NSF

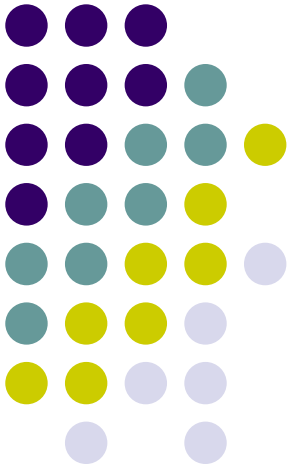
Office of Cyberinfrastructure

(OCI)

Abani K. Patra

Program Director, Office of Cyberinfrastructure

apatra@nsf.gov



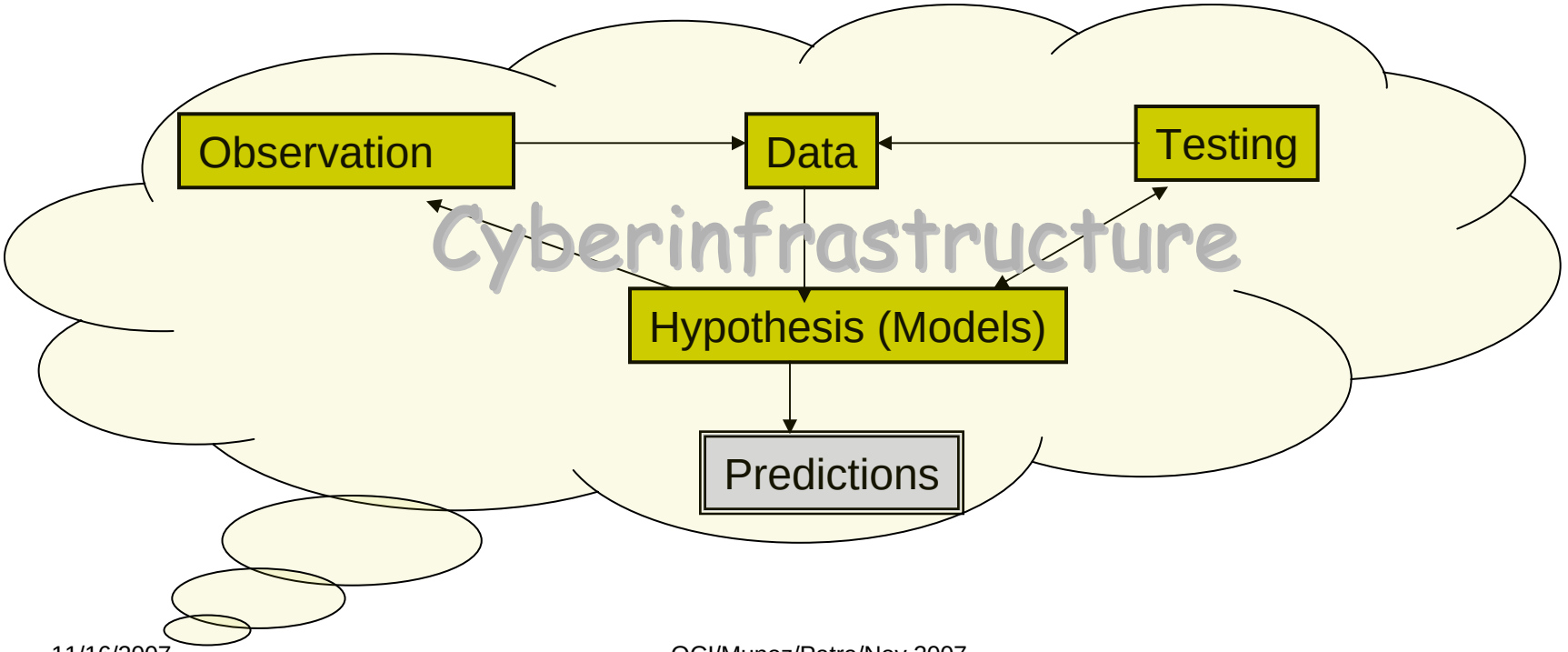
OCI People

- D. Atkins
- J. Munoz
- S. Meacham
- C. Greer
- K. Thompson
- L. Nowell
- D. Rhoten
- A. Patra

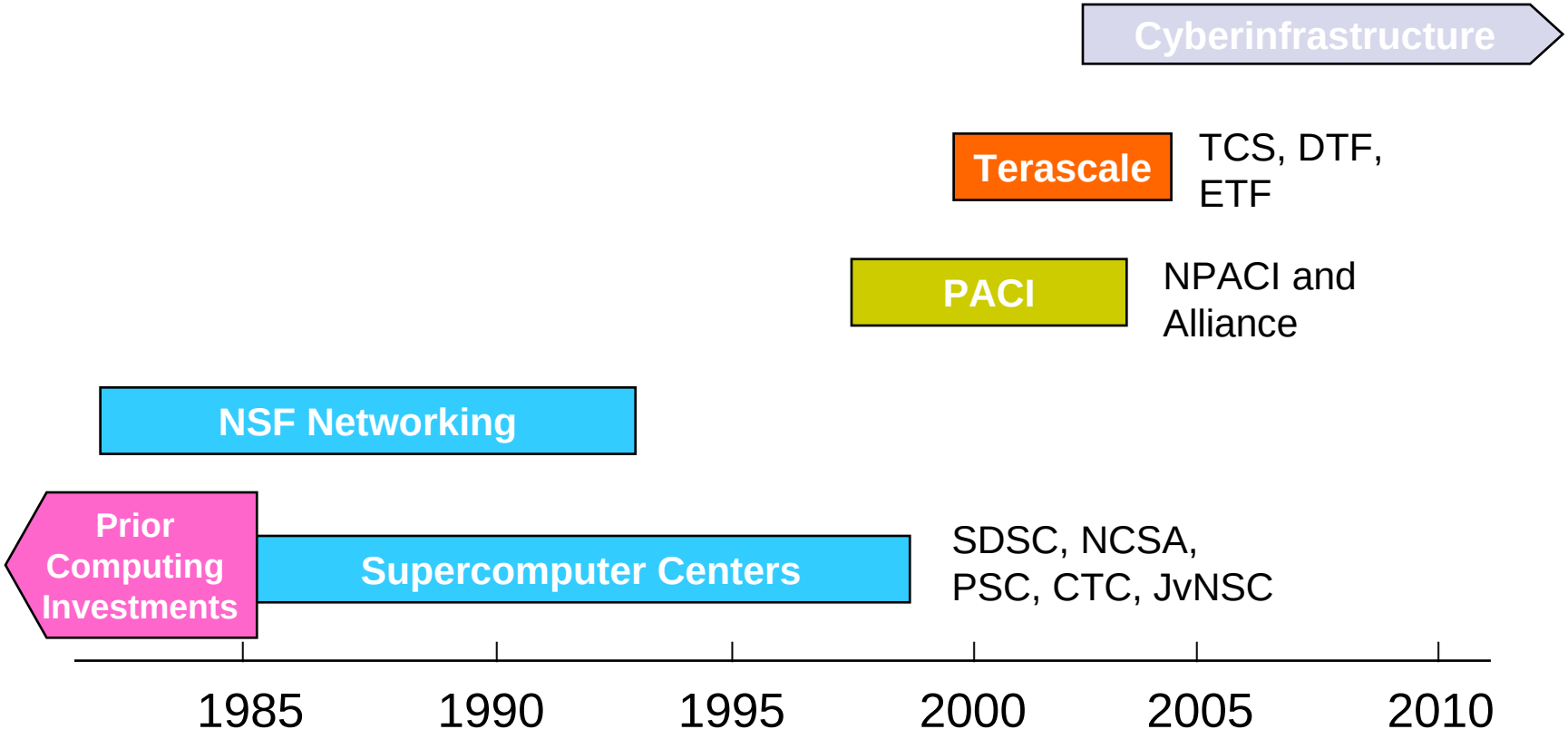
Cyber enabled Science & Eng

Scientific method

- systematic exploration of phenomena



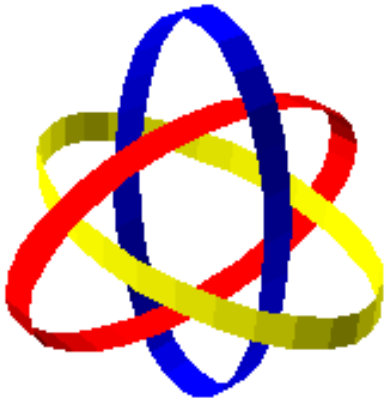
What is CI: The History



Slide from S. Iacono

Achieving the NSF CI Vision requires synergy between 3 types of activities

Transformative Application - to enhance discovery & learning



catalyzes

Provisioning -Creation, deployment and operation of advanced CI

provides shared and connecting CI

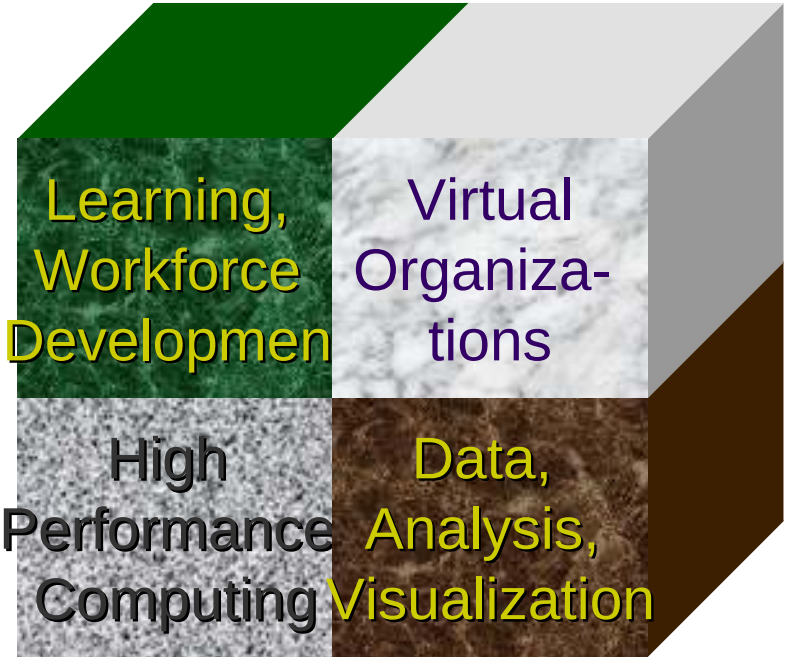
Office of Cyberinfrastructure

Borromean Ring: The three rings taken together are inseparable, but remove any one ring and the other two fall apart. See www.liv.ac.uk/~spmr02/rings/

R&D to enhance technical and *social* effectiveness of future CI environments

Slide from D. Atkins

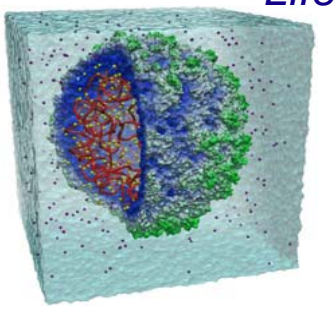
Office of Cyberinfrastructure



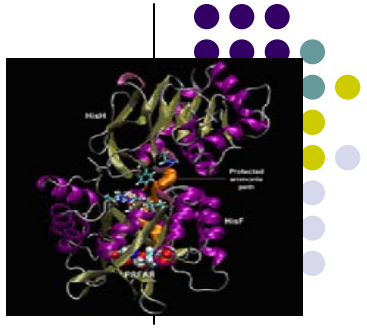
High Performance Computing

increasingly important tool for understanding

Life



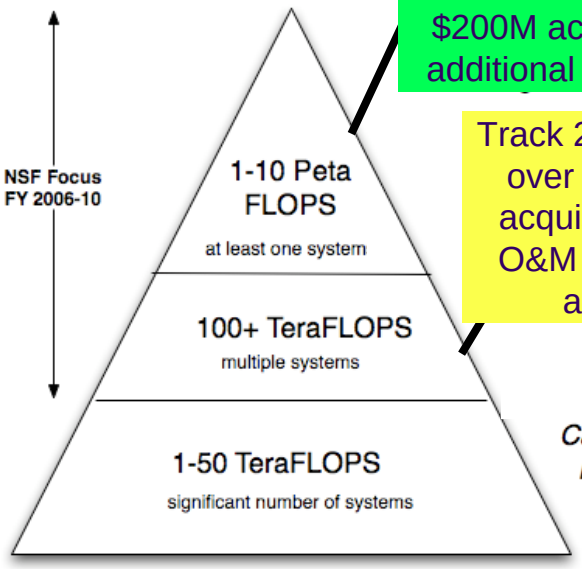
Satellite tobacco mosaic virus, P. Freddolino et al.



Aldehyde dehydrogenase, T. Wymore and S. Brown

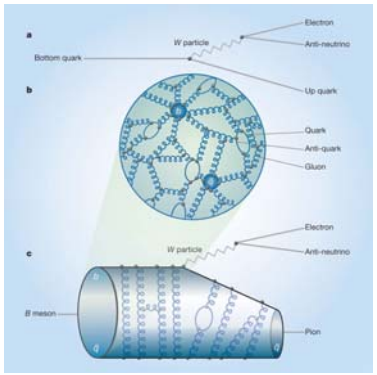
Track 1: One solicitation funded over 4 years: $200M acquisition + additional O&M cost.

Track 2: Four solicitations over 4 years: $30M/yr acquisition + additional O&M cost. First track 1 approved 8-07



Campus Level

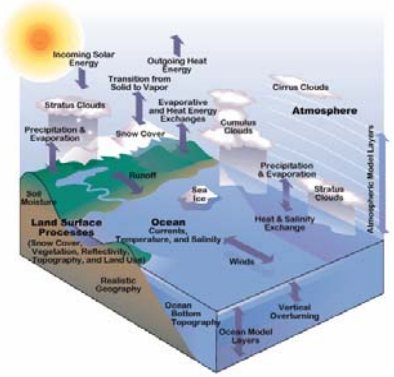
Matter

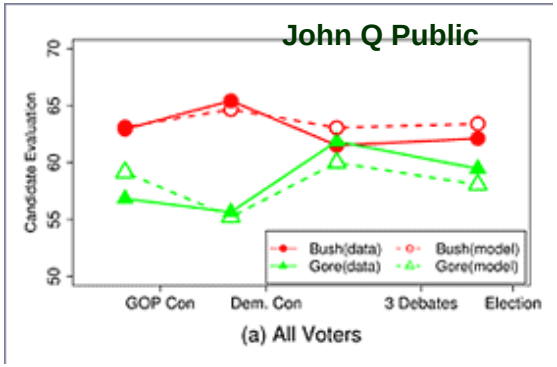


I. Shipsey

Society

The Environment





S.-Y. Kim, M. Lodge, C. Taber.

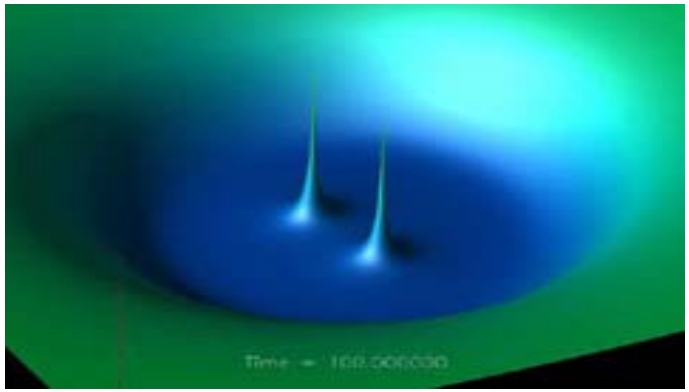
Hardware Acquisitions

NSF building a portfolio of high end systems

Equipment and 4/5 years of operations

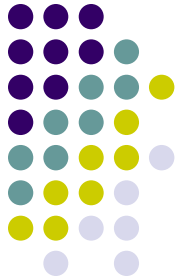
First Acquisition at Texas Advanced Computing Center → over 500 TF, 100 TB memory, 1.7PB disk, using 50K+ cores, Infiniband interconnect;

Proposed Ranger System at TACC



Campanelli et. al. produced the first simulations of generic, highly-spinning black holes with unequal masses by solving Einstein field equations using TeraGrid supercomputers and a specialized AMR technique.

Hardware Acquisitions



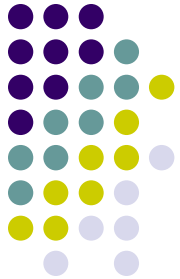
Two further acquisitions in progress

- "Track I" Leadership class machine - capable of sustaining PF/s on a range of problems with commensurate memory and communication frameworks -- awarded to University of Illinois for system named "Blue Waters" to be operational in 2011

- "Track II" – second competition awarded to Univ. of Tennessee JICS -- two more are planned

- HPCOPS

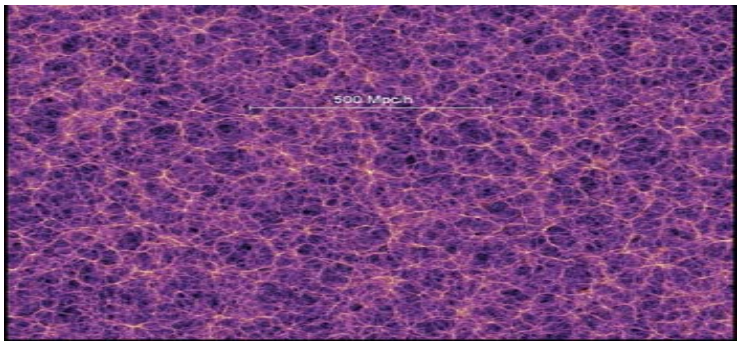
PetaScale Software Development

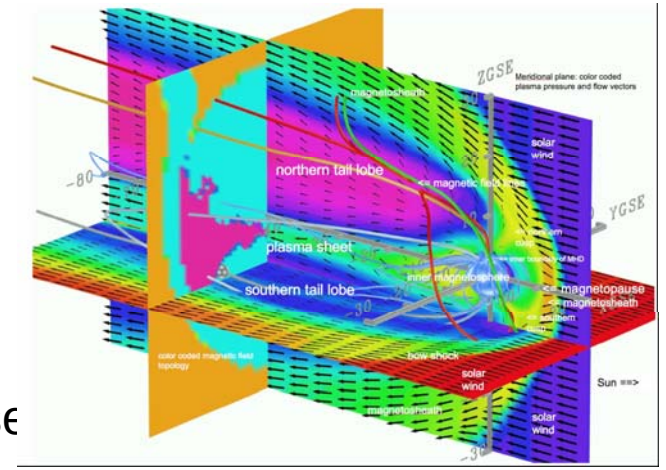


NSF 07-559 "PetaApps" –

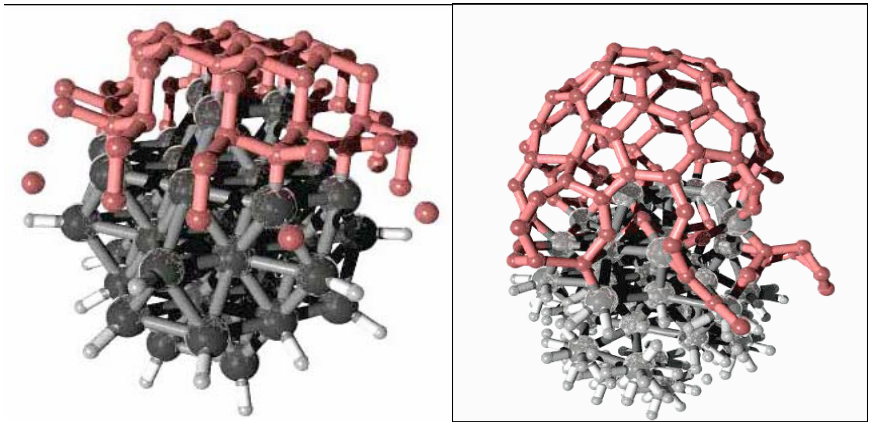
- "to develop the future simulation, optimization and analysis tools that can use petascale computing" … "implementation and exploitation of forefront techniques"
- Cross Foundational program with participation from OCI, MPS, CISE, Engineering and GEO
- $26M in 18 awards each of less than $2M over 3-5 years. Wide range of science/engineering Applications funded
- Subsequent years in discussion currently

Significant Science Applications Funded

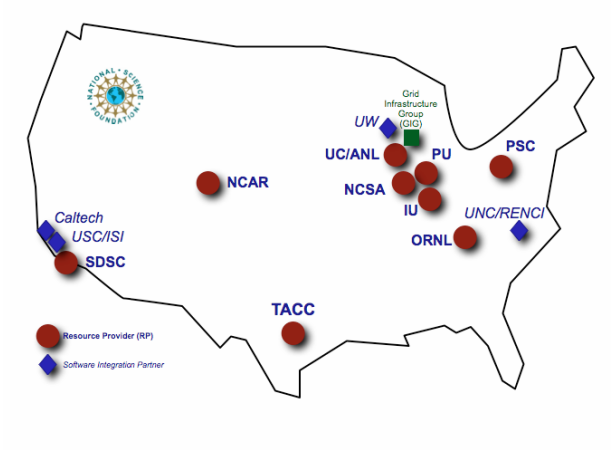


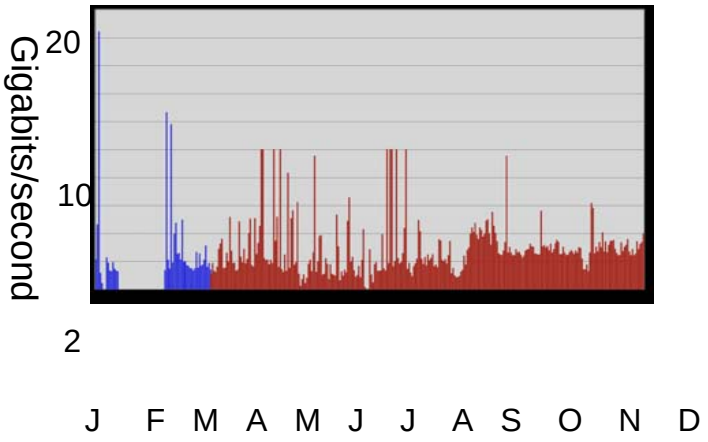


- Climate Change
- Earthquake Dynamics and structural response
- Nanoscale Transistor Models
- Supernovae Simulations
- High Reynolds Number Turbulent Flows
- Particulated Flows
- First principle molecular dynamics
- Quantum Chromo Dynamics …



TeraGrid (TG)





- Nine resource providers (RPs), four software integration partners, and the Grid Infrastructure Group (GIG), LSU to join in FY'08

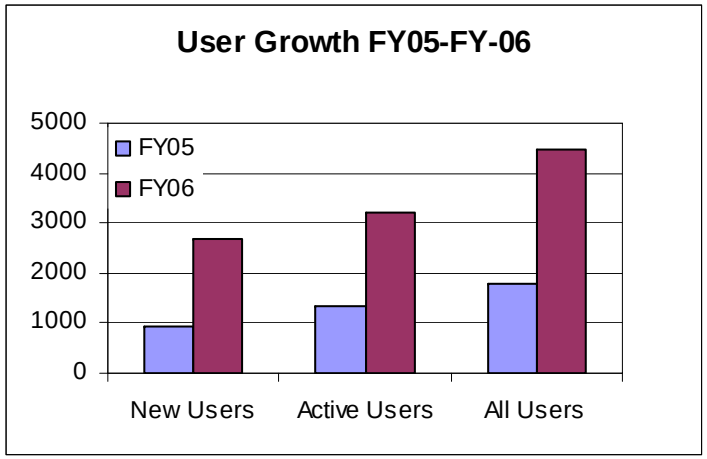
- Common user environments, global file systems, cross-site workflows, coordinated support …

- Peer reviewed merit based common allocation process for Computing, Data and as of this year sustained user support.

- 875* user publications acknowledged TG in CY2006 up from 485 in CY2005

* Publication counts are probably underreported –

Usage and Science Impact



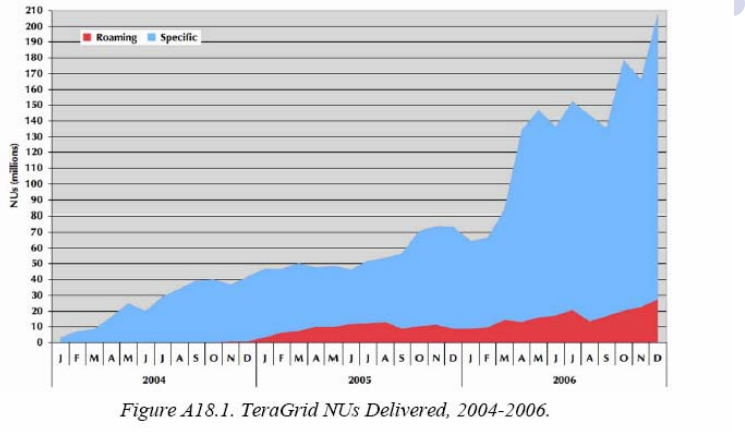
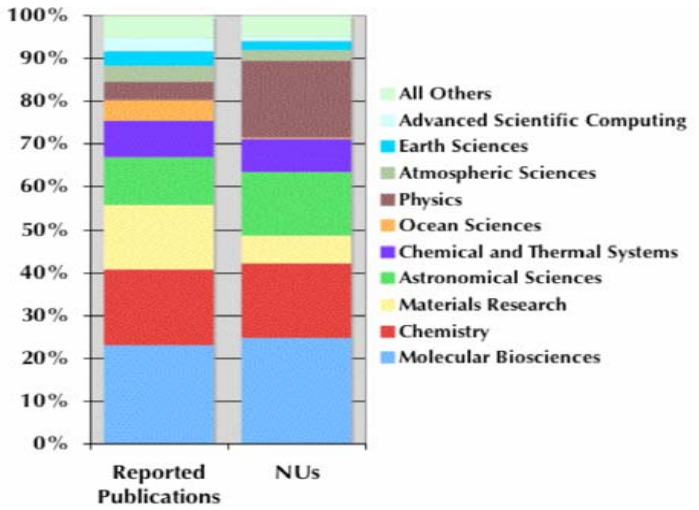


Figure A18.1. TeraGrid NUs Delivered, 2004-2006.

Users are from 265 institutions and 47 states.

Table A18.2. FY06 Total NUs Available and Delivered

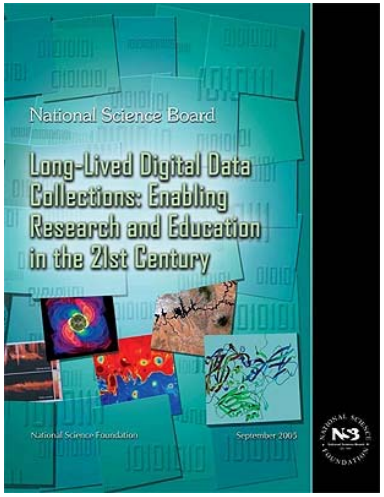
	NUs	%
NUs Available	2.23 billion	
NUs Requested	2.96 billion	132.7
NUs Allocated	1.917 billion	86.0
NUs Delivered	1.281 billion	57.4
Usage by Category (% of NUs Delivered)		
Academic (allocated)	1.257 billion	98.1
Academic (discretionary)	13.5 million	1.1
TeraGrid Staff	10.2 million	0.8



[1] 1 NU is the computational equivalent of 1 CPU-hour on the Cray X-MP.

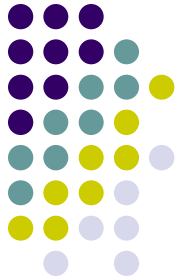
Data, Data Analysis & Visualization

Digital Universe Created 161 Exabytes Of Data Last Year -- expected to grow to 988 EB in 2010



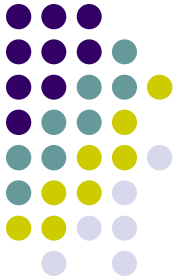
- Challenges: increased scale, heterogeneity, and re-use value of digital scientific information and data. Inadequate digital preservation strategy of long-lived data.
- Taking initial steps to catalyze the development of a federated, global system of science and engineering data collections that is open, extensible, evolvable, (and appropriately curated and long-lived.)
- Complemented by a new generation of tools and services to facilitate data mining, integration, analysis, visualization essential to transforming data into knowledge.

Community Based Data Interoperability Networks

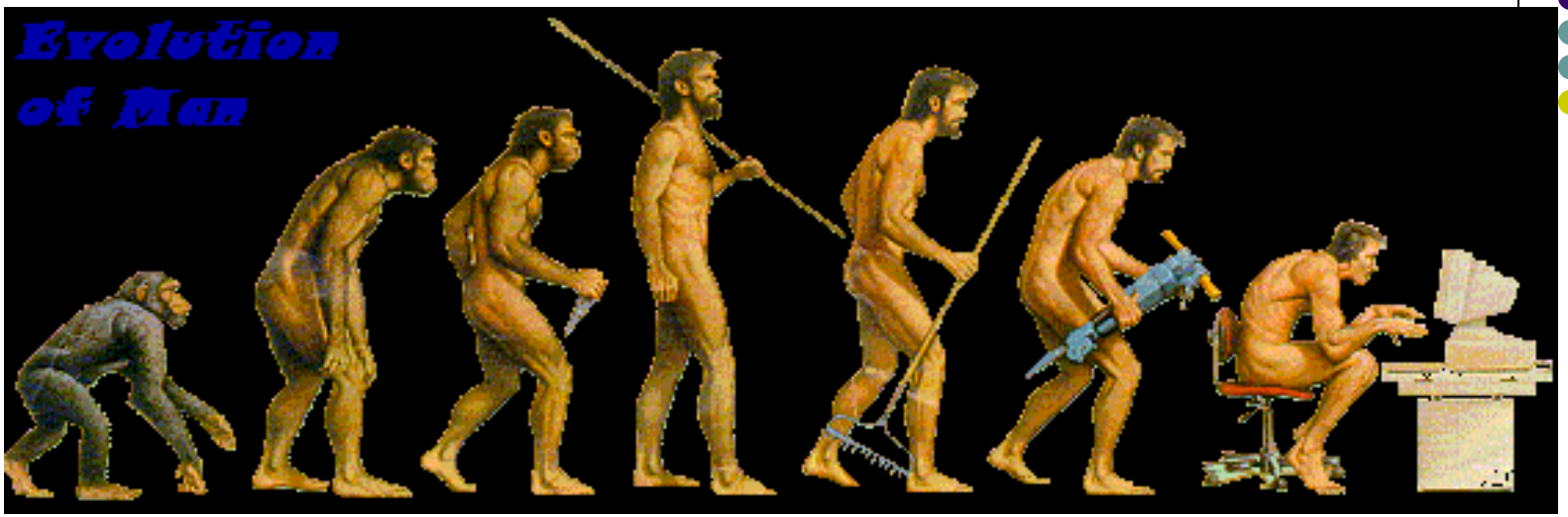


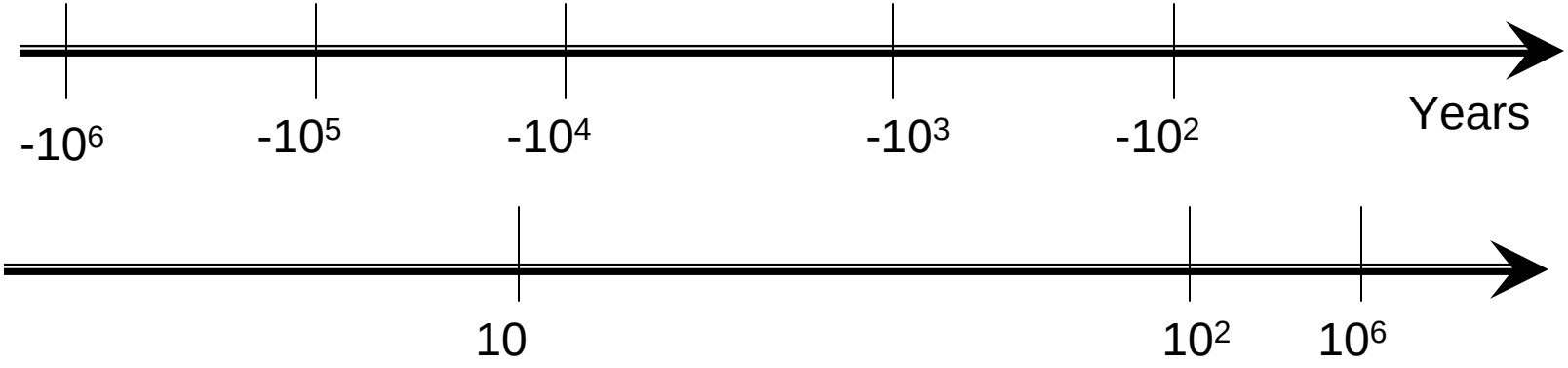
- **NSF 07-565 “INTEROP”**
  - Several NSF Directorates participating
- Support community efforts to provide for broad interoperability through the development of mechanisms such as robust data and metadata conventions, ontologies, and taxonomies
- Each project shall have **two goals**:
  - Develop community consensus (e.g. workshops, task groups, community websites, etc.)
  - Turn consensus into technical standards with implementation tools (e.g. ontologies, taxonomies, software tools, web resources, etc.)
- Approximately 10 $250K/yr. awards (3-5 yrs.)
- August 2007

Cyberinfrastructure for Environmental Observatories



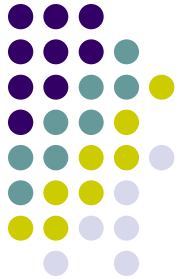
- Observatory Initiatives
  - WATERS …
- CEO:P Program
  - OCI, OCE, ENG, BIO
  - ...help insure that the information infrastructure technologies needed to support the widespread use, for cutting-edge research, of large environmental observing systems are available [...
  - 34 proposed projects, 8.5 M, 5 Awards
  - Research Questions include: Forecasting coastal conditions and erosion, Seasonal hypoxia in coastal waters, Impacts of novel climate conditions on ecosystems





Size of personal network / power of computing and communication

CI enabled Science & Eng



- Collaborative efforts are bedevilled by a host of social, cultural and technological issues …
  - CI must also enable the power of many – multidisciplinary science and the loose dynamic federations required to do this
- "Discover, codify, disseminate" – classical paradigm for science comfortably done by simple rigidly structured organizations.
- "Engage, Explore, Apply, Share"– new paradigm for scientific efforts.

→ Community Driven Science Gateways, …

Virtual Organizations

- To catalyze the development, implementation and evolution of a national cyberinfrastructure that integrates both physical and cyberinfrastructure assets and services.
- To promote and support the establishment of world-class VOs that are secure, efficient, reliable, accessible, usable, pervasive, persistent and interoperable, and that are able to exploit the full range of research and education tools available at any given time
- To support the development of common cyberinfrastructure resources, services, and tools that enable the effective, efficient creation and operation of end-to-end cyberinfrastructure systems for and across all science and engineering fields, nationally and internationally.

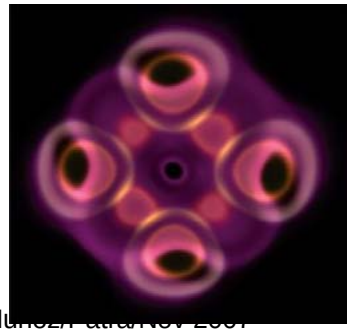
NVO

LEAD

iVDgL

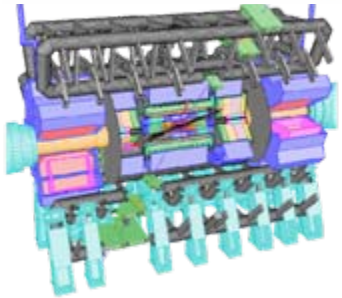
TeraGrid

NEES



NanoHub

CMS



ATLAS

Open Science Grid

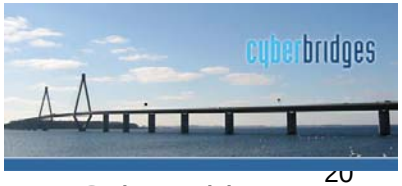
Learning & Workforce Development

- Learning **supported by** CI. (cyber-enabled learning).
- Workforce development **to create and use** CI for S&E research and education.

Broadened participation: Exploit the new opportunities that cyberinfrastructure brings for … people who, have been excluded from the frontiers of scientific and engineering research and education.

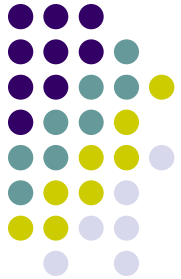
- Explore CI support for **integrated research and education**.
- Effective, Transferable, Sustainable, Scalable

EPIC



BIOINFORMATICS CI INSTITUTE MARIACHI CyberBridges

Software Development for CyberInfrastructure



- **NSF 07-503 (closed), FY'08 deadline in April**
- Develop, deploy and sustain a set of reusable and expandable software components and systems that benefit a broad set of science and engineering applications
  - **software activities for** enhancing scientific productivity and for facilitating research and **education collaborations** through sharing of data, instruments, and computing and storage resources. The program requires open source software development
- Three focus areas in '07: HPC, Middleware, Digital Data
- Pending funding will be re-issued in 2008

Strategic Technologies for Cyberinfrastructure

PD 06-7231 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").*

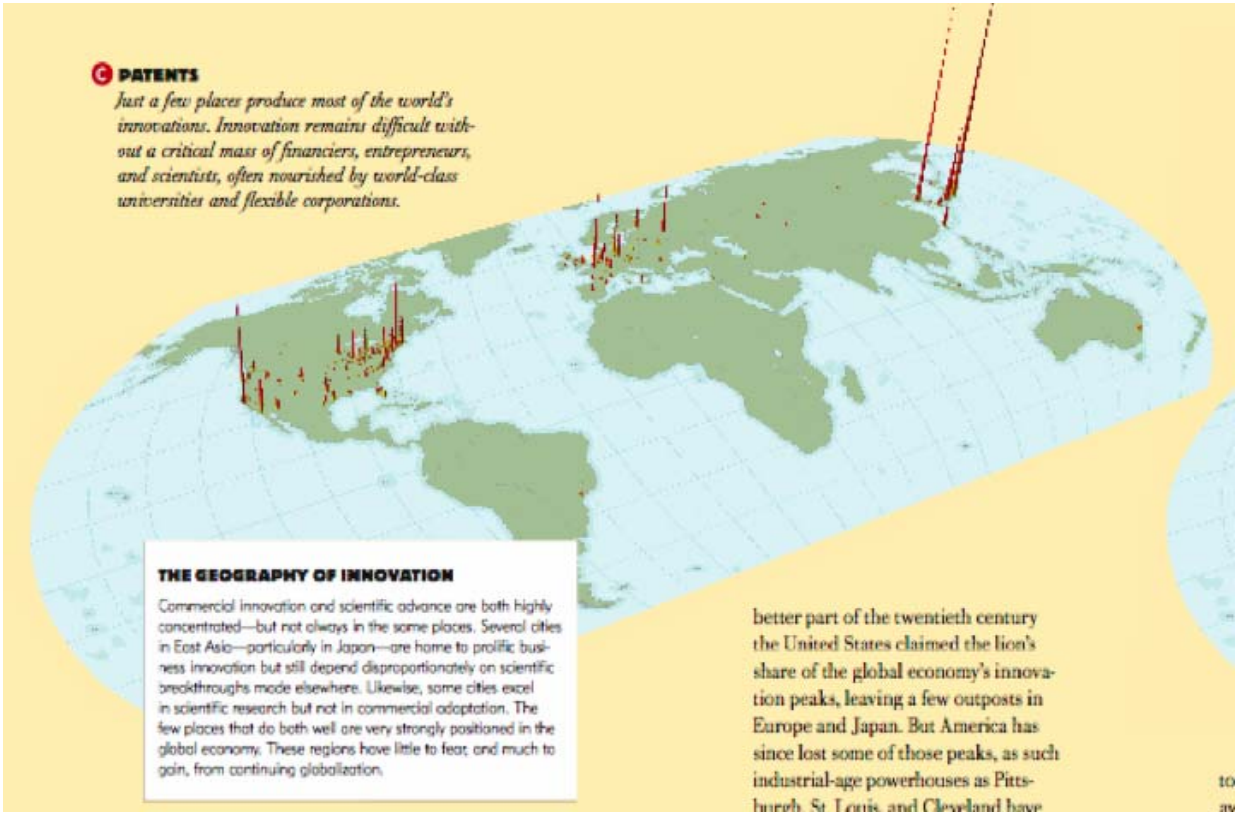
Two dates each year

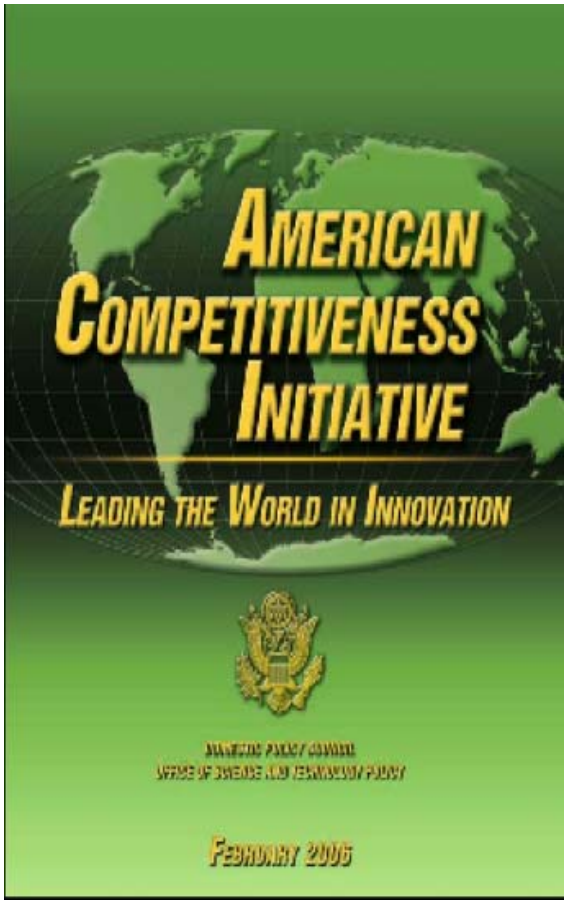
August and February

$2.5M

Data, Models, Predictions

- Friedman's Flat World vs Florida's Spiky World
  - "Industrial production" is increasingly global
  - Scientific Discovery and Innovation remain geographically local

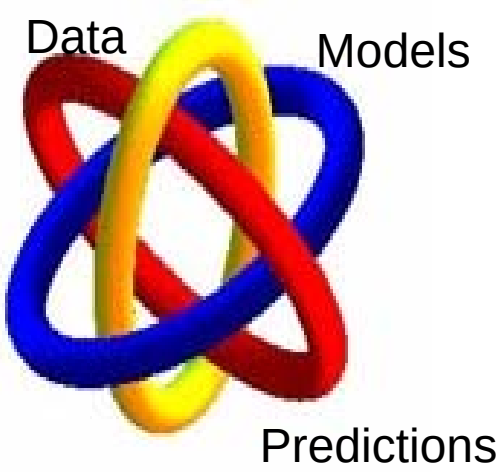


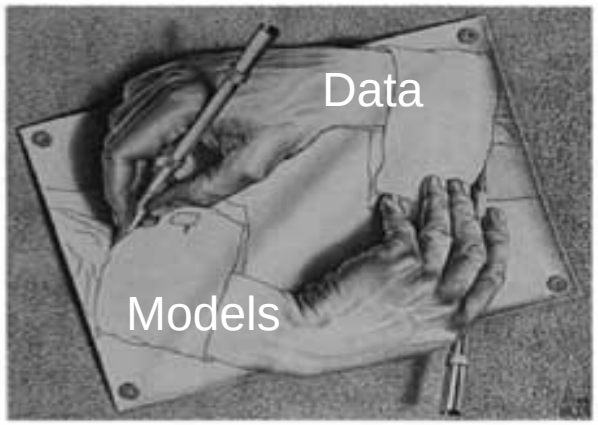


Data, Models, Predictions

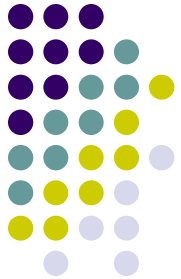
- Predictive simulations integrate the best models and **all available data** in a rigorous framework that enables meaningful extrapolation.
  - *Shark behavior in gulf predicted Katrina effects well*!

→ PREDICTIVE SCIENCE requires that all DATA and MODELS be seamlessly accessible with minimal barriers – technological and intellectual.





Barriers

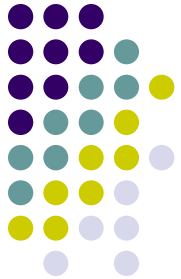


- *"But HPC and other high end CI tools are still only accessible to the few computational scientists who can master a domain science, program parallel, distributed algorithms, and use/manage a supercomputer[1]."* – R. Waite on "Why MicroSoft has gotten into HPC?"

- High Performance Computing, simple access to ALL RELATED data from observation, experiment and earlier simulations and good visualization must become **routine** – not heroic efforts!

[1]*whose architecture and programming change every 18 months!*

Cyber-Enabled Discovery and Innovation (CDI)

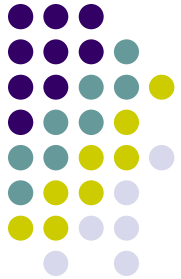


Objective of CDI:

Enhance American competitiveness by enabling innovation through the use of computational thinking.

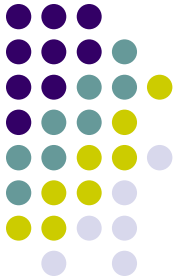
- Multi-disciplinary research seeking contributions to more than one area of science or engineering, by innovation in, or innovative use of computational thinking
- Computational thinking refers to computational…
  - …Concepts, Methods, Models, Algorithms & Tools

CDI is Unique within NSF



- five-year initiative;
- minimum of $26M in FY 2008 in a common solicitation with up to an additional $26M available to the divisions
- Budgets projected to increase each year to a total investment of $750 M
- to create *revolutionary* science and engineering research outcomes made possible by innovations and advances in computational thinking
- emphasis on bold, multidisciplinary activities

Transformative Research



- NEW in NSF Review Criteria:
  - To what extent does the proposed activity suggest and explore creative, original, or potentially transformative concepts?
- Additional Review Criteria in solicitation
  - Special emphasis will be placed on proposals that promise to enhance competitiveness, innovation, or safety and security in the United States.
- "Science/Innovation First"
- High Risk and High Promised Impact
- Peer Review process is conservative

Three CDI Themes

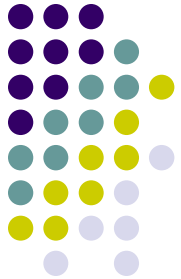
CDI seeks transformative research in the following general themes, via innovations in, and/or innovative use of, computational thinking:

- **From Data to Knowledge:** *enhancing human cognition and generating new knowledge from a wealth of heterogeneous digital data;*

- **Understanding Complexity in Natural, Built, and Social Systems:** *deriving fundamental insights on systems comprising multiple interacting elements;* and

- **Building Virtual Organizations:** *enhancing discovery and innovation by bringing people and resources together across institutional, geographical and cultural boundaries*.

Types of Projects

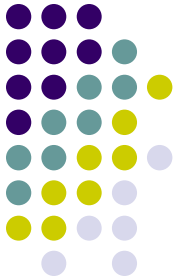


- Project size not measured by $$ but by magnitude of effort
- Three types are defined: Types I (~2 PI, 2 GRA), II (~3 PI, 3 GRA, 1 post-doc), and III (center scale).
- Type III, center-scale efforts, will not be supported in the first year of CDI

Key Dates and URL:

- Letters of Intent (required) due: Nov 30, 07
- Preliminary Proposals due: Jan 8, 08
- Full proposals due: April 29, 08
  - Full proposals by invitation only!
- Awards: no later than October 2008

- For more information:
  - Solicitation: http://www.nsf.gov/pubs/2007/nsf07603/nsf07603.htm
  - FAQ, examples, resources: http://www.nsf.gov/crssprgm/cdi .



Questions?