Cyberinfrastructure Framework for 21st Century Science & Engineering (CF21)

NSF-wide Cyberinfrastructure Vision

People, Sustainability, Innovation, Integration

Alan Blatecky
Acting Deputy Director
Office of Cyberinfrastructure

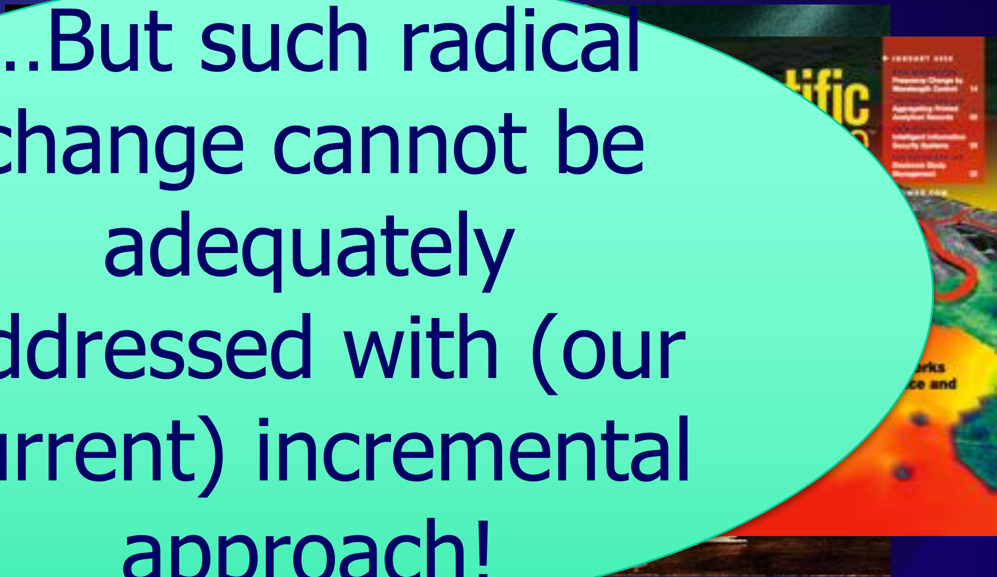
Framing the Question

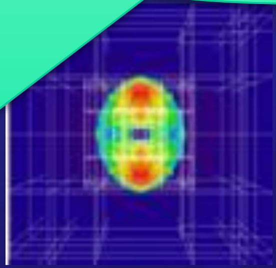
Science is Radically Revolutionized by CI

- Modern science
  - Data- and compute-intensive
  - Integrative
- Multiscale Colla
  for Com
  - I
- Mu
  appr
  - Integr ve, multiscal
  - 4 centuries of const ncy, 4 decades 10^{9-12} change!

We still think like this…

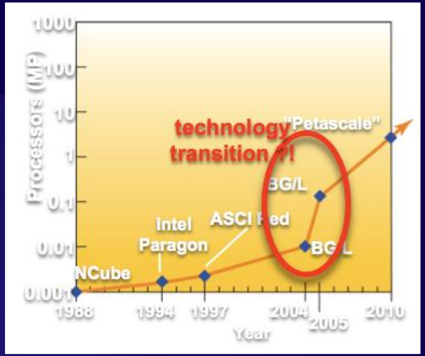
…But such radical change cannot be adequately addressed with (our current) incremental approach!

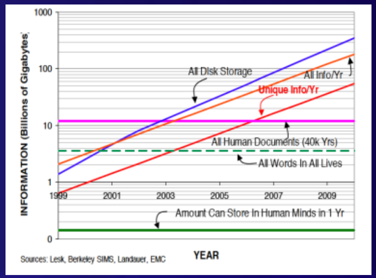




Five Crises

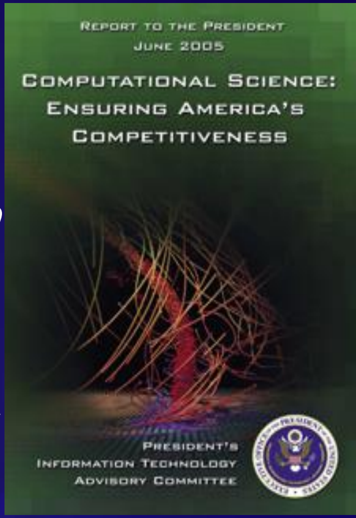
- Computing Technology
  - Multicore: processor is new transistor
  - Programming model, fault tolerance, etc
  - New models: clouds, grids, GPUs,… where appropriate
- Data, provenance, and viz
  - Generating more data than in all of human history: preserve, mine, share?
  - How do we create "data scientists"?
- Software
  - Complex applications on coupled compute-data-networked environments, tools needed
  - Modern apps: 10^6+ lines, many groups contribute, take decades





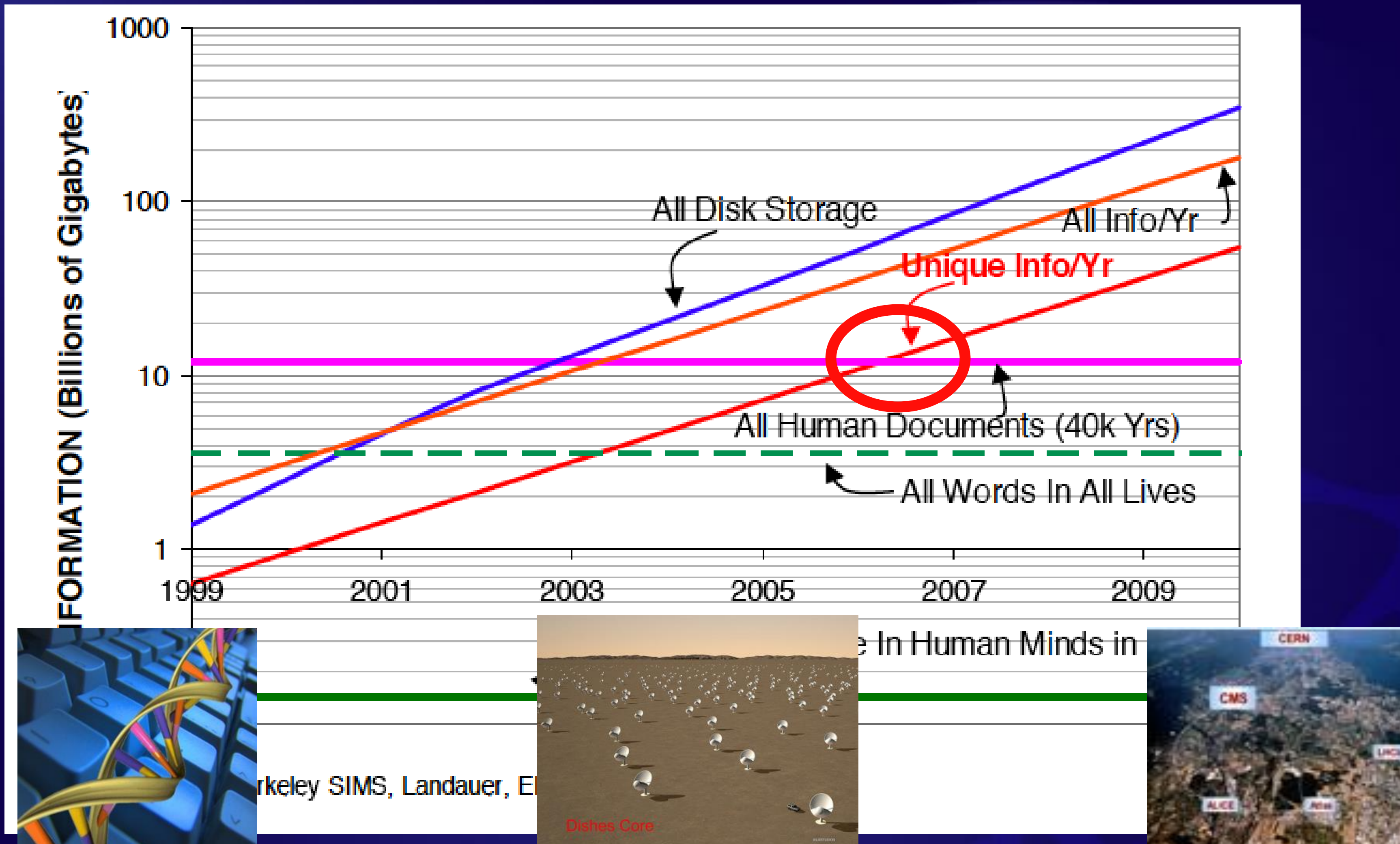
Five Crises con't

- Organization for Multidisciplinary Computational Science
  - "*Universities must significantly change organizational* structures*: multidisciplinary & collaborative research are needed [for US] to remain competitive in global science*"
  - *"Itself a discipline, computational science advances all science…inadequate/outmoded structures within Federal government and the academy do not effectively support this critical multidisciplinary field"*



- Education
  - The CI environment is becoming more complex and more fundamental for research
  - How do we develop a workforce to provide leadership, expertise and support ?

Aside on Data



Some observations

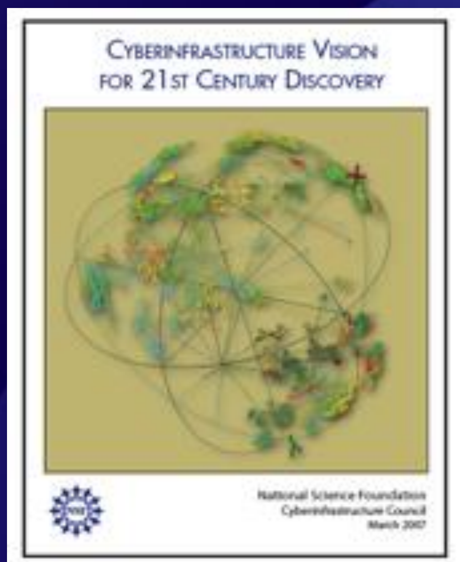
- Science and Scholarship are team sports
  - Competitiveness and success will come to those who can put together the best team, and can marshal the best resources and capabilities
- Collaboration/partnerships will change significantly
  - Growth of dynamic coalitions and virtual organizations
  - International collaboration will become even more important
- Ownership of data plus low cost fuels growth and number of data systems
  - Growth in both distributed systems and local systems
  - More people want to access more data
  - Federation and interoperability become more important

More observations

- More discoveries will arise from search approaches
  - Mining vast amounts of new and disparate data
  - Collaboration and sharing of information
- Mobility and personal control will continue to drive innovation and business
- Gaming, virtual worlds, social networks will continue to transform the way we do science, research, education and business
- The Internet has collapsed six degrees of separation and is creating a world with two or three degrees.

What is Needed?

An ecosystem, not components…



NSF-wide CI *Framework* for 21st Century Science & Engineering

People, Sustainability, Innovation, Integration

Cyberinfrastructure Ecosystem

Maintainability, sustainability, and extensibility

CF21

- A goal of Virtual Proximity – as though you are one with your resources
  - Continue to collapse the barrier of distance and remove geographic location as an issue
  - ALL resources (including people) are virtually present, accessible and secure
    - Instruments, HPC, Vis, Data, Software, Expertise, VOs, etc

End-to-End Integrated Cyberinfrastructure

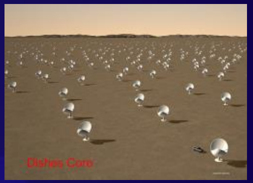
Science, discovery, throughput and usefulness become the metrics

Driving Forces

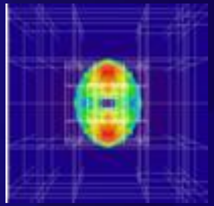
- Need to support the efficient pursuit of S&E
  - Multi-domain, multi-disciplinary, multi-location
  - Leading edge CI network capabilities
  - Seamless integration
- Need to connect Researcher to Resource
  - Access to major scientific resources and instruments
  - CI resource availability – at speed and in real-time
    - (HPC, MREFC, Data Center, Vis center, Clouds, etc)
  - Campus environment including intra-campus
  - State, regional, national and international network and infrastructure transparency

CF21: Cyberinfrastructure Framework…

- **High-end computation, data, visualization, networks** for transformative science; *sustainability, extensibility*
  - Facilities/centers as *hubs of innovation*



- **MREFCs and collaborations** including large-scale NSF collaborative facilities, international partners



- **Software, tools, science applications, and VOs** critical to science, integrally connected to instruments

- **Campuses** fundamentally linked end-to-end; clouds, loosely coupled campus services, policy to support

- **People**. Comprehensive approach workforce development for 21st century science and engineering

CF21 and HPC

- Critical metric is sustained APPLICATION performance
  - Tera/Peta/Exascale applications
  - Lead and follow the science and engineering researchers
  - Partnerships with other US and international agencies
- How to address sustainability, user requirements
  - Balance and integration of:
    - Leadership, production and experimental systems
    - Different types or classes of resources (compute, data; national, local)
    - Alternate models of computing:
      - Clouds, grids, commercial providers? Pay per service?
- How to build on existing and planned CI?
  - Development and integration of advanced services
  - Connect out to MREFCs, across to campus and regional resources
- Developing the computational science program at scale

HPC example

UIUC Petascale Facility

- Leadership: Petascale-to-Exascale
  - Sustainable facilities
  - Multi-agency cooperation discussions
  - Long term
- Advanced capabilities
  - Sustainable hubs of Excellence/Innovation; a focus on people and expertise
  - Sustainable data and compute-intensive centers (major machines every 4 years), broader array of services
  - Medium term
- Experimental
  - 1-2 awards every year to explore new architectures, coupled with application/software dev

Networking Infrastructure Example

- Major Scientific Facility Interconnects
  - Networking infrastructure focus
- High Performance End-User Access
  - Address at-speed connection at desktop
  - Usefulness and User throughput
  - Pilots and prototypes
- Experimental Research Networks
  - Multi layer, hybrid networks including cybersecurity
  - Applications with end-to-end focus
- Digital Divide issues
  - Geographically remote, rural areas, community colleges, etc
  - On campus, off campus

The Shift Towards Data
Implications

- All science is becoming data-dominated
  - Experiment, computation, theory
- Totally new methodologies
  - Algorithms, mathematics
  - All disciplines from science and engineering to arts and humanities
- End-to-end networking becomes critical part of CI ecosystem
  - Campuses, please note!
- How do we train "data-intensive" scientists?
- Data policy becomes critical!

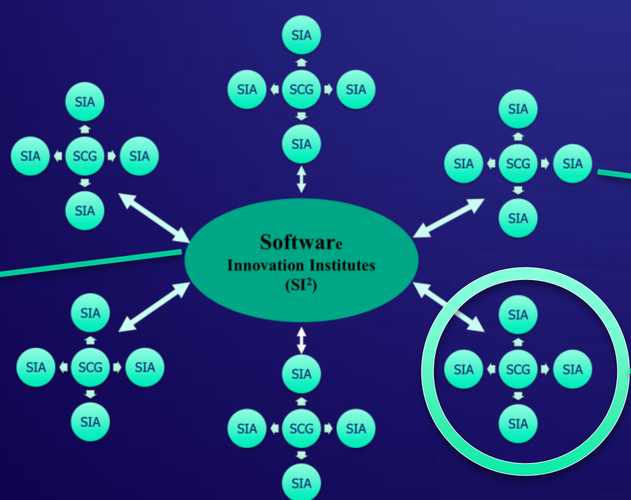
Data Example

- Provide reliable digital preservation, access, integration, and analysis capabilities for science/engineering data over decades-long timeline
- Achieve long-term preservation and access capability in an environment of rapid technology advances
- Create systems and services that are economically and technologically sustainable
- Empower science-driven information integration capability on the foundation of a reliable data preservation network

Software Institutes Example

- Transform innovation into sustainable software
- Significant multiscale, long-term program
  - Perhaps $200-300M over a decade
    - $10M identified in FY10
    - $14M+ annual in OCI in future years
  - Connected institutes, teams, investigators
  - Integrated into CF21 framework with Directorates

3-6 centers, 5+5 years, for critical mass, sustainability



Many individuals w/short term grants, funded by OCI and directorates

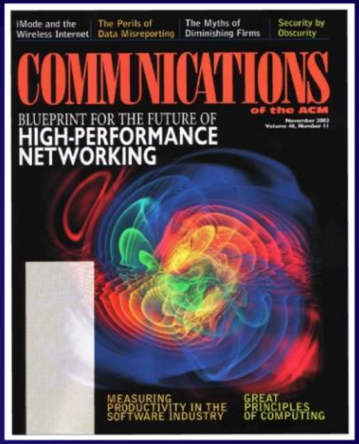
Numerous teams of scientists and computational and computer scientists with longer term grants

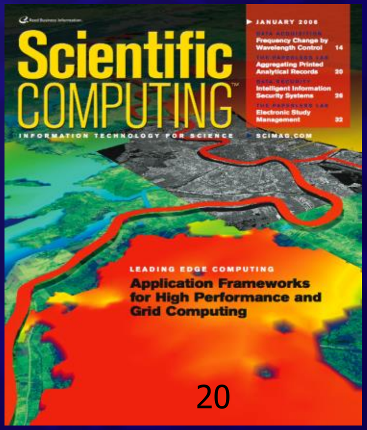
Virtual Organizations Example

- What constitutes effective virtual organizations? How do they enhance research and education; what about production and innovation?
- Multi–disciplinary
  - Anthropology, complexity sciences, CS, decision and management sciences, economics, engineering, organization theory, organizational behavior, social and industrial psychology, public administration, sociology
- Broad variety of qualitative and quantitative methods
  - Ethnographies, surveys, simulation studies, experiments, comparative case studies, network analyses.
- Grounded in theory, rooted in empirical methods

Critical Factors

- Science and society profoundly changing
- Comprehensive approach to CI needed to address complex problems of 21st century
  - All elements must be addressed, not just a few
  - Many exponentials: data, compute, collaborate
- Data-intensive science increasingly dominant
  - Modern data-driven CI presents numerous crises, opportunities
- Academia and Agencies must addressed
  - New organizational structures, rebalanced investments, educational programs, policy
- End-to-end; researcher to resources





ACCI Task Forces

Campus Bridging: Craig Stewart, IU (BIO)

Data & Viz: Shenda Baker, Harvey Mudd (MPS); Tony Hey, (CISE)

- Completion by end of year
- Advising NSF
- Conducting Workshop(s)
- Recommendations
- Input to NSF informs CF21 programs, 2011-12 CI Vision Plan

Software: David Keyes, Columbia/KAUST (MPS)

Computing: Thomas Zacharia, ORNL/UTK (DOE)

Education & Workforce: Alex Ramirez, CEOSE

Grand Challenge Communities/VOs: Tinsley Oden, Austin (ENG)

21

CF21 DRAFT Plan

- Incorporate and integrate existing ACCI Task Forces Recommendations and input
- CF21 Colloquium (C^2) – in process
- Plan to establish CF21 group at NSF
- Development of a CI requirements template
- Develop a CF21 budget building exercise for FY12
- Complete draft by 1st Quarter, 2011

CI requirements Template (Draft)

- CI requirements to meet science and engineering challenges over the next 10 years
  - Scientific instruments, facilities, MREFCs, etc
  - Data, collections, libraries, access, curation, preservation, etc
  - Software, application, middleware, etc
  - Compute resources, data centers, clouds, grids, etc
  - Networking, end-to-end, cybersecurity, campus, national, international, etc
  - Organizations, labs, universities, VOs, communities, etc.
  - Expertise, education, workforce development, collaboration

Challenges

- Foundational substrate for research and education
- Involves user interacting with CI capabilities
  - Data, software, visualization, HPC, clouds, organizations, etc
- Involves entire CI ‘stack’ – transparency of desktop to world capabilities
- Throughput and usefulness to the end-user is the metric
- New “world view” and culture for research infrastructure