Korea e-Science Project

Dr. Sang Boem Lim

Supercomputing Center at KISTI

What is Grid?

Technologies and infrastructure that support sharing and coordinated use of diverse resources in dynamic, distributed virtual organizations (VO's) – Ian Foster

Ref. Paul Avery
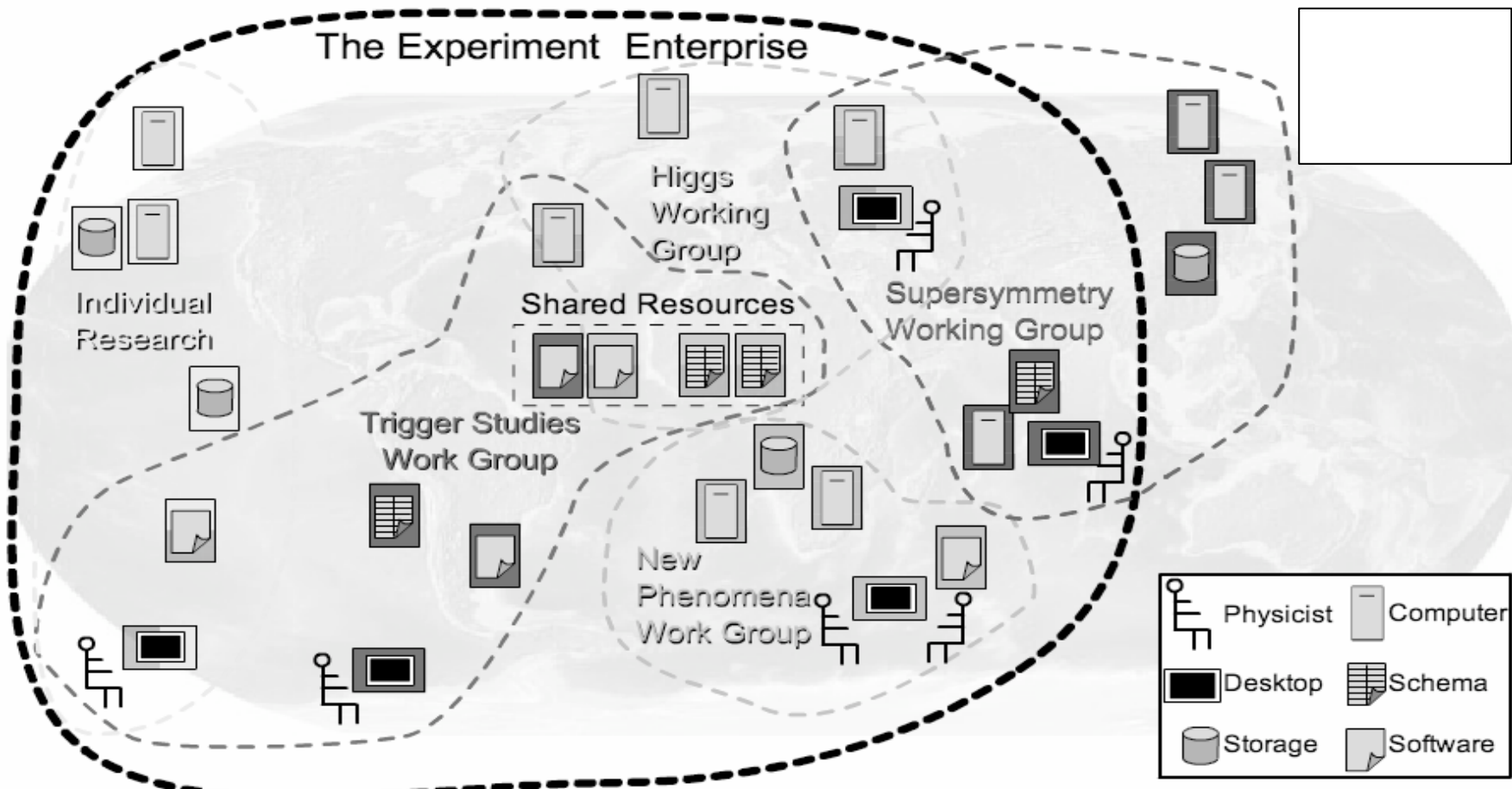
Korea e-Science Project

Definition

National e-Science Project in Korea is about providing innovatively enhanced research infrastructure which enables researchers to use nationwide distributed R&D resources such as, high performance computing, databases, scientific instruments, and also human resources in research areas as a linkage system in cyberspace by using high-end information technology.

Purposes

- **Innovation of R&D environment and international competitive power in Science**
  - Rapid changes in paradigm of S&T into the global collaboration by e-Science
  - To participate in the standardization of global e-Science environment near future

- **Infrastructure of e-Science helps the improvement of the various fields of R&D activities in Korea**
  - Effective investment of R&D activities by sharing the nationwide R&D resources
  - A balanced development of national infrastructure for the S&T by the advancement of the traditional R&D activities by using e-Science

- **To achieve successful e-Science project by the use of the highly developed information technolgies and network infrastructure in Korea and to do the good outreach to the benefits of common society**

Korea e-Science Project

Goals

- To advance research infrastructure and national competitiveness by effectively using distributed resources using information technology
    - Making the best use of Korea's IT infrastructure
- To providing advanced collaborative environments to researchers distributed over the country
    - Helping the decentralization

Korean Situation

- Construction of Grid infrastructure is being constructed through "National Grid Project" of Ministry of Information and Communication
    - Limited to the construction of next generation internet
    - Support for application projects are limited, only small scale projects are supported
    - National scale application projects are not possible
- Sharing and remote use of expensive equipments in government labs and academic institutions are very limited

Korea e-Science Project

Goals

Improving productivity of R&D by e-Science

| 2010 | To deploy the world-class e-Science infrastructure |

2007	"Anytime, anywhere" accessible e-Science infrastructure
2006	Full-scale developing common SWs for e-Science
2005	Preparation for building the e-Science infrastructure

Activities of e-Science

Finding e-Science models

Summary

- Goal : To develop and deploy common SWs for e-Science
- Periods : 1st Phase (KFY 2005 ～ 2007)
  : 2nd Phase (KFY 2008 ～ 2010)
- Budget : 100 M USD for developing common softwares
- Supervisor : MOST (Developer: KISTI)

Korea e-Science Project

e-Science Application Development in 2005

- Currently, we selected 5 areas of application for e-Science project.
- Each application will be developed based on common software and each group will contribute to the common software.

	Title	Organization	Area
1	Development of Molecular Simulation E-Science Research Environment and e-Glycoconjugates	Konkuk University	BT, NT
2	Development of e-Science Environment for HG2C (Human Gene to Chemical) based on Service-Oriented Architecture	Soongsil University	BT
3	Construction of Numerical Wind Tunnel on the e-Science Infrastructure	Seoul National University	Aerospace
4	Establishment of e-Science environment using the high voltage election microscope	KBSI(Korea Basics Science Institute)	Equipment control
5	Construction of e-Science Environment for Weather Information System	Pukyoung University	Meteorology

e-AIRS (Aerospace Integrated Research Systems)

* The main goal of e-AIRS is to establish the powerful and user friendly collaboration environment to aerospace researchers.

* Through the user friendly e-Science grid portal system, the e-AIRS provides remote CFD(Computational Fluid Dynamics) calculation and experiment management system .

* KISTI, SNU,
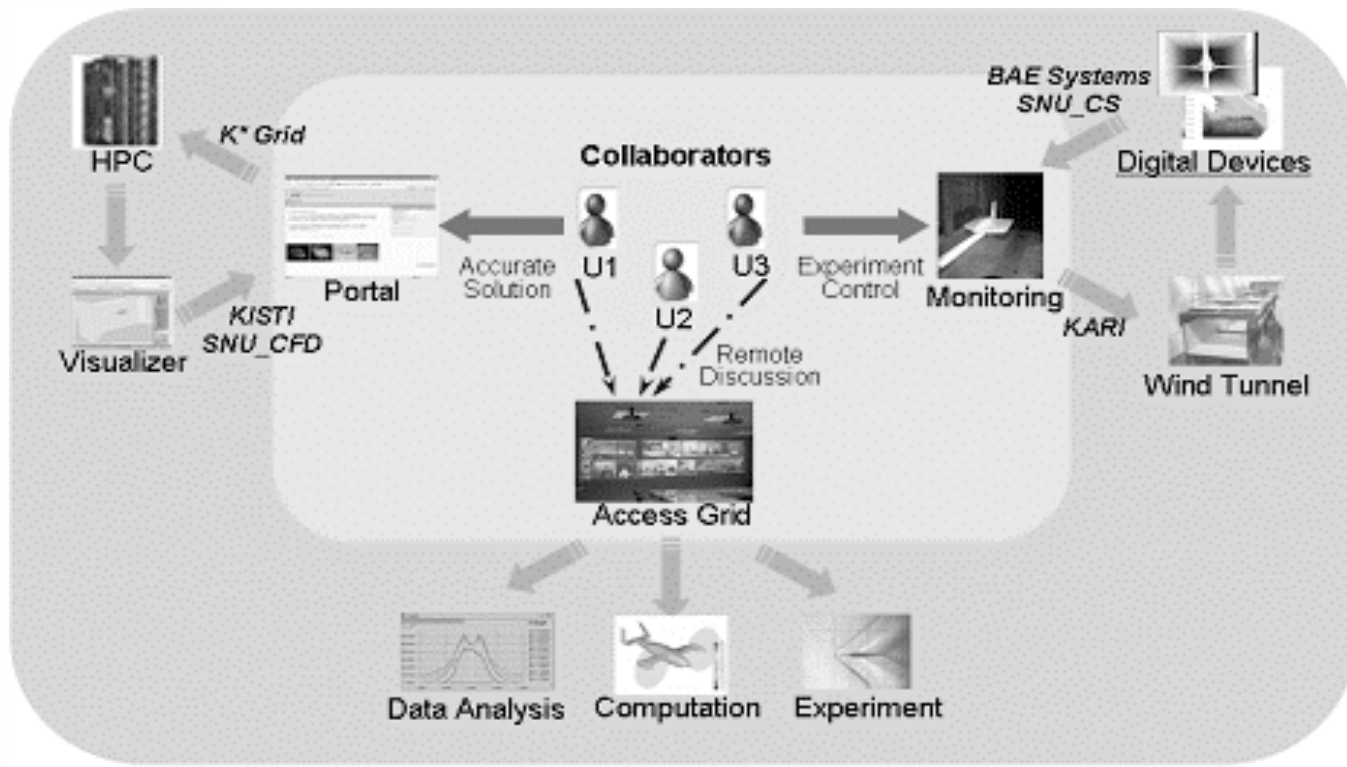
e-AIRS (Aerospace Integrated Research Systems)

CFD Framework

- Fortran / Cactus CFD solvers
- Multi-block & automatic parallel computation
- Accurate and efficient numerical schemes
- Euler / N-S / turbulence problems
- Design optimization

Remote Wind Tunnel Exp. Management

- Wind tunnel tests by accurate PIV(Particle Image Velocimetry) system
- Portal interface : remote request and monitoring service on the portal service

Collaboration System on the Access Grid

- Access Grid : Video and audio conference
- Data-sharing system between individual / group researchers

Portal Service

Integrated Research Environment

- CFD computation service by fortran / Cactus solver
- Portal interface : independent of time & place
- Remote management system of wind tunnel experiment
- CFD/experimental data-comparison system
- Collaboration environment using Access Grid

e-AIRS (Aerospace Integrated Research Systems)

❖ Portal Service

❖ CFD calculation service – mesh generation and job monitoring

❖ Remote experimental service – management of requests and monitoring service of the wind tunnel experiment

❖ Simulated result correctness verification system

e-AIRS (Aerospace Integrated Research Systems)

◈ Expected Results

  ◈ Because the expensive parallel computing resources and wind tunnel can be shared and any researcher

  ◈ can access these resources, the boundary of research topics will be enlarged.

  ◈ The R&D time and cost will be reduced and the efficiency will be increased.

  ◈ With the more powerful collaboration system, many individual and group researchers are able to

  ◈ establish large size projects and share their research products.

  ◈ Web portal service will provide the convenient research environment independent of time and space.

  ◈ Portal GUI will make it easy to access the research devices. This will make the non-experts produce thei
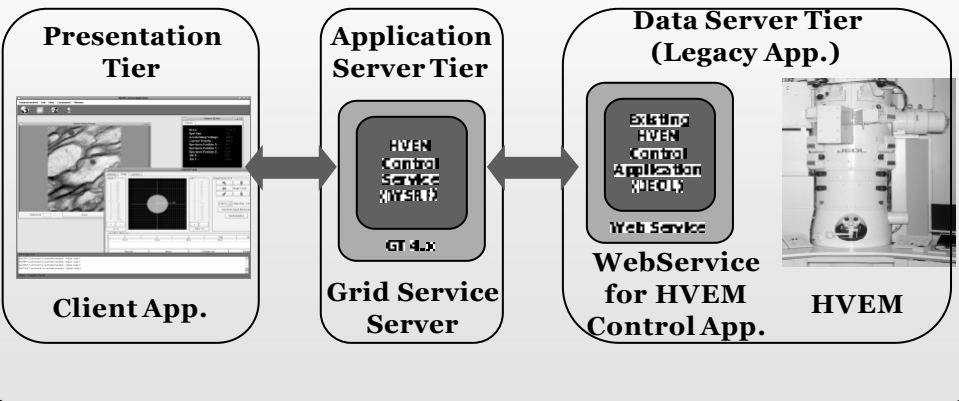
     own research data more conveniently.

High Voltage Electron Microscope (HVEM) Grid System

* To improve research performance using the global research network

* To share and use the costly High Voltage Electron Microscope (HVEM) efficiently

* To max the synergy effect achieved by databases of experimental data

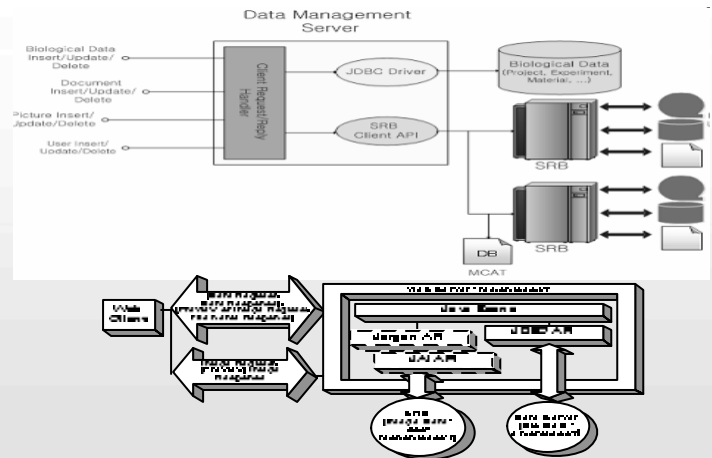HVEM Grid Infrastructure
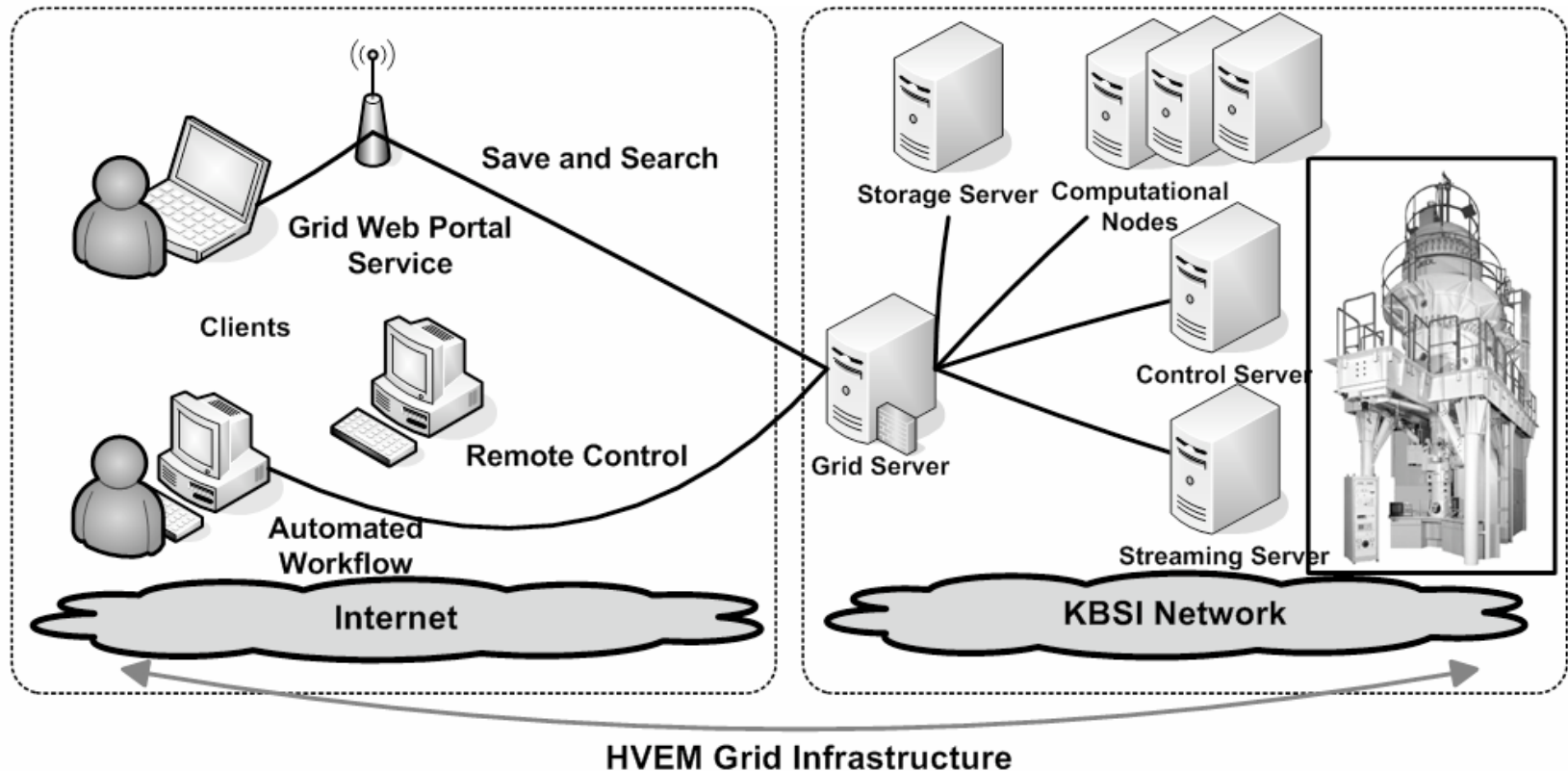
High Voltage Electron Microscope (HVEM) Grid System

● **Remote Control :** HVEM control service provides remote control of HVEM, goniometer, and CCD camera via encapsulated web service.

Presentation Tier

Client App.

Application Server Tier

HVEM Control Service XML(SRB) GT 4.x

Grid Service Server

Data Server Tier (Legacy App.)

Existing HVEM Control Application XDEO(X) Web Service

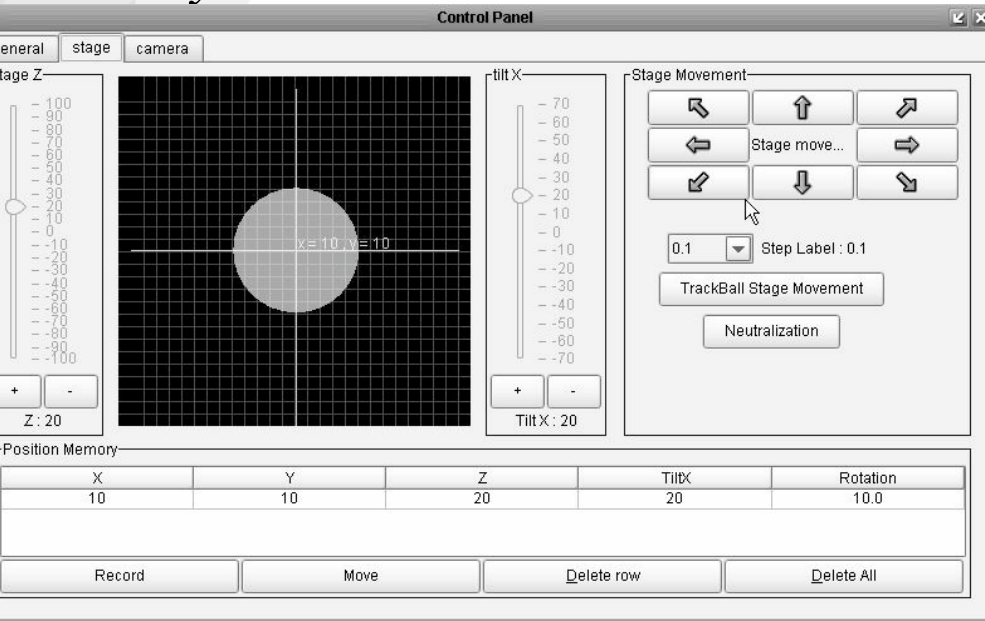WebService for HVEM Control App.

HVEM

❋ **Image Handling :** Streaming service provides real-time streaming from CCD camera and capturing service enables the images streamed to be captured and stored.
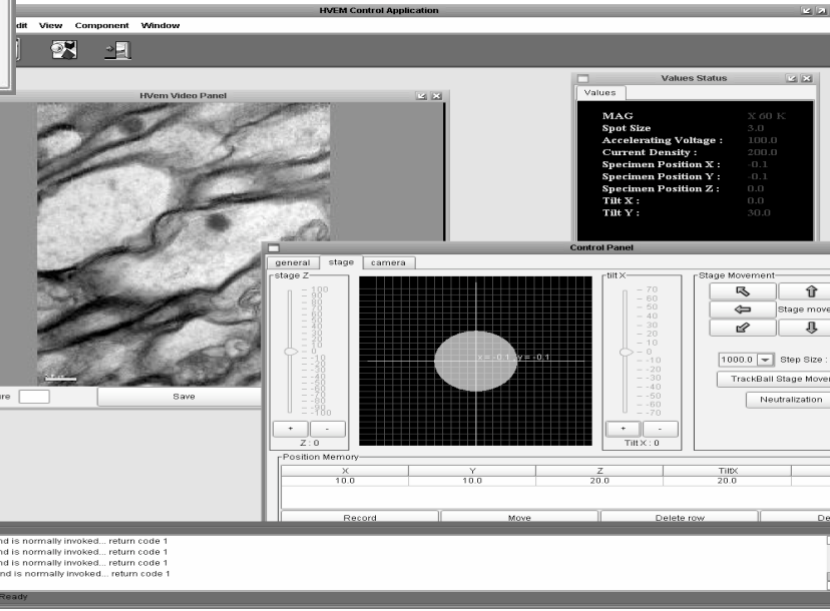
<Client Program>

Streaming

Data Turbine (RBNB)
Streaming Server

GT4
Capture Service
DM Library

Capturing

Digital Micrograph

❋ Data Grid: Data are stored in both a database and SRB, the former has bio-information and the latter has the 2-D images from CCD camera, the 3-D images processed from the 2-D images and related documents.

Data Management Server

Biological Data Insert/Update/Delete
Document Insert/Update/Delete
Picture Insert/Update/Delete
User Insert/Update/Delete

Client Request/Reply Handler

JDBC Driver

SRB Client API

Biological Data (Project, Experiment, Material, ...)

SRB

SRB

DB

MCAT

High Voltage Electron Microscope (HVEM) Grid System
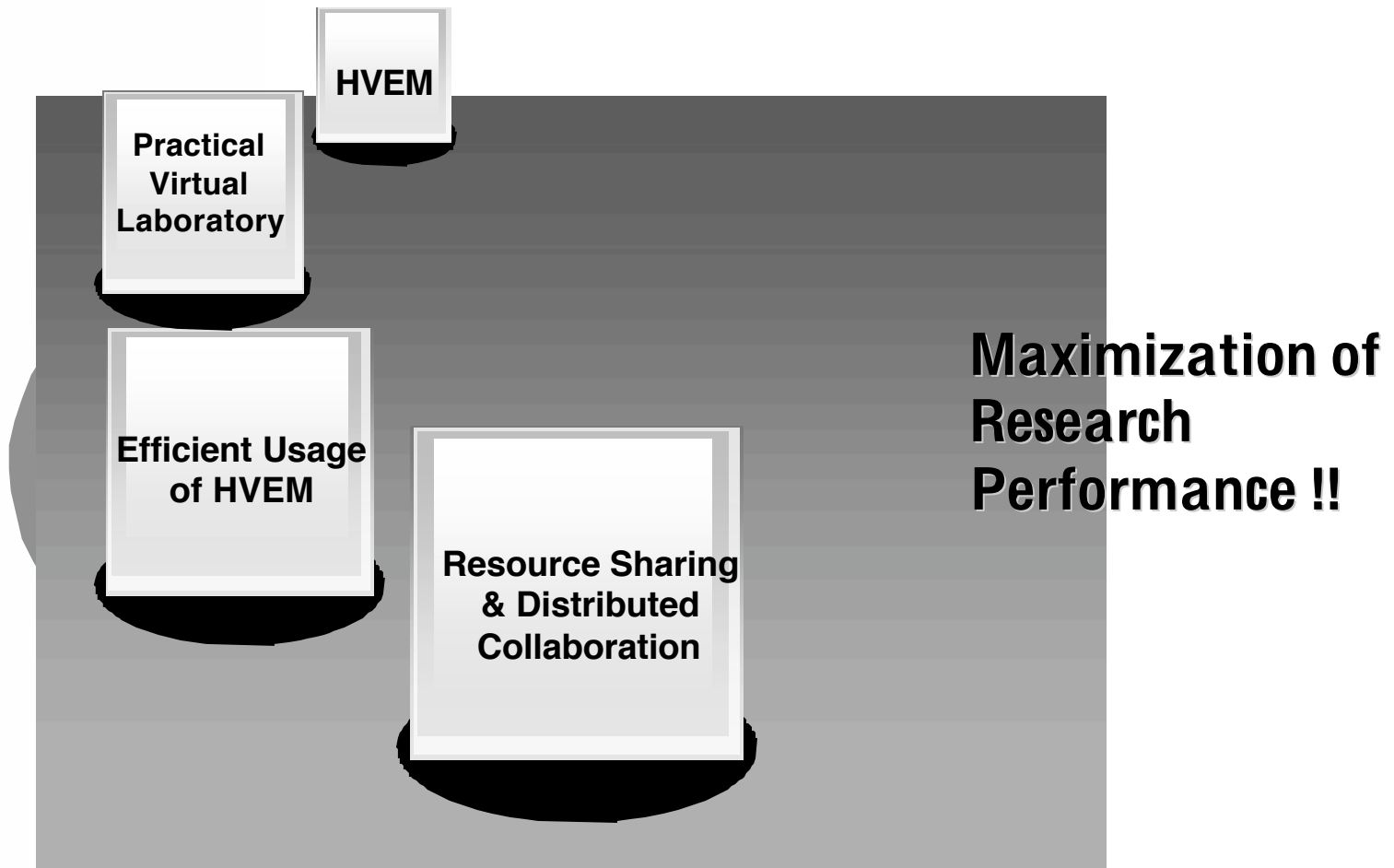
HVEM Remote Access System GUI

HVEM Goniometer Remote Control GUI

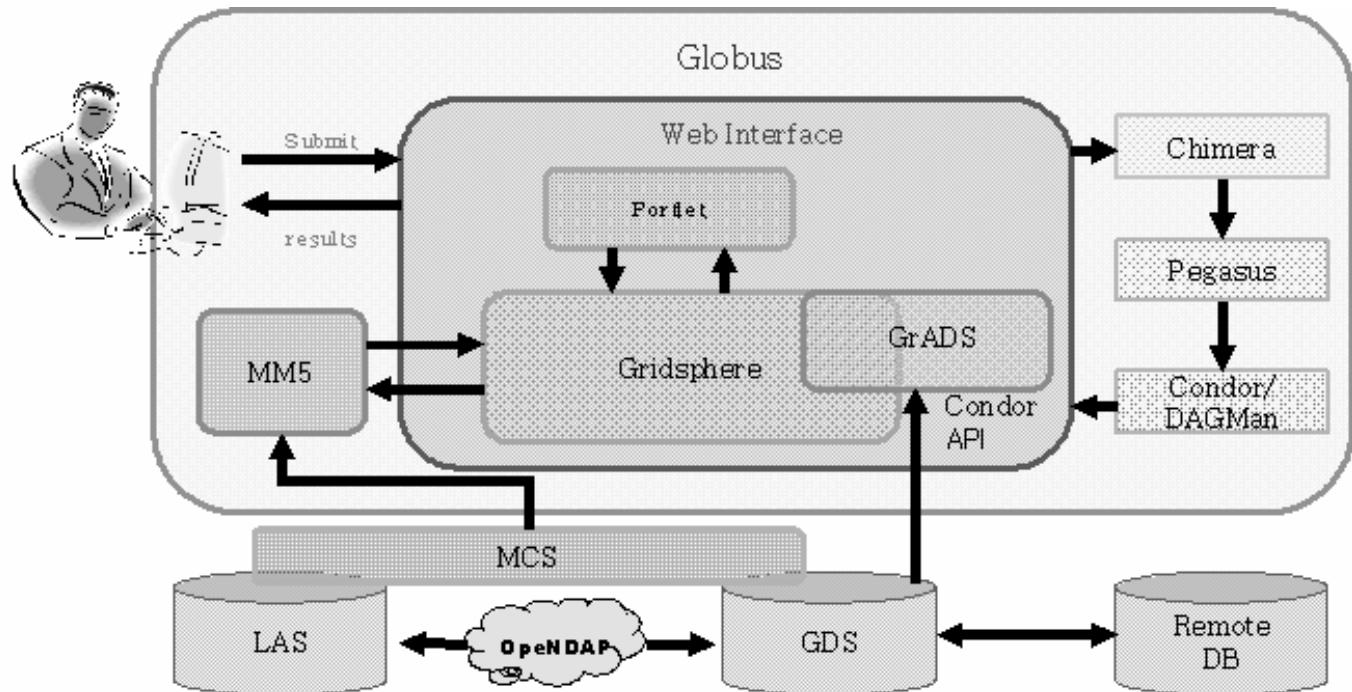# *High Voltage Electron Microscope (HVEM) Grid System*

◈ Expected Results

HVEM

Practical
Virtual
Laboratory

Efficient Usage
of HVEM

Resource Sharing
& Distributed
Collaboration

**Maximization of
Research
Performance !!**

Meteorology e-Science

◈ Efficient and Standardized Data Access

◈ Intelligent Data Management of SAM file Pool

◈ Dynamic Cataloging

Meteorology e-Science

- **Data Management Service**
  - Compliant with Data Grid
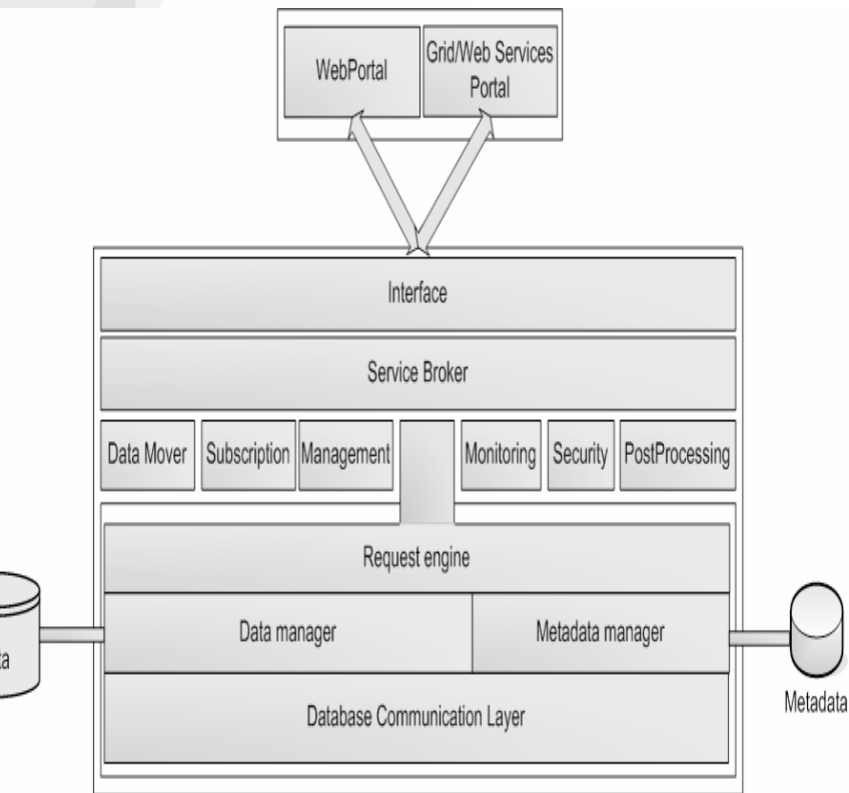 - On-Demand Cataloging on GDS (Grads Data Server)

- **MM5 Workflow**

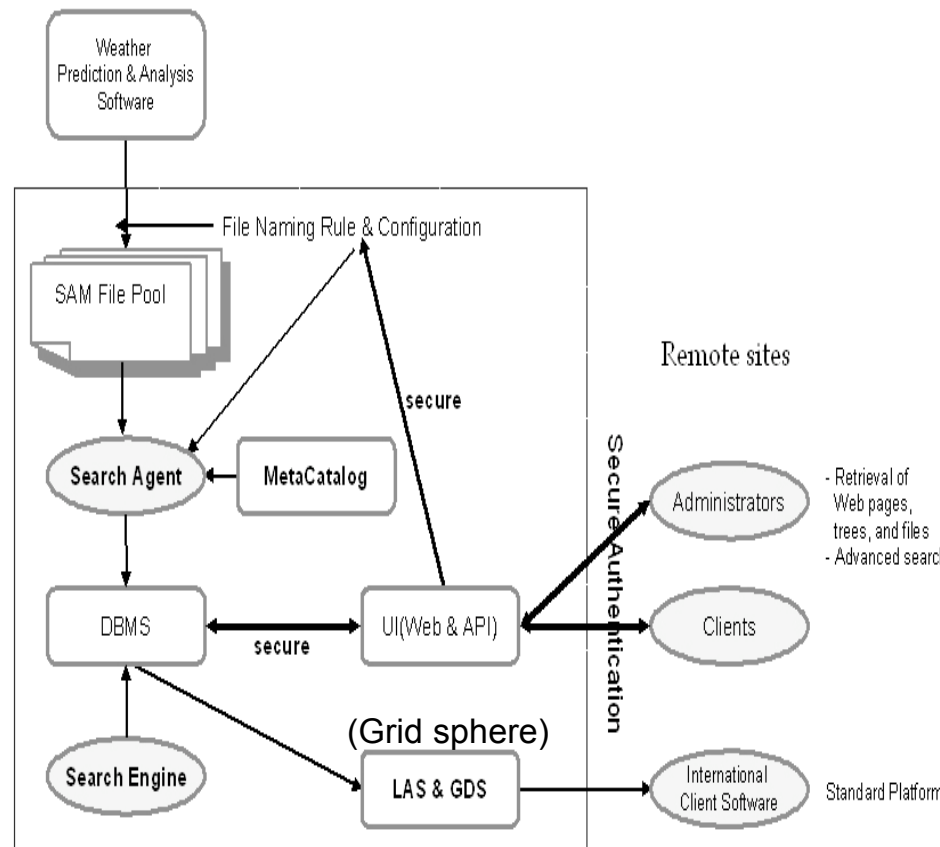- **Output Display Service**
  - GrADS Plug-in on Grid Sphere

- **Resource Management Service**
  - Computing Resource　monitoring & discovery
- **Peta-Bytes Data Pool Handling**
  - Meaningful Data Representation from huge size of files

Meteorology e-Science

System Architecture
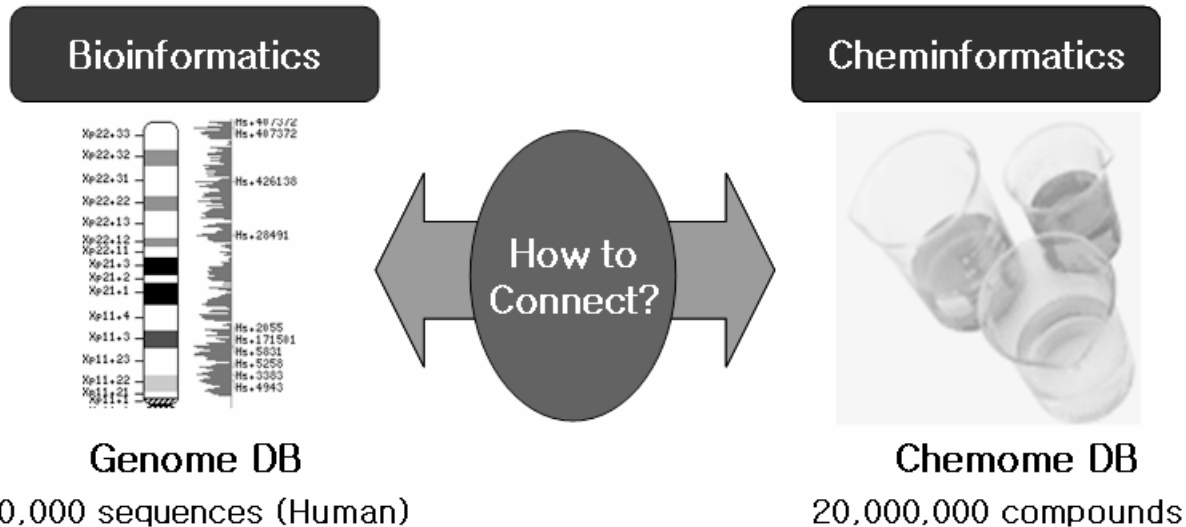

Data Pooling Flow

Meteorology e-Science

◈ Expected Results

  ◈ Efficient Data Management (File & DB combination)

  ◈ Efficiency of Storage using Dynamic Cataloging

  ◈ Easy Access & Quick Searching to Peta-Bytes

  ◈ Weather Data

  ◈ Interoperability between LAS(Live Access Server) and

  ◈ GDS (Grads Data Server)

  ◈ Changing to Semantic Data Management

HG2C Project
(Human Genome to Chemicals for Drug Discovery)

❑ Construct an HG2C Portal that integrates various BT applications
for efficient drug discovery

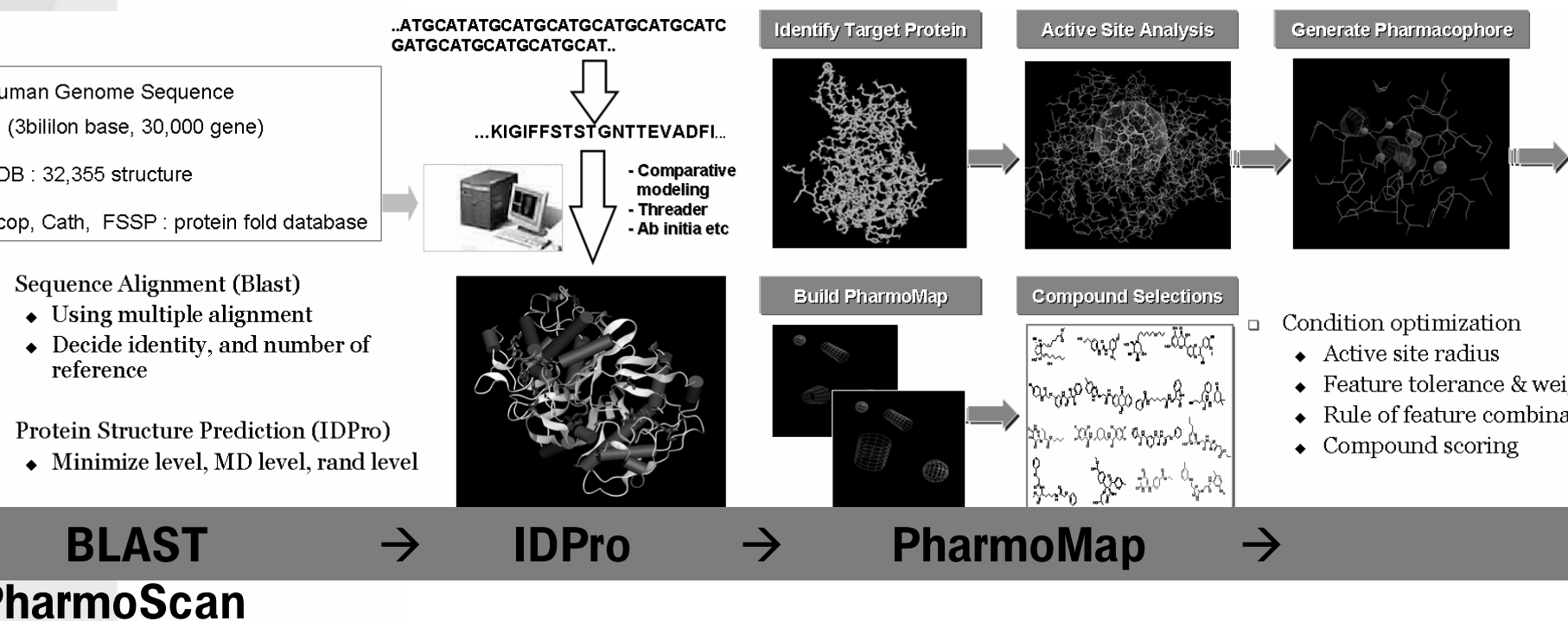| Bioinformatics		Cheminformatics
Genome DB	How to Connect?	Chemome DB
30,000 sequences (Human)		20,000,000 compounds

❑ BT Application S/W
- ◆ IDPro
  - ▪ Predict protein structure
- ◆ IDPharmo (PharmoMap + PharmoScan)
  - ▪ PharmoMap: Define active sites and generate multiple 3d feature maps
  - ▪ PharmoScan: Virutal screening

HG2C Project
(Human Genome to Chemicals for Drug Discovery)

◈ HG2C Workflow

..ATGCATATGCATGCATGCATGCATGCATC
GATGCATGCATGCATGCAT..

...KIGIFFSTSTGNTTEVADFI...

- Comparative
  modeling
- Threader
- Ab initia etc

| Identify Target Protein | Active Site Analysis | Generate Pharmacophore |

uman Genome Sequence

(3bililon base, 30,000 gene)

DB : 32,355 structure

cop, Cath, FSSP : protein fold database

Sequence Alignment (Blast)
- Using multiple alignment
- Decide identity, and number of reference

Protein Structure Prediction (IDPro)
- Minimize level, MD level, rand level

| Build PharmoMap | Compound Selections |

❑ Condition optimization
- Active site radius
- Feature tolerance & wei
- Rule of feature combina
- Compound scoring

BLAST → **IDPro** → **PharmoMap** →

PharmoScan

# HG2C Project
(Human Genome to Chemicals for Drug Discovery)

◈ Meta Services

Define a part of a workflow as a new service
- Workflow instance is declared as a workflow unit in the service description
- By overriding some attributes of a workflow unit,
  - Pass parameters of a service to the workflow unit
  - Setup service specific information
- The new service can be wrapped to a Web service or a Grid service, and reuse it easily

Specify service specific information
- Restrict resources to allocate a specific service (user preference)
- Scheduling priority
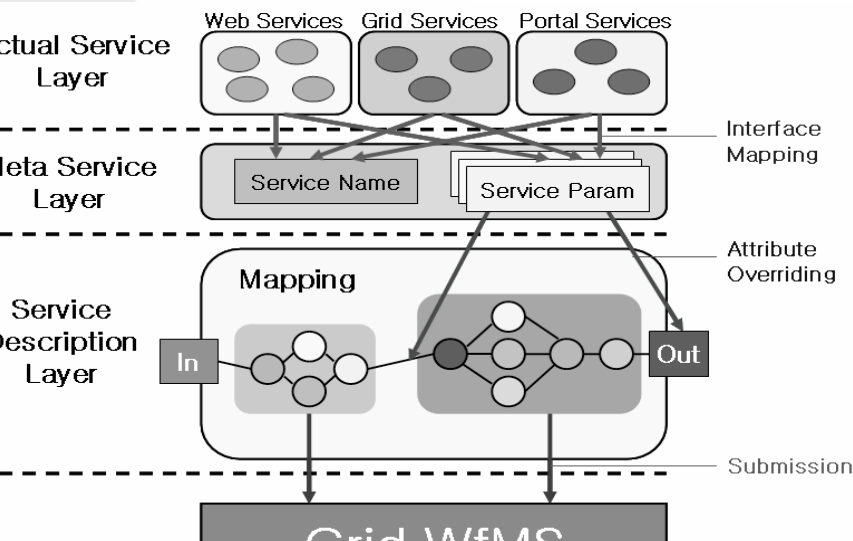
◈ MSF (Meta Services Framwork)

- Meta services concept is combined with WfMS system
  - Provide more reusable and adaptable workflow management environment
  - Adapt Meta services to various service environments such as Web services, portal services, and Grid services

- Workflow description is divided into MSF_Service, MSF_Flow, and MSF_Task

- Consist of five small agents
  - AM, RM, EM, OM, and SM
  - Increase flexibility and adaptability

- Ontology concept is applied to describe services, flows, and tasks
  - Advanced searching and sharing of the description is possible
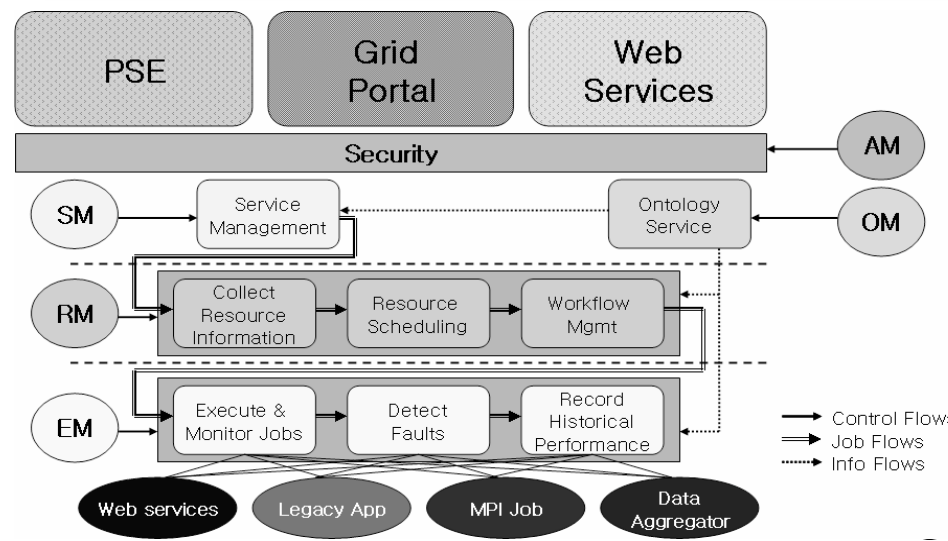
HG2C Project
(Human Genome to Chemicals for Drug Discovery)

◈ MAGE (Modular and Adaptive Grid Environment)

❑ Modular and Adaptive Grid Environment
  - ◆ Provide API for easy development of Grid application
  - ◆ Provide transparency to end-users and developers
    - ▪ Protocol transparency
    - ▪ Running location transparency
    - ▪ Implementation details transparency
 - ◆ Provide flexible and reliable layers

❑ Consist of 3 major components
  - ◆ Serivce Manager
  - ◆ Component Manager
  - ◆ Reconfiguration Manager
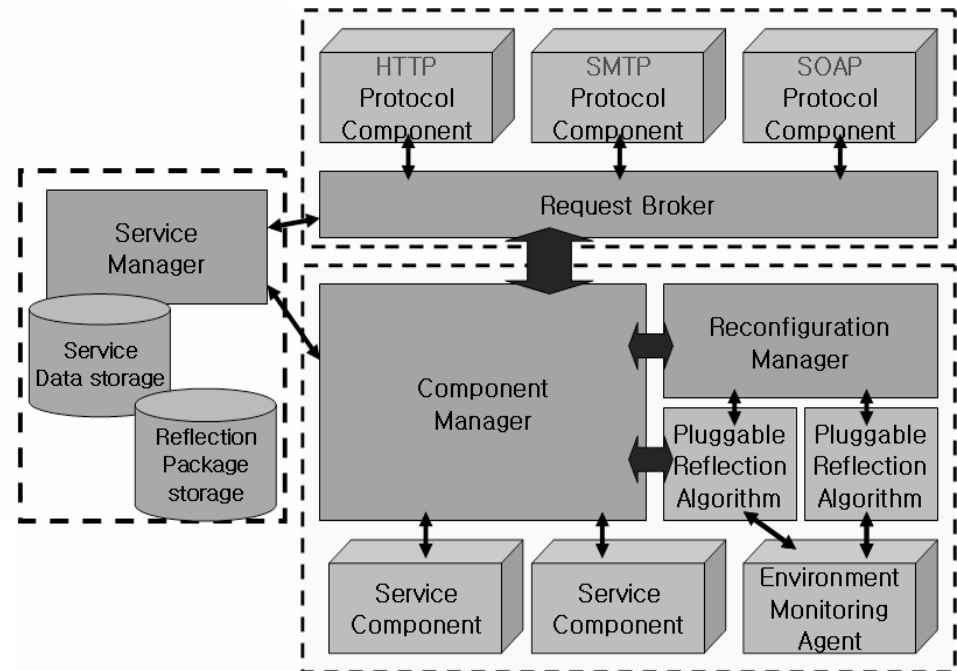
e-Glycoconjugates: Integrated Grid Portal for the Molecular Simulations of Glycoconjugates

* The e-Science Grid portal system of e-Glycoconjugates was motivated to study whole range of glycoconjugates with web-interface.

* It employs an innovative architecture to execute the molecular simulation, to analyze the simulation trajectory and to share the simulation data with all other world users
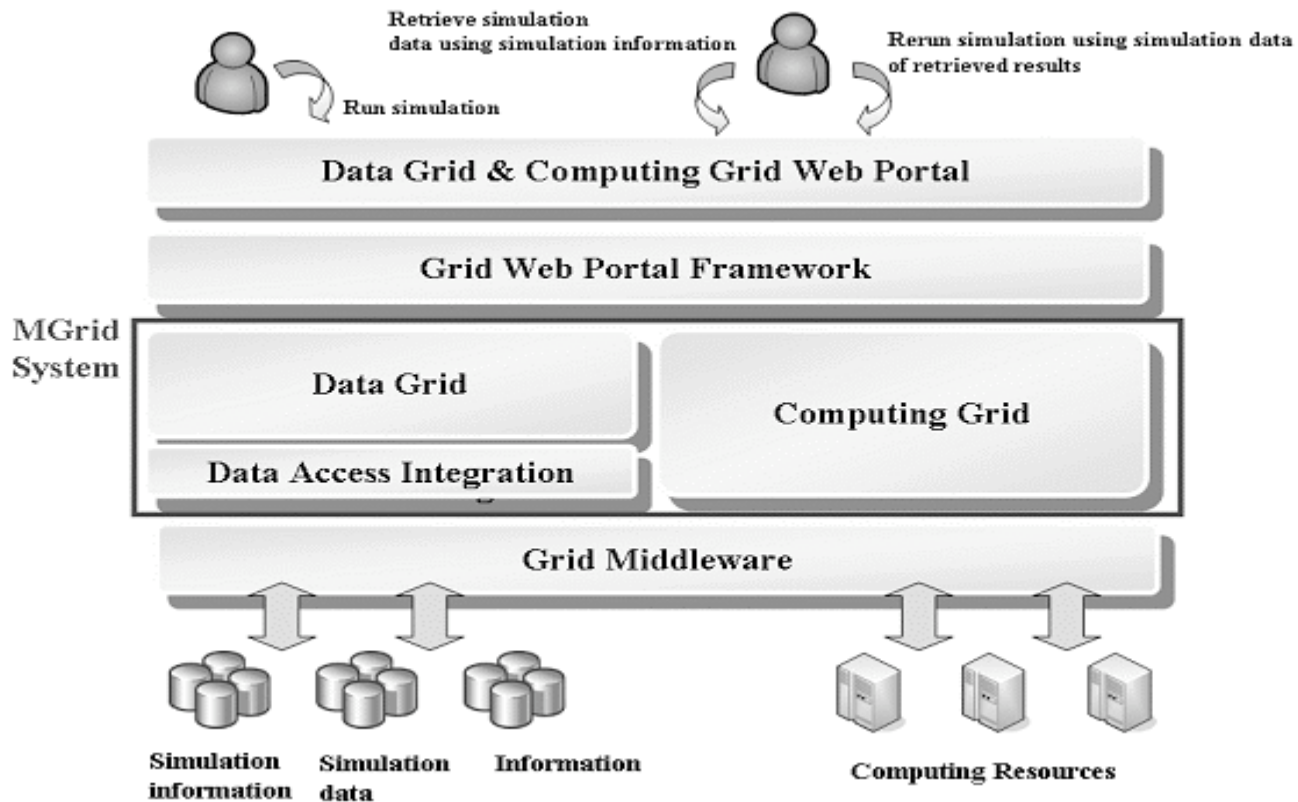
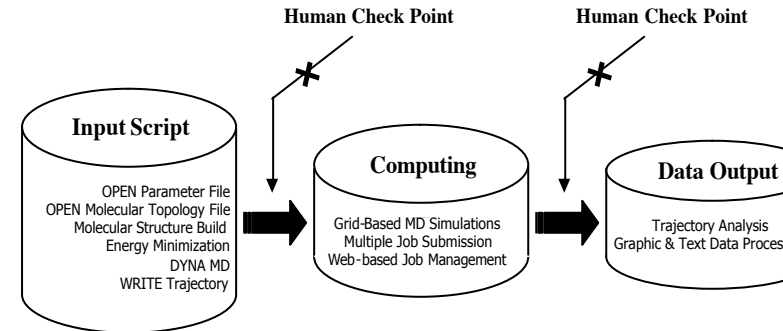# *e-Glycoconjugates: Integrated Grid Portal for the Molecular Simulations of Glycoconjugates*

Process WorkFlow

Our problem solving process can be classified into three steps as follows.

- 1. Input script programming & Job submit
- 2. Multiple job simulations on the Grid computing system
- 3. Trajectory analysis & Text data processing

Trajectory Processing

The analysis facility of e-Glycoconjugates will support the automatic trajectory processing to obtain valuable information on the hydrogen bonding, hydration number, translational diffusion, rotational motion and radial distribution of water molecules around carbohydrates or glycoconjugate molecules.

Process workflow. Human check points were minimized to run a automatic job simulation
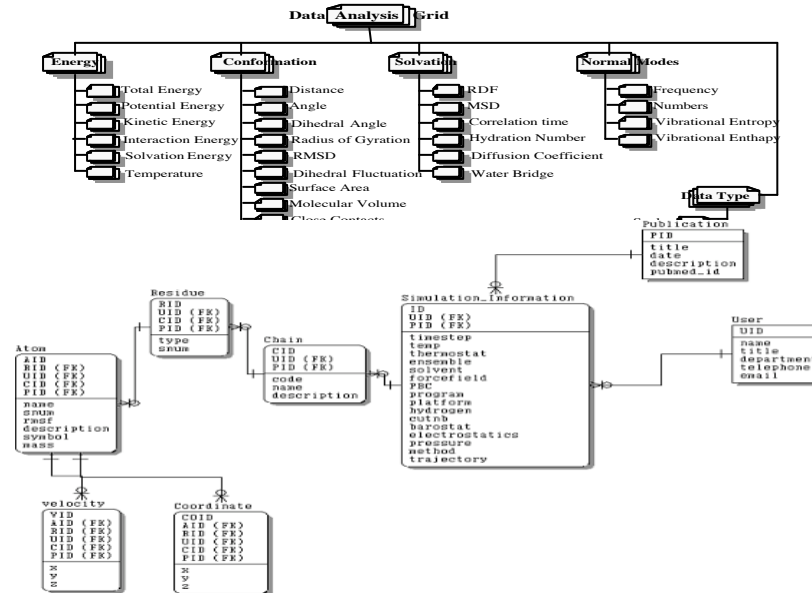
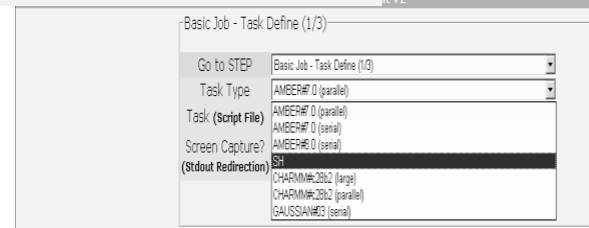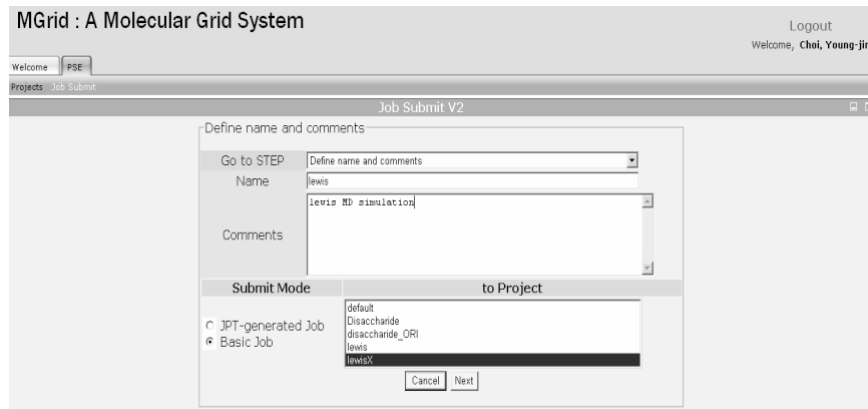Trajectory output and E,R-diagram for DB construction

e-Glycoconjugates: *Integrated Grid Portal for the Molecular Simulations of Glycoconjugates*

- A web portal for e-Glycoconjugates has a graphic user interface (GUI)

  - create a job,

  - submit a job to the computing grid,

  - save and retrieve data to the database on the data grid

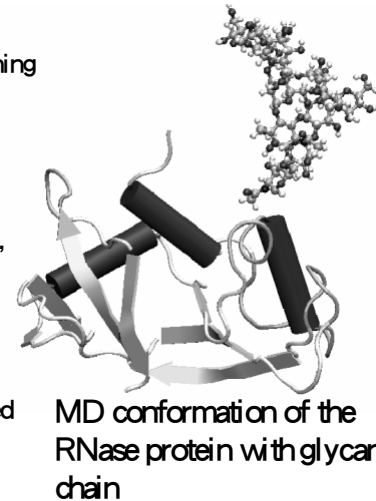- Supports various search method on the molecular simulations.

  - queries by keyword, structure, classification or by combining or all these factors.

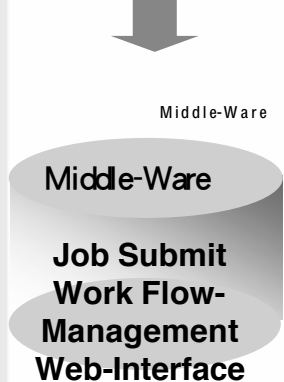 - Full supports for the name-searching and automatic analyses are now in progress.

e-Glycoconjugates: Integrated Grid Portal for the Molecular Simulations of Glycoconjugates

Bio-Information Production

❖ We have been carrying out conformational mapping of more than 1,000 carbohydrate chains to serve dynamic structure DB containing wide range of glycans of which three-dimensional structures are known or unknown.

❖ The effects of glycan chain structure on the protein folding and stability are also going in a progress.

❖ The e-Glycoconjugates performs the molecular simulations on the more than 100 biologically important glycoprotein such as prion, collagen, antifreeze peptide or ribonuclease.

❖ Understanding for the structural dynamics of glycoconjugates is a prerequisite for the opening in personalized medicine. We are convinced that valuable systematic information from e-Glycoconjugates is able to offer theoretical background to the structure-based design of novel glycoconjugates.

MD conformation of the RNase protein with glycan chain

Biosimulation Job   -   Grid Computing Engines -   Information

Middle-Ware

Middle-Ware

Job Submit
Work Flow-
Management
Web-Interface

Molecular Simulations & Data Storage, Analysis

Computing Grid

Data Grid

Analysis Grid
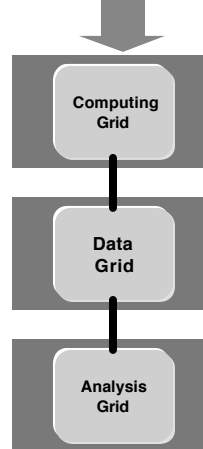
Analyzed Database Sharing Mining
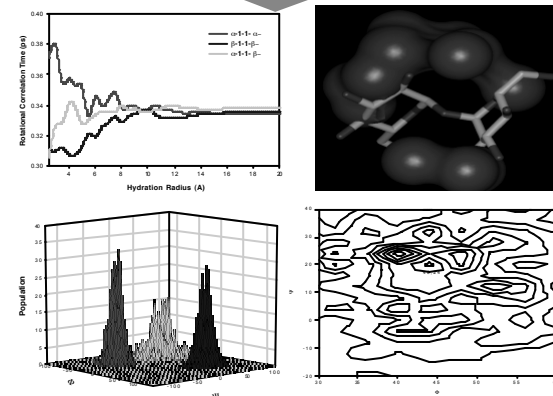
Korea e-Science Project

Domestic Research Network

- Area : High-speed Research Network (KREONET) Giga backbone deployment and services

- Characteristic: Seoul-Daejeon(10Gbps), 12 local network centers (5 and 2.5Gbps)

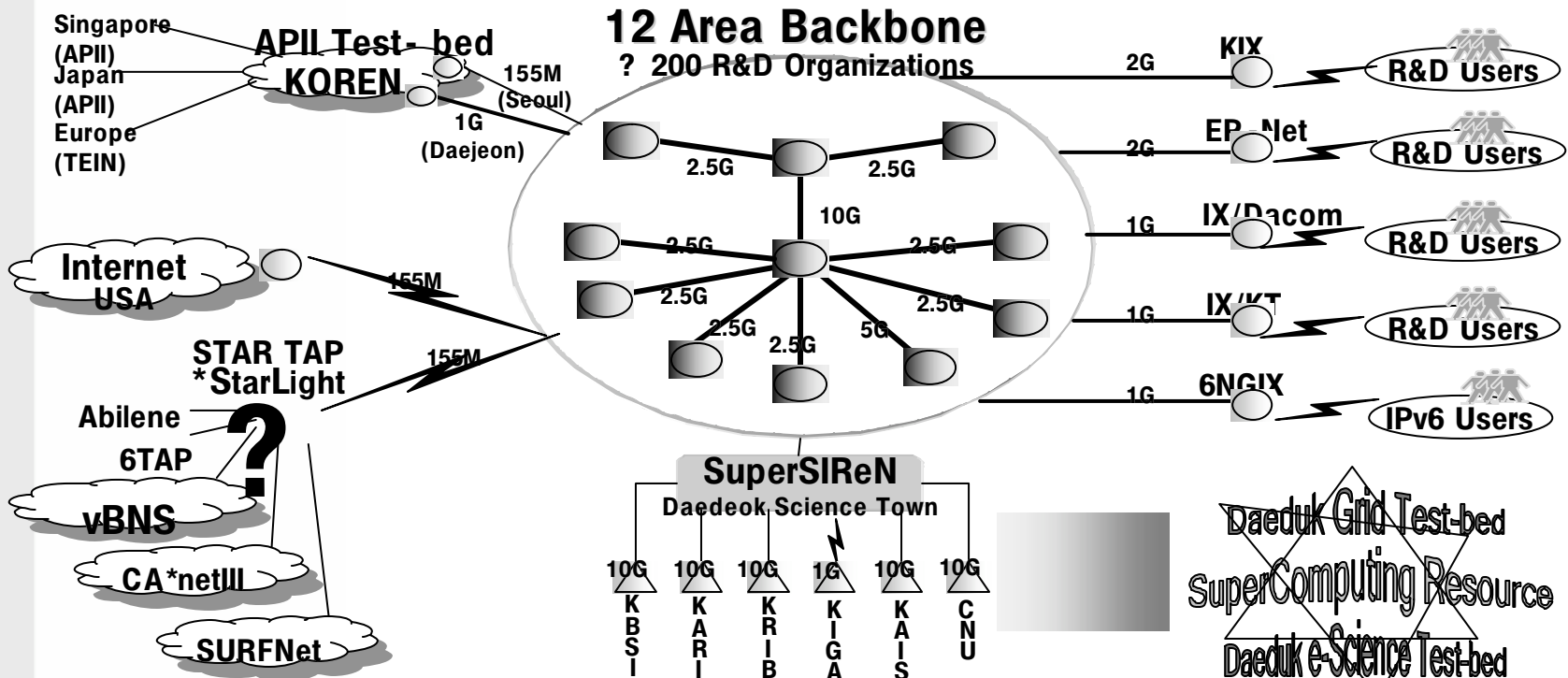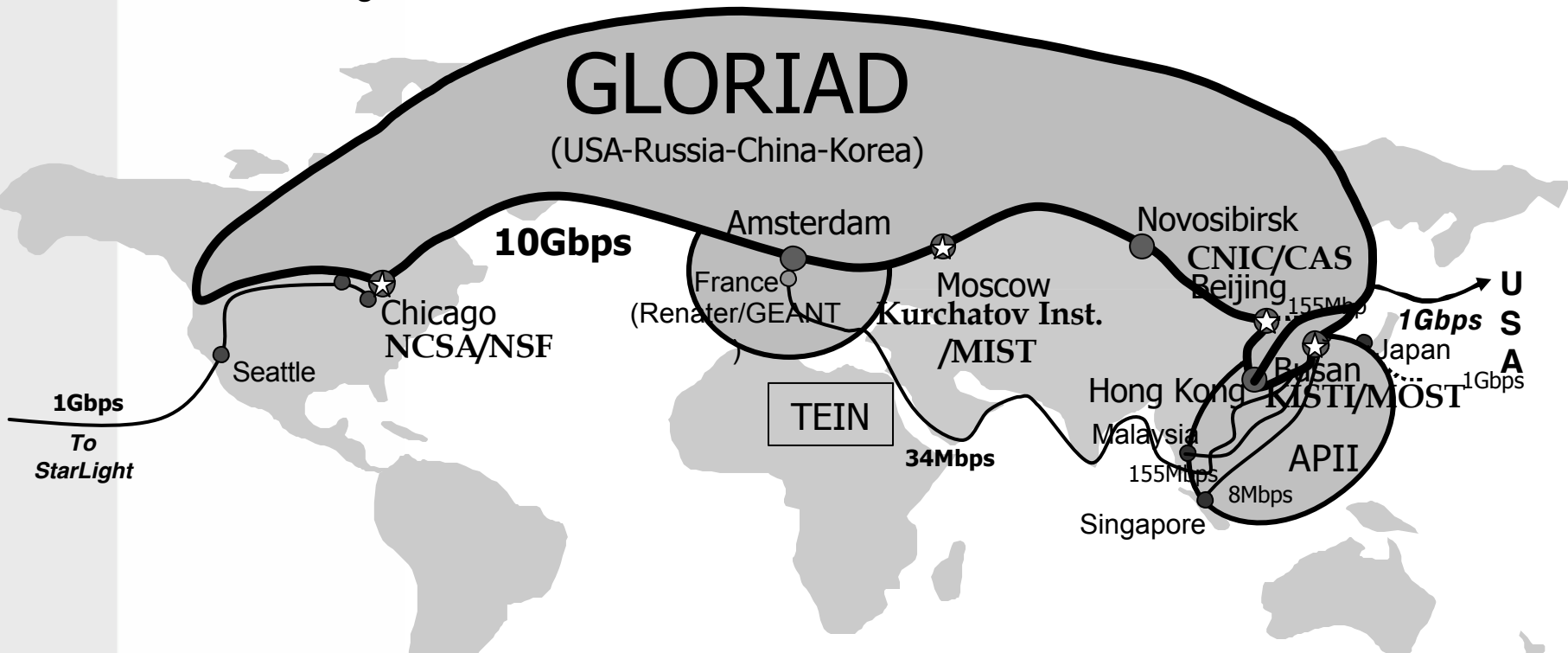- Integrated monitoring system service (24 hours), Apply next-generation technology (IPv6, QoS, Multicast)

- Results: High-end application (e-Science, Grid) needs based Infrastructure

12 Area Backbone
? 200 R&D Organizations

Singapore (APII)
Japan (APII)
Europe (TEIN)

APII Test‑bed
KOREN

155M (Seoul)
1G (Daejeon)

2.5G　2.5G
10G
2.5G　2.5G
2.5G
2.5G　2.5G　5G

Internet USA
155M
STAR TAP *StarLight
155M
Abilene
6TAP
?
vBNS
CA*netIII
SURFNet

2G　KIX　R&D Users
2G　EP-Net　R&D Users
1G　IX/Dacom　R&D Users
1G　IX/KT　R&D Users
1G　6NGIX　IPv6 Users

SuperSIReN
Daedeok Science Town

10G　10G　10G　1G　10G　10G
K B S I　K A R I　K R I B　K I G A　K A I S　C N U

Daeduk Grid Test-bed
SuperComputing Resource
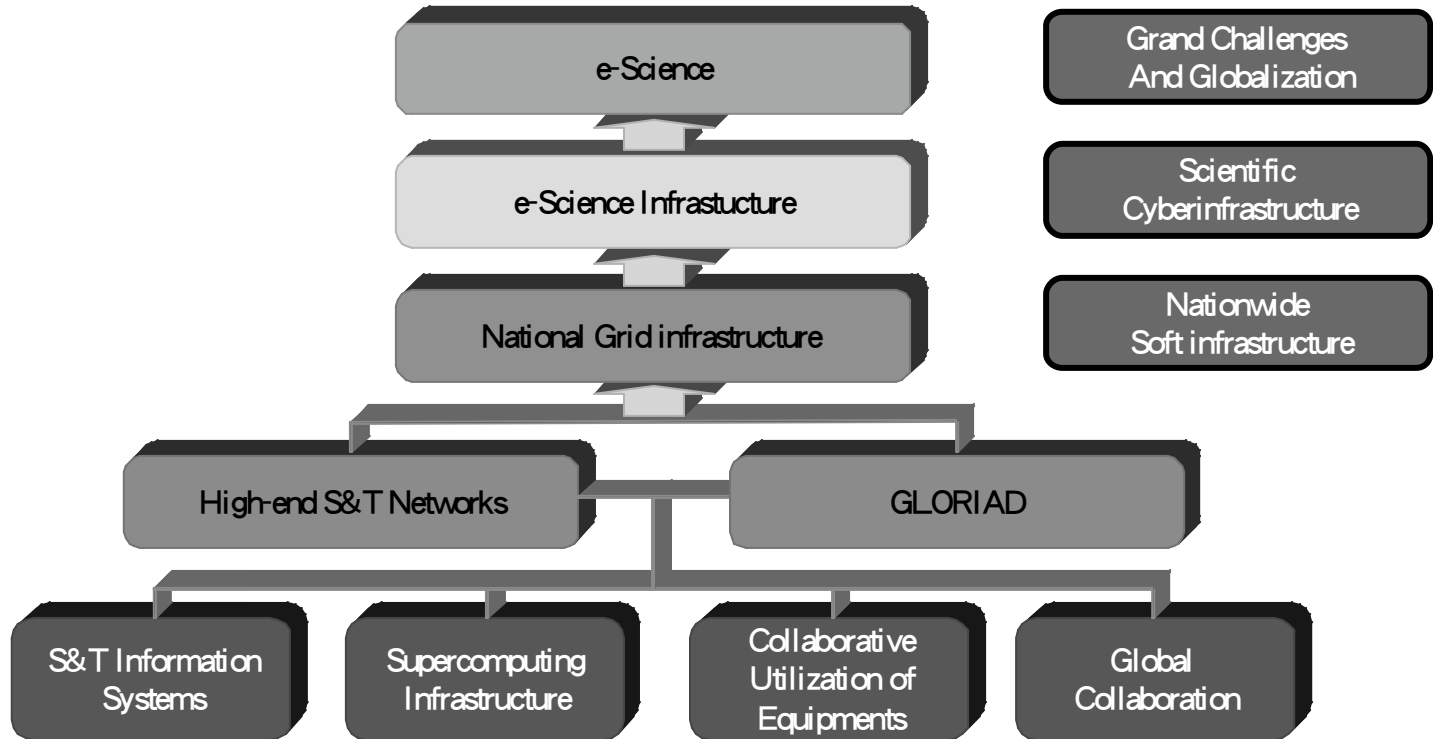Daeduk e-Science Test-bed

Korea e-Science Project

Global Ring Network for Advanced Application Development

- Definition : The world's first global science & technology network connecting the contin
  with 10Gbps ring-type lambda network (Funded by Korea, US, Russia, and China)

- Application : Serve as international collaborative network for high-end science and
  technology areas such as high-energy physics, bioscience and nuclear fusion that requ
  real-time large scale data transfers

GLORIAD
(USA-Russia-China-Korea)

10Gbps

Amsterdam
France
(Renater/GEANT)

Novosibirsk
CNIC/CAS
Beijing

Moscow
Kurchatov Inst.
/MIST

Chicago
NCSA/NSF

Seattle

1Gbps
To
StarLight

TEIN

34Mbps

Hong Kong

Malaysia

155Mbps

Singapore

8Mbps

Busan
KISTI/MOST

APII

Japan

USA

1Gbps

1Gbps

155Mbp

Korea e-Science Project

Cyber R&D infrastrure

- e-Science
- e-Science Infrastucture
- National Grid infrastructure
  - High-end S&T Networks
  - GLORIAD
    - S&T Information Systems
    - Supercomputing Infrastructure
    - Collaborative Utilization of Equipments
    - Global Collaboration

- Grand Challenges And Globalization
- Scientific Cyberinfrastructure
- Nationwide Soft infrastructure