

A Generalised Application Deployment Framework for An Abstracted, Service-oriented Scientific Grid focusing on Life Science Applications

Assoc. Prof Tan Tin Wee

National University of Singapore

Ong Guan Sin

INET-One. Pte Ltd

In collaboration with KOOPrime Pte Ltd

In a nutshell

A general deployment framework which provides:

- ease of deployment and management of applications and software packages
- applications abstraction of grid resources
- transparent existence of grid services
- seamless integration of grid-enabled applications into work processes

Goal

- To create an impact on acceptance and endorsement of grid technology and service as a way of life for general scientific workers to develop and use grid-based applications with ease
- To provide valuable experience and feedback to the general grid computing technology researchers

Aspects of Research/Challenge

1. Grid implementation and middleware deployment strategy with life science applications
2. User perception and expectation of grid as a service or utility with high level abstraction of applications
3. Depth and breadth of service abstraction in relation to Life Sciences applications
4. Constructive and conflicting aspects of service abstraction vs. policy control and service discovery vs. registry control

The Framework

1. A grid architecture for Life Sciences applications in a seamless worldwide workflow framework
2. A deployment, distribution, management system of the necessary software components
3. A harmonized integration standard of various software layers and services
4. A consistent, intuitive interface to the service/application abstraction
5. A powerful, flexible policy definition, control and negotiation mechanism for a collaborative but autonomous grid environment

Current Research Projects pave the groundwork for this framework to be built

1. APBioBox Project funded by International Development Research Centre (IDRC) for integrating Globus Grid software with Bioinformatics Applications (2002-2003)
2. APBioNet-Asia Pacific Science and Technology Centre joint project on SunBioBox (2003)

Provides Rapidly Deployable Package of Frequently Updated SuperPackage of Packages of Biocomputational Software which can be deployed in this framework

Current Research Projects pave the groundwork for this framework to be built

3. APBioNet-APAN BioMirror project with Indiana Univ, AcSys ANU, Peking U, AFFRC Japan, KAIST, KU Thailand, UPM, NYMU Taiwan, NUS (1998 to date)
4. Biochemistry Dept/BIC Official mirror site of Protein DataBank with NIST, Rutgers etc (1999-date)
5. Biochemistry Dept, NUS collaborative project on life science omnidatabase integration of about 10 separate database projects – Project MOAD (2003) over a mysql integrated platform

Provides the biological and bioinformatics, genomics, proteomic datasets of >200Gbytes which is doubling every 12 to 18months

Current Research Projects pave the groundwork for this framework to be built

6. APBioGrid and KOOPrime collaboration on BioWorldWide Workflow at NUS (previous collaborators include Cray and Lion Biosciences) (2001-2003)

Provides the basis for developing a higher level abstraction of Grid Resources, Grid Datasources, Grid BioComputational resources and BioInstrument Interfacing.

Current Research Projects (cont.)

7. S* Singapore-Stanford-SouthAfrica-UCSanDeigo-Sydney-Sweden Alliance Access Grid initiative for online education in life science informatics training and collaboration spanning four continents (2000-date)

Provides the freely accessible and regularly organised 10 week online bioinformatics training course to conduct online tutorials to get users worldwide for trial-usage of this biogrid framework

Current Research Projects (cont.)

8. FastDNAML global phylogenetics joint project led by Indiana University featured at SC2003 "Global analysis of arthropod evolution" using computers in a grid spanning six continents, and at SC'97 and SC'98 on mitochondrial COP protein.

Channel the experience of several SC demos and the network of contacts for the global trial deployment of this biogrid framework

Current Research Projects (cont.)

9. Biochemistry Dept, NUS bioinformatics

research projects on:

- MGALIGN, rapid EST-genomic sequence alignment application.
- Large scale analysis of Intron-Exon-Structure
- siRNA prediction software application
- Software on automated NMR peak assignment
- Peptide bioinformatics databases on Small Disulphide proteins, Antimicrobial peptides, Scorpion toxins and predictive software

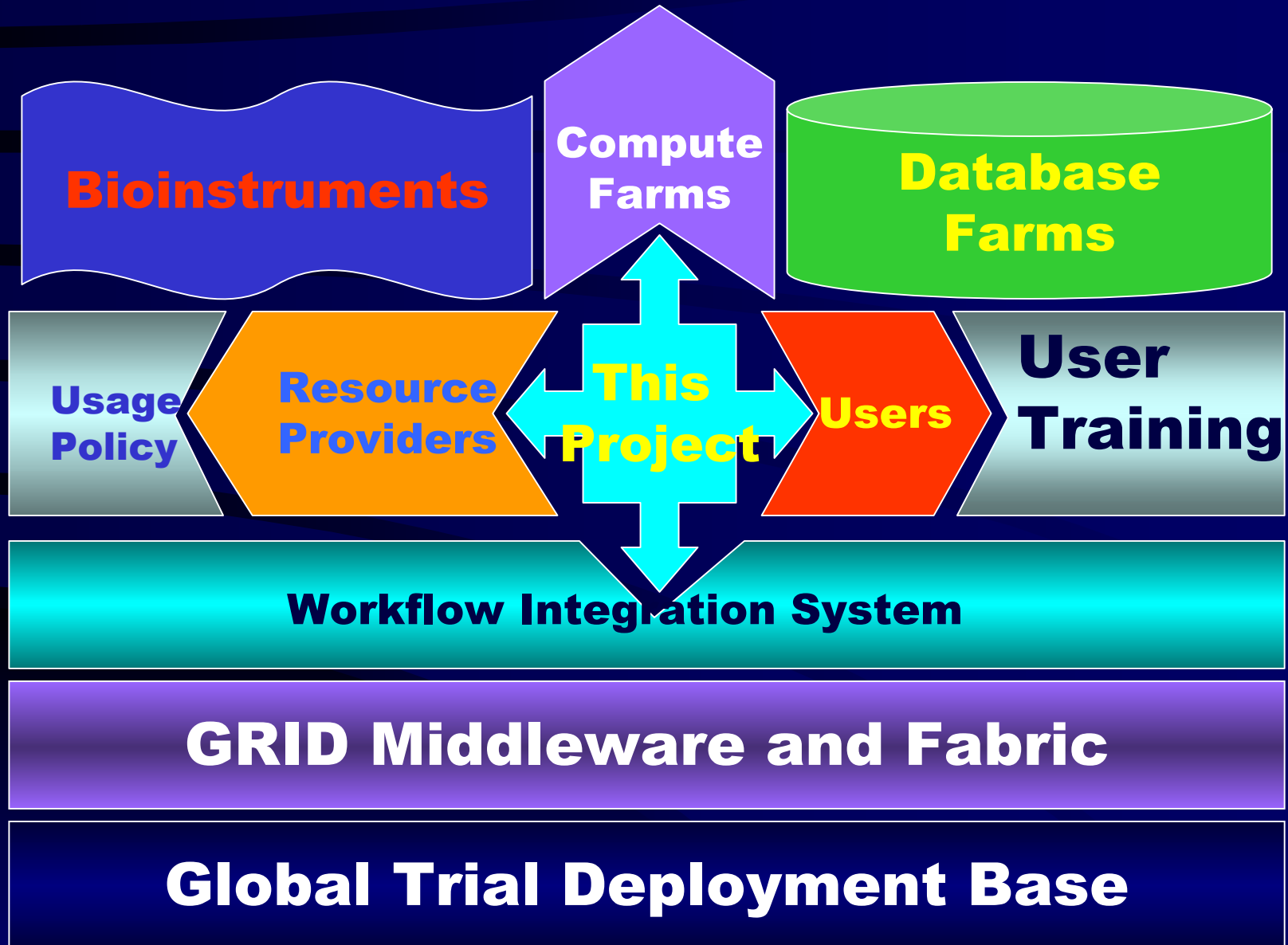
Provide continual supply of new Datasets and Computational tools to attract new and retain old users and sustain a strong user base

Prospective Research Projects

10. Medical Faculty, NUS new bioinformatics research projects with University of California San Diego UCSD and SDSC on their Joint Centre on Structural Genomics (currently in negotiation) and previous collaboration with Stanford Synchrotron Radiation Lab for realtime data gathering from Xray crystallographic beamline (1998-2000)

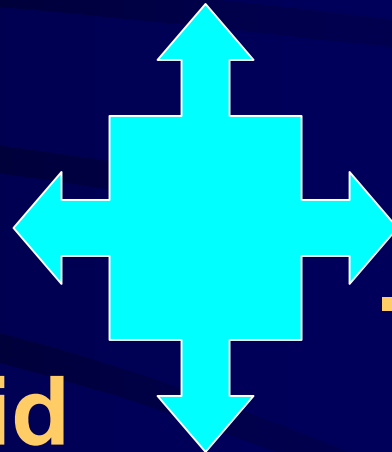
Provide interface to high throughput 3D structural determination of protein folds using synchrotron light source

Adding the central pieces to the framework



This Project

**A Generalised
Deployment
for An Abs-
Service-
Scientific Grid**



**Application
Framework
-traced,
- Oriented
focusing on
Life Science Applications**

Conclusion

- The plethora of applications in life sciences which are ready for integration into the Grid Framework positions this project very well for success.
- The framework allows a comprehensive platform for life science users to deploy their projects on the grid without worrying about the middleware and fabric layers.
- The wide range of project partners who are willing to be involved include academic, industry software vendors, industry hardware vendors and international parties in USA and beyond.