# Grid Technology

MSI (CI)^2 Meeting, Jan 30-31, 2006

Karan Bhatia, PhD, Grid Middleware Group Leader

# Agenda

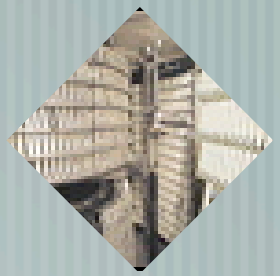
- What is Grid Computing?
- Components of a Grid
- Services-oriented Computing (current state)

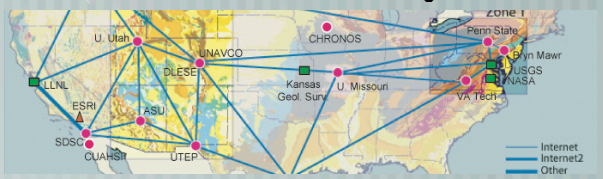
# What is Grid Computing?

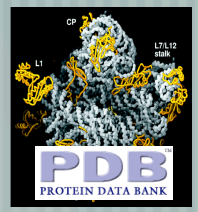
- Start with Raw Hardware,
- Add data
- and Networks,
- Mix in Scientific Datasets.
- Question: how do you manage, provision, schedule, authenticate, monitor, program, and access these resources?



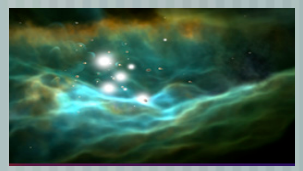




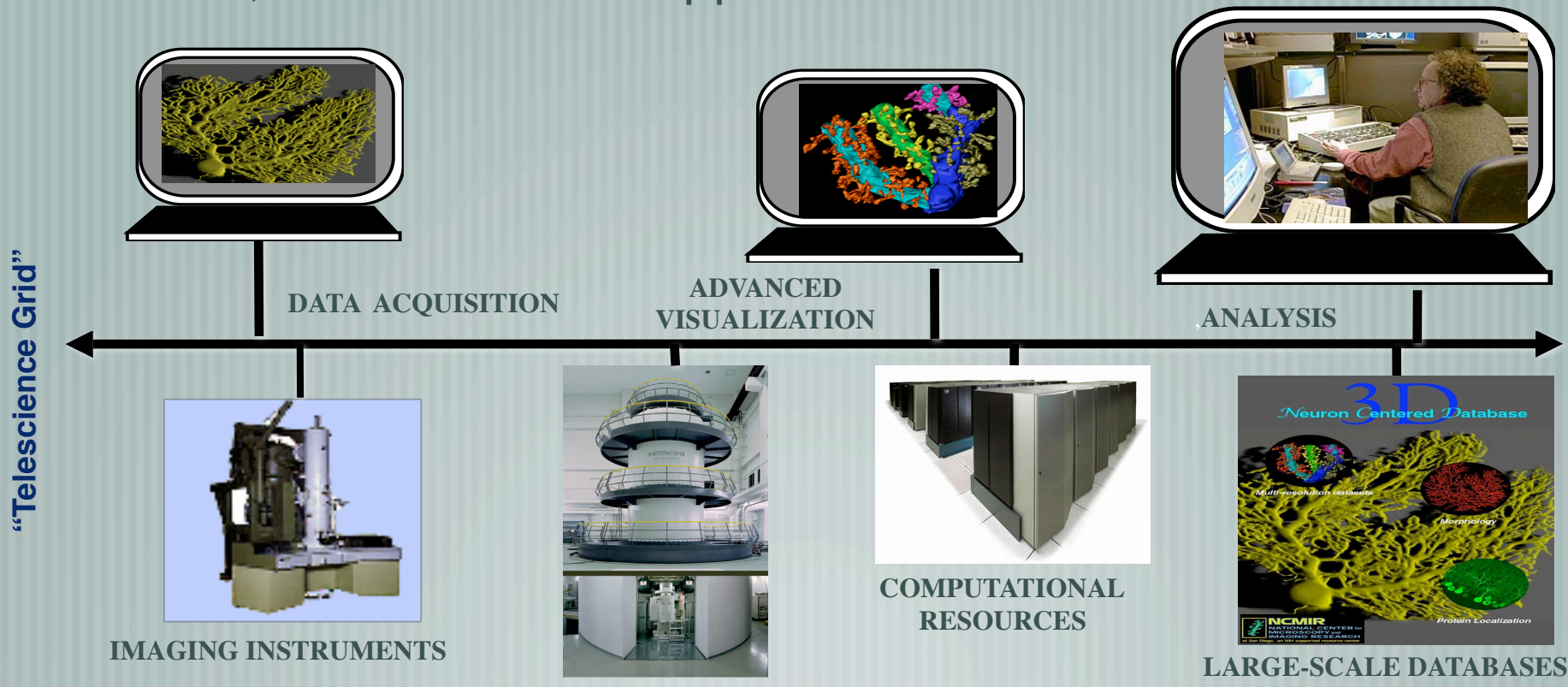






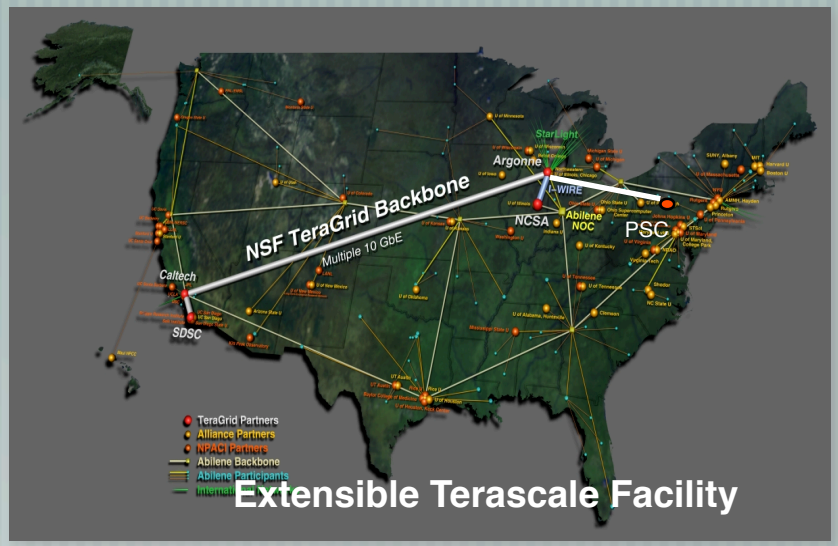
# example: Telescience Grid

**Grid Computing:** The ability to dynamically link resources together as an ensemble to support the execution of large-scale, resource-intensive, and distributed applications



# example: Teragrid

- > 20 TF distributed at 9 sites (SDSC, NCSA, ANL. PSC, Caltech, Indiana U., Purdue U., Oak Ridge National Lab/U of Tennessee, UT Austin)
- > 1 PB storage distributed at 5 sites
- Fast national network with 40 Gb/sec between hubs
- Linux-based SW environment, uniform administration
  - Focus is a national, production Grid
  - Globus enabled
- Basic and Advanced services




TeraGrid is a “top-down”, planned Grid

# example: PRAGMA

Goal: Establish sustained collaborations and advance the use of Grid technologies for applications around the Pacific Rim

Activities:

- Collaboration on development of grid SW, apps, and technologies
- Resource sharing
- Multi-site training and people exchange
- Outreach


PRAGMA:
Pacific Rim Applications and Grid Middleware Assembly

# Distributed Computing in the "Real World"

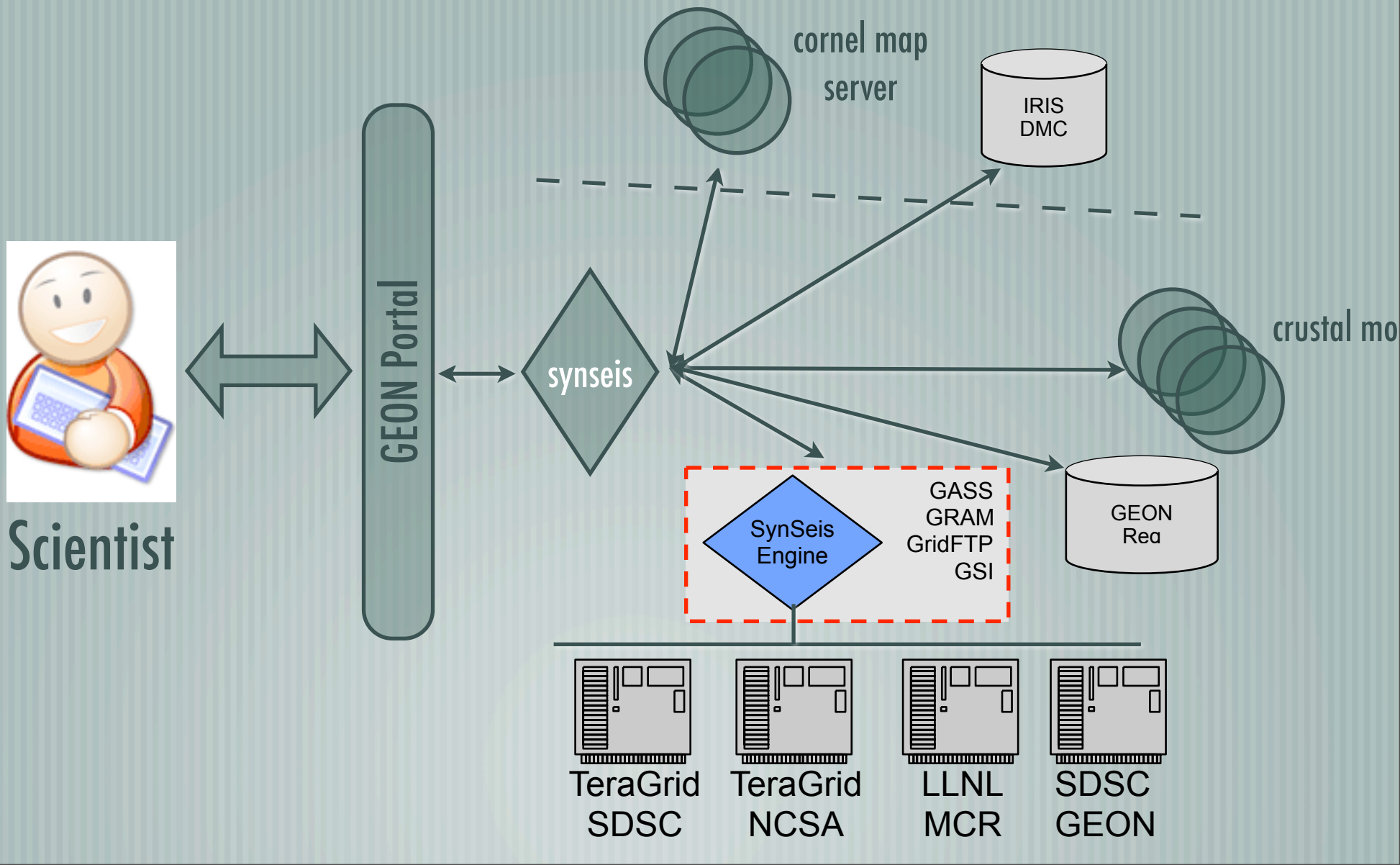

## Walmart Inventory Control

- Satellite technology used to track every item
- Inventory adjusted in real time to avoid shortages and predict demand

- *Data management, prediction, real-time, wide-area synchronization*

- **Everquest**
  - Clusters used as servers for 40,000+ users
    - *Real-time interaction, individualized database management, communication between players*
  - Data management adapted to span both client PC and server to mitigate communication delays

# example: Synseis Application



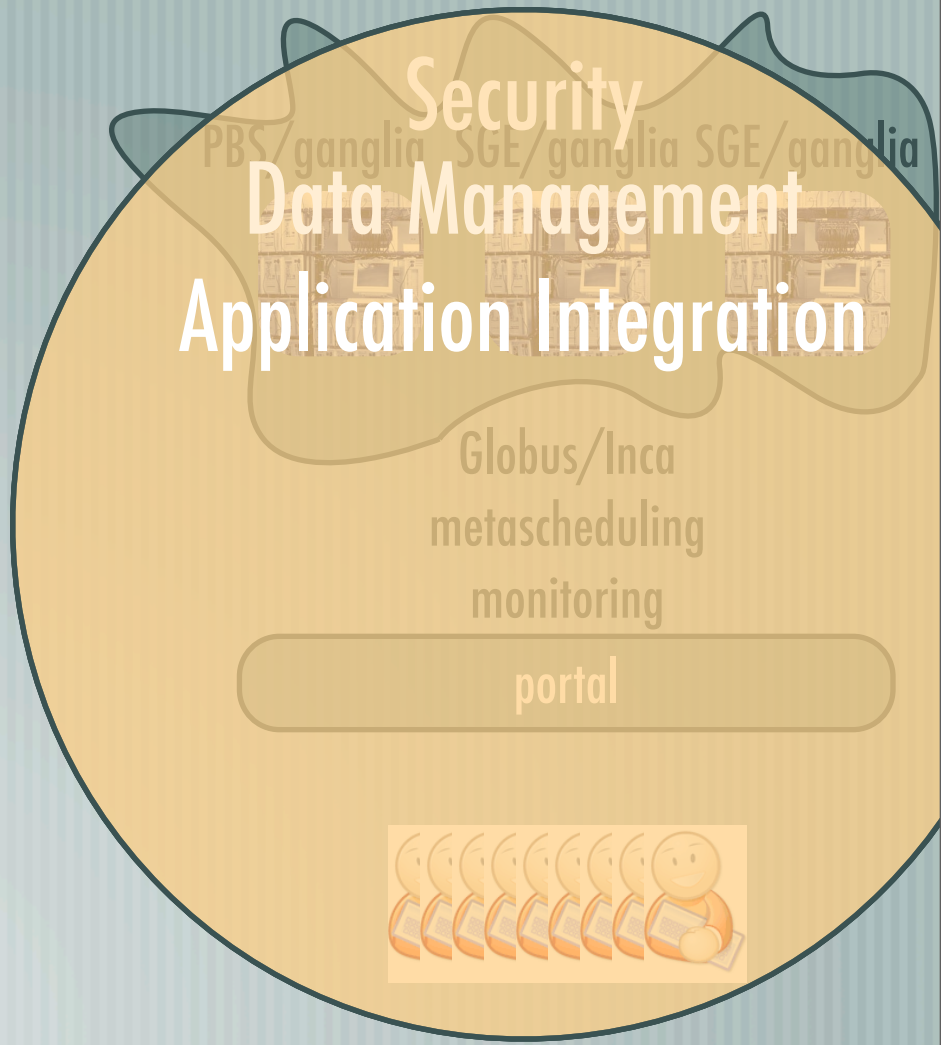
# Summary

- Grids aggregate a set of resources,
  - compute, data, applications, instruments, etc.
- Resources inherently geographically distributed in nature and often transient,
- Grid middleware abstracts away details and provides a consistent logical interface to resources.

- Specific Challenges:
  - decentralized resource management and scheduling,
  - data management,
  - interoperability among components and protocols,
  - security within and across grids,
  - application integration,

# Components of the Grid

- Cluster-level resource/job management
  - Rocks, Ganglia, SGE, PBS
- Intra-cluster resource/job management
  - Inca, Globus, Condor, CSF
- End-user Portal
  - GridSphere, JetSpeed, GridPort ...



# component: Security

- Certificate Authority
  - CACL, naregi
- x509-based public key infrastructure
  - alternatives (kerberos, shiboleth)
  - users have public/private key pair
  - host certificates to identify trusted hosts
- Proxy certificate
  - short-term self-signed certificate
- Proxy Repository (MyProxy)
  - stores certificates for later use
- Authentication based on correctly decrypting the private key
  - support for delegation
  - proxy renewal
- Authorization
  - Security Assertion Markup Language (SAML)
  - Community Authorization System (Globus CAS, VOM)
- support for portals, rich clients, applications, clusters, databases, etc.

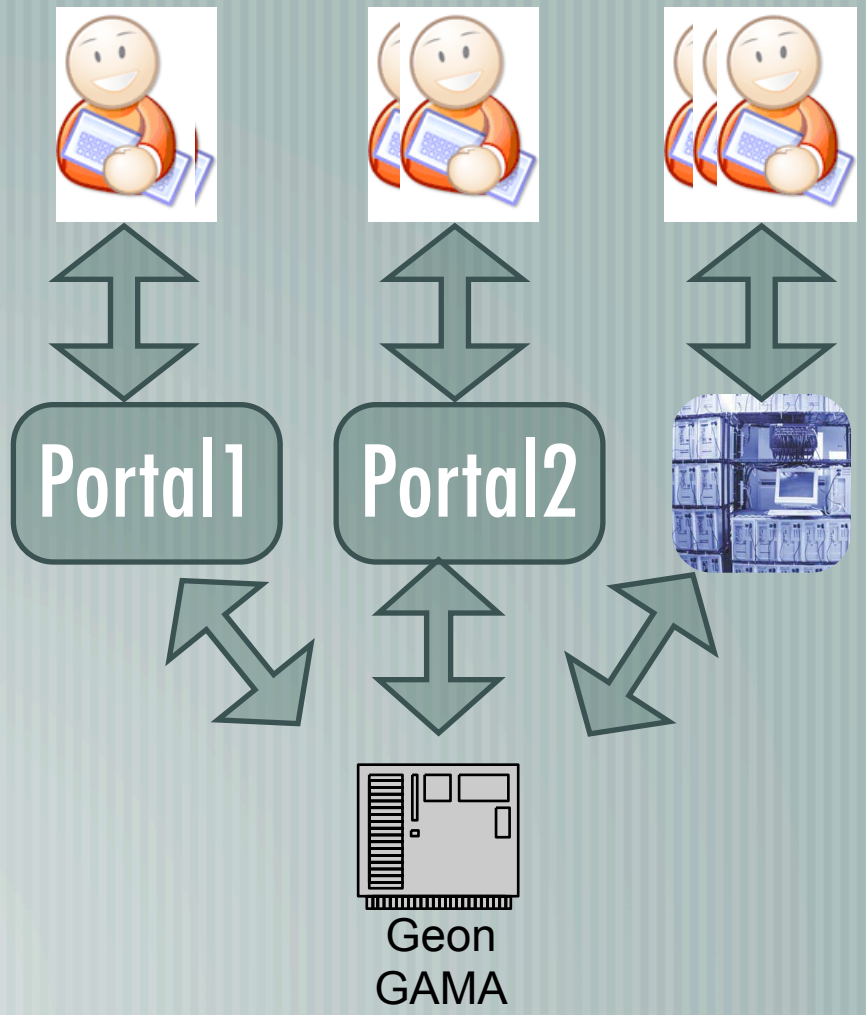
Or

Use Grid Account Management Architecture (GAMA)

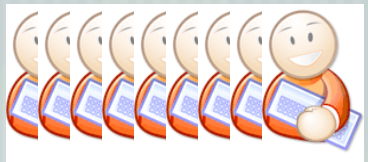
Turnkey solution

# GAMA 2.0

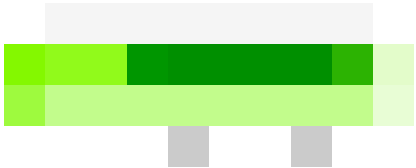
- Available as Rocks Roll, easy deployment
- Wraps certificate authority (CACL or naregi), MyProxy, basic authorization Web Services
- Supports multiple portals, clusters, rich clients or any other resource
- Support multiple sites on one GAMA server
  - Each site has local admin(s) responsible for managing user accounts for their site in GAMA
  - GAMA users may be given access from only one site or multiple



# component: Portals

# Geon Portal



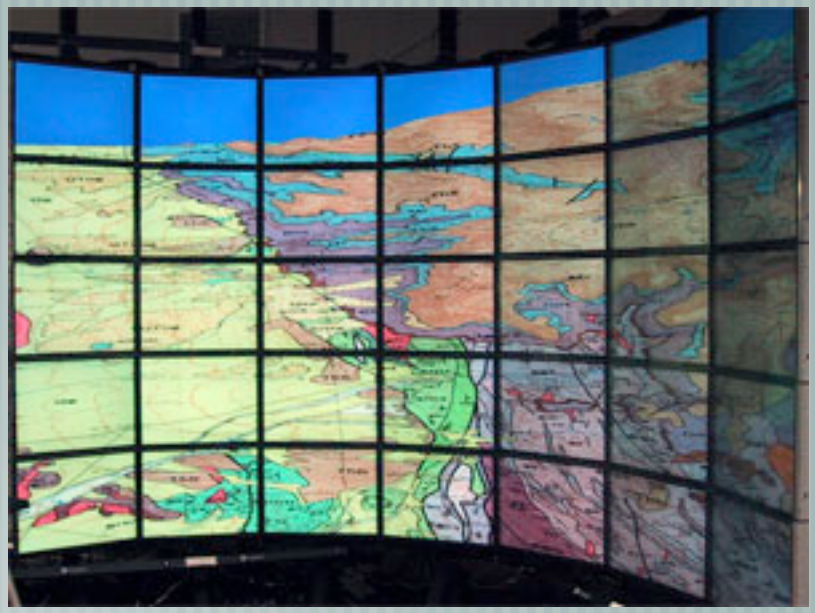
# Portal Technology

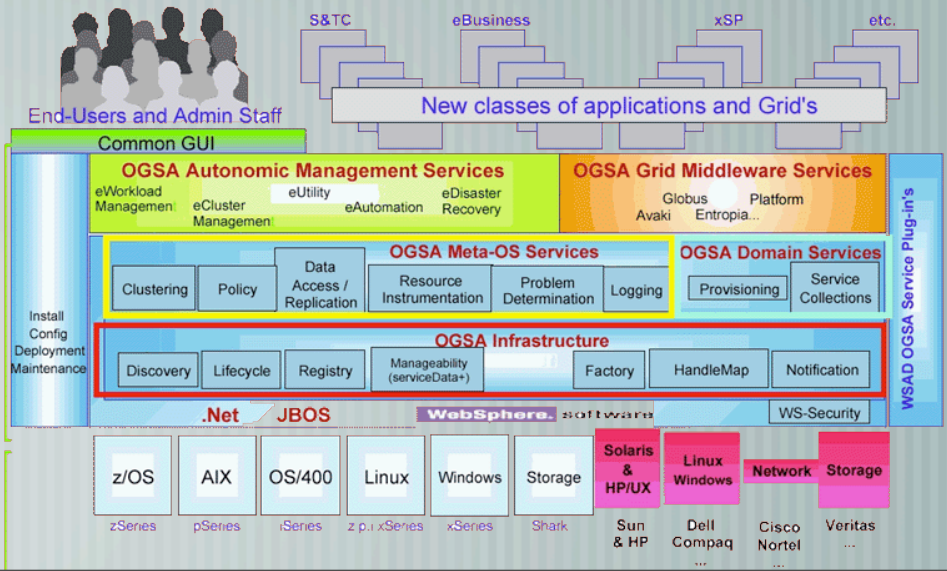
powered by gridsphere

- GridSphere Portal Framework
  - JSR 168 compliant
  - built in "grid"/GAMA support
  - runs in Apache Tomcat/JBoss
  - supported by OGCE
- Alternatives include jetspeed, Chef, Sakai, GridPort, uPortal

# other components ...

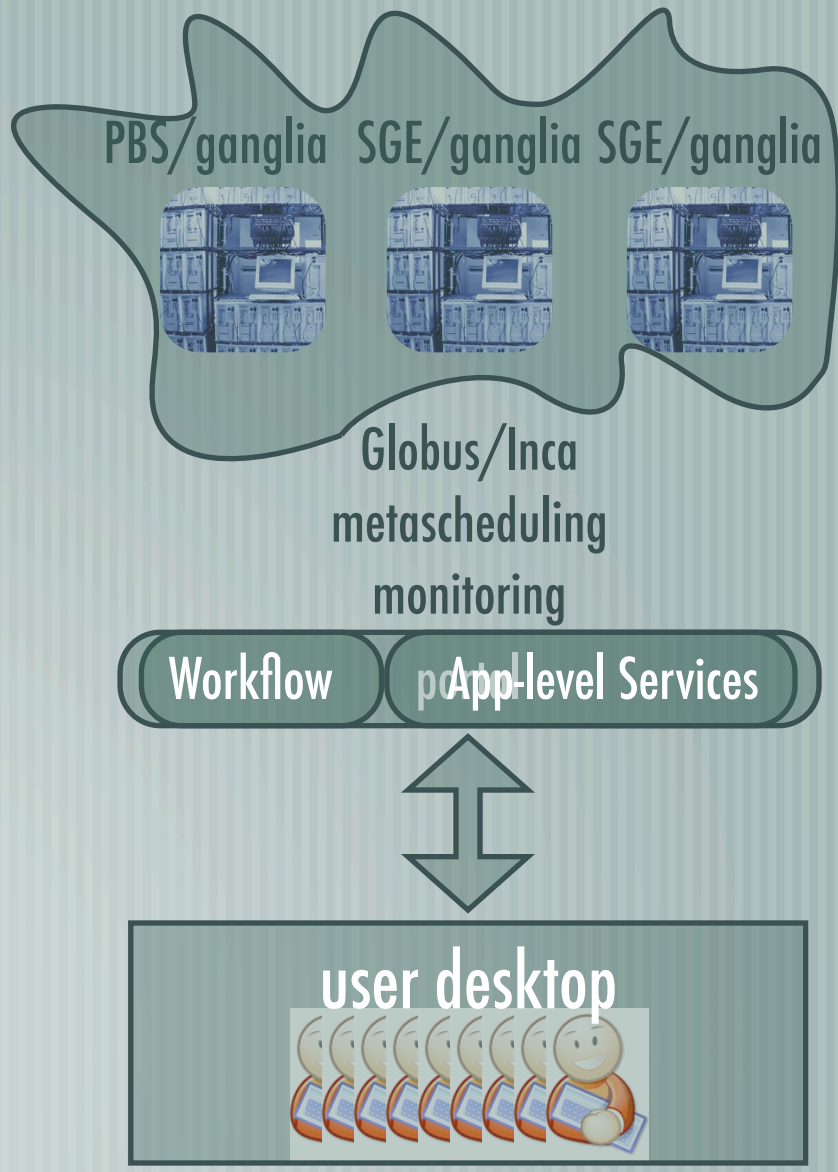
- Data Management
  - GridFTP, Storage Resource Broker, Globus RLS ... etc
- Visualization
  - What can you do with 100 MegaPixel Display?
- Networking
  - dynamic optical fiber provisioning
  - Quality of Service
  - replication/consistency



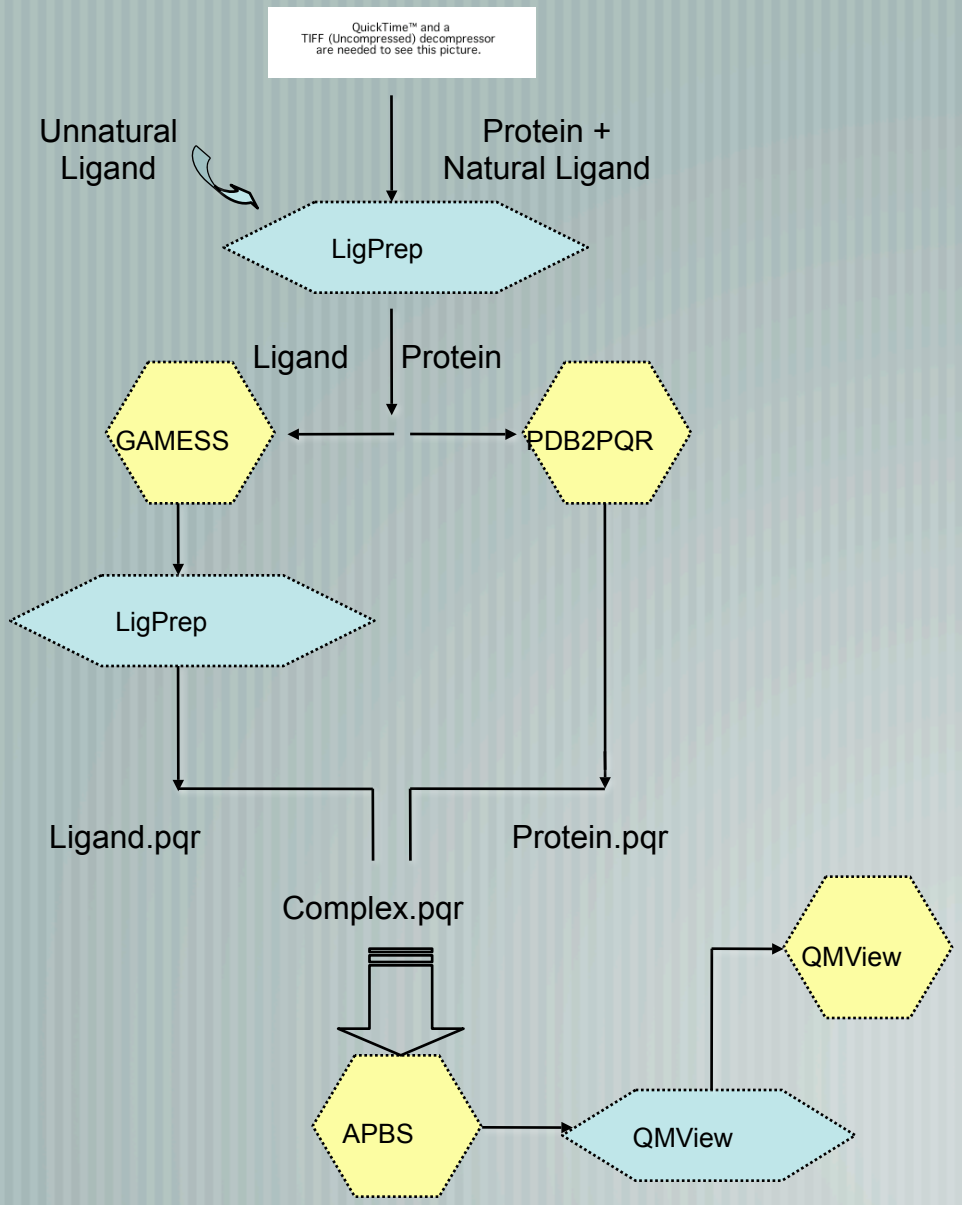


# Services Oriented Computing

- Strongly-typed Services
- Rich clients
  - integrated with user's desktop/laptop
  - vizualization
  - discovery of new services



# Workflows and Strong Data Typing




## Ligand-Protein Interaction

- Baldridge, Greenberg, Amoreira, Kondric
- GAMESS Service
  - More accurate Ligand Information via GAMESS-XML
  - Generation of Conformational Spaces
  - Assignment of parameters for APBS
- PDB2PQR Service
  - Protein preparation
- APBS Service
  - Generation of electrostatic information
- QMView Service or VMD Service
  - Visualization of electrostatic potential file
- Applications:
  - Electrostatics and docking
  - High-throughput processing of ligand-protein interaction studies
  - Use of small molecules (ligands) to turn on or off a protein function

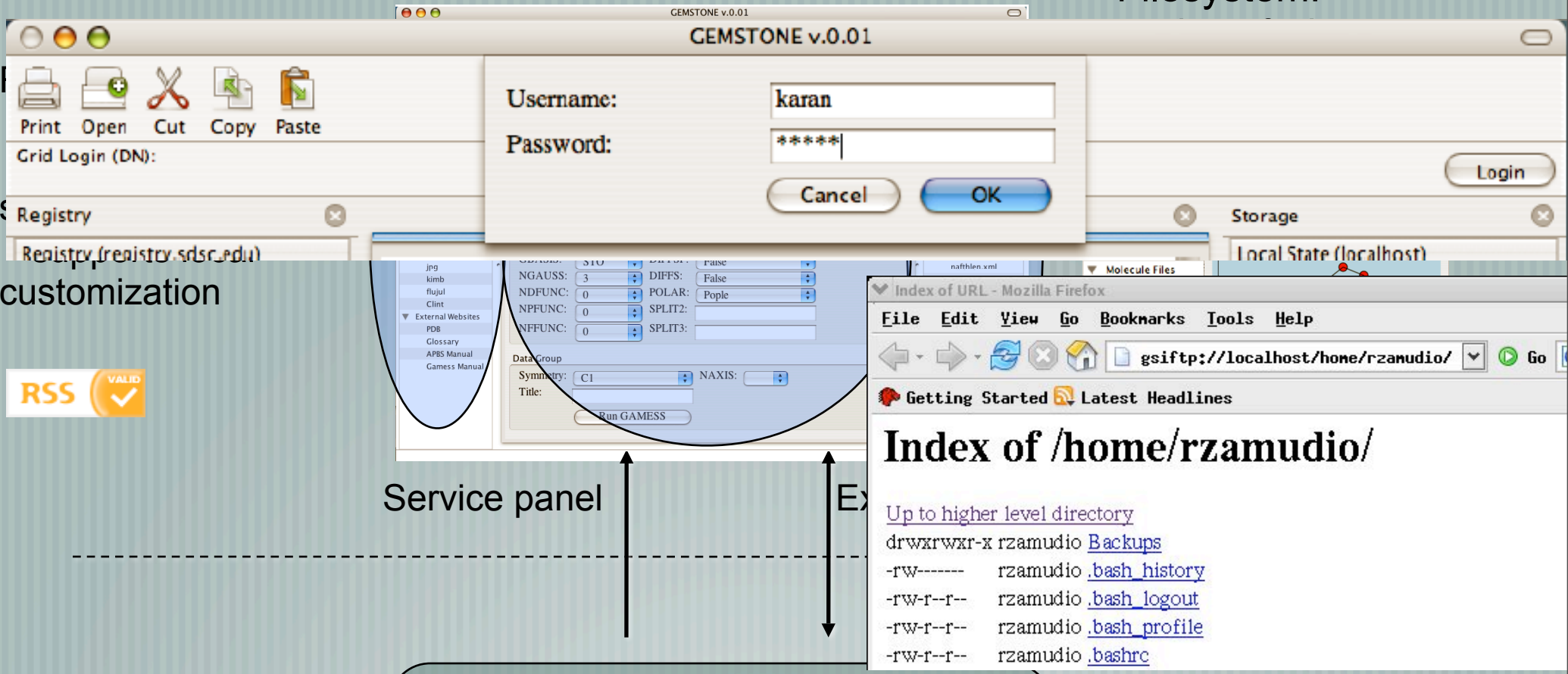
# GEMSTONE Frontend

- XML User Interface Language (XUL)
  - "XUL (pronounced "zool") is Mozilla's XML-based User interface Language that lets you build feature-rich cross platform applications that can run connected or disconnected from the Internet."
  - Standard user interface elements
- XPCom extensions to support new protocols, capabilities (gridftp)
- Provides a UI for the remote Web Service APIs
- Platforms
  - Windows, mac, linux, solaris, *bsd, BeOS, AIX etc. etc. etc.


mozilla

Security:
- supports GSI-based authentication

Filesystem:




customization

RSS VALID

Service panel

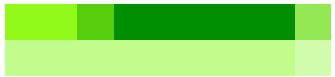
APBS.xul

WSDL

APBS Application WS

# GEMSTONE

"Grid Enabled Molecular Science Through Online Networked Environments"



# Notes on future direction

- Globus Web Service Resource Framework (GT 4.0)
  - convergence between Web and Grid standards
  - interface-level compatibility among competing components
- Technologies everyone must know and understand:
  - xml, xml Schema, wsdl, xpath,jsp/servlets