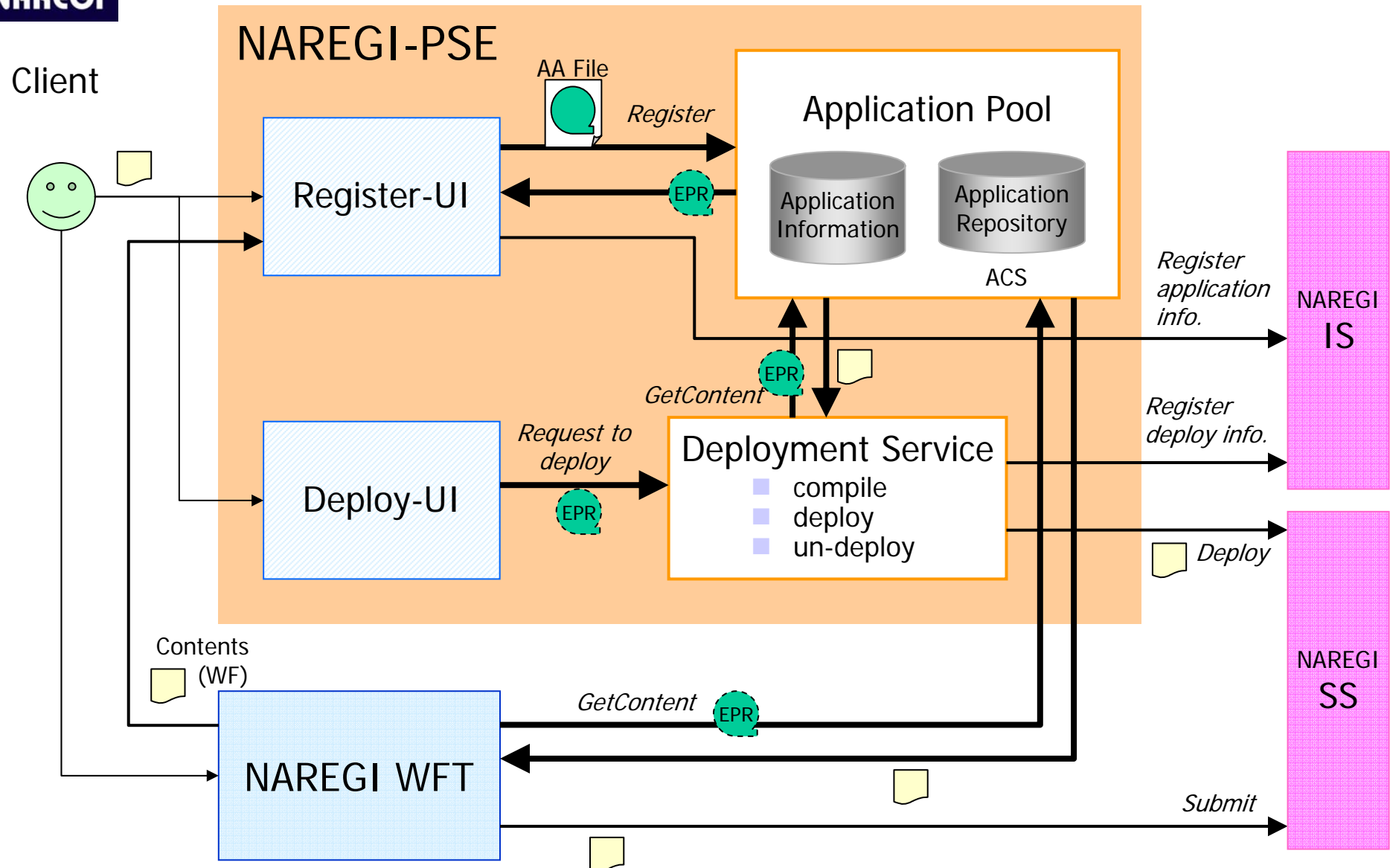


# Architectural Positioning of NAREGI PSE (WP3)





# NAREGI -PSE Architecture

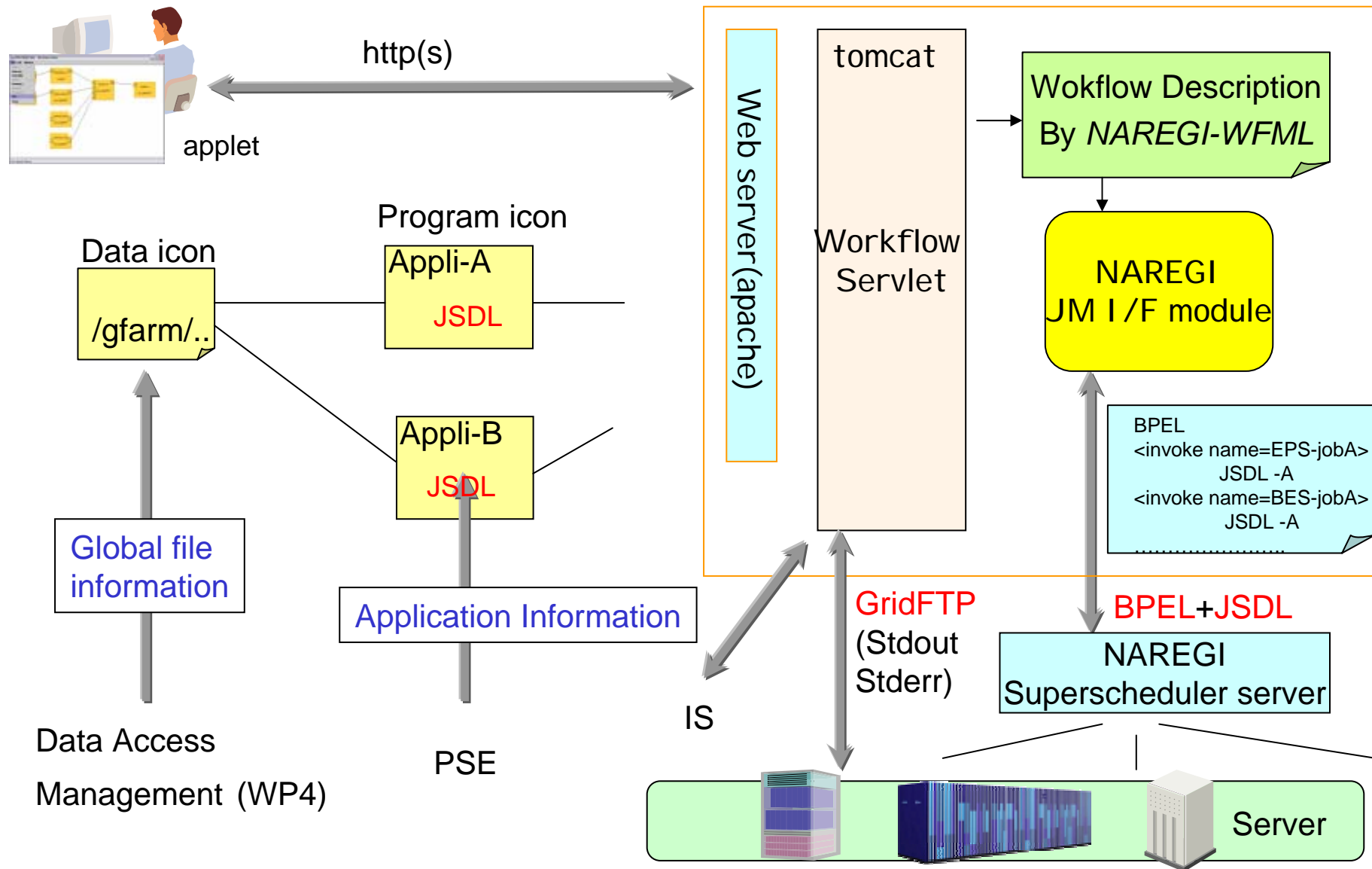




# NAREGI PSE (WP3) Standards

- GGF Standards we help set within a WG
  - ACS (Application Contents Service) WG
- GGF and related Standards we employ
  - CDDLM
  - 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)

# NAREGI WP3 Workflow Tool



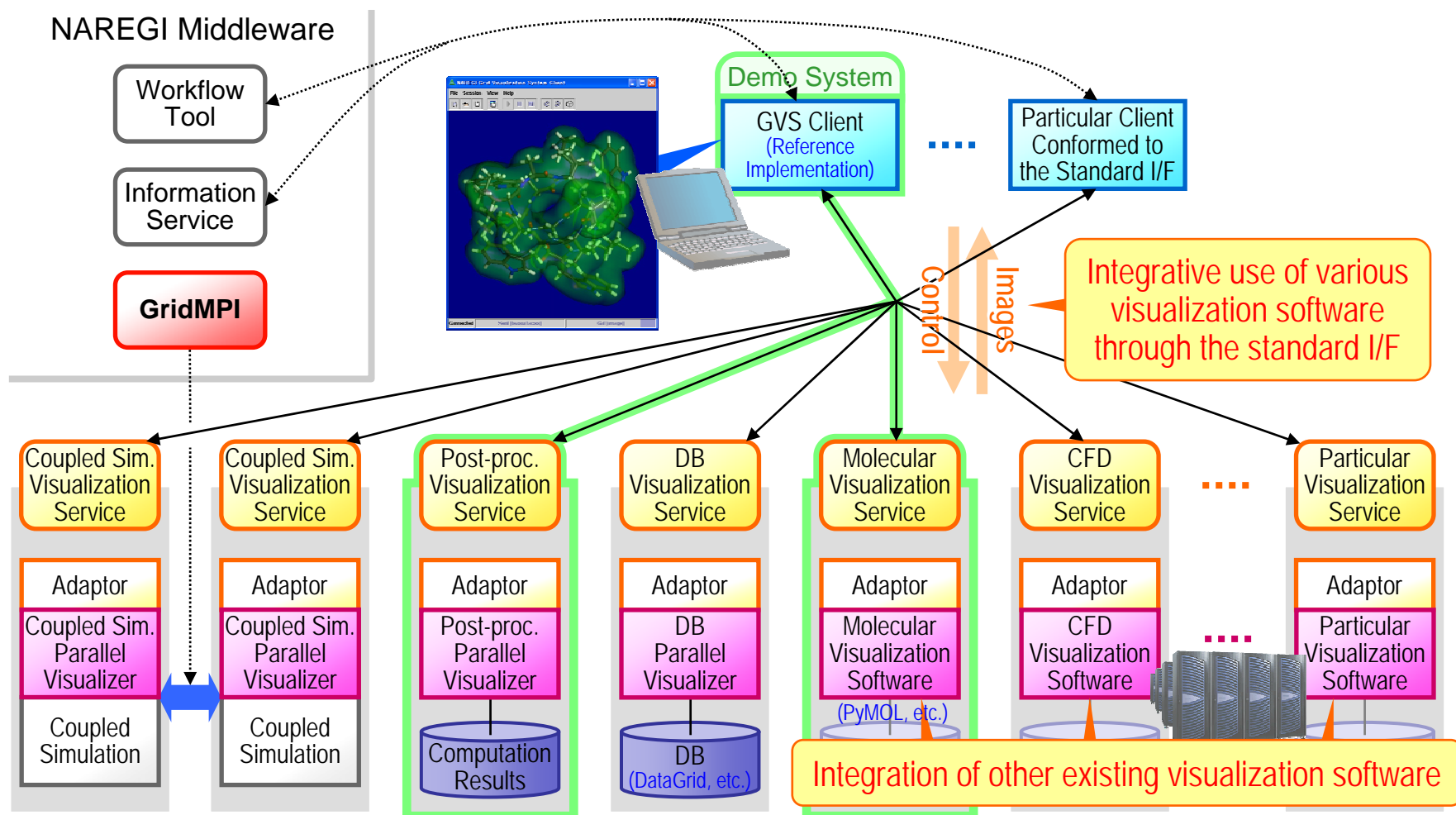


# NAREGI WFT Standards

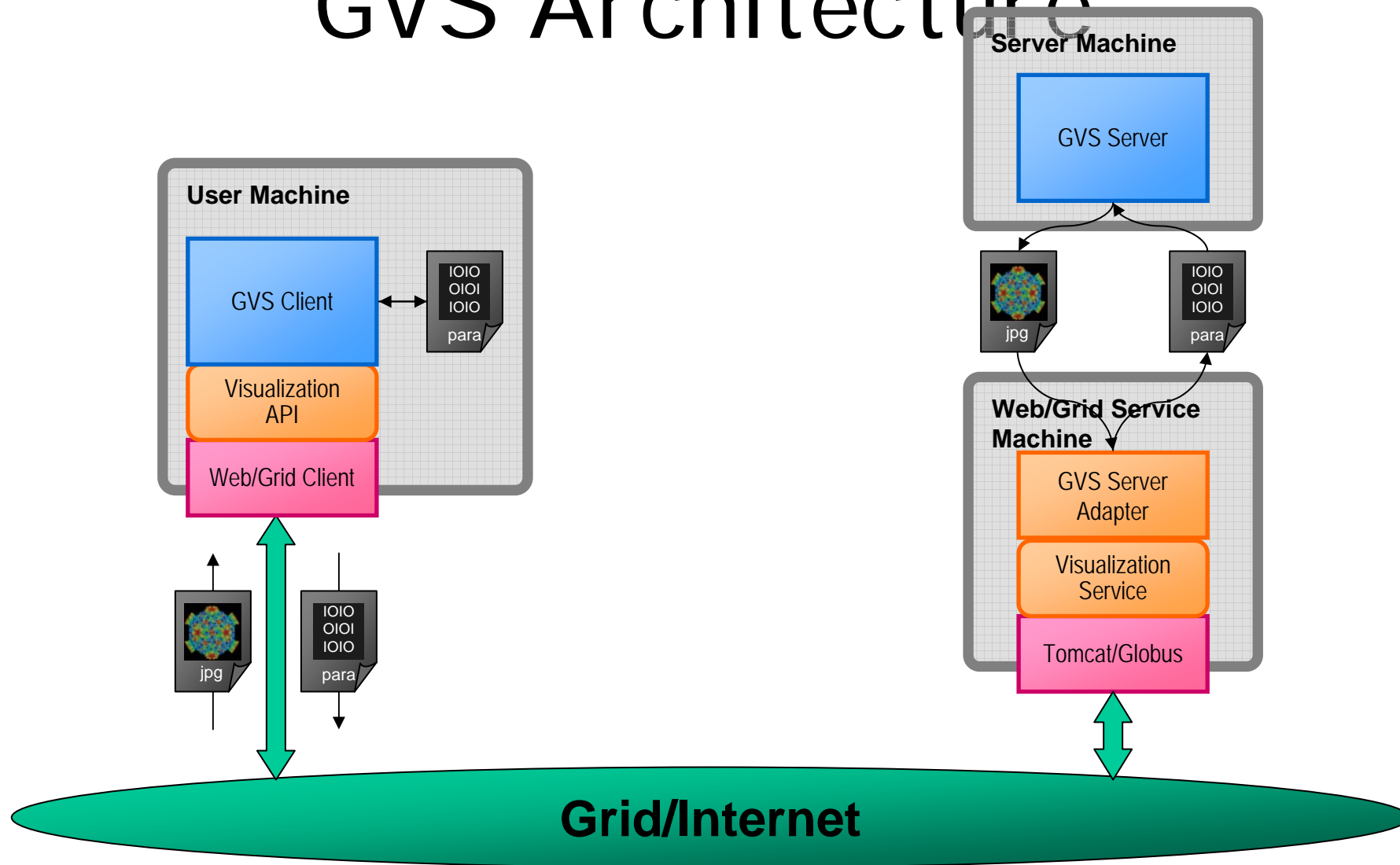
- GGF Standards we help set within a WG
  - Workflow Management Research Group WFM\_RG (currently observing)
- GGF and related Standards we employ
  - GridFTP
  - JSDL
- Other industry standards we employ
  - BPEL+JSDL
    - NAREGI WFML has roots in Unicore WF but is heavily extended, and is translated into BPEL + JSDL
- Other de-facto “standards” we employ
  - Apache+Tomcat
  - Unicore

# GVS Framework

Integrated WSRF-based remote visualization environment



# GVS Architecture





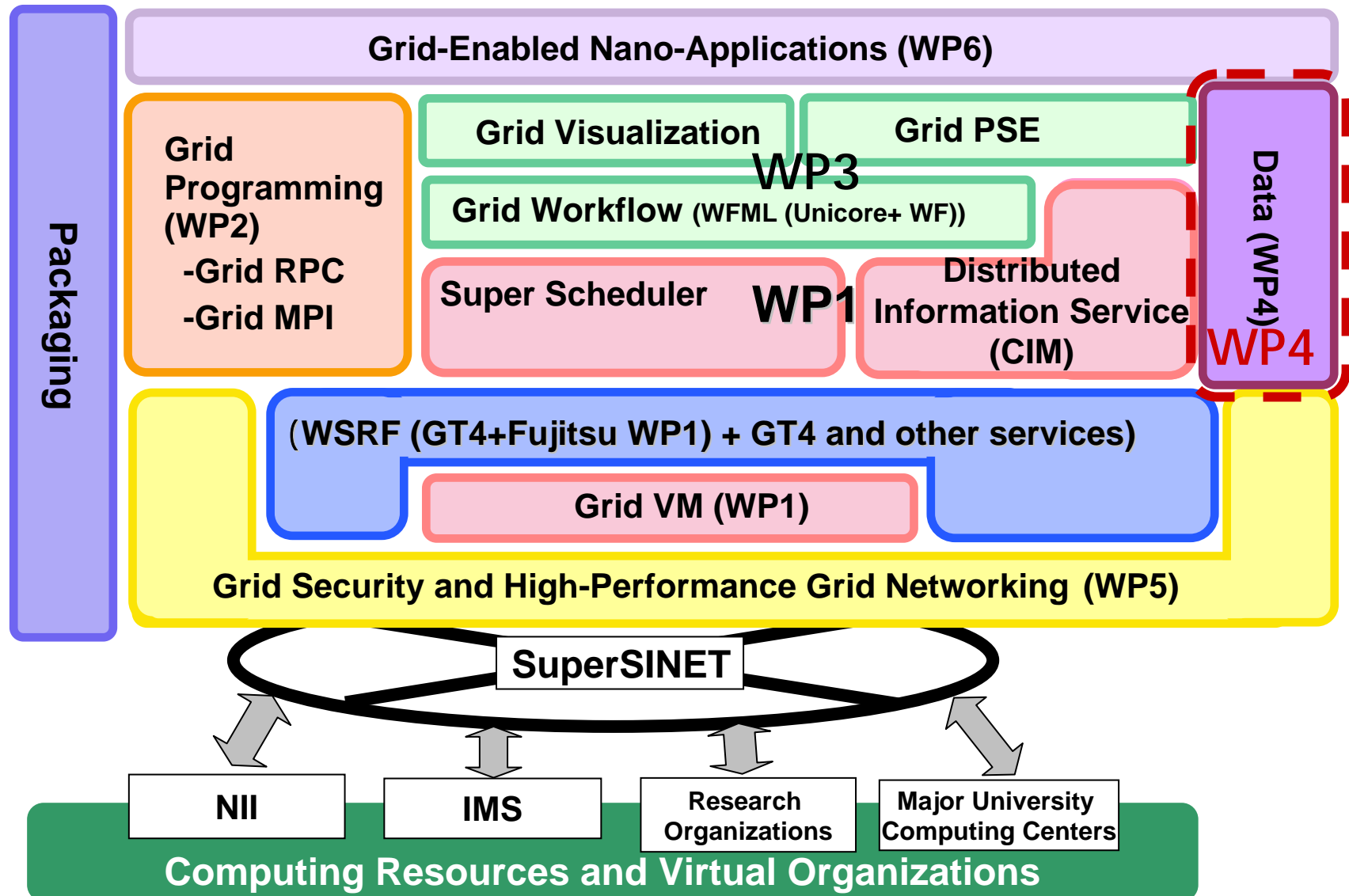


# GVS Standards

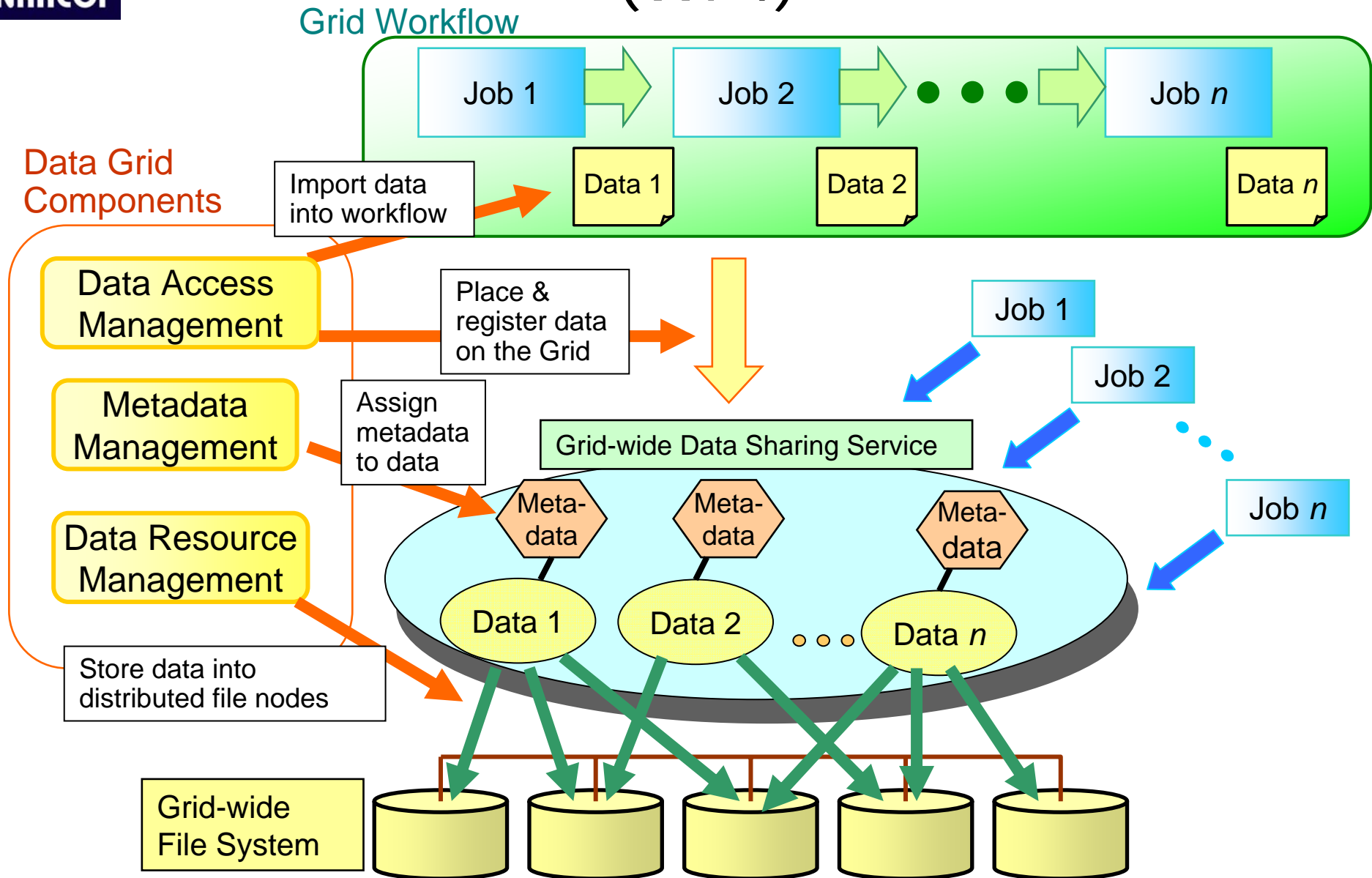
- GGF Standards we help set in WG/RG
  - SAGA W/RG & OGSA-Byte-I O WG
- GGF and related Standards we employ
  - **GridMPI**
  - **WSRF 2.0**
  - **JSDL**
- Other de-facto “standards” we employ
  - Globus 4
  - Tomcat (and associated WS/XML standards)



# NAREGI Software Stack (Beta Ver. 2006)

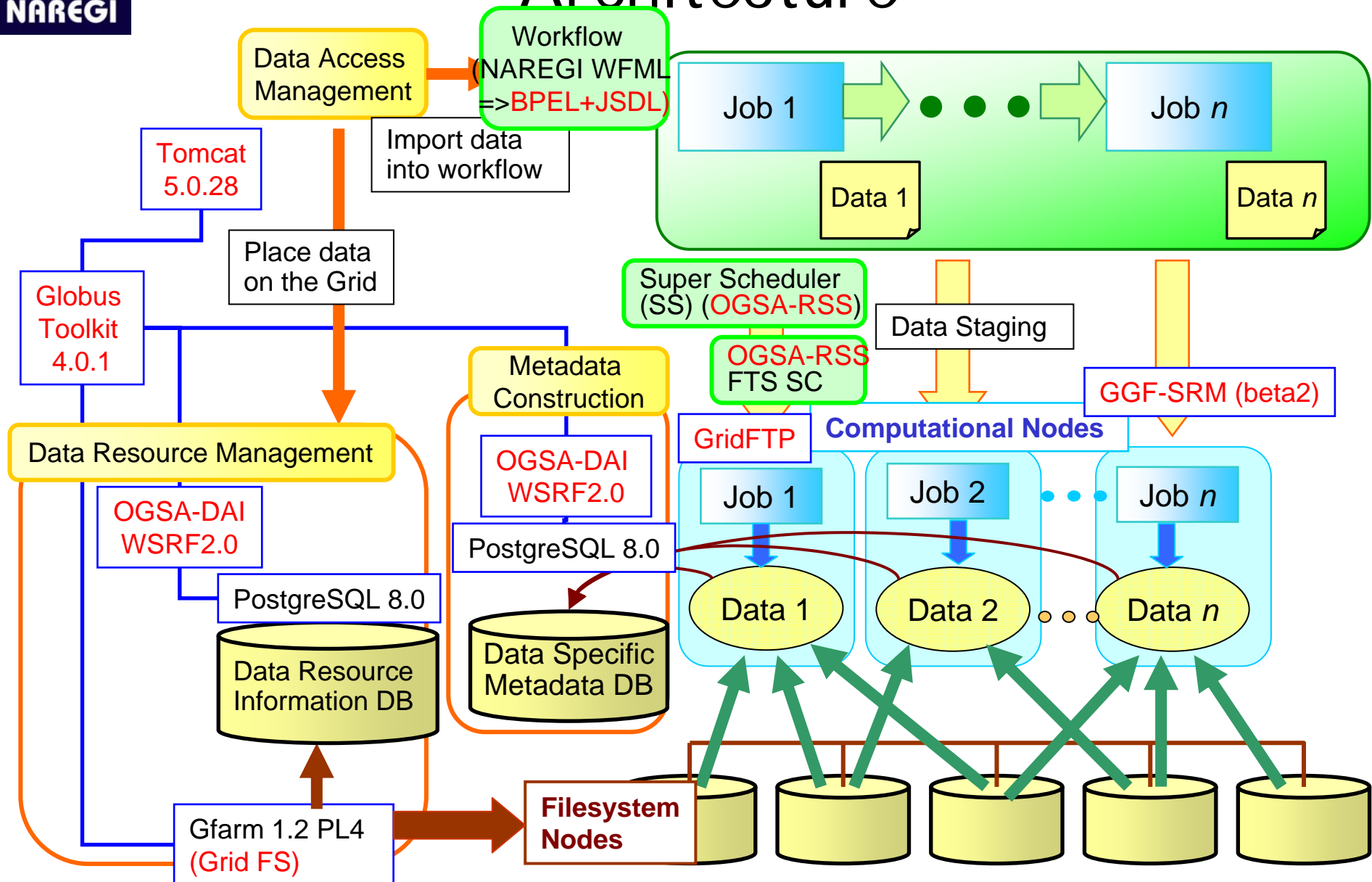


# NAREGI Data Grid beta1 Architecture (WP4)





# 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)

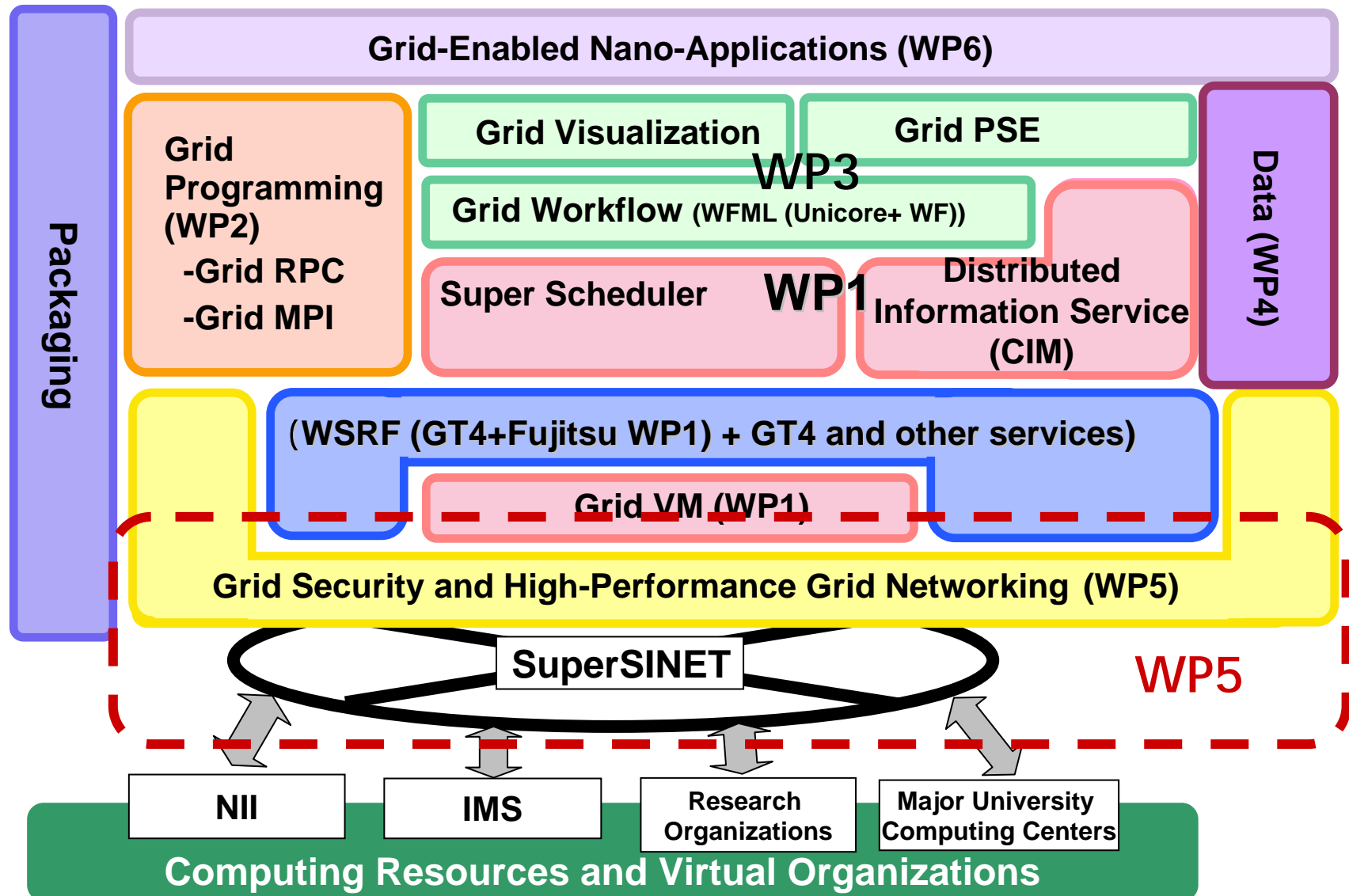


# Roadmaps, Future Plans

- 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 ByteI O?)
- Virtualization and/or Integration of Metadata
  - For Data Exchange among different storage/file-systems (SRM, SRB, ...)
  - Logical Namespace Service (OGSA RNS, WS-Name)

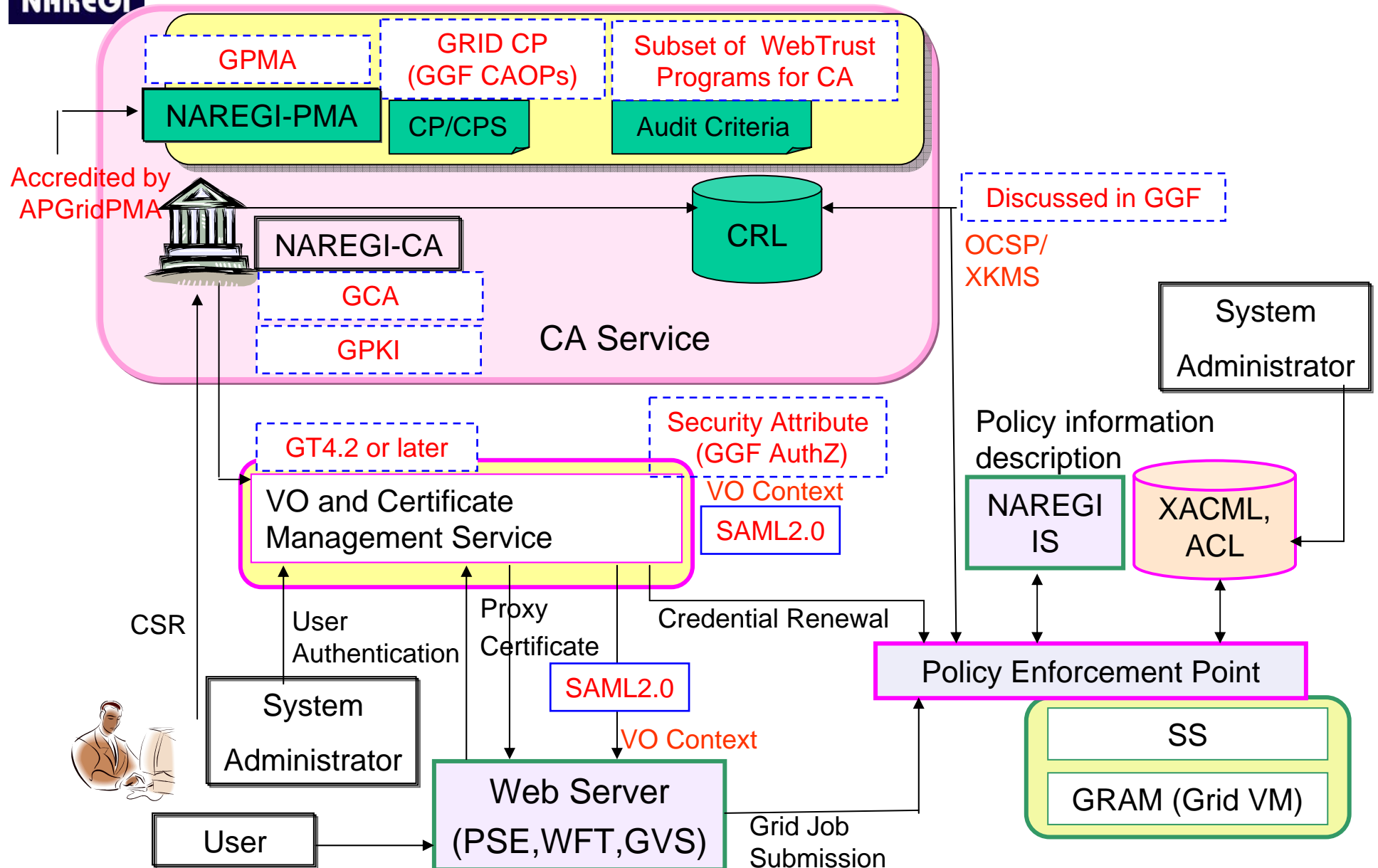


# NAREGI Software Stack (Beta Ver. 2006)



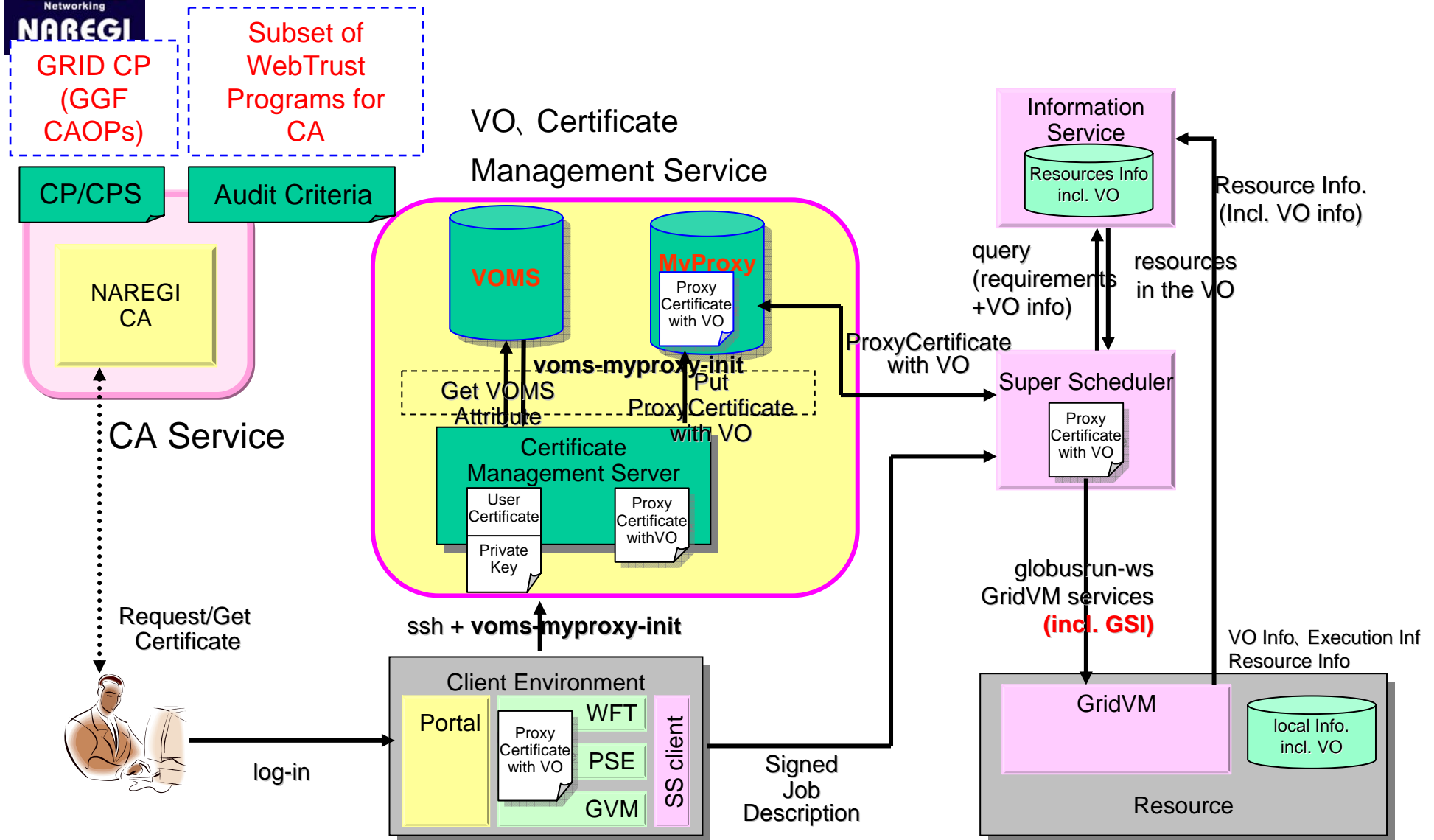


# NAREGI -1.0 Security Architecture





# NAREGI -beta1 Security Architecture



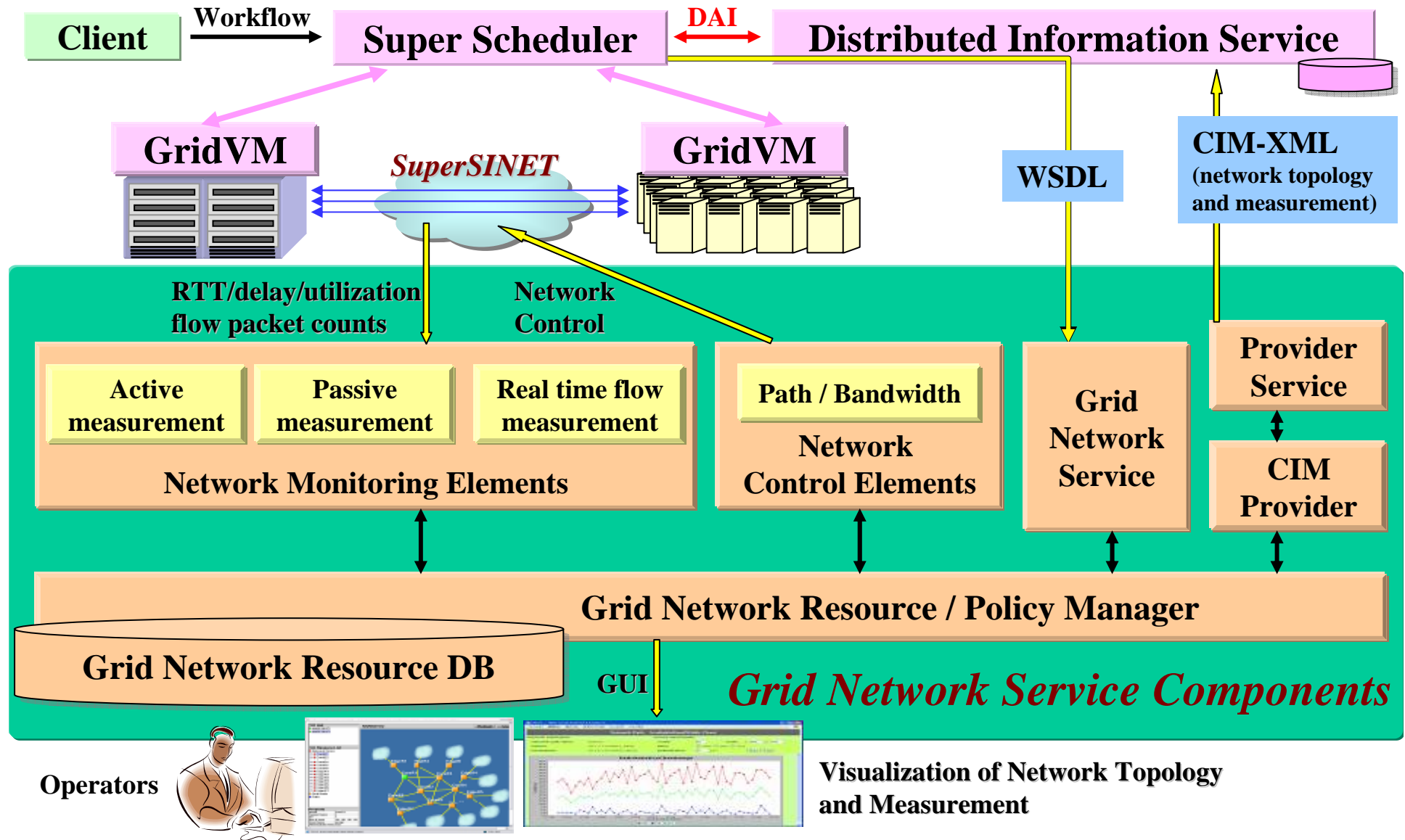


# WP5 Security

## Roadmaps, Future Plans

TOPIC	Standardization Status	Standardization		
		To GGF	From GGF	Comment
Attribute Propagation	-			
Definition of VO	-			
NAREGI Security Architecture	-			
ID federation	SAML2.0 x Grid	-		
Security Attribute (OGSA-AuthZ WG)				
Policy Description	-			

# NAREGI WP5 Network Architecture





# NAREGI WP5 Network Standards

- GGF Standards we help set within a WG
  - Grid High-Performance Networking (We presented network measurement, control and measurement architecture as NAREGI activities)
  - Network Measurement (in Future)
- GGF and related Standards we employ
  - Grid network service use-case scenario between NAREGI Grid middleware and grid network services (discussion currently and contribution in Future)
  - IETF Network Standards such as OCSP
- Other de-facto “standards” we employ
  - Apache + Tomcat

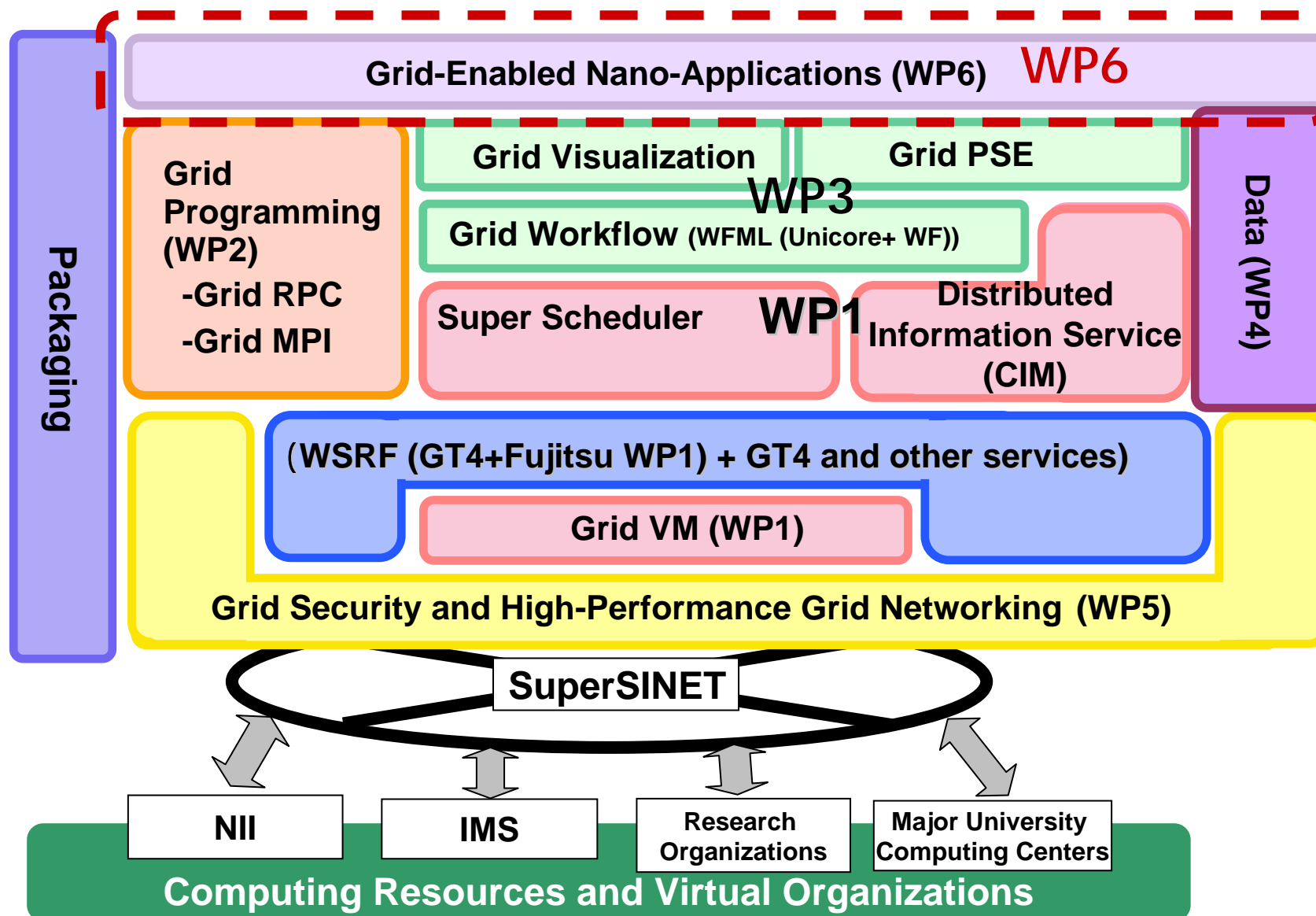


# NAREGI WP5 Sec Standards

- GGF Standards we help set within a WG
  - CAOPs, AuthZ (in Future)
- GGF and related Standards we employ
  - GRID CP[CAOPs]
  - OCSP (in Future)
  - Security Attributes[AuthZ] (in Future)
  - Grid Policy Management Authority
  - Grid Certificate Authorities
- Other industry standards we employ
  - WebTrust Programs for CA (referred to define the grid audit criteria)
- Other de-facto “standards” we employ
  - GT4.2 or later (in Future)
  - MyProxy, VOMS
  - SAML 2.0 (in Future)



# NAREGI Software Stack (Beta Ver. 2006)







# NAREGI Application Mediator (WP6) for Coupled Applications

## Mediator Components

*Support data exchange  
between coupled simulation*

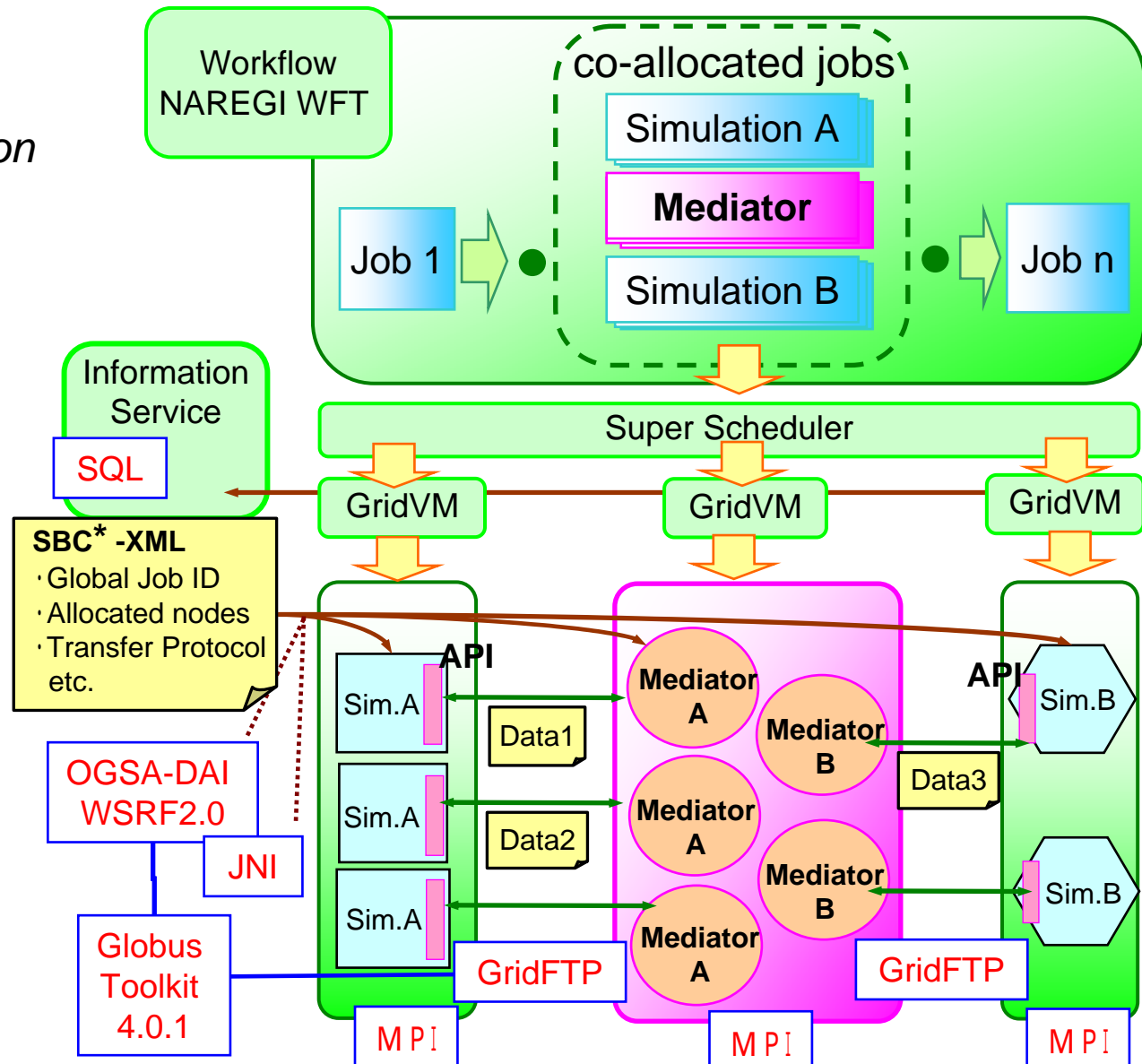
### Data transfer management

- Synchronized file transfer
- Multiple protocol GridFTP/MPI

### Data transformation management

- Semantic transformation libraries for different simulations
- Coupled accelerator

\*SBC: Storage-based communication





# NAREGI WP6 Standards

- GGF and related Standards we employ
  - GridFTP
  - OGSA-DAI
  - WSRF 2.0
- Other industry standards we employ
  - MPI Standard 1.1
  - SQL
- Other de-facto “standards” we employ
  - JNI 1.2 (Java™ Native Interface)
  - Globus Toolkit 4.0.1



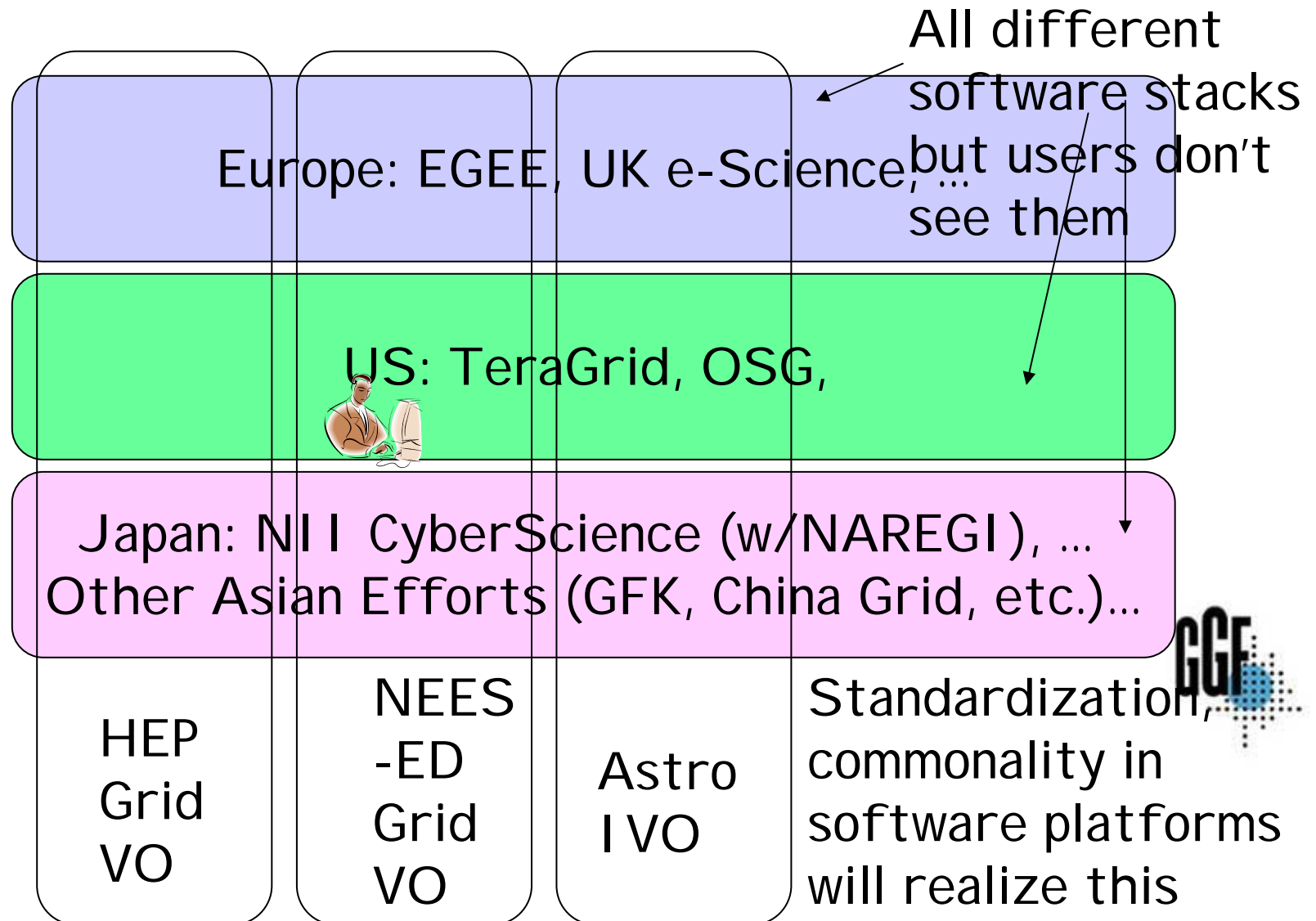
# NAREGI and Standards: Summary

- Interoperability is the key for grid adoption
- NAREGI believes, as a major national grid project, that committing to standards is the best bet for interoperation and industry adoption
  - Will be working next 2 years to set and adopt more standards
- Will be working with other parties, within and outside GGF to achieve common standards



# Grid Interoperations---Ideal World: Ubiquitous VO & user management for international e-Science

Grid Regional Infrastructural Efforts  
Collaborative talks on PMA, etc.



# The Reality: Convergence/Divergence of Project Forces

(original slide by Stephen Pickles, edited by Satoshi Matsuoka)

