

New NSF initiatives in cyberinfrastructure

David W. Lightfoot

Assistant Director, National Science Foundation
Social, Behavior & Economic Sciences
UCGIS: 7 February 2008

Why is cyberinfrastructure important and why now?

- ▶ We are at a unique moment in the history of science – scientists from different disciplines are sharing methodologies and tools
- ▶ The Atkins Report urges a “third way” for science – a path that will allow us to make better use of:
 - Intensive numerical computation
 - New types of computer-assisted meta-analysis
 - CI-enabled collaboration undermines barriers of time and space

How do the behavioral and social sciences contribute to the development of cyberinfrastructure?

- ▶ Contributions are significant and many – covered in detail in 2005 report from joint CISE-SBE Airlie House conference
- ▶ Here, we will look at social and behavioral science contributions to cyberinfrastructure with respect to four key NSF investment areas:
 - Adaptive Systems Technology
 - Environment
 - Science of Science Innovation & Policy
 - Cyberinfrastructure

Adaptive Systems Technology

- ▶ Much has been learned about how the brain functions in the past generation due to various imaging tools
- ▶ However, brain science is still in its infancy – new tools and technologies are necessary to help us better understand the anatomy, development, and physiology of the brain. These tools and technologies include:
 - More powerful computationally based imaging devices
 - Tools for gathering coordinated, simultaneous data from different monitoring devices (SBE/CISE/OCI Next-Generation Cybertools award to U. Chicago)
 - High performance computers capable of storing and analyzing massive data sets
- ▶ AST aims to develop new technologies mimicking brain processes.

Environment

- ▶ NSF has recently launched the Coupled Natural and Human Systems program – a vehicle for future investments in environmental matters, including work on climate change
- ▶ GIS – the ability to combine geospatial data with data gathered by social and behavioral scientists has allowed sophisticated research on environmental change, resource inequality, business networks, criminal justice, health and disease
- ▶ Disasters – a recent NSTC report urges integration of climate, environment, and social science data to enable better prevention, preparation, and mitigation
- ▶ Simulations of societies – current simulations are too simplistic to capture social processes in even small groups, so substantial high-speed computing resources are required
- ▶ Observatories enable fine-grained multidimensional recording of natural and human-built assets over time

Science of Science & Innovation Policy

- ▶ Goal of this initiative is to investigate how national R&D systems work, how to measure and nurture innovation, and how to direct our investments
- ▶ One specific area of research is how cyberinfrastructure impacts scientific research and scientific culture. Cyberinfrastructure has:
 - Undermined disciplinary barriers,
 - Increased access to digital data, and
 - Created new mechanisms for sharing computational tools
- ▶ New cyberinfrastructure is needed for:
 - New data extraction
 - New collaboratories

Cyberinfrastructure

- ▶ Behavioral and social scientists study human behavior in many domains, including science, and have much to contribute to developing the infrastructure associated with new computational capacities.
- ▶ Three areas of areas of interest for NSF are:
 - Developing data-oriented cyberinfrastructure while maintaining confidentiality of data
 - Broadening participation
 - Developing a cyber-savvy workforce

Cyberinfrastructure – Data and confidentiality

- ▶ SBE is committed to continue developing and deploying data-oriented cyberinfrastructure, through investments in:
 - Upgrading the existing “gold standard” surveys
 - New data infrastructure projects
 - Facilities and technology for accessing confidential social and behavioral data resources
 - Toolkits for facilitating data integration, mining, analysis, and validation
 - Facilities for preserving data over the long term
- ▶ One example is the Documenting Endangered Languages project – goal is to establish sustainable repositories for many languages faced with extinction and this entails new annotation techniques
- ▶ Many opportunities for building and using data sets require access to confidential micro-data, therefore data confidentiality is a serious and ongoing concern

Cyberinfrastructure – Broadening participation

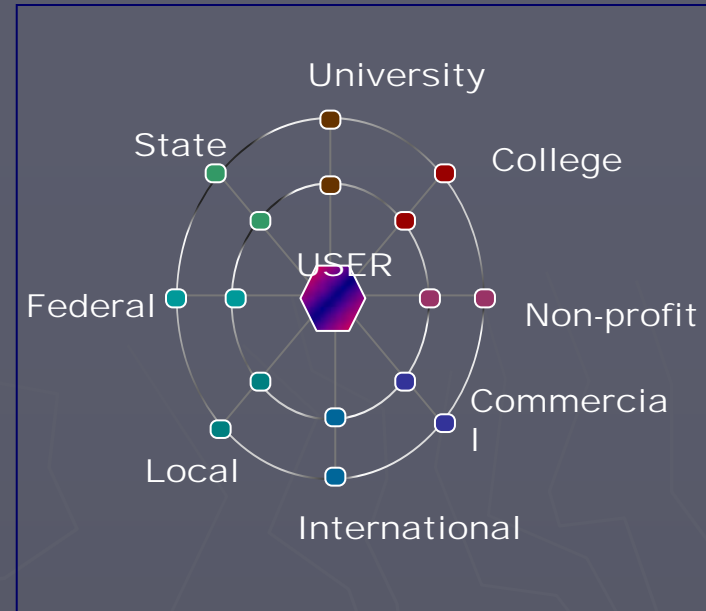
- ▶ Cyberinfrastructure tools have the potential to improve access to data needed for making informed decisions. However, many questions have yet to be answered, including:
 - Do innovative methods for improved accessibility narrow the digital divide?
 - Do participatory practices via the cybersphere enhance democratic processes?
 - Does cyberinfrastructure improve participatory opportunities and experiences?
- ▶ Another Next-Generation Cybertools award is to a team of researchers who are mining 40-billion Web pages to identify and analyze patterns of innovation and diffusion of ideas. The tools will be made available to the research community, but also to individuals and community groups.

Cyberinfrastructure – Cyber-savvy workforce

- ▶ Support for education and training opportunities for the development of the next-generation, cyber-savvy scientific workforce and the re-tooling of existing scientists is critical for advancing these research areas, including activities to broaden workforce participation
- ▶ Learning and workforce development initiatives are crucial to using and extending cyberinfrastructure – one of NSF's Science of Learning Centers is bringing together researchers from several disciplines to improve cyberlearning

DataNet

- ▶ Vision: “Science and engineering digital data are routinely deposited in a well-documented form, are regularly and easily consulted and analyzed by specialists and non-specialists alike, are openly accessible while suitably protected, and are reliably preserved.”
 - NSF Cyberinfrastructure Vision for 21st Century Discovery, Chapter 3
- ▶ Goals:
 - Catalyze development of a system of science & engineering data collections that is open, extensible & evolvable.
 - Support development of a new generation of tools & services facilitating data acquisition, mining, integration, analysis, & visualization.
- ▶ Status: Preliminary proposals under review. Full proposals due March 21. Anticipate 2 awards of ~\$4M/year for 5 years (\$20M each) in FY08. FY09 DataNet competition expected to make 3 awards.



- ✓ User-centric
- ✓ Multi-Sector
- ✓ Sustainable
- ✓ Extensible
- ✓ Evolvable
- ✓ Nimble
- ✓ Reliable



INTEROP

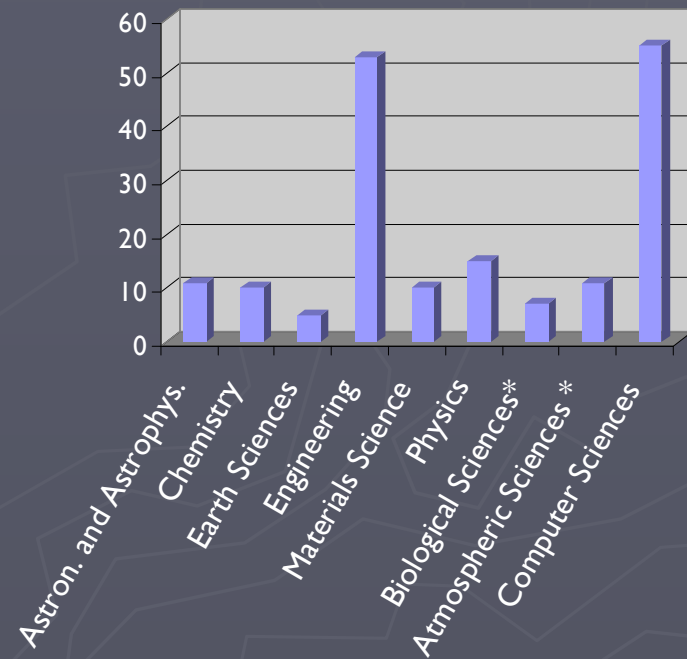
- ▶ Re-purposing data – using it in innovative ways & combinations not envisioned by its creators – requires that it be possible to find & understand data of many types & from many sources.
- ▶ Interoperability – ability of two or more systems or components to exchange information & use information that has been exchanged.
- ▶ Data Interoperability Networks: responsible for consensus-building activities & for providing expertise necessary to turn consensus into technical standards with associated implementation tools & resources.
- ▶ Preference for providing for broad interoperability across a wide variety of disciplinary domains, topic areas, &/or data types & sources.
- ▶ Status: 35 proposals reviewed; 6-8 likely awards pending.

PetaApps

Solicitation sought proposals that

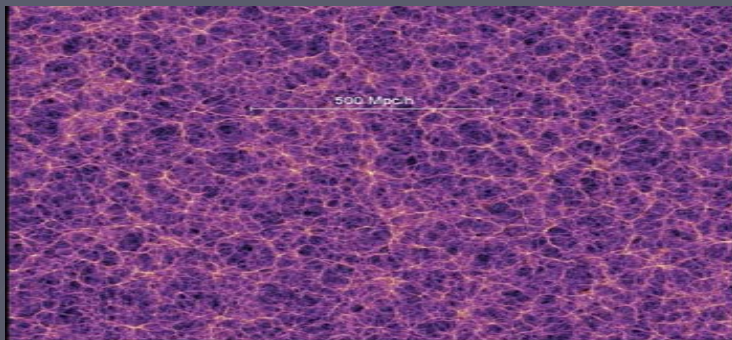
- ▶ develop the future simulation, optimization and analysis tools that can use emerging petascale computing to advance the frontiers of scientific and engineering research;
- ▶ have a high likelihood of enabling future transformative research;
- ▶ 133 distinct project proposals received;
- ▶ 18 awards ~\$26M (50% funding from OCI, 50% from CISE and domain sciences/engineering)

PetaApps Proposal Distribution

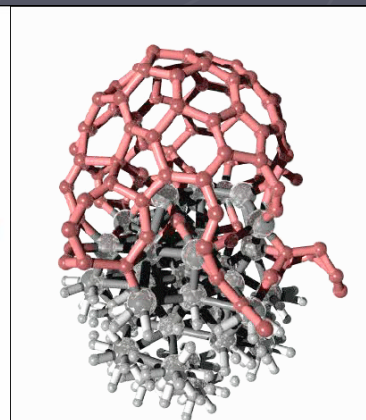
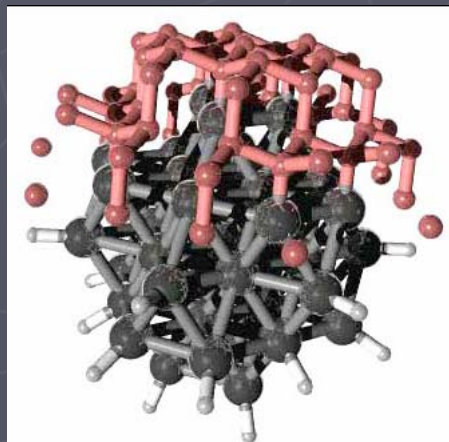
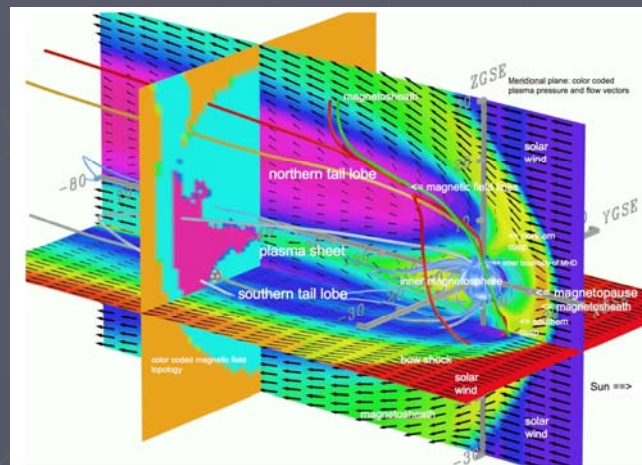


PetaApps Was Very Competitive

Significant Science Applications Funded



- Climate Change
- Earthquake Dynamics and structural response
- Nanoscale Transistor Models
- Supernovae Simulations
- High Reynolds Number Turbulent Flows
- Particulated Flows
- First principle molecular dynamics
- Quantum Chromo Dynamics ...



Strategic Technologies for Cyberinfrastructure

PD 06-723 I Standing program Strategic Technologies for CI – Core OCI
program

support work leading to the development and/or demonstration of
innovative cyberinfrastructure services for science and engineering
research and education that fill gaps left by more targeted funding
opportunities

consider highly innovative cyberinfrastructure education, outreach and
training proposals that lie outside the scope of targeted solicitations --
(*“the crazies”*).

Two dates each year

August and February

~\$2.5M per solicitation

Cyberinfrastructure Teaching, Education and Mentoring

- ▶ The CI-TEAM program supports projects that promote, leverage, and optimize cyberinfrastructure technologies, tools, and services.
- ▶ CI-TEAM awards will:
 - Prepare current and future generations of scientists, engineers, and educators to design and develop as well as adopt and deploy cyber-based tools and environments for research and learning, both formal and informal.
 - Expand and enhance participation in cyberinfrastructure science and engineering activities of diverse groups of people and organizations, with particular emphasis on the inclusion of traditionally underrepresented individuals, institutions, and communities as both creators and users of cyberinfrastructure.

Conclusions

- ▶ Cyberinfrastructure is fundamentally changing the way that scientists build and test theories of social, behavioral and economic phenomena
- ▶ SBE scientists are pushing the CI envelope in order to advance their understanding and their sciences
- ▶ The human sciences study Cyberinfrastructure and its consequences for science and society
- ▶ Science and technology co-exist, interact, and evolve interactively