

# ***Cyberinfrastructure and Collaboration***

***Dr. Fran Berman***

Director, San Diego Supercomputer Center and NPACI  
Professor and High Performance Computing Endowed  
Chair, UCSD



SAN DIEGO SUPERCOMPUTER CENTER

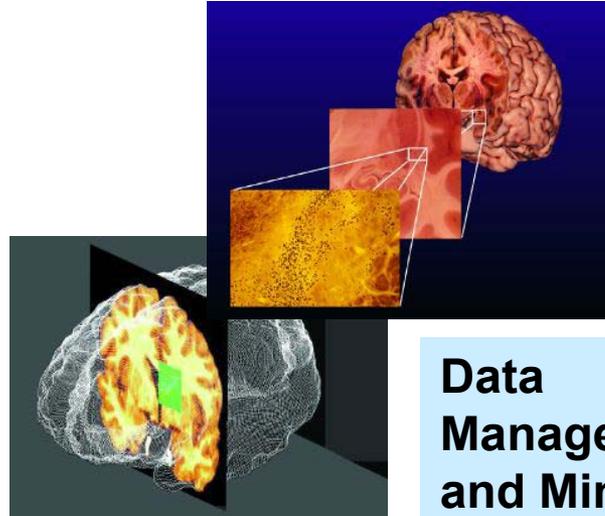
University of California, San Diego



# Science is a Team Sport



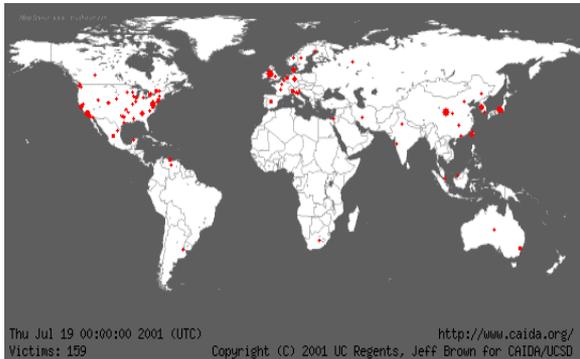
Astronomy



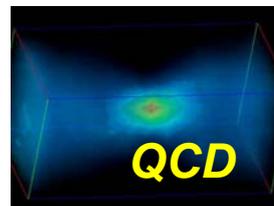
Data Management and Mining



Geosciences



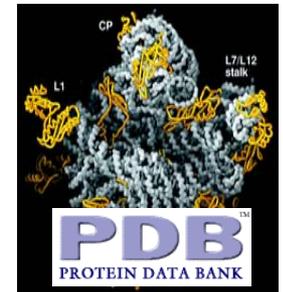
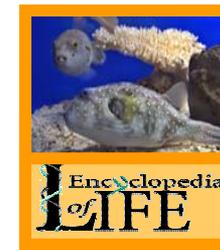
Modeling and Simulation



Life Sciences



Physics



SAN DIEGO SUPERCOMPUTER CENTER

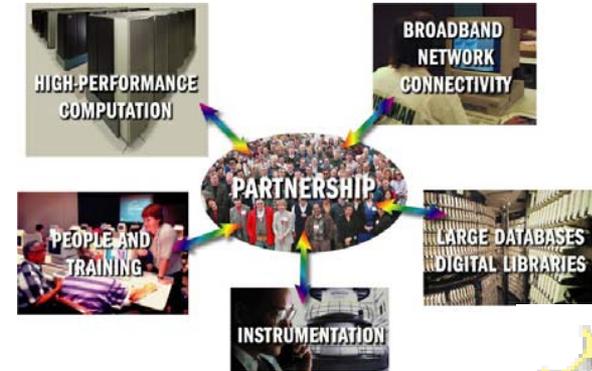
University of California, San Diego



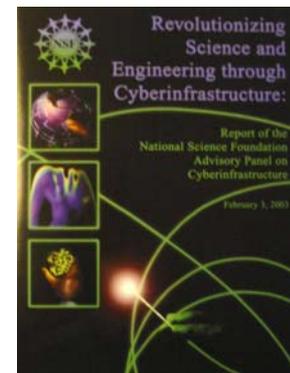
# Integrated Cyberinfrastructure

**Cyberinfrastructure** is the organized aggregate of technologies that enable us to access and integrate today's information technology resources – **data, computation, communication, visualization, networking, scientific instruments, expertise** – to facilitate science, engineering, and societal goals.

- National Science Foundation coined the term and is developing a coordinated and comprehensive set of projects in Cyberinfrastructure



National Science  
Foundation's  
Cyberinfrastructure



**NSF Blue Ribbon Panel (Atkins) Report provided compelling and comprehensive vision of an integrated Cyberinfrastructure**



SAN DIEGO SUPERCOMPUTER CENTER



# Cyberinfrastructure requires broad participation

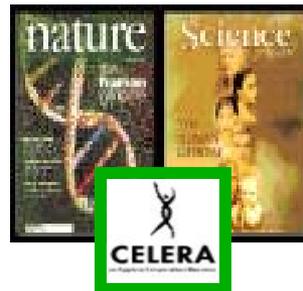
*The success of Cyberinfrastructure will require unprecedented coordination between the academic and commercial environments*

- **Science has been a key driver for commercial technologies**

- Internet
- Email
- Portal technologies
- Open source
- Data visualization,
- etc.

were first developed by and for the scientific community to address its problems

Arpanet  
1969



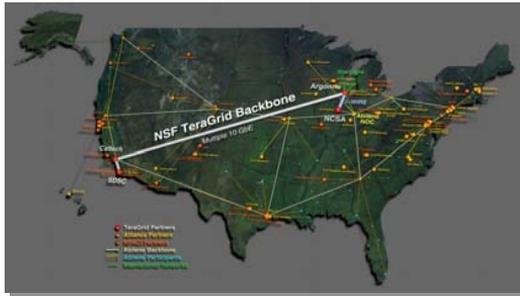
- **Driving technology Innovations coming out of the commercial sector**

- In the race to decode the complete human genome, **Celera** used a sophisticated combination of innovative hardware, software and data technologies.
- **Walmart** inventory control is one of the largest distributed programs on the planet
- **Airlines** daily use scheduling systems much more sophisticated than those used on supercomputers.



SAN DIEGO SUPERCOMPUTER CENTER

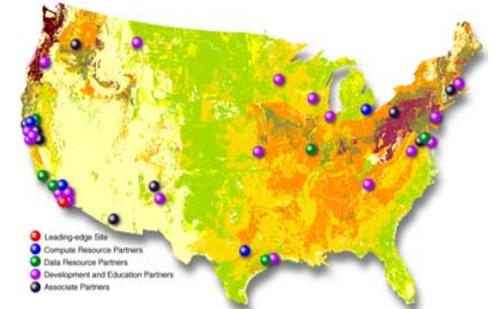
# Community Cyberinfrastructure Projects (point of the spear, tip of the iceberg)



**TeraGrid:** “Top down”  
Grid infrastructure



**PRAGMA:** Pacific Rim Grid  
Middleware Consortium



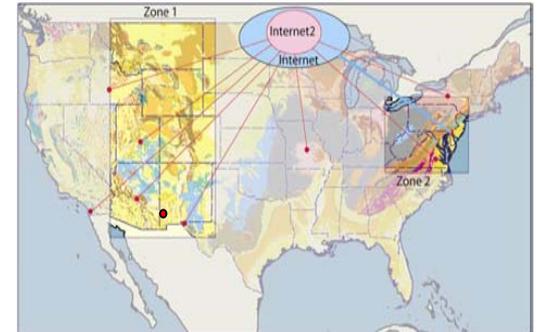
**NPACI Grid:** “Bottom up”  
Grid infrastructure



**NEES:** Earthquake  
Engineering infrastructure



**BIRN:** Biomedical Informatics  
Research Network

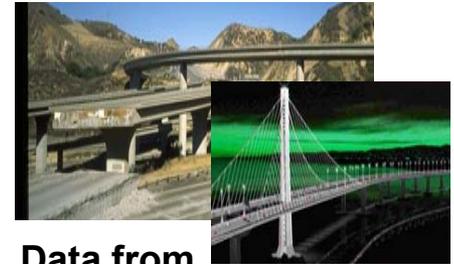
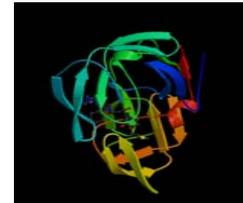


**GEON:** Geosciences  
Grid infrastructure

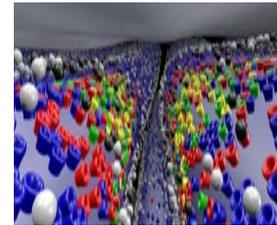


# Data Cyberinfrastructure – pushing the limits

- **Data comes from everywhere**
  - Scientific instruments
  - Experiments
  - Sensors and sensornets
  - New devices (personal digital devices, computer-enabled clothing, cars, ...)
- **And is used by everyone**
  - Scientists
  - Consumers
  - Educators
  - General public
- **Software environment must support unprecedented diversity, globalization, integration, scale, and use**



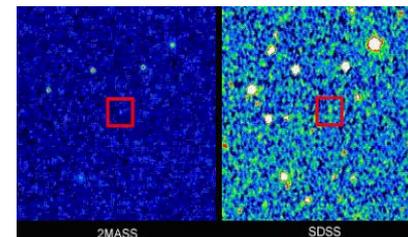
Data from sensors



Data from simulations



Data from instruments

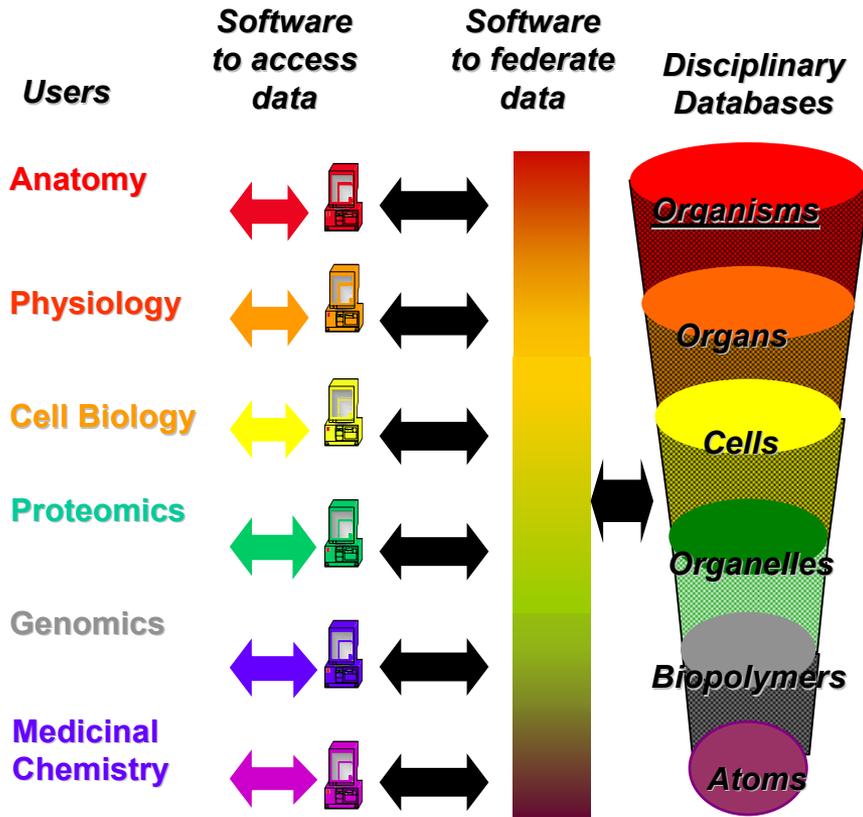


Data from analysis



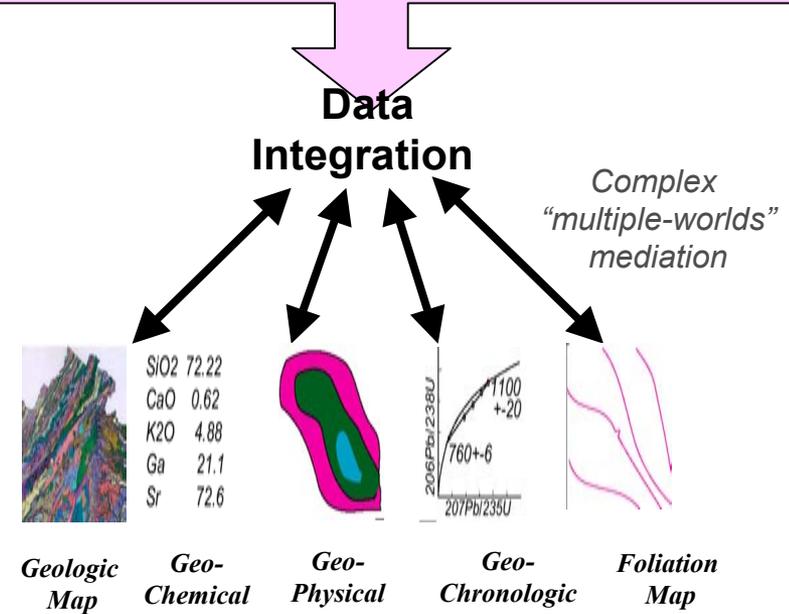
# Data Integration from a Science Perspective

- Data Integration in the Biosciences**



- Data Integration in the Geosciences**

Where can we most safely build a nuclear waste dump?  
 Where should we drill for oil?  
 What is the distribution and U/ Pb zircon ages of A-type plutons in VA?  
 How does it relate to host rock structures?

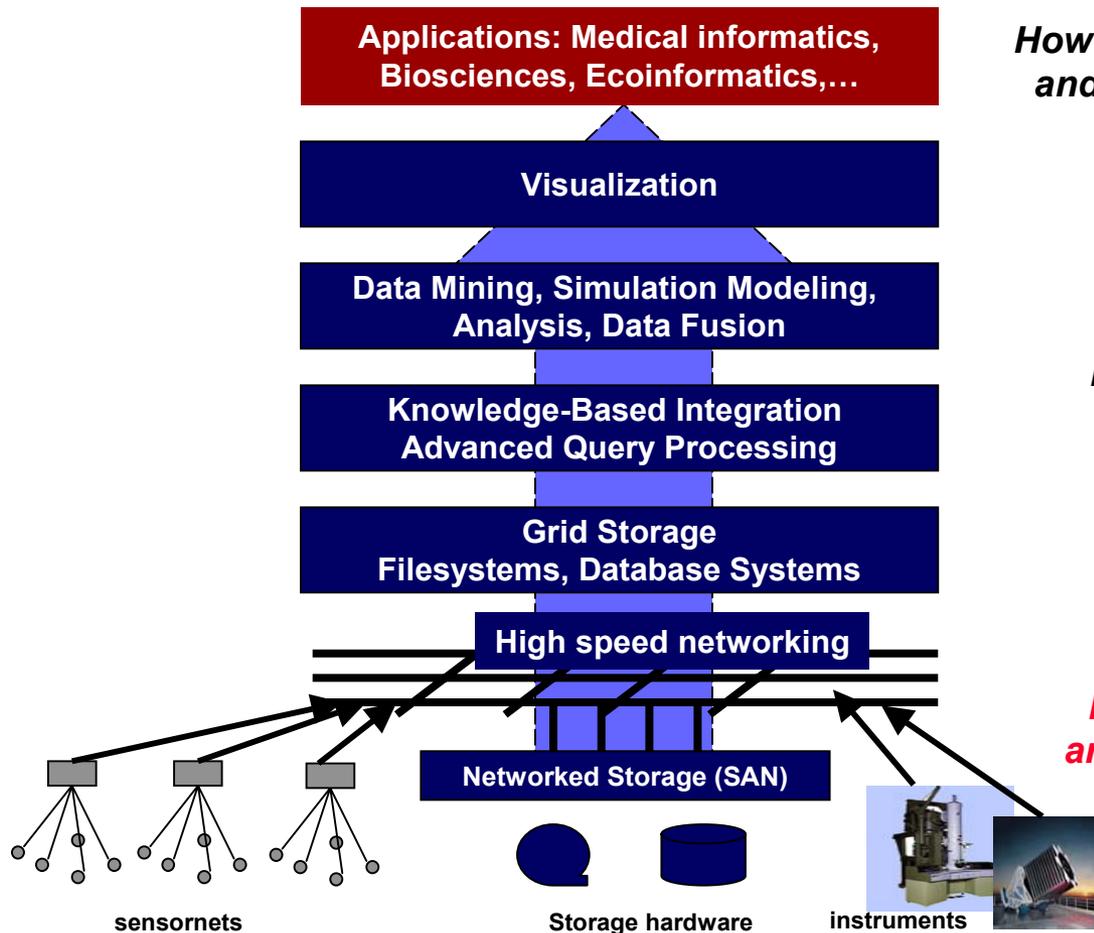


SAN DIEGO SUPERCOMPUTER CENTER

University of California, San Diego



# Data Integration from a Technology Perspective



*How do we combine data, knowledge and information management with simulation and modeling?*

*How do we represent data, information and knowledge to the user?*

*How do we detect trends and relationships in data?*

*How do we obtain usable information from data?*

*How do we collect, access and organize data?*

*How do we configure computer architectures to optimally support data-oriented computing?*



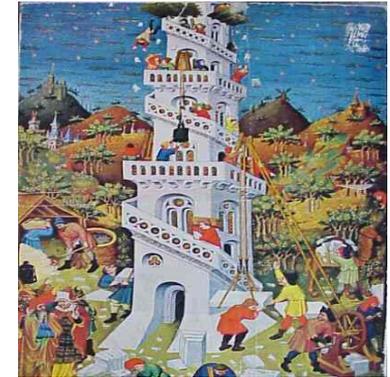
# Cyberinfrastructure's Many Faces



*Cyberinfrastructure as core infrastructure*



*Cyberinfrastructure as a balancing act*



*Cyberinfrastructure as software  
Tower of Babel*

**NetworkMagazine.com**

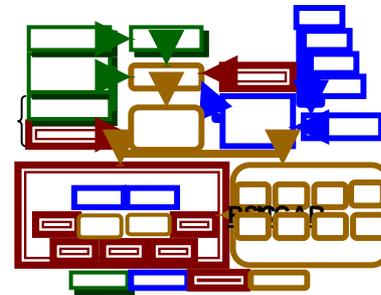
**BUILDING GRIDS: HYPE MEETS REALITY**

The senior executives from Compaq (now part of the new HP), Platform Computing and renowned analysts from IDC invite you to attend an Exclusive Executive event on Grid Computing.

**Insanely Overhyped Technology of the Month- Grid Computing**

Distributing executable processing capacity across thousands or millions of CPUs isn't the right approach for most of the problems that computers

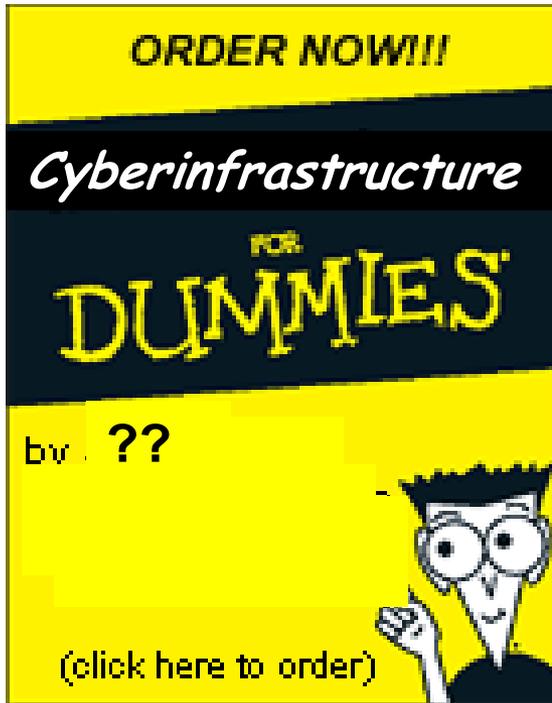
*Cyberinfrastructure vision as over-hyped abstraction*



*Cyberinfrastructure as an organizational challenge*



# Cyberinfrastructure is a work-in-progress



*What users ultimately  
want to see*

- **Current challenges include**
  - **Technical challenges** (What went wrong?)
  - **Ideological challenges** (What should be distributed? What should be centralized?)
  - **Political challenges** (What if I want an account in the U.K. What if I want an account in Pakistan?)
  - **Logistical Challenges** (Who maintains the SW? Who fixes the bugs?)



# Broad-Usage Challenges

**Social, economic, political policies for Cyberinfrastructure are largely uncharted territory**



*Cyberinfrastructure-wide policies will be key to avoiding resource instability, thrashing, and “blackouts”*



*Aggregate user behavior must not destabilize the system*

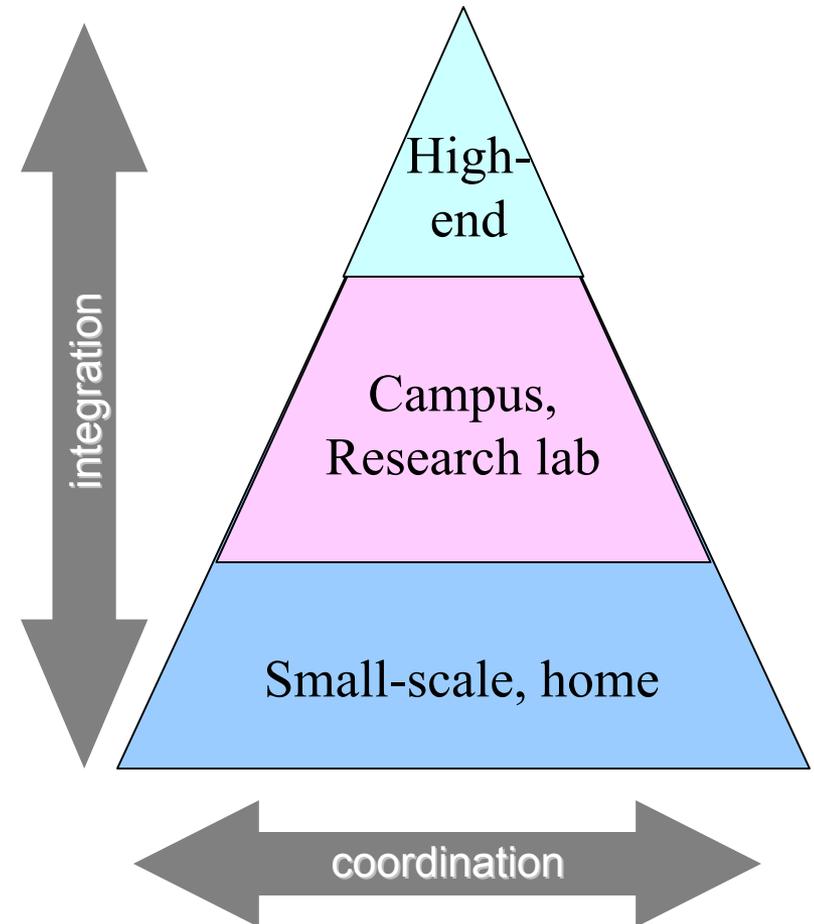


*Cyberinfrastructure “economics”: Balancing cycles, bandwidth and storage presents a “multiple currencies, multiple users” problem*



# *Broad Usage of Cyberinfrastructure will require unprecedented linkage between “local” environments*

- Users want to work within an environment that provides what they need, when they need it
- **Goals:**
  - It should be easy to go from small-scale to large-scale and back as needed for an application
  - It should be reasonable to pull together a customized “personal cyberinfrastructure” that supports individual, project, and community applications
  - The environment should provide what the users need, not just what the technologists want

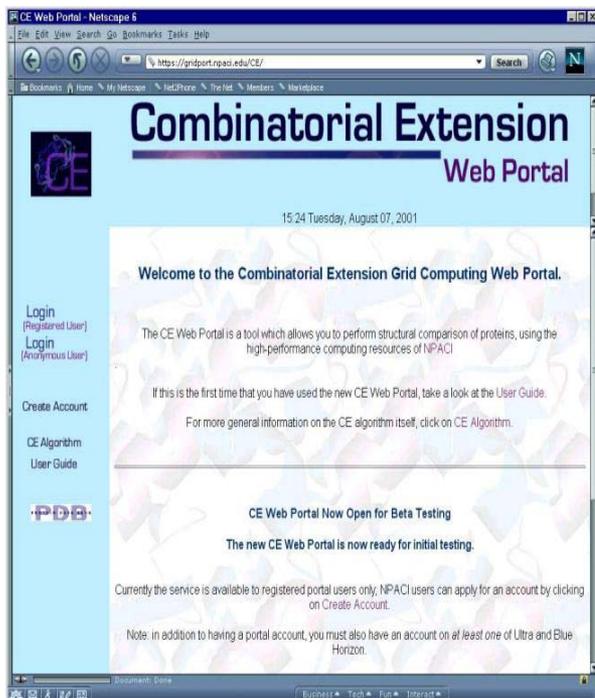


# User and Collaborative Environments

Every 14 year old  
gamer regularly uses  
more sophisticated  
user environments  
than most scientists

Today's RPG games  
provide robust,  
interactive, dynamic,  
distributed environments  
with 10's of thousands  
of users

Gaming technologies  
demonstrate  
sophisticated adaptation  
to available resources  
(home PCs, cable  
modems)



Today's "high-tech"  
user portals

Today's "low-tech"  
RPG games



SAN DIEGO SUPERCOMPUTER CENTER

University of California, San Diego



# Using Gaming Technologies to Improve Science User Interfaces

- “Back-end” custom gaming server(s) can be replaced with more powerful technologies (high-end computers, grid environments, community data bases, etc.)
- Re-usable “grid services” could link sophisticated user environments to a wide spectrum of computational and data management environments
- **Opportunity: Use Gaming technologies to create a dynamic, flexible, rich, and evolutionary environment for interaction and collaboration**



# Getting There

- **What should our aggregate Grid/Cyberinfrastructure/integrated technology platforms look like in 10 years?**
  - What kind of research and development should we be encouraging **now**?
  - What are the building blocks? What are the goals? What are the metrics of success? How should we evaluate where we're at?
  - **Lessons learned from Grid Computing:**
    - **Don't oversell the vision.**
    - **Don't underestimate what it takes to get there.**
    - **It's not successful infrastructure until the users say it is.**



*Coordination  
of efforts  
will be  
critical for  
success*



---

*Thank you*

*www.sdsc.edu*



SAN DIEGO SUPERCOMPUTER CENTER

---

University of California, San Diego

