

## R&amp;D Project Report

# Overview of Japanese science Grid project NAREGI

Kenichi MIURA  
*National Institute of Informatics*

## ABSTRACT

This paper outlines the National Research Grid Initiative (NAREGI), which started as a five-year project from fiscal 2003 as one of the major Japanese national IT projects currently being conducted. Collaboration among industry, academia, and the government will play a key role in its success. The Center for Grid Research and Development has been established at the National Institute of Informatics as a center for R&D of high-performance, scalable Grid middleware technologies, which are aimed at enabling major computing centers to host grids over high-speed networks to provide a future computational infrastructure for scientific and engineering research. As an example of utilizing such Grid computing technologies, the Computational Nanoscience Center at the Institute for Molecular Science is conducting research on leading-edge, grid-enabled nano-science and nanotechnology simulation applications, which will lead to the discovery and development of new materials and next-generation nano-devices. The NAREGI Grid middleware is expected to be utilized as one of the software layers in the Cyber Science Infrastructure (CSI) framework which has been newly initiated at the National Institute of Informatics.

Starting April 2006, this project will be integrated into the new, seven-year “Development and Application of Advanced High-performance Supercomputer” project, in order to provide the Grid middleware for the peta-scale computing era, in addition to the current research and development efforts toward CSI.

## KEYWORDS

Computational Grid, research Grid, meta computing, virtual organization, hosting service, nano-science, nanotechnology, Cyber Science Infrastructure

## 1 Introduction

Research and development of the Grids for scientific use, which we call “research Grids,” have been conducted mainly by universities and national research institutes in both Japan and other countries. The Japanese national projects for the research on Grids preceding the project to be described in this paper are the Information Technology Based Laboratory (ITBL) project [1] (initiated in fiscal 2001), Grid research in the Super Science Information Network (SINET) project [2] (ini-

tiated in fiscal 2001), BioGrid project [3] (initiated in 2001), and VizGrid project [4] (initiated in fiscal 2002), all of which have been supported by the Ministry of Education, Culture, Sports, Science and Technology (MEXT).

At the start of the National Research Grid Initiative (NAREGI), [5] the computational research environment was expected to be at the level of several hundred TFLOPS (tera-floating-point instructions per second) within five years. This project conducts research and development of the practical Grid middleware required for an operational and robust computational research environment. It is also intended to prove the usefulness of the Grid Computing environment in ac-

Received February 3, 2006; Revised February 28, 2006; Accepted March 5, 2006.

kenmiura@grid.nii.ac.jp

DOI: 10.2201/NiiPi.2006.3.7

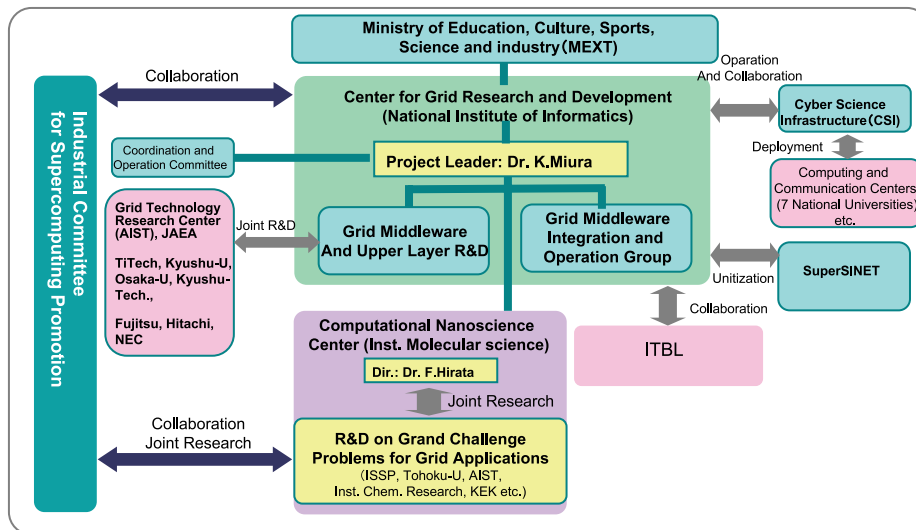


Fig. 1 Research organization of NAREGI.

tual applications (e.g., in the application software and large-scale simulation in the cutting-edge nano-science and nano-technology fields), which are useful for developing new materials and next-generation ultrafine devices, and finally establish a Grid Computing environment that is also accessible to users in industrial communities. The major purposes of the project are to, 1) conduct research on the Grid middleware and integrate the developed software, 2) conduct research on the application of high-speed network technology to the Grid environment, and 3) conduct research and development on the large-scale simulation programs for use in nano-science fields and study their optimization in the Grid environment.

## 2 Organization of NAREGI

To attain the intended goals, NAREGI has a research organization based around two core centers: 1) the Center for Grid Research and Development (National Institute of Informatics), in charge of research and development of Grid middleware, Grid network, and Grid-enabled nano-science applications and 2) the Computational Nanoscience Center (Institute for Molecular Science, National Institutes of Natural Sciences), in charge of research and development of the most advanced application software in the nano-science/nano-technology (Fig. 1). Each center also has collaborating organizations in industry, academia, and the government. The project plans to use SuperSINET, which is managed by the National Institute of Informatics, as the network infrastructure (Fig. 2).

### 2.1 Subjects of research at the Center for Grid Research and Development

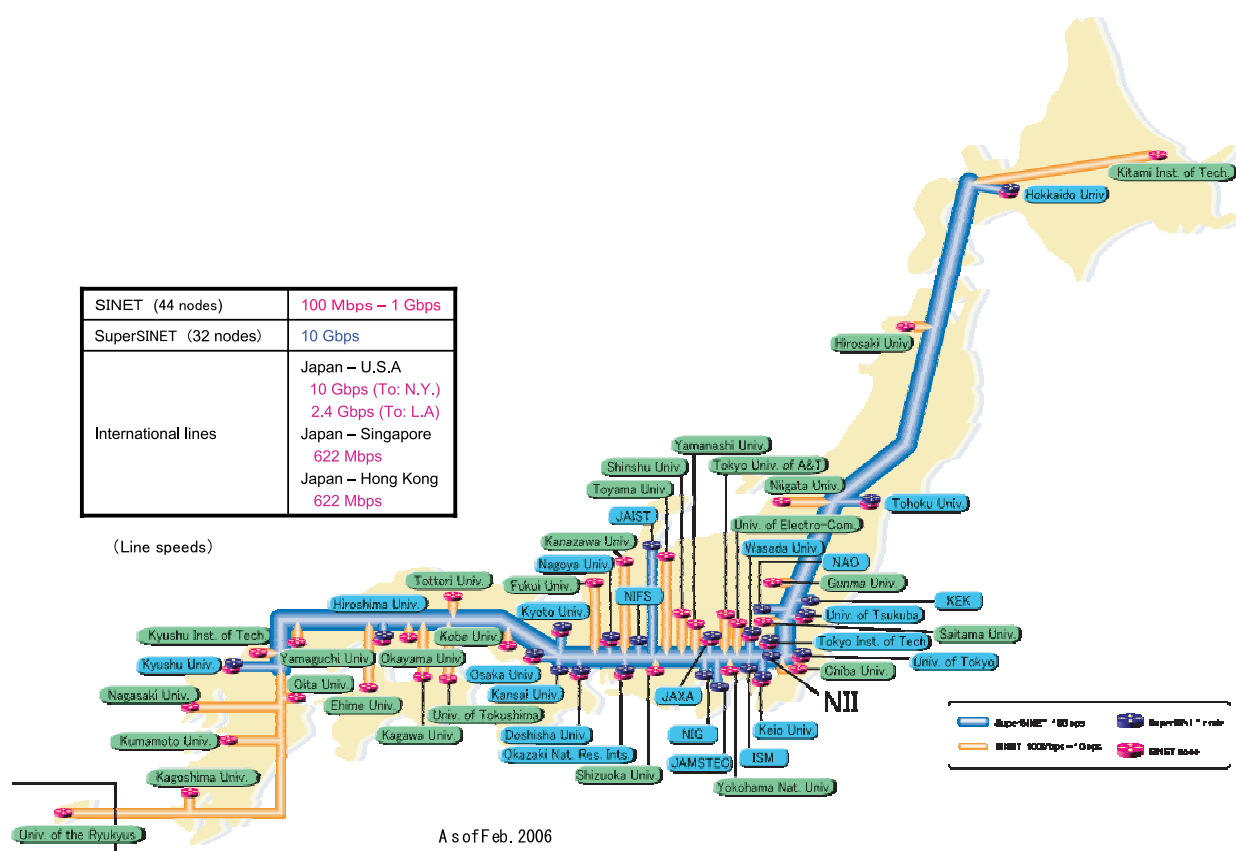
NAREGI does not aim at developing ultrafast computer hardware but focuses on research and development of the software technology (Grid middleware) that links existing heterogeneous computer resources with each other and presents them as seamlessly integrated resources. Under NAREGI, the following research subjects are to be pursued at the Center for Grid Research and Development:

#### 2.1.1 Resource management in a Grid environment (WP1)

Super scheduler brokers computing resources for submitted jobs, based on the job requirements such as the number of CPU's, memory size, and urgency. GridVM controls and preserves computing resources and manages local job scheduling. The Distributed Information Service (DIS) module collects information on a Grid, such as configurations and status of grid resources and usage records, and serves it in an integrated way.

#### 2.1.2 Grid programming environment (WP2)

The Grid RPC System enables grid applications to be efficiently and easily developed. It is based on a model in which a main program calls library functions at remote computers comprising a cluster of clusters, each of which has 10 to 100 CPUs. In the Grid MPI System, the communication library at the TCP/IP layer and the MPI layer enables efficient and interoperable communication while taking account of transmission latency



on the Grid.

### 2.1.3 Software and environment for development of Grid applications (WP3)

To realize a user-friendly Grid environment, Grid PSE (Problem Solving Environment) provides the functionalities to deploy and register applications developed by users into the grid with ease. To help users execute applications distributed over the Grid, Grid Workflow Tool (GWT) enables users to create workflows graphically, to execute them on the Grid, and to control the execution via easy-to-use GUIs, and the Grid Visualization System (GVS) enables users to visualize massive computational results distributed over the grid remotely with a light network load.

#### 2.1.4 Data Grid environment (WP4)

In this environment, widely-distributed and large-scale data resources are virtually integrated. The Access Management System enables such data resources to be imported into the environment. The Data Re-

source Management System provides an interface to manage the resources with globally-unique names. The Metadata Construction System enables users to construct metadata based on ontologies in various scientific fields.

### 2.1.5 Network communication infrastructure (WP5)

To ensure high-performance and secure Grid communications on the NAREGI network infrastructure, fundamental networking middleware adaptive to the wide-area Grid computing will be developed and verified, which include authentication mechanisms supporting secure Grid communications; measurement, management, and control of underlying networks; and communication protocols for Grid applications. Another area in the networking is the development of NAREGI-CA, a certificate authentication software.

### 2.1.6 Grid-enabling nano-simulation software (WP6)

To realize multi-scale and/or multi-physics simulations in the large-scale Grid environment, application-

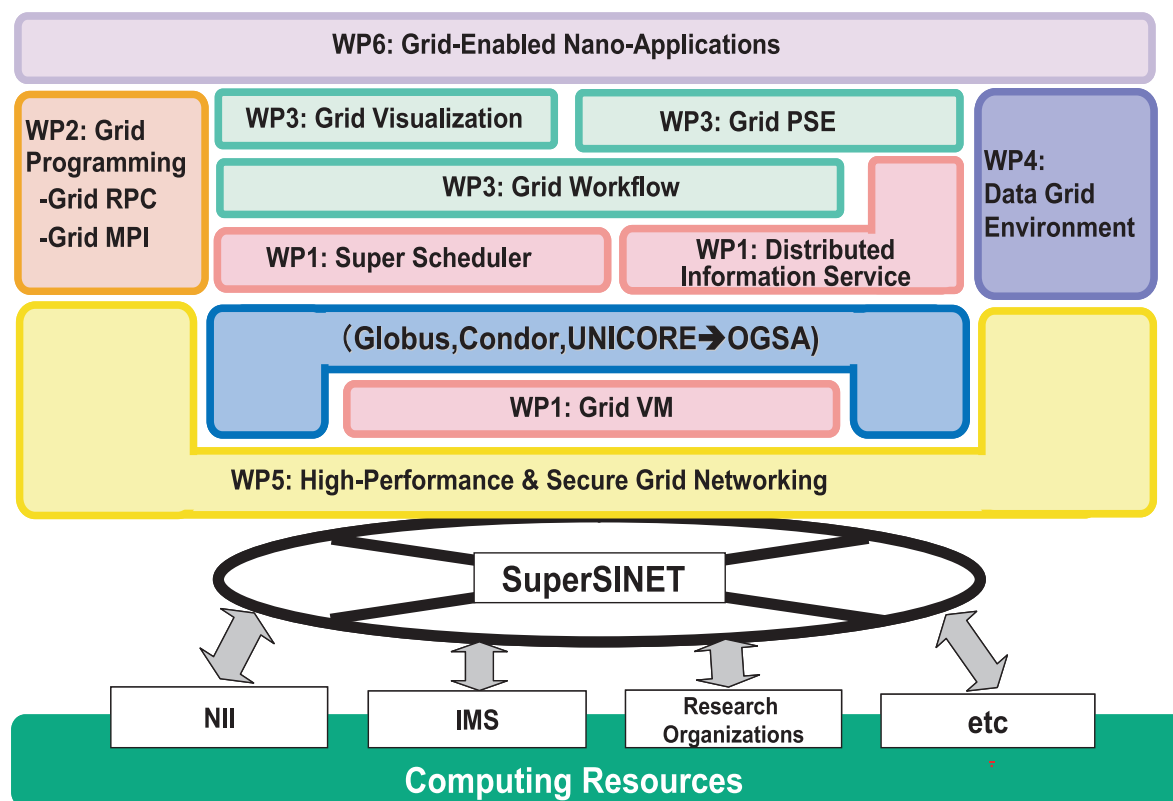


Fig. 3 Research topics of NAREGI Grid Middleware R&D and software stack.

specific middleware, named Mediator is being developed. Mediator provides high-level transparency in automatically transferring and transforming data between heterogeneous application components. To take the greatest advantage of the Grid application environment, we are also developing methodology for grid-enabling, such as modularization of large granularity computational components.

The subjects also include development of the technology to integrate, verify, and deploy the software deliverables of the project. Fig. 3 shows the relationships (hierarchical structure) of the above research subjects and corresponding software components.

In this project, research and development of Grid middleware were not started from scratch. In the first phase (so-called  $\alpha$  version), the Grid middleware, for example, UNICORE, [6] Globus Toolkit, [7] and Condor, [8] which are already popular as de facto standards, have been used appropriately as the basis; research and development have been conducted concerning the upper-layer of the middleware. In the second phase, the base software for NAREGI middleware will evolve into the Open Grid Services Architecture (OGSA) framework, [9] which is commonly regarded as the future di-

rection by the Grid community.

## 2.2 Subjects of research at the Computational Nanoscience Center

From the viewpoint of nano-science, proving the concept that Grid computing is useful as a research tool for nano-science is one of the major objectives of this project. This computational aspect of nano-science is regarded as complementary to the experimental aspects in the “wet-lab.”

It is to be hoped that, by taking advantage of computational science technology such as large-scale simulations on the Grids, a new methodology can be established to solve the types of nano-scale problems that cannot be solved by the conventional theories and methodologies.

The research and development of most advanced nano-science applications are led and conducted by the Institute for Molecular Science in collaboration with research institutions such as the Institute for Solid State Physics of Tokyo University, Institute for Materials Research of Tohoku University, the Institute for Chemical Research of Kyoto University, and the National Institute of Advanced Industrial Science and Technology,

as well as with industry. The research subjects are as follows:

#### 2.2.1 Functional nano-scale molecule

This research subject involves the establishment of technology to practically design and analyze the nano-scale molecules that have desired structures, physical properties, and functions. This research will be conducted with simulations based on quantum chemistry computations such as the molecular orbital (MO) method, density functional theory (DFT), and the fragment molecular orbital method (FMO) [10] as a combination of MO and DFT. This research will include suggested applications of the computational technology to develop new high-efficiency catalysts, molecular-scale electronics, and molecular magnetism.

#### 2.2.2 Nano-scale molecular assembly

This involves the clarification of the structural formation of nano-scale molecular assemblies and single nano-scale molecules by self-organization. It also involves the theory of function representation based on the structures by use of the molecular dynamics method, generalized-ensemble algorithms, and statistical mechanics computation based on the reference interaction site model (RISM). The research subjects here include the establishment of a simulation methodology to estimate the structures and functions of proteins and bio-membranes.

#### 2.2.3 Nano-scale electronic system

This involves the execution of high-accuracy, large-scale numerical computation based on the solid-state electron theory for the purpose of controlling the internal degrees of freedom (charge, spin, and orbit) of electrons, by first analyzing their behavior at each degree of freedom individually then analyzing the correlations among charge, spin and orbit. The research subjects here include the establishment of a basis for developing next-generation electronic technologies, for example, nano-scale electronic devices that have the functionality of optical switches and other new functions.

#### 2.2.4 Nano-scale magnetism

This involves the development of a new computational method concerning 1) the spin state based on the non-empirical calculation of electron state and quantum Monte Carlo method, 2) the clarification of the mechanism of presentation of the magnetic and dielectric properties of nano-scale substances by large-scale numerical computation, and 3) the establishment of a fundamental research platform for developing functionally new devices, especially nano-scale magnetic devices, including superdense magnetic memory.

#### 2.2.5 Design of nano-scale complex system

This involves the clarification of the properties of the individual nano-scale components of a complex system that are required to apply nano-technology by using a simulation based on the fragment molecular orbital (FMO) and the first principle molecular dynamics methods as well as the properties of a complex system that contains a macro circuit substrate and nano-scale components, for example, quantum wires, quantum dots, and a phase-separated alloy system.

#### 2.2.6 Grid nano-simulation system

In order to enhance collaboration among users in a grid environment, where the researchers may develop application software in different fields at separate sites, providing the means to combine those programs together over the Grid for coupled simulations is highly desirable. The Grid nano-simulation system, currently under development, consists of a data input module called IGNITION and a module called GIANT (General Inter-Application Translator). GIANT, together with the Grid Workflow Tool, allows coupling of the programs and data of such applications as the non-empirical molecular orbital method, the density functional theory, the molecular dynamics method, and the statistical mechanics computation, to be used for nano-science and nano-technology research in this project.

#### 2.2.7 Industrial participation — an early exposure —

Research proposals from the industry have been solicited with applications that meet the demands from the industrial sector, in order to verify the concept and obtain useful feedback to the above research and development. Through such efforts, it will be possible to construct a Grid-enabled simulation system based on highly efficient methodologies and easy-to-use software.

Fig. 4 outlines the research subjects at the Computational Nanoscience Center and their expected contributions to industry.

### 3 Grid and nano-science/nano-technology simulation

Applying the Grid technology to computations in nano-science and nano-technology is expected to enable ultra large-scale distributed simulations, large-scale linked computations and coupled simulations requiring complicated workflows, and simulations using extensively distributed data, which have been difficult to accomplish on a single supercomputer because of limited computing power and storage capacity. Some examples of expected applications of Grid technology are as follows.



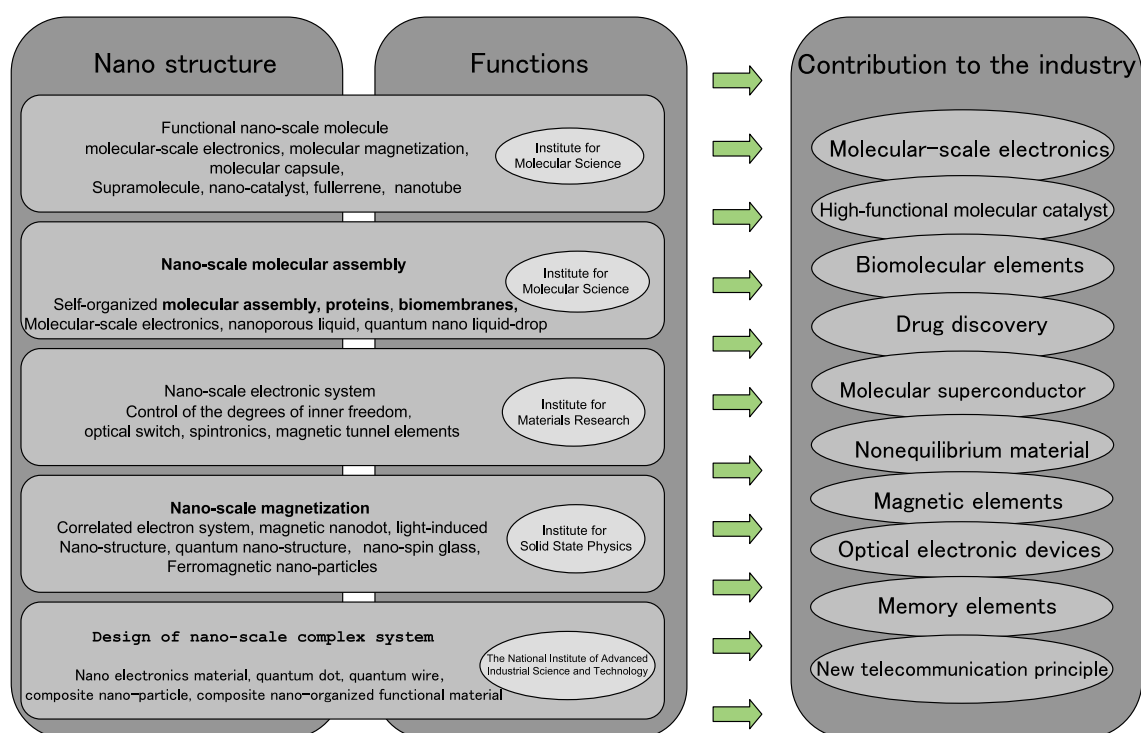


Fig. 4 Research topics of NAREGI Computational Nanoscience R&D and their targeted application in industry.

### 3.1 Ultra large-scale distributed simulation on the Grid

Tools will be developed based on the high-performance communication libraries on the Grid, such as Grid RPC and Grid MPI. These tools will enable large-scale simulations to be executed utilizing the distributed resources on the Grid.

### 3.2 Seamless data exchange and multi-scale linked/coupled simulations

The Grid middleware as a common platform will be developed to efficiently exchange data between nano-science application programs running on different computers on a Grid.

Complex, large-scale linked computation and large-scale coupled simulations can be performed when the Grid middleware is built in with the data conversion capabilities, which are applicable to the wider multi-scale problems and multi-physics problems. Multi-scale problems analyze physical systems consisting of various models with different scales in time and length. Multi-physics problems, on the other hand, analyze physical systems that are described as a combination of different physical models (e.g., the combination of classical mechanics and quantum mechanics models or of fluid and structural models). In solving multi-physics problems, to satisfy the conservation laws, the physical

quantities at the boundary surfaces must be kept consistent as the numerical simulations proceed.

Fig. 5 illustrates one example of coupled simulation.

In this case, two application programs, the fragment molecular orbital method (FMO) and the reference interaction site model (RISM) [11] are coupled, where the quantum behavior of a molecule under the influence of a solvent is simulated by using the Grid middleware called Mediator.

### 3.3 Workflow scenario in the Grid environment

When conducting large-scale parameter surveys and linking computing resources to form pipelines, and also when conducting collaborative research with remotely located researchers, the PSE will greatly enhance the efficiency of research and development activities. The PSE is a simulation environment providing flexible access, linkage, and deployment of the nano-science application programs, computing modules, and libraries that have already been registered by researchers on a Grid.

## 4 International cooperation and standardization

Grid research activities have been spreading worldwide not only in the US and Europe but also in the Asia

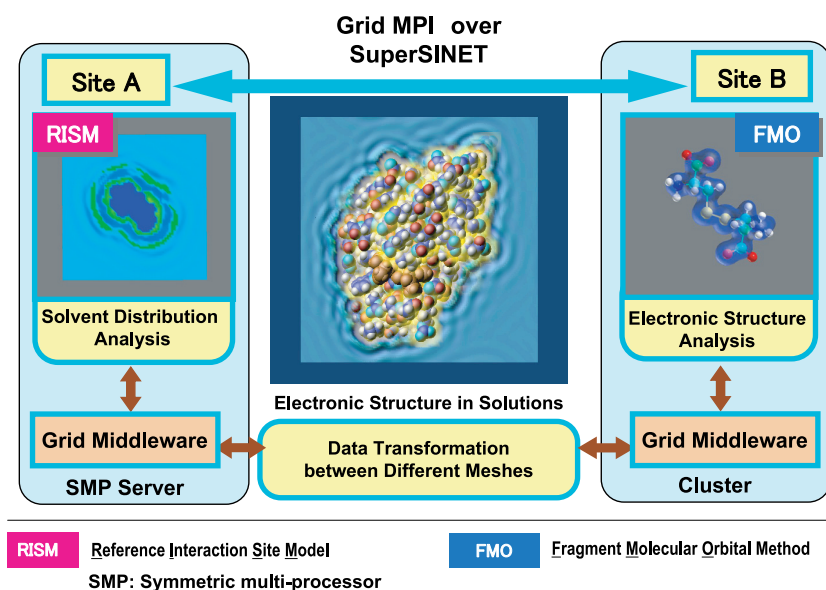


Fig. 5 Example of coupled simulation.

Pacific region, where a research organization named the Asia Pacific Partnership for Grid Computing has been formed. The leading international organization for Grid research is the Global Grid Forum (GGF), [12] which has various working groups to hold discussions about necessary Grid functions and perform the tasks and documentation required for Grid standardization. GGF is based on the idea that the software components for Grids should be open and interoperable and individual Grid projects should not independently develop all the necessary software from scratch, but should share useful products with each other. Therefore, NAREGI emphasizes active participation in the working group meetings held by GGF and other organizations; NAREGI seeks to deepen its cooperative relations with international Grid projects, for example, the Teragrid project [13] in the US as well as the UNICORE forum [6] and the EGEE (Enabling Grids for E-science) [14] and DEISA (Distributed European Infrastructure for Supercomputer Applications) [15] projects in Europe. NAREGI also intends to disclose the developed software as open sources.

## 5 Current status and future directions

As shown in Fig. 6, NAREGI has entered the third year and is now fully engaged in the software development phase for integration and delivery of the NAREGI  $\beta$  version middleware (scheduled in May 2006). The research projects under NAREGI have made considerable progress in the development of Grid middleware, and some of them have already produced initial results, including Grid RPC, Grid MPI, and NAREGI-

CA. These achievements were demonstrated at the international conference on supercomputing and networking, SC2005, [16] held in November 2005 in Seattle, Washington in the US and also at the NAREGI Symposium in Tokyo in February 2006.

As for the computational resources for this project, a 10-TFLOPS-class computer facility was installed at the Computational Nanoscience Center of the Institute for Molecular Science and a 5-TFLOPS-class computer facility was installed at the Center for Grid Research and Development of the National Institute of Informatics in March 2004. These computer facilities are connected to Super-SINET to constitute the NAREGI testbed for the purpose of middleware and application software development as well as for testing.

As for the future direction of the NAREGI, two new developments are noteworthy. Firstly, the National Institute of Informatics initiated the "Cyber Science Infrastructure (CSI)" project in FY 2005, [17] and the NAREGI Grid middleware is regarded as one of the essential components of the Cyber Science Infrastructure, which aims to provide wider access of computational and network resources to the academic and educational communities in Japan (Fig. 7). Secondly, the NAREGI will join the project named "Development and Application of Advanced High-performance Supercomputer" in April 2006, which will also be sponsored by MEXT for seven years. NAREGI's role in this project is to enhance the functionality and scalability of NAREGI grid middleware to provide the computational research environment in the forthcoming peta-scale computing era.

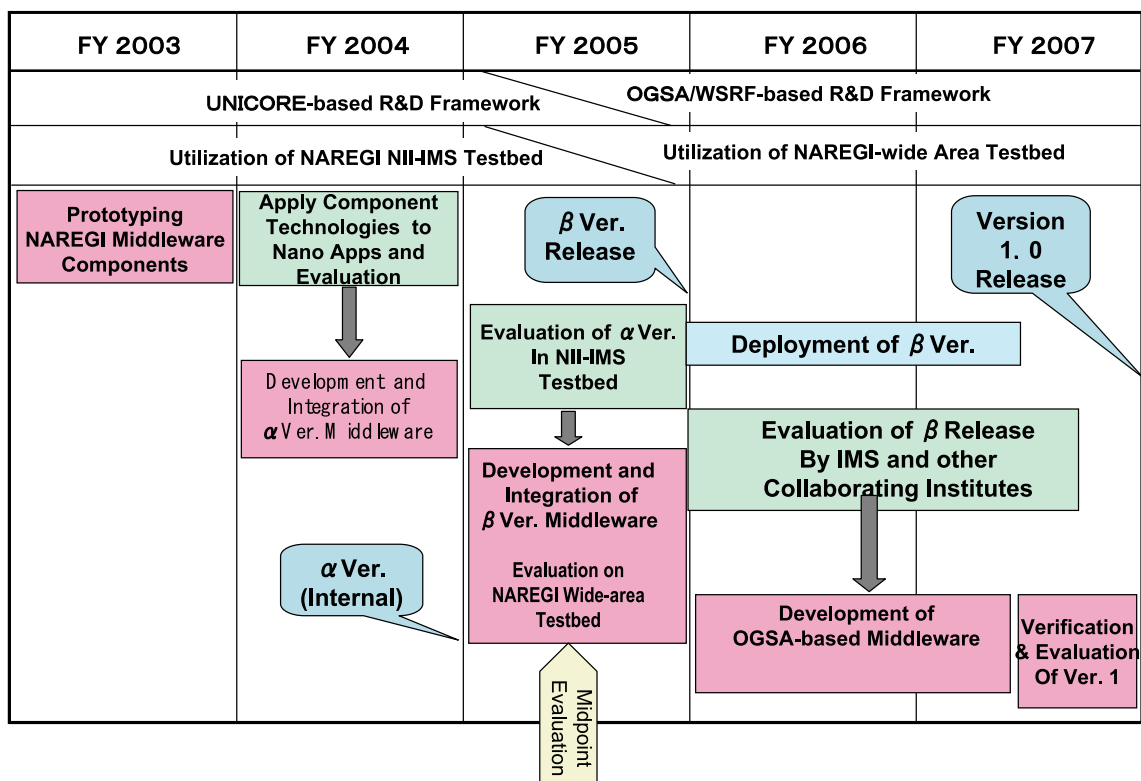


Fig. 6 Schedule and milestones of NAREGI.

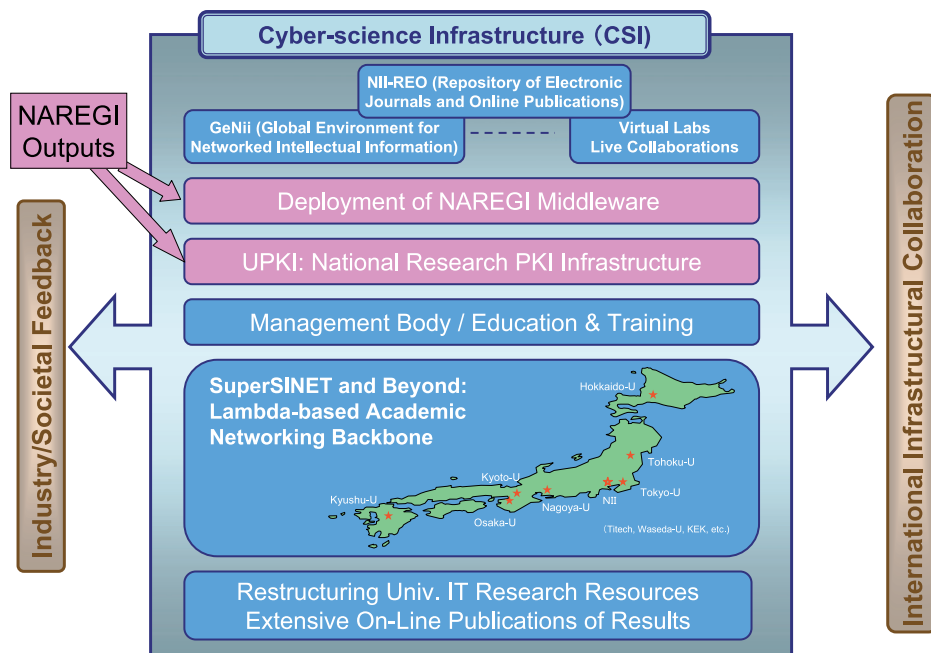


Fig. 7 Towards cyber-science infrastructure for Research and Development.



## 6 Conclusion

In NAREGI, development of the Grid middleware for seamless federation of heterogeneous resources is the primary objective. Some Grid research and development topics are deliberately not covered by NAREGI, which focuses on the computational and data aspects of the Grids hosted by a federation of R&D centers. These topics include utilization of desktop resources (desktop Grids), and collaborative human interface Grids, such as the Access Grid. NAREGI will certainly stay informed on the research and development in such areas, and we may collaborate with such projects and/or include their efforts as part of the final software distribution.

Finally, we regard Grid as one of the fundamental technologies of the IT infrastructure in the 21st century, and expect that the outputs of NAREGI will promote research and development in scientific fields, improve Japan's international competitiveness, and have a major economic impact.

## Acknowledgements

The author expresses his thanks to the researchers in NAREGI for providing valuable information and materials for this paper.

## References

- [1] ITBL: Information Technology Based Laboratory  
<http://www.itbl.riken.go.jp>
- [2] SuperSINET: Super Science Information Network  
<http://sinet.ad.jp>
- [3] BioGrid: <http://www.biogrid.jp>
- [4] VizGrid: <http://www.vizgrid.org>
- [5] NAREGI: National Research Grid Initiative  
<http://www.naregi.org>  
(National Institute of Informatics)  
<http://nanogrid.ims.ac.jp/nanogrid/>  
(Institute for Molecular Science)
- [6] UNICORE: Uniform Interface to Computing Resources  
<http://www.unicore.org>
- [7] Globus Toolkit: <http://www.globus.org>
- [8] CONDOR: <http://www.cs.wisc.edu/condor/>
- [9] OGSA: Open Grid Services Architecture  
<http://www.globus.org/ogsa/>
- [10] K. Kitaura, E. Ikeo, T. Asada, T. Nakano, and M. Uebayasi, "Fragment molecular orbital method: An approximate computational method for large molecules," *Chem. Phys. Lett.*, vol.313, pp.701–706, 1999.
- [11] H. Sato, F. Hirata, and S. Kato, *J. Chem. Phys.*, vol.105, p.1546, 1996.
- [12] GGF: Global Grid Forum  
<http://www.globalgridforum.org>
- [13] Teragrid: <http://www.teragrid.org>
- [14] EGEE: <http://www.egee.org>
- [15] DEISA: <http://www.deisa.org>  
<http://www.einfrastructures.org/>
- [16] SC2005: Supercomputing Conference 2005  
<http://www.sc-conference.org/sc2005/>
- [17] Cyber Science Infrastructure: <http://www.nii.ac.jp>



### Kenichi MIURA

Dr. Miura is a professor in high-end computing at the National Institute of Informatics (NII). He is the project leader of the Japanese National Research Grid Initiative (NAREGI). Dr. Miura received the Ph.D. degree in computer science from the University of Illinois at Urbana-Champaign in 1973. He joined Fujitsu in 1973, where he has been engaged in high-end computing. From 1992 to 1996, he was Vice President and General Manager of the Supercomputer Group at Fujitsu America, Inc., where he was responsible for supercomputing-related activities in the U.S.A. Since June 2002, he has been a Fellow of Fujitsu Laboratories, Ltd. He also served as a visiting professor at the Computer and Communications Center of Kyushu University from 1990 to 1993, and has been a visiting professor at the National Astronomical Observatory since April 2005.