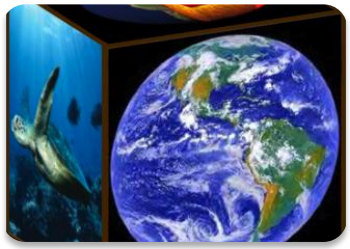
Transforming the Conduct of Research and Education in the Geosciences through EarthCube

Dr. Timothy Killeen, Assistant Director,
Directorate for Geosciences (GEO)
National Science Foundation

Outline

Context

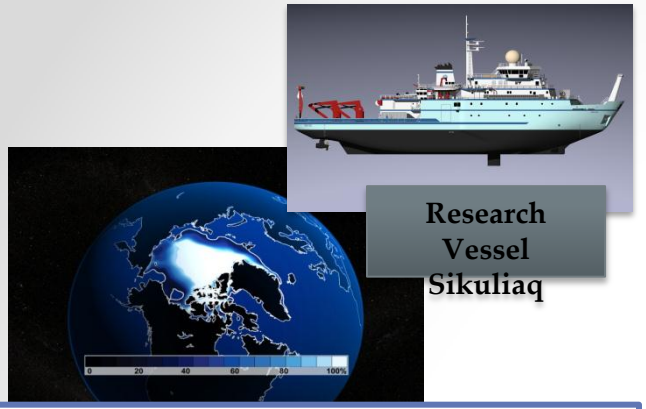


EarthCube Vision

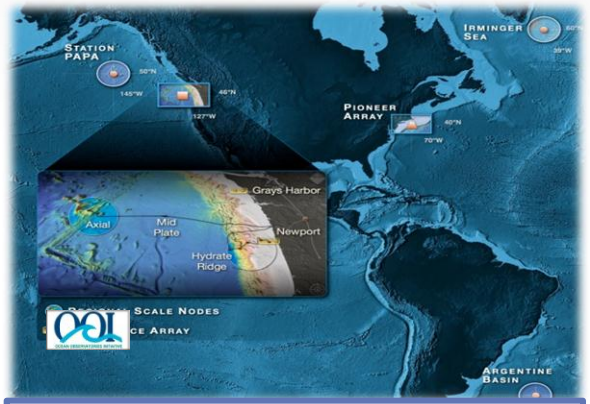
Early Efforts

Context

The age of observation, simulation, and data



Arctic Sea Ice

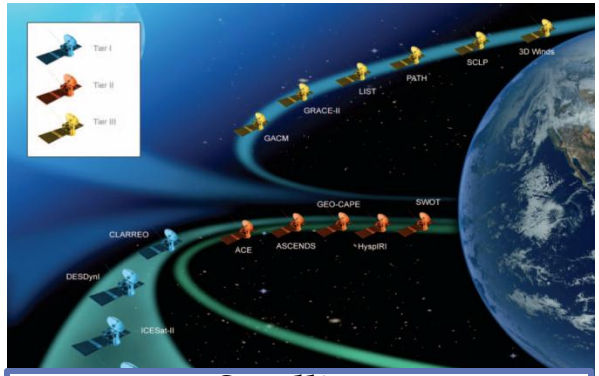


Oceans

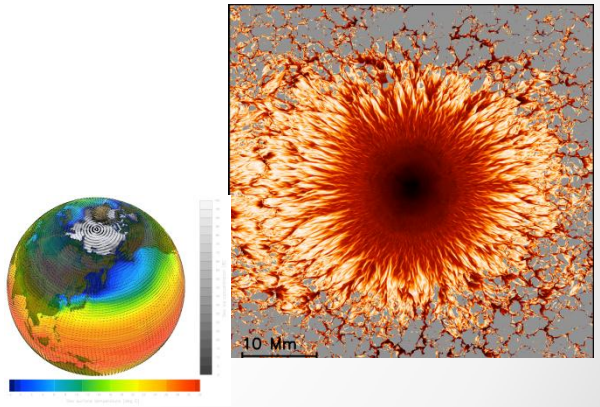
EarthScope Observatory Network

Era of Observation and Simulation

Water



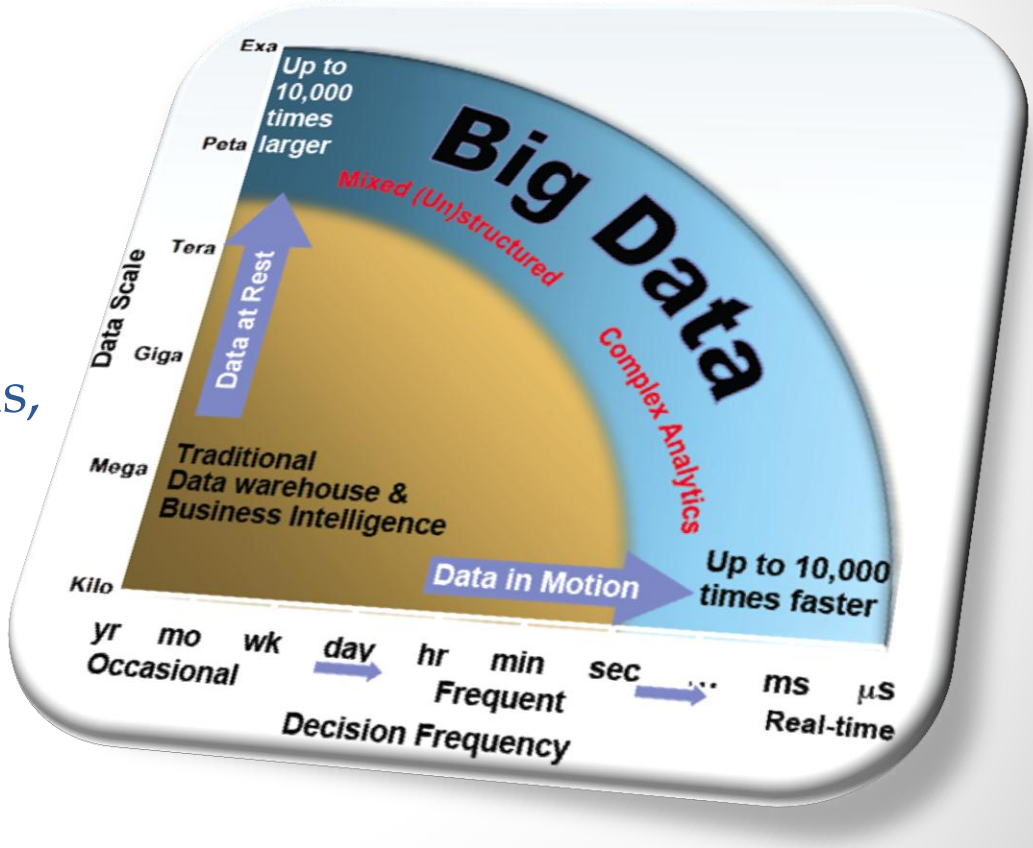
Satellites



Earth System Modeling

Science and Society Transformed by Data

- **Modern geoscience**
  - Data- and compute-intensive
  - Integrative, multi-scale
- **Multi-disciplinary collaborations to address complexity**
  - Individuals, groups, teams, communities
- **Sea of Data**
  - Age of Observation
  - Distributed, central repositories, sensor-driven, diverse, etc



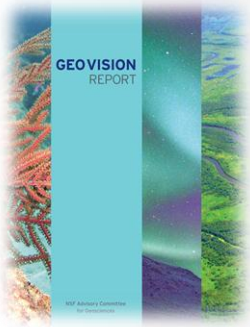
The Science Context

"Fostering a sustainable future through a better understanding of our complex and changing planet."

NSF's GEO Vision report, 2009

Science foundations for EarthCube

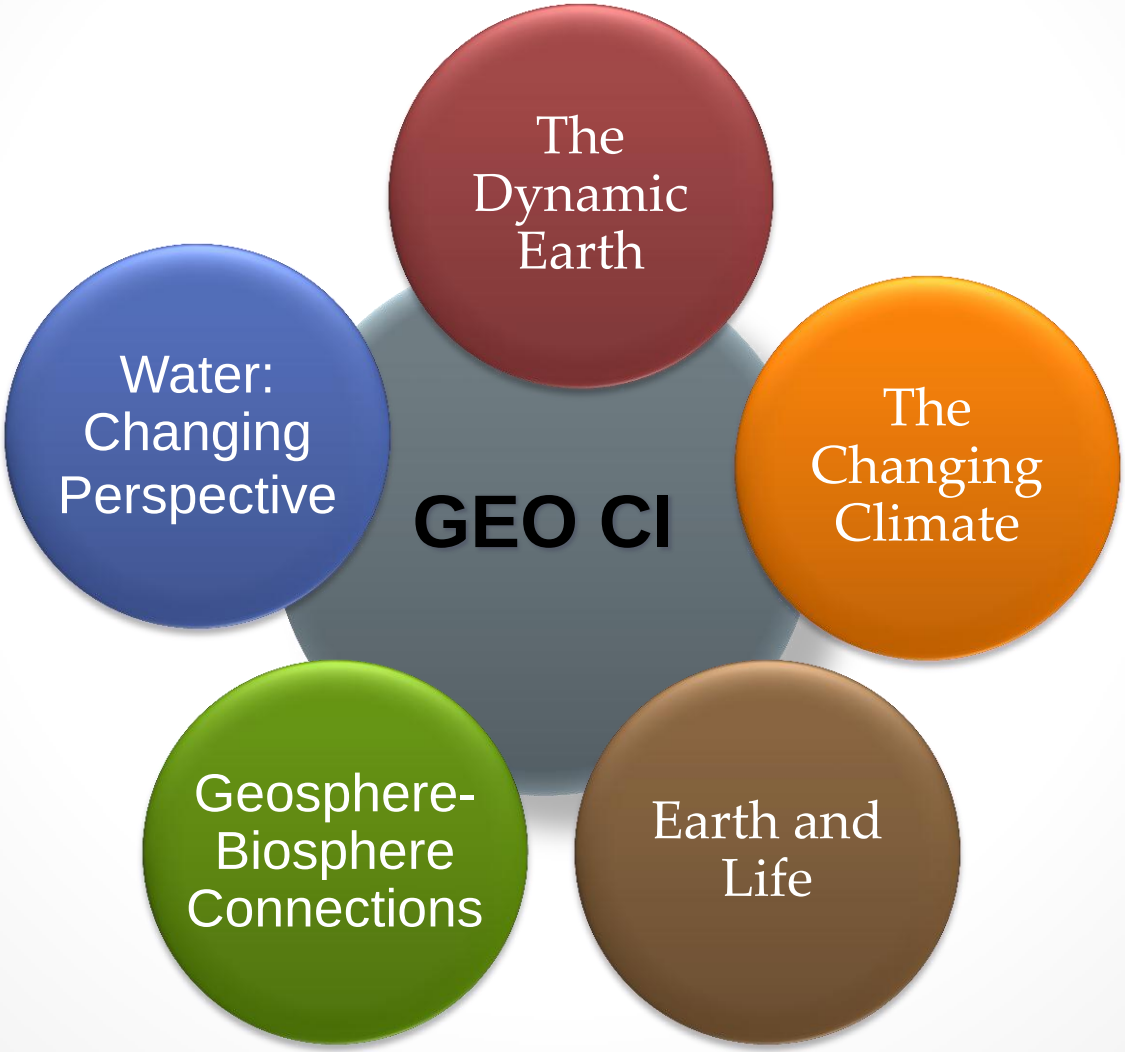
PURPOSE: "To understand more deeply the planet and its interactions will require the geosciences to take an increasingly holistic approach, exploring knowledge coming from all scientific and engineering disciplines."



CALL TO ACTION: "Over the next decade, the geosciences community commits to developing a framework to understand and predict responses of the Earth as a system—from the space-atmosphere boundary to the core, including the influences of humans and ecosystems."

NSF's GEO Vision report

Crossroad Challenges of GEOvision



EarthCube Vision

- An alternative approach to respond to daunting science and CI challenges
- EarthCube is an outcome and a process
- EarthCube will require broad community participation

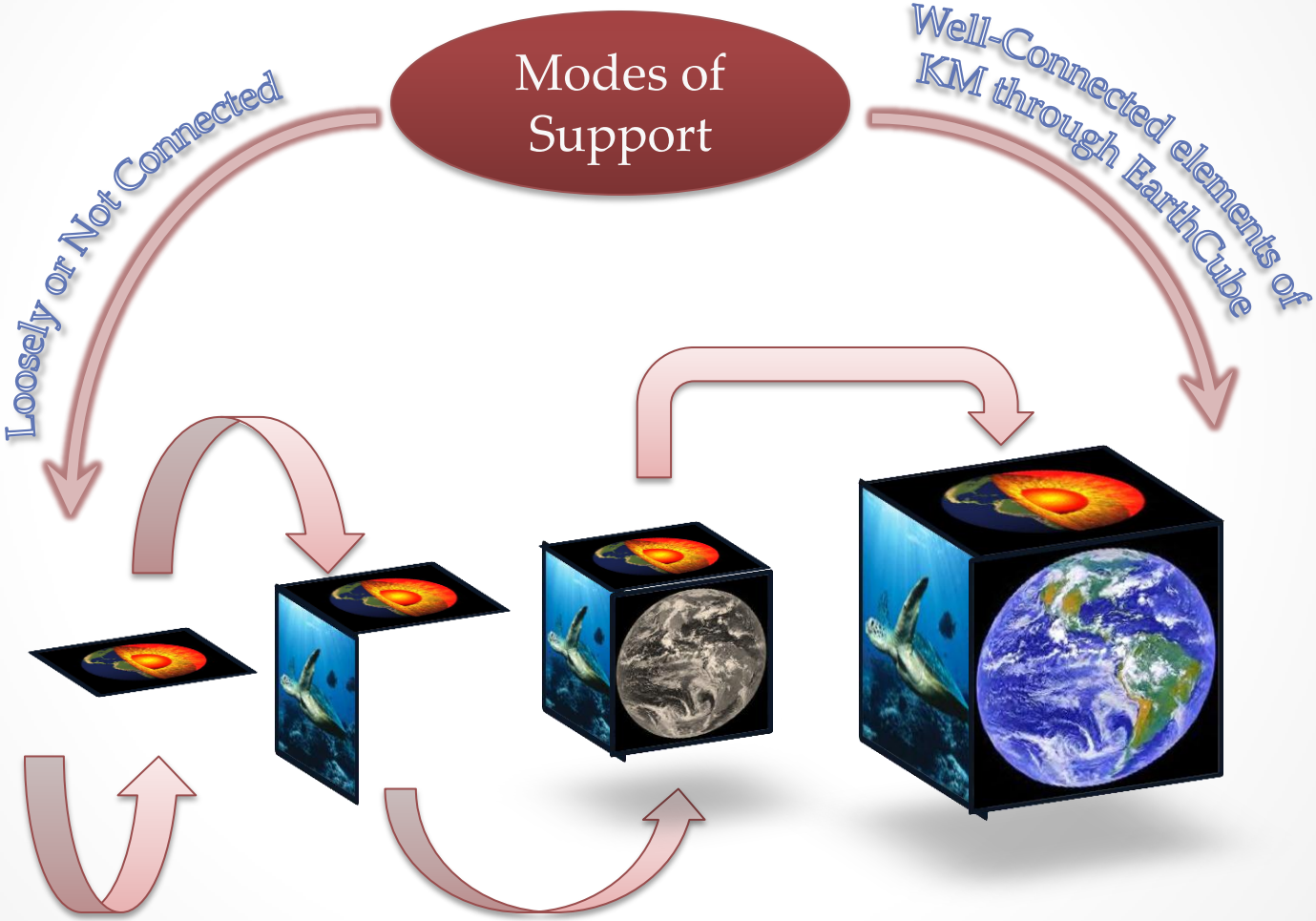
Goal of Earth Cube

to transform the conduct of research in geosciences by supporting the development of community-guided cyberinfrastructure to integrate data and information for knowledge management across the Geosciences.

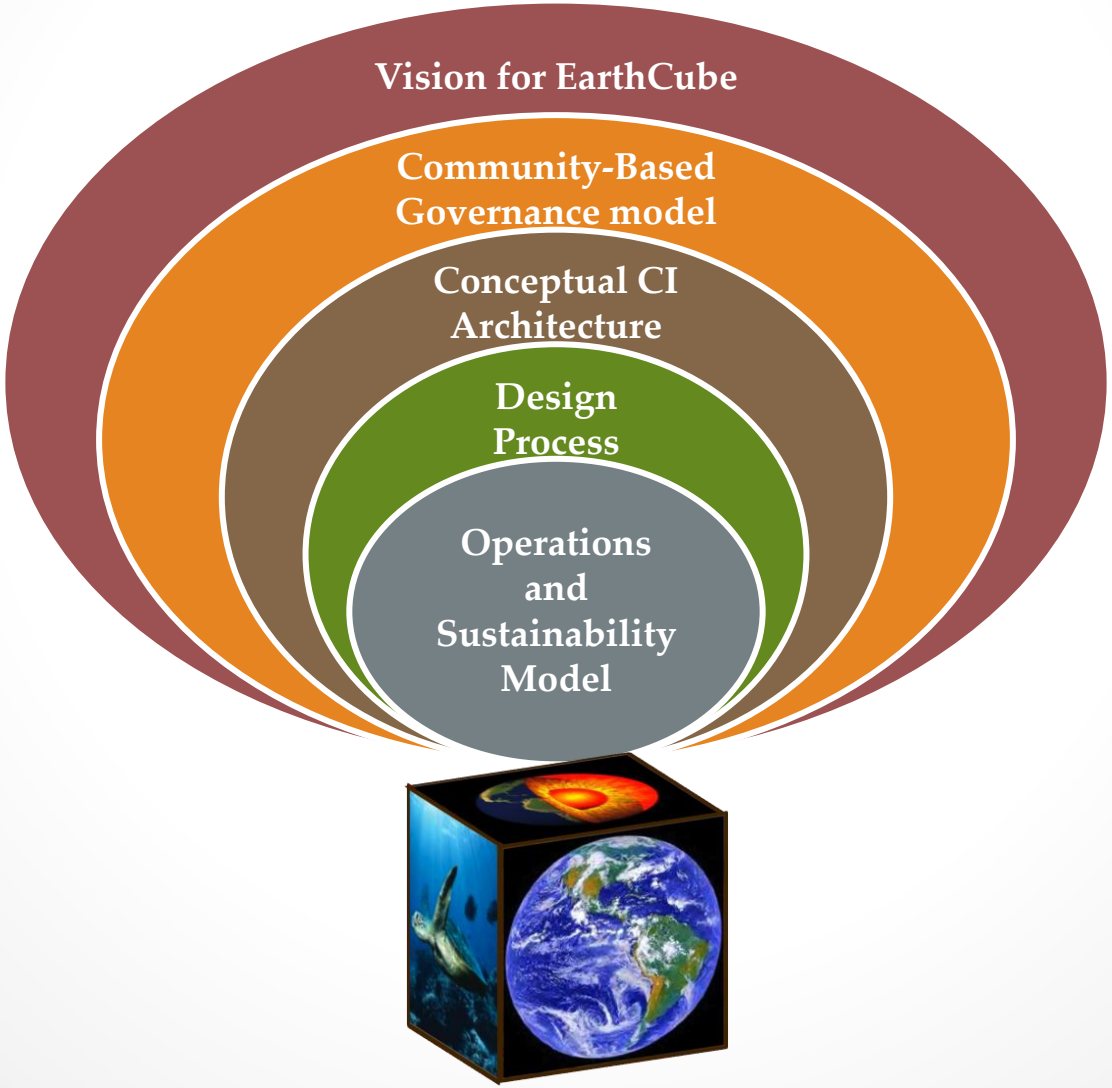
Long-term Goals for EarthCube

- Transform practices within the geosciences community spanning over the next decade
- Provide unprecedented new capabilities to researchers and educators
- Vastly improve the productivity of community
- Accelerate Earth system research
- Provide a uniform knowledge management framework for all of the geosciences
- Facilitate creation of networks of excellence and new partnerships
- Train and "liberate" the next generation of researchers

Strategic Convergence Toward EarthCube



EarthCube Design Elements



Initial EarthCube efforts

Unanticipated community engagement and contribution

Milestones

• Release of first DCL	June 2011
• WebEx presentations	June & July 2011
• Establish EarthCube Social Network site	August 2011
• Charrette	November 2011
• Second DCL	December 2011
• EAGER and Supplement Submissions	December 2011 – April 2012
• Second Community event	May 2012
• Continued community opportunities	May 2012 - ?

Engage

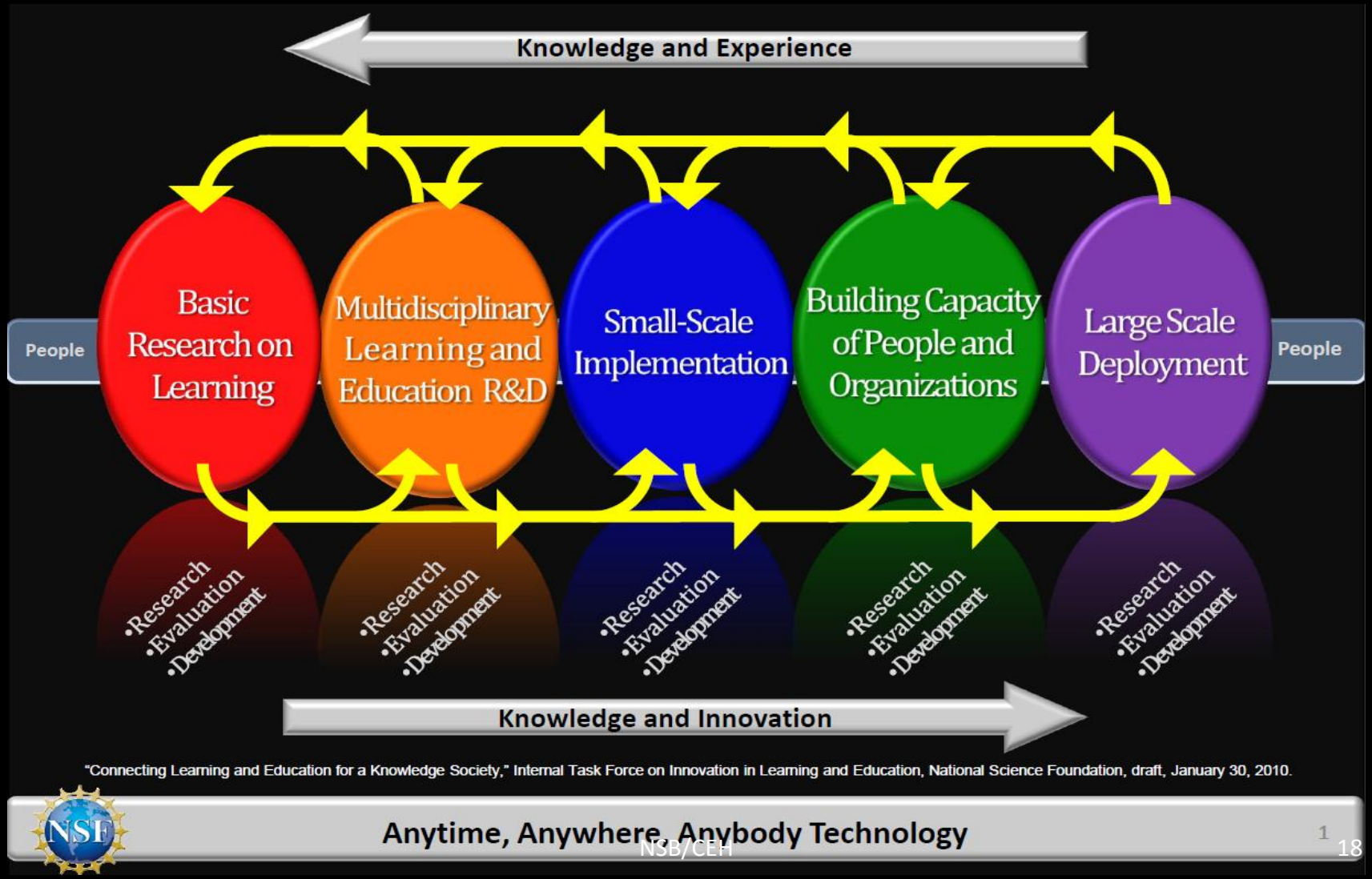
Empower

Education - Why NSF?

Energize

Anytime, Anywhere, Anybody

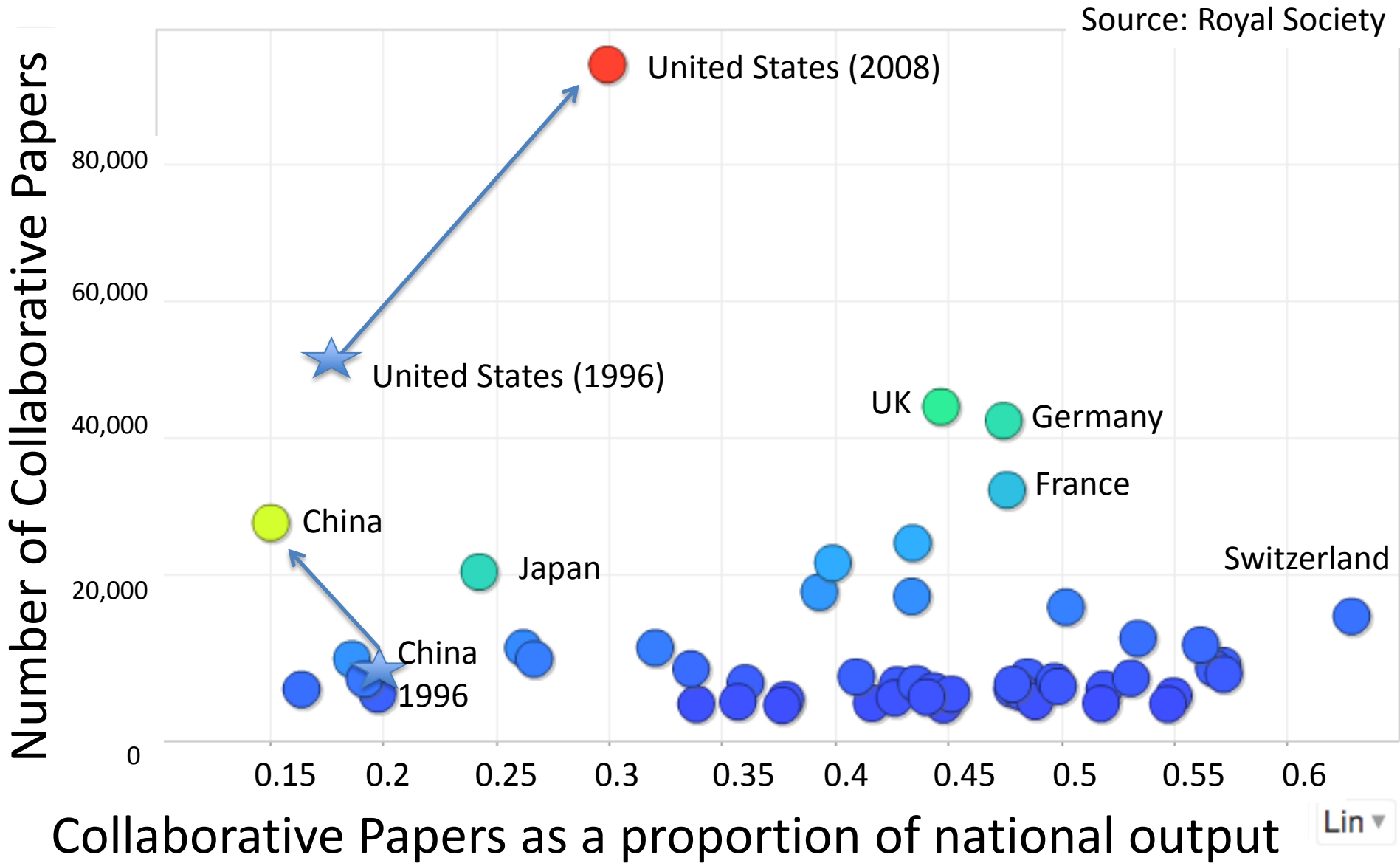
R&D Cycles in STEM education



An Example Expedition: CYBER-SEES^{E2}

- Goals
  - **Connect** Cyberinfrastructure, Facilities, data sets, research activities: a wide range of learners in a variety of settings via technology
  - **Access** – real time data, widely publicized and available
  - **Enrichment** - Engage, Empower, Energize
  - **Current Content -** Relevance and timeliness: local to global
- Delivery mechanisms and partners *exist*
- How do we learn?

Internationally Collaborative Papers (2008)



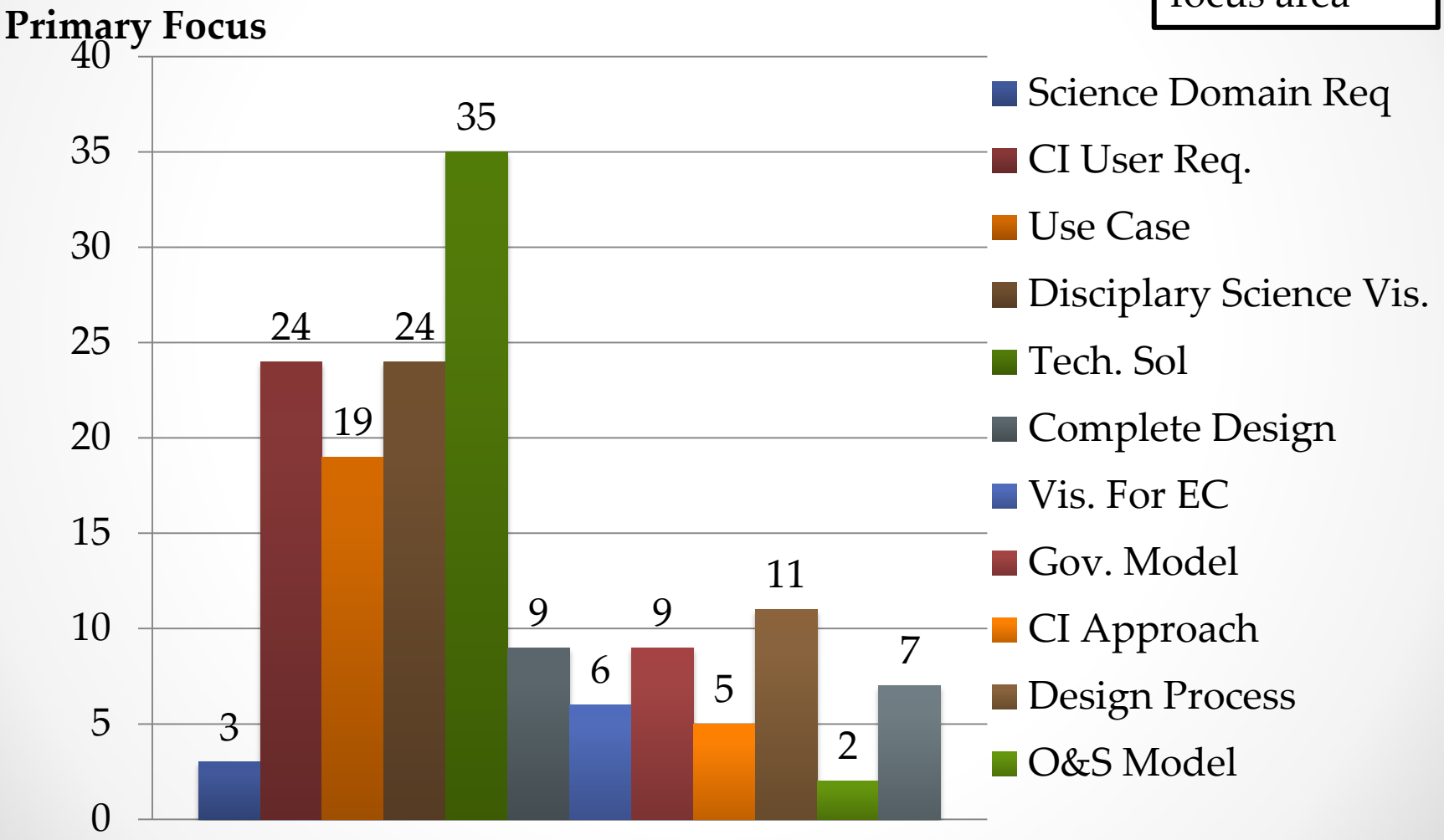
Who is Collaborating with Who?

EarthCube Website

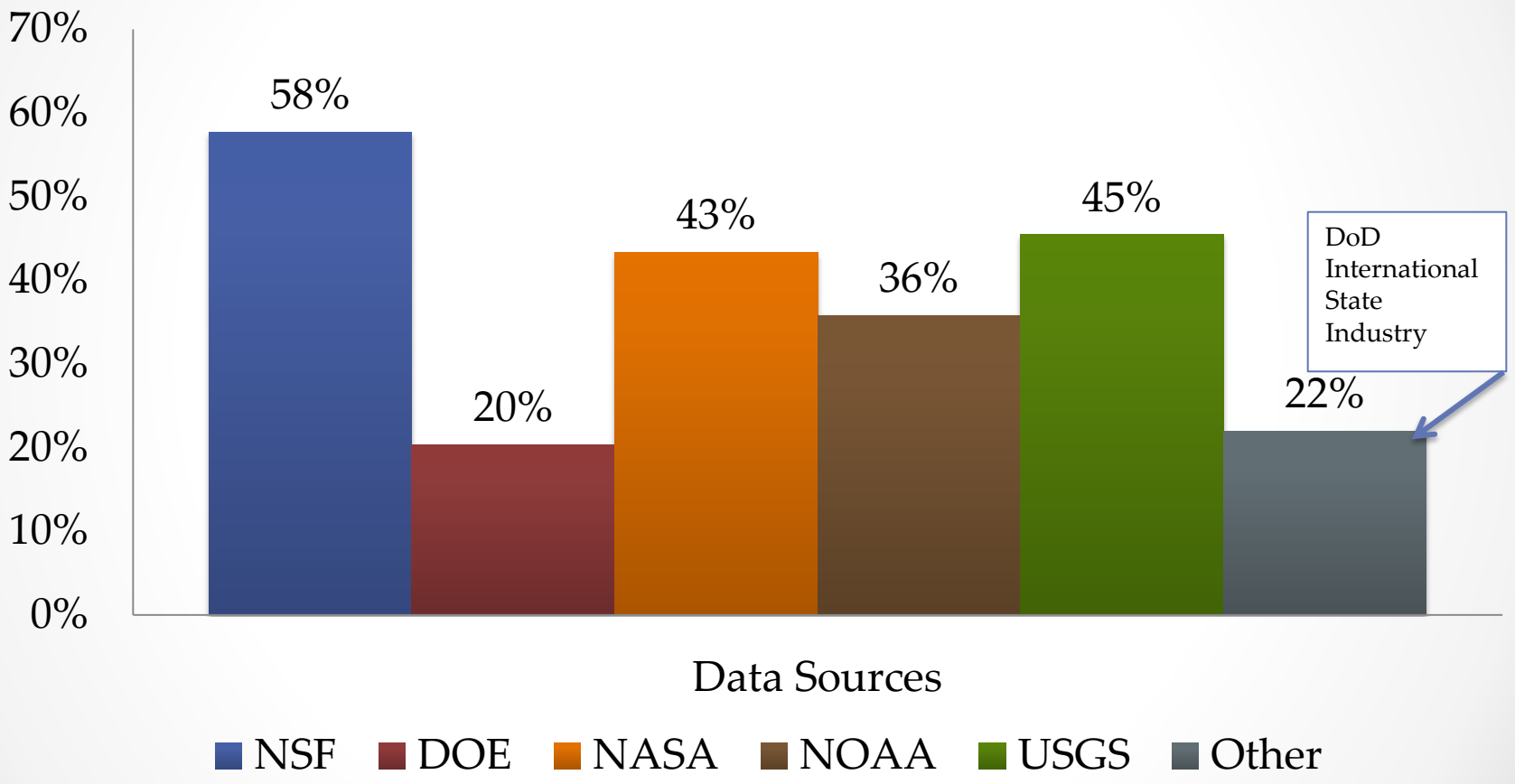
- ~600 members to the EarthCube website
- 113 white papers submitted – all online
- 185 respondents to user survey
- Unknown number of hours of pro bono contributions by community
- Unprecedented view of the pulse of the geoscience community

White Papers

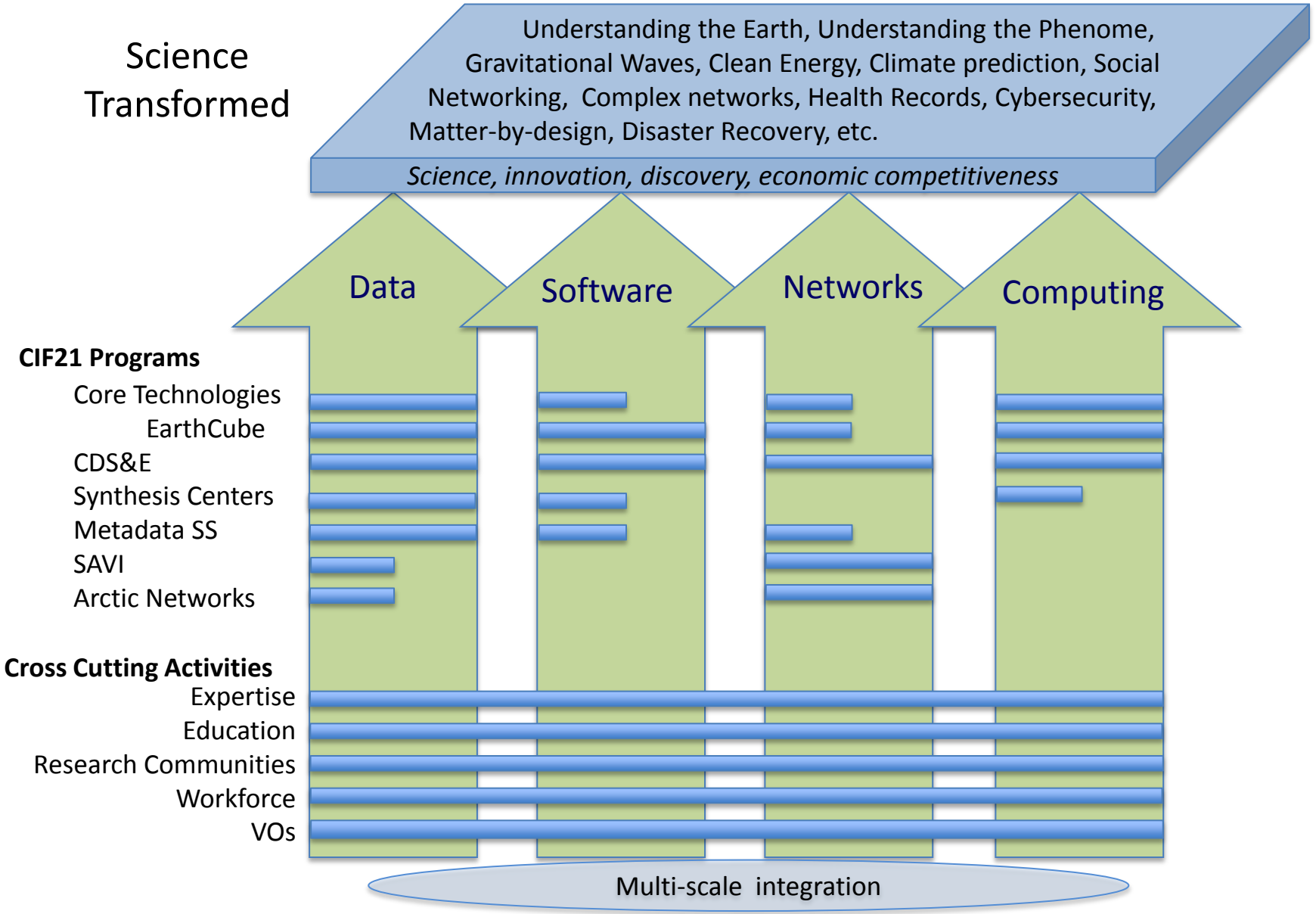
Some papers had more than one focus area



Where do you obtain your data?



NSF-Wide Perspective (CIF-21)



Data

- Do your require real-time data?
  - 34% Yes
  - 34% No
- Do you require data or data products from outside your immediate discipline?
  - 52% Yes
  - 14% No
- Is there a place for products of your own work to be archived and accessed?
  - 45% Yes
  - 25% No

Charrette Attendance

- On-site 146
  - Not all for the entire 3 ½ days
  - Attendance was high even on last morning
- Virtual ~ 100 (varied by day)
- Broad representation of geosciences research community
  - Including several graduate and post-docs
- Representation also included
  - Other federal agencies
  - International
  - Industry
  - Publishers
  - Societies and science organizations
  - Organizational specialists
- ~20 GEO & OCI POs (not at same time)
  - Representatives from BIO and CISE also in attendance

Immediate Outcomes

- Identifying the most important capabilities to be embodied in EarthCube
  - Defining the initial view of EarthCube
- Many "2-pager" papers on new capabilities
  - Resulted from work at the Charrette
- New collaborations enabled by the Charrette
- Instructions NSF's interest in Eager awards post-Charrette
  - Strategic organizational framework
  - Development of new capabilities
  - Progress on critical milestone (developed at Charrette)
- EarthCube website was and is an effective tool for community dialog
  - Labor intensive to monitor and maintain
- Bottom line: a new starting place for creative and collaborative ideas, partnerships, and building CI for the geosciences

Future Plans

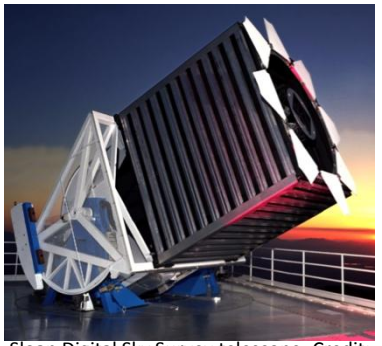
Future directions will be shaped by the community

NCAR-Wyoming Supercomputing Center

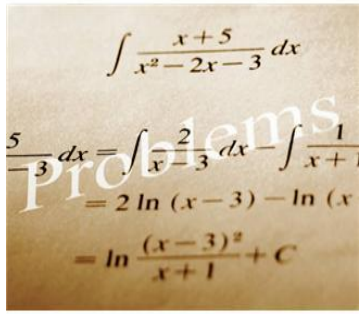
- Construction began June 2010 and proceeding on schedule
- Yellowstone: inaugural computing resource
- 1.6-petaflops IBM iDataPlex cluster.
- 11 PB useable disk capacity
- Data analysis and visualization systems
- 600 million core hours available for allocation
- Greater than 120 PB mass storage
- Production operations expected in summer 2012

Foundational Research in Large-Scale Data Management and Analysis

- **Collection, Storage, and Management of "Big Data"**
  - Data representation, storage, and retrieval
  - New parallel data architectures, including clouds
  - Data management policies, including privacy and access
  - Communication and storage devices with extreme capacities
  - Sustainable economic models for access and preservation
- **Data Analytics and Machine Learning**
  - Computational, mathematical, statistical, and algorithmic techniques for modeling high dimensional data
  - Learning, inference, prediction, and knowledge discovery for large volumes of dynamic data sets
  - Data mining to enable automated hypothesis generation, event correlation, and anomaly detection
  - Information infusion of multiple data sources
- **Research in Data Sharing and Collaboration**
  - Tools for distant data sharing, real time visualization, and software reuse of complex data sets
  - Cross disciplinary information and knowledge sharing
  - Remote operation and real time access to distant data sources and instruments



Sloan Digital Sky Survey telescope. Credit: Fermilab Photoz.



Future Plans

- Fund EAGER and Supplements that prompt community convergence
- Prepare for the next community event in mid-2012
- Continued engagement and reconnaissance of the community to identify future funding opportunities
- Prototype development efforts
- Facilitate community events that prompt convergence
- Continue outreach nationally and internationally

Questions and/or Comment