Cyberinfrastructure for 21st Century Science & Engineering

EPSCoR Meeting
May 16, 2011

Alan Blatecky
Director OCI

1

Framing the Challenge:
Science and Society Transformed by Data

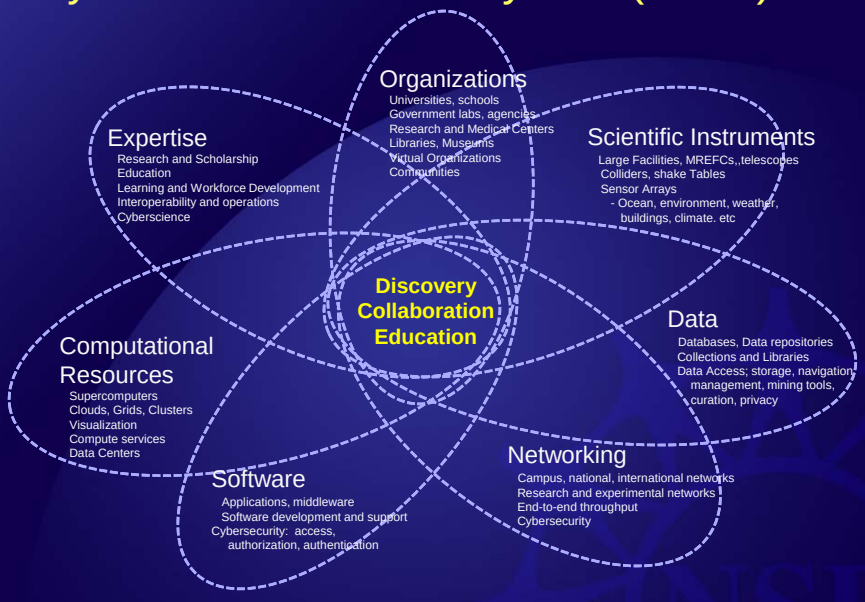
- Modern science
  - Data- and compute-intensive
  - Integrative, multiscale
- Multi-disciplinary Collaborations for Complexity
  - Individuals, groups, teams, communities
- Sea of Data
  - Age of Observation
  - Distributed, central repositories, sensor-driven, diverse, etc

ACCI Task Force Reports

- Final recommendations presented to the NSF Advisory Committee on Cyberinfrastructure Dec 2010
- More than 25 workshops and Birds of a Feather sessions and more than 1300 people involved
- Final reports

  http://www.nsf.gov/od/oci/taskforces/

Cyberinfrastructure Ecosystem (CIF21)



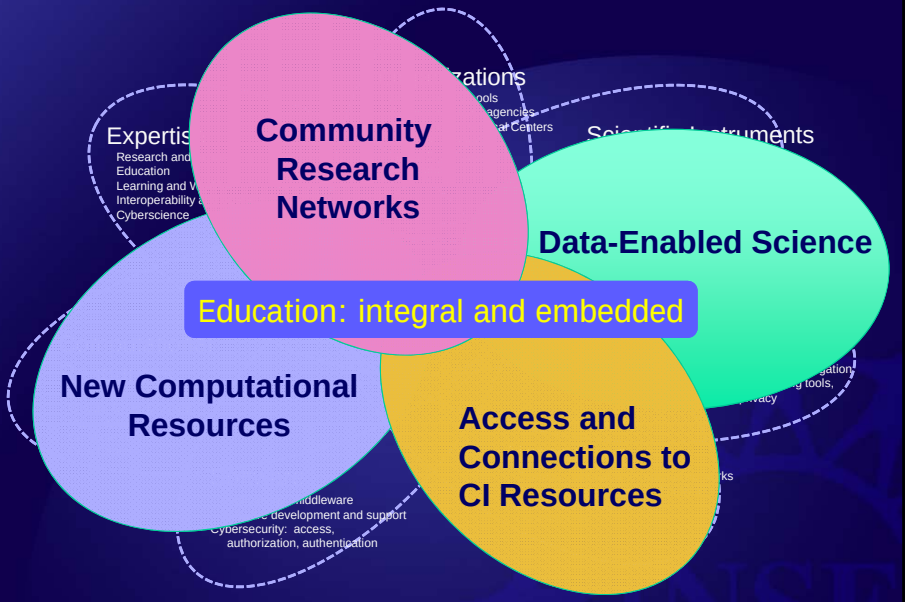
Maintainability, sustainability, and extensibility

CIF21 – a metaphor

- A goal of Virtual Proximity –--
  " 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
  - End-to-end integrated resources
  - Science, simulation, discovery, innovation, education are the metrics

An organizing fabric and foundation for science, engineering and education

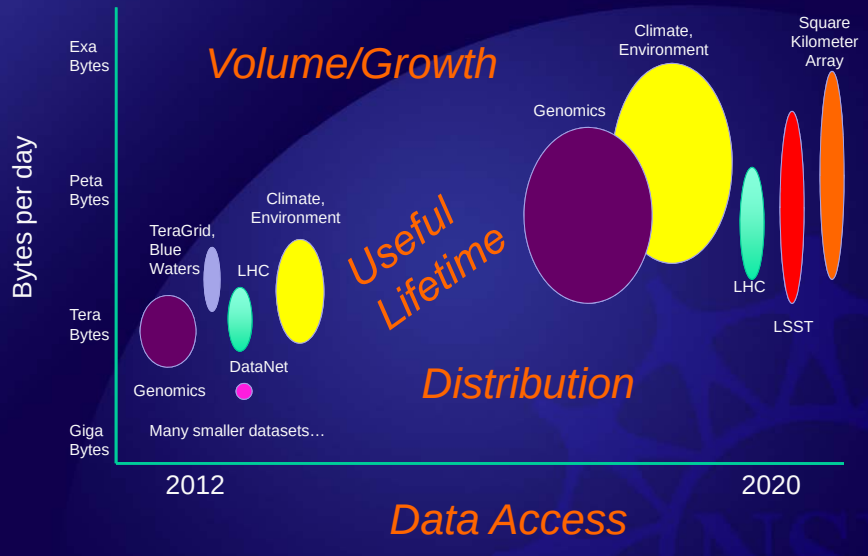
Four Thrust Areas



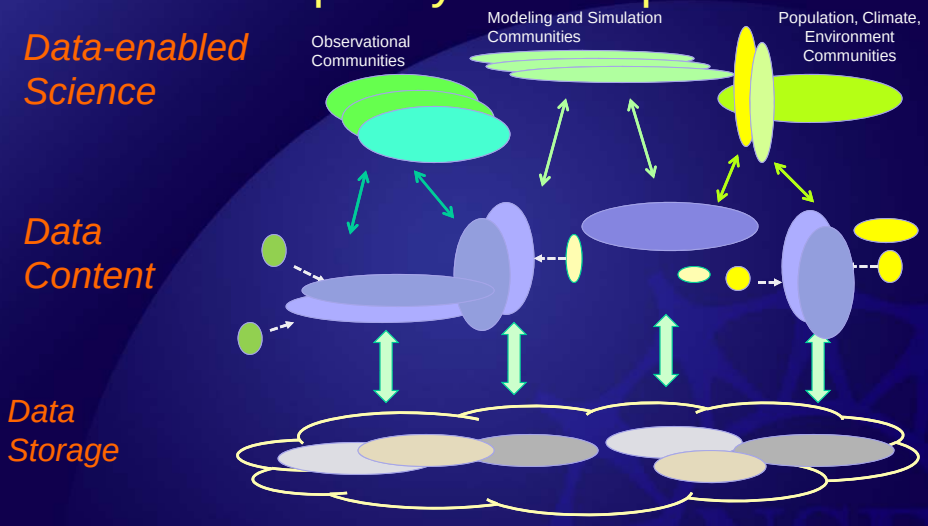
Broad CIF21 Principles

- Builds national infrastructure for S&E
- Leverages common methods, approaches, and applications – focus on interoperability
- Catalyzes other CI investments across NSF
  - Provides focus and is a vehicle for coordinating efforts and programs
  - Is a "force multiplier" across NSF
- Based upon a shared governance model involving every directorate and office
- Managed as a coherent program

Scientific Data Challenges



DataNET: A Multi-tiered and Multi-Disciplinary Landscape



Data-Enabled Science

- Data Services/Storage *(data)*
  - Provide reliable digital preservation, access, integration, and analysis capabilities for science and/or engineering data over a decades-long timeline
- Data Content *(information)*
  - Data mining, data analysis, manipulation, modeling, visualization, decision-making systems, tools
- Data-enabled Science *(knowledge)*
  - Intensive disciplinary efforts
  - Simulation, modeling
  - Multi-disciplinary S&E

THE CHRONICLE of Higher Education

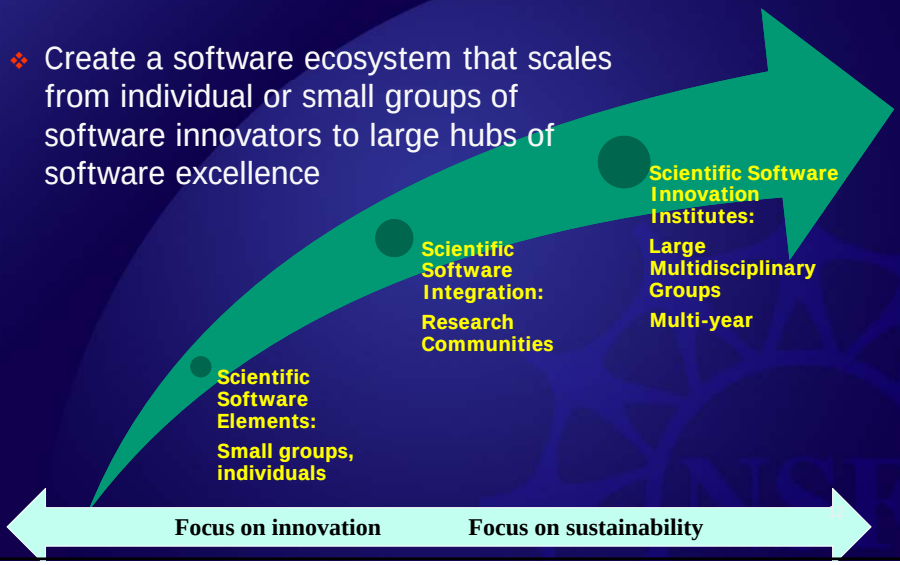
Dumped On by Data: Scientists Say a Deluge Is Drowning Research

New Computational Infrastructure

- Computational and Data-enabled resources
  - HPC, Clouds, Clusters, Data Centers
  - Computation capabilities
  - Modeling, simulation, visualization
- Long-term software for science and engineering
  - Sustained software development and support
- Discipline-specific activities
  - Services, tools, compute, simulation environments that serve specific research efforts and communities

Creating Scalable Software Development Environments

- Create a software ecosystem that scales from individual or small groups of software innovators to large hubs of software excellence



Community Research Networks

- New multidisciplinary research communities
  - Address challenges beyond individuals and disciplinary research communities
  - Support and optimize collaboration across small, mid-level and large community networks
  - Support SEES and new research communities
- Advanced research on community and social networks
  - Structures, leadership, fostering and sustainability
  - "virtuous cycle" providing feedback through formal evaluation and program iteration

Access and Connectivity

- Network connections and engineering program
  - Real-time access to facilities and instruments; Begins to tie in MREFC activities
  - Integration and end-to-end performance to provide seamless access from researcher to resource
- Cybersecurity – from innovation to practice
  - Deployment of identity management systems
  - Development of cybersecurity prototypes

Challenges for EPSCoR (1)

- Adopt the new world of cyberinfrastructure
  - CI is the foundation for research and education
  - Users need to interact with CI capabilities
- Become a data provider rather than just a consumer
- Diversity is a strength
  - sensor and personal networks
  - social networking, crowd sourcing
  - climate, environment, genomics
  - "long tail" efforts contribute significantly

Some suggestions for next steps

- Need to get ready
  - Access essential
  - Data policies in place to enable sharing
  - Institute authentication, authorization
- Develop pinnacles of excellence rather than deploy general approaches which are thin
- Cyberinfrastructure education, expertise and experience is critical
- Take the opportunity to "leap frog" capabilities by investing smart
  - Jump to next generation approaches

Thank you