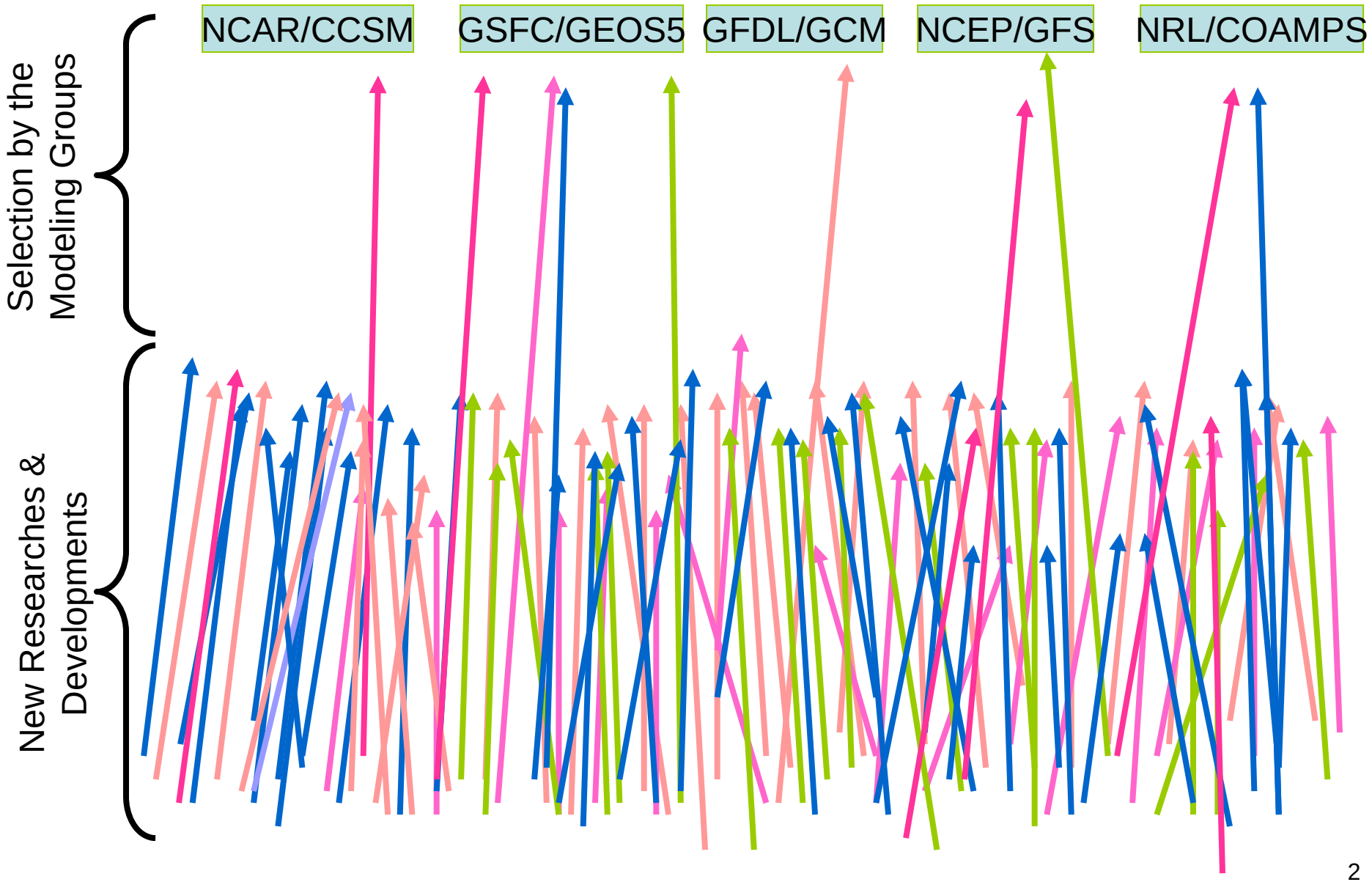
# Climate-in-a-Box

- A NASA Computational Climate Modeling Campaign

http://climateinabox.nasa.gov


**Tsengdar Lee**
**Tsengdar.j.lee@nasa.gov**
**(202)358-0860**

# The Situation



# The Transition Challenge

- The cost to transition is never appreciated by the R&D community.
  - Recent NSF cyber-infrastructure study estimated 1:10 ratio for R&D vs. transition costs
- Operating centers never have enough money to transition many R&D results to the operation environment.
- R&D PIs are frustrated by not getting their innovation accepted.

# Challenges in Climate Model Development

- It's a natural tendency for climate models to become more and more "elaborate."
- It is difficult to verify and validate the complex models.
- Climate model code development is tightly controlled by selected few organizations.
- Hierarchical structure inhibits community inputs into the core model.
- As the community grows, the challenge becomes unmanageable.
- Need an **agile governance model** and a **reward system** that encourage community engagement and allow a community selection process.

# The Industry Model

1. Announce a development plan with multi-phase release schedule.
2. Release “developer’s tool kit” to set the boundary conditions.
3. Set the condition (benchmark, test procedure, certification) for inclusion.
4. Frequent update of the schedule, tools, and benchmarks.

# Campaign Goals

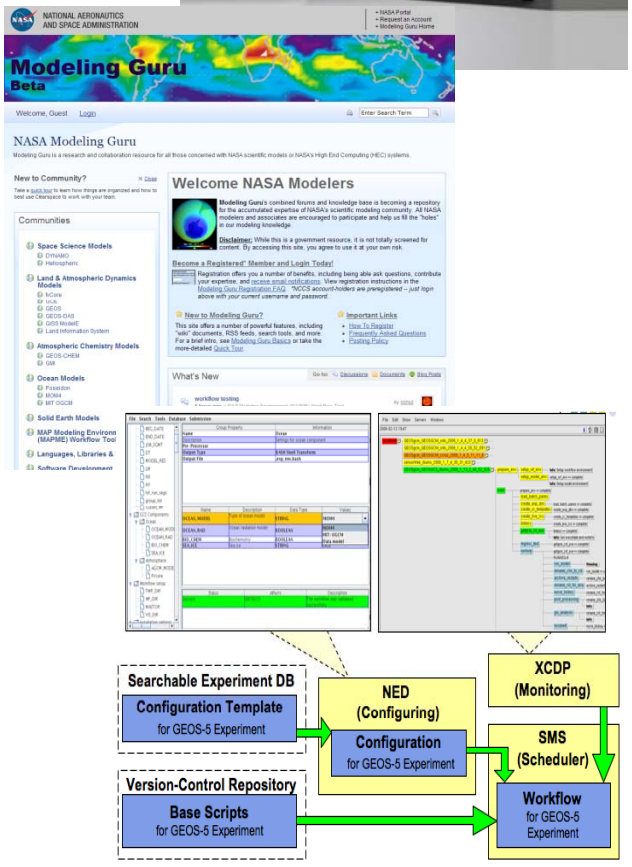
- Remove bottlenecks from climate model development life cycle.
- Evolve climate models using Darwinian natural selection processes.
- Build and maintain the climate modeling knowledge base.
- Broaden the base of climate modeling developer and user community.
- Advance science and science applications of satellite data assimilation and computational modeling for climate, weather, water and carbon cycles.

# Solution

- Lower the bar for entry
  - Build and distribute low cost and turn key system packaged with HW, SW, data, scenarios, and productivity tools.
- Establish common baseline and boundary condition
  - Provide a development framework and a consistent climate modeling software architecture.
  - Provide startup models, configurations, and data analysis system
  - Establish standardized tests.
- Build a social network
  - Climate modelers
  - Model users
  - Data providers
  - Application users (water managers, energy & insurance sector, agriculture sector)
- Create incentives for modelers and users to participate and to volunteer the knowledge
- Create reward systems for long-term sustainment

# Lower the Bar for Entry – Model Developers

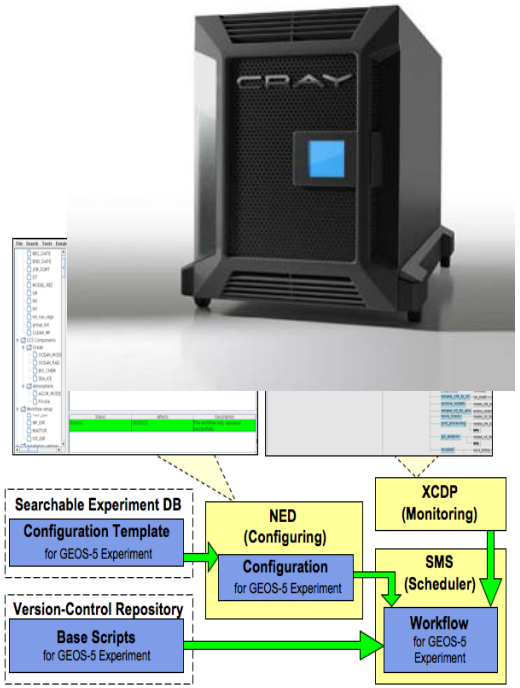
- Low cost computing platform (e.g. Cray CX1 w/ MS HPC 2008 or Redhat Linux)
- Atmosphere/Ocean models* (Model-E, WRF, GEOS-5, CCSM, GFS)
- Earth System Modeling Framework & MAP Library (ESMF/MAPL)
  - Componentized architecture to reduce software engineering complexity
- Data (MERRA, SST, NCEP/NCAR reanalysis)
- Modeling Workflow
  - Model configuration, experiment design, and input/output data management
  - Tracking of experiments
  - Share experiment designs
- Development environment with compiler and debugger (e.g. eclipse, MS Visual Studio)
- Visualization Software (e.g. IDL, MatLab)
- Startup AMIP, CMIP, weather, and S/I runs
- MERRA scout run, ECMWF nature run
- Scenarios, OSSE, OSE

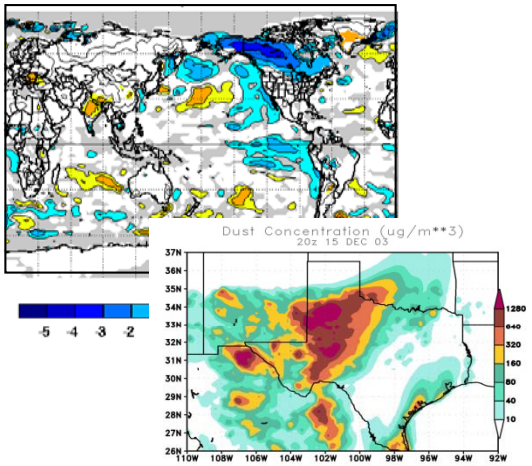




* gray color indicates future implementation.

# Lower the Bar for Entry- Application Users

- Low cost computing platform (e.g. Cray CX1 w/ MS HPC 2008 or Redhat Linux)
- Pre-configured Model Interfaces
  - Global Climate to Regional to Local If-Then Applications
  - Chesapeake Bay Is Single Use Case (other examples: energy, agriculture insurance, transportation, etc.)
- Pre- Configured Modeling Workflow Hides Model Complexity
  - Interfaces Validated for Applications
  - Collaborative Hind-Casting Testing versus MERRA data
  - Share experiment designs
- Visualization Products via WMS/WCS/GIS Type Interfaces
- Remote Link to Large Scale Ensembles Runs on Large Scale Computing Facilities (e.g. Larger Numbers of CPUs needed for Ensemble Global Runs, Driving Cray WRF/Regional Models)
- Demonstration Project Taken to Applications Community to Identify Additional Specific Use Cases

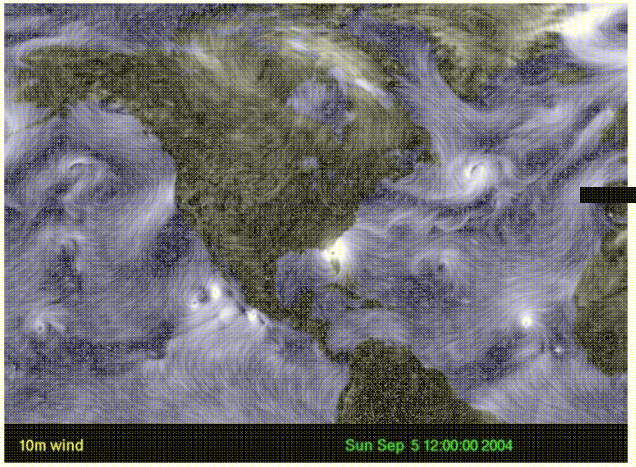


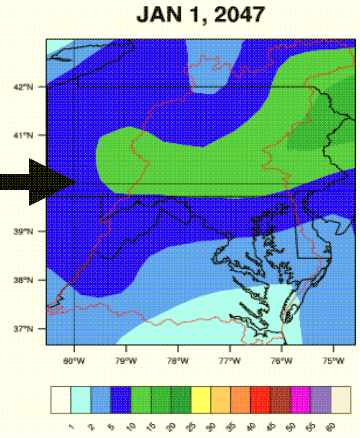
# Climate-In-A-Box: Application Users

## Dynamic Downscaling: Scales That Matter to Decisions

GLOBAL MODEL




REGIONAL MODEL




REGIONAL DECISIONS

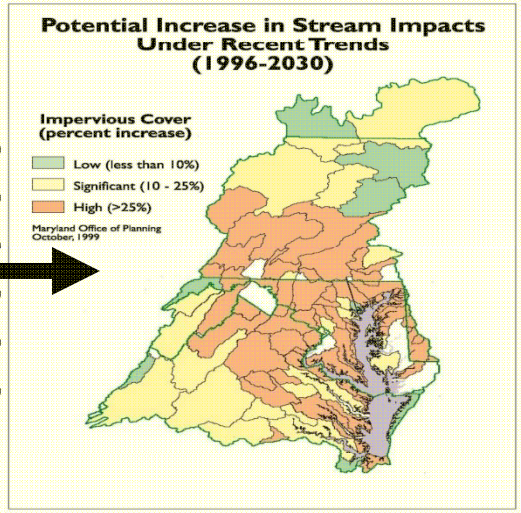
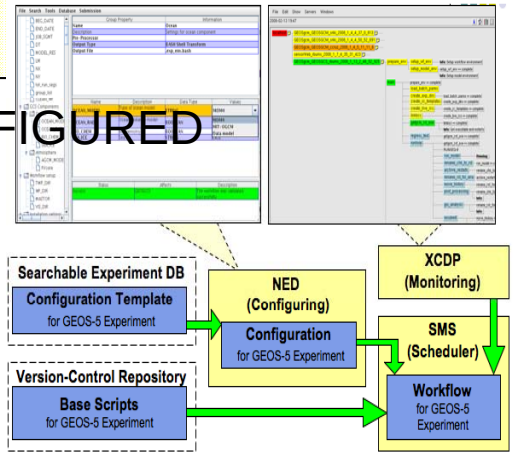



*Figure 5-8. Great stretches of the Cheapeake Bay watershed will likely see more areas covered by impervious surfaces—roads, highways, driveways, rooftops, and parking lots. The areas most acutely affected (see map) will experience increases of 25 percent or more in impervious cover, if recent trends persist.*

PRE-CONFIGURED




# Cray CX1

- Personal “turn-key” supercomputer
- Plug to the wall - No additional power and cooling required
- Starting $25K
  - 4 socket, 16 compute cores
- Fully populated at $90K
  - Up to 8 compute nodes
  - Up to 64 compute cores
  - 16 gigabytes of memory
  - 4 terabytes of disks

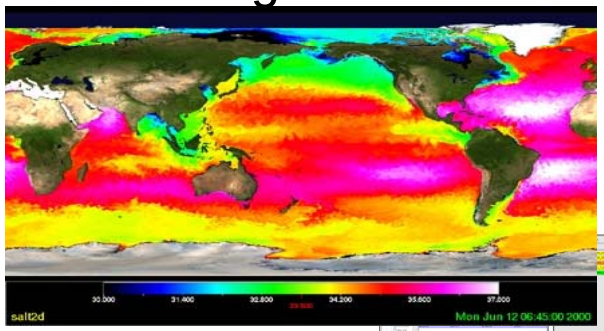

- ¼ x ¼ degree global atmosphere model run for hurricane forecast will fit in this machine
  - 5 day hurricane forecast may be done in two hours

http://www.cray.com/products/CX1.aspx

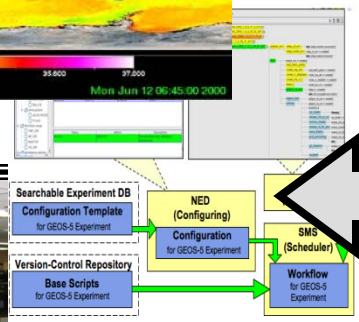
# Advanced Concept

*Collaborating Computing via "Sharing Workflow"*

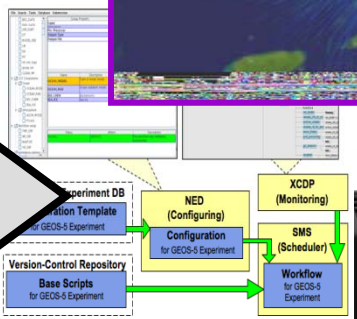
- Shift from Local Model Runs to Larger Facility (Cloud or Grid computing) to Obtain Greater Numbers of CPUs, or
- Collaborate among Ensemble, or
- Use Global Model Output to Drive Regional Models
- Using Pre-Configured "Workflow Sharing" Support Global and Regional Modeling

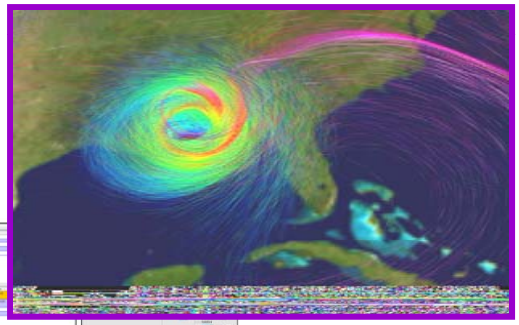


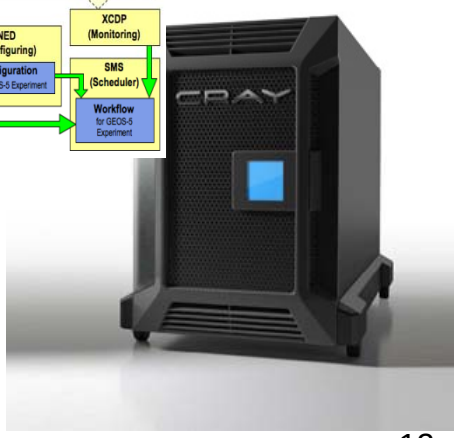





Share pre-configured Workflows



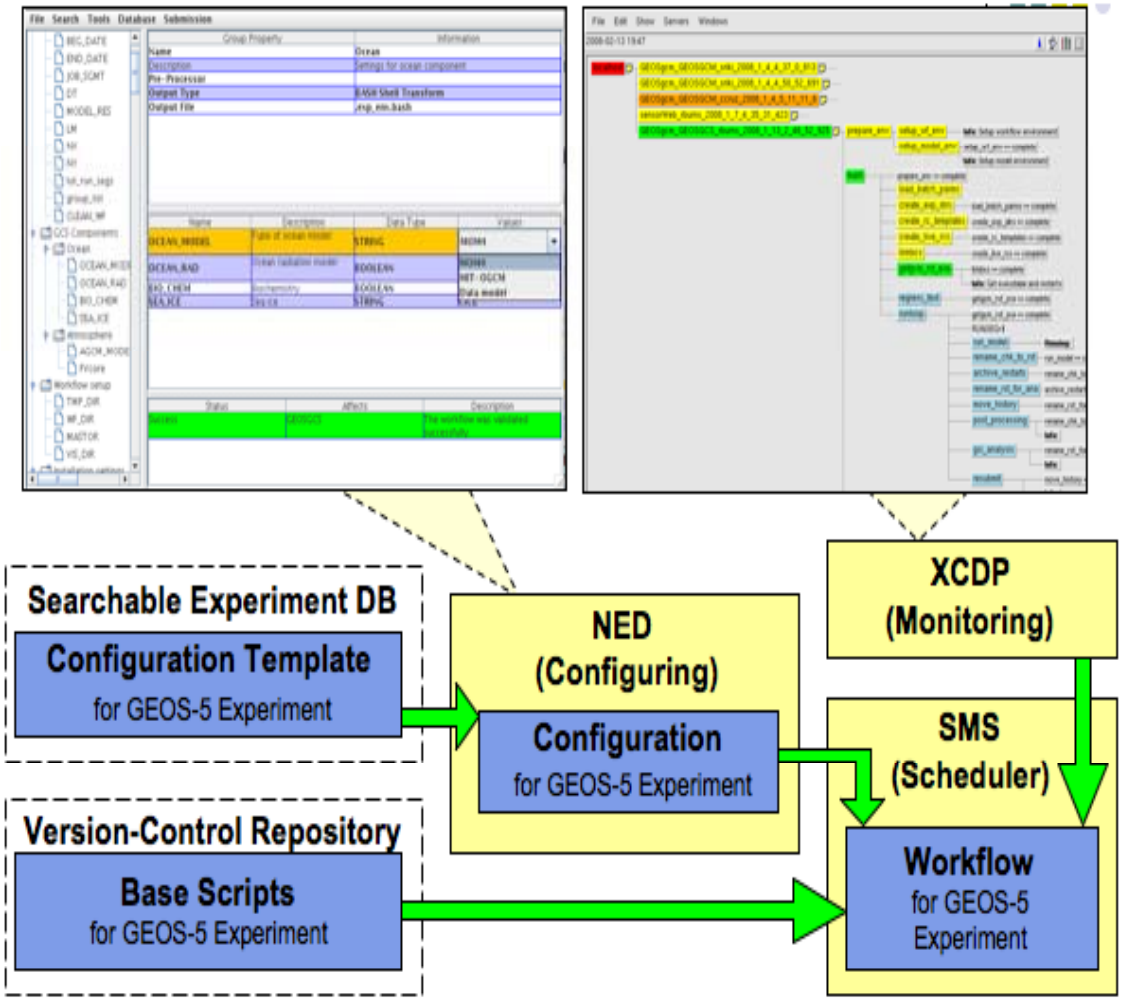


# Modeling Guru Social Network

- Web 2.0 based modeler's social networking site
- Knowledge management tool
- Exchange of model components and blog about modeling experiences
- Ranking by natural selection

http://modelingguru.nasa.gov

# Workflow & NASA Experiment Designer (NED)




- Model configuration, experiment design, and input/output data management
- Tracking of experiments
- Tracking and maintaining of I/O data
- Version control
- Repeatable experiments
- Sharing experiment designs

# Schedule and Resource

Immediately:

- Target: Roll out at AGU 2009 with workshop and tutorial sessions
- 1 year, 2 phases delivery schedule with complete functional system before June 2009 and production ready by September 2009
- 2-3 FTE, 5 HW/SW systems, $ for workshop and tutorial
- Possible external industry partners: Cray, Microsoft, Intel

Near Term (within 5 years):

- Enhanced capability to include ocean model, carbon model, ecology model
- Seamless transition from the entry level platform to large computing centers
- Create incentives for scientists/engineers to work in this paradigm
- Build a self-organizing consortium to maintain knowledge base, data source, workflow, and scenarios