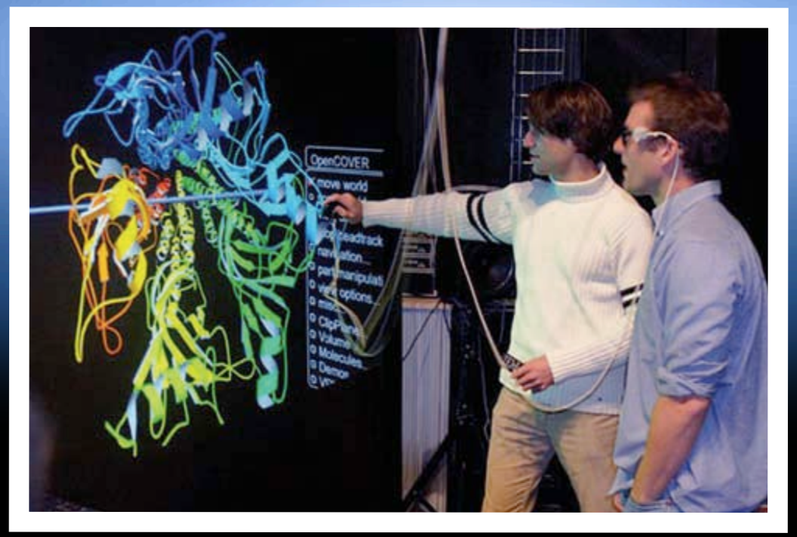
Scientific Digital Data as Cyberinfrastructure:

Sustainable Digital Data Preservation and Access Network Partners

Lucy Nowell, Program Director, OCI



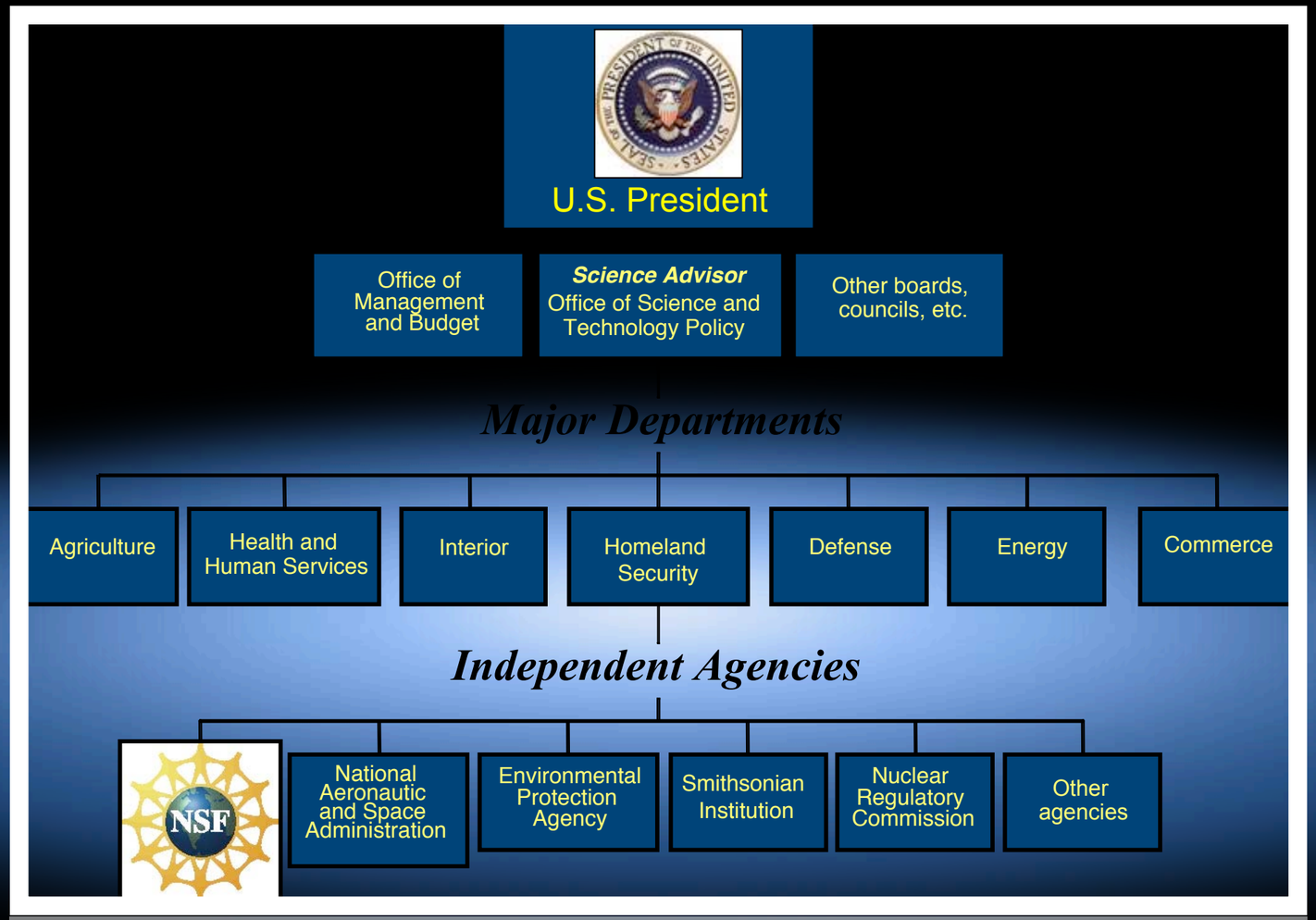
National Science Foundation
Where Discoveries Begin

Office of Cyberinfrastructure

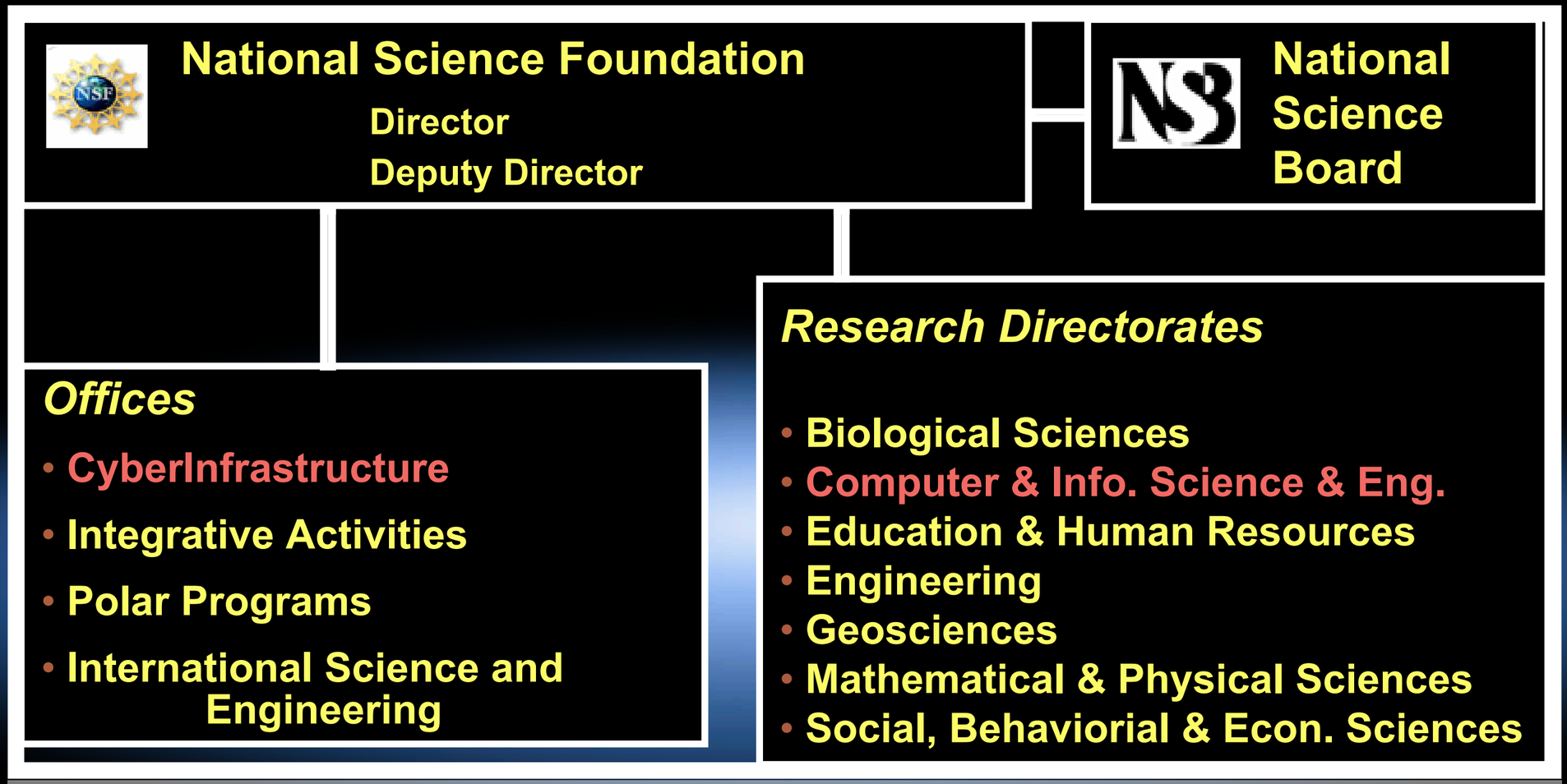
- About NSF
- Background for the DataNet Program
- DataNet Partners
  - Goals
  - Characteristics
- Status

Today's Presentation

Lucy Nowell
lnowell@nsf.gov

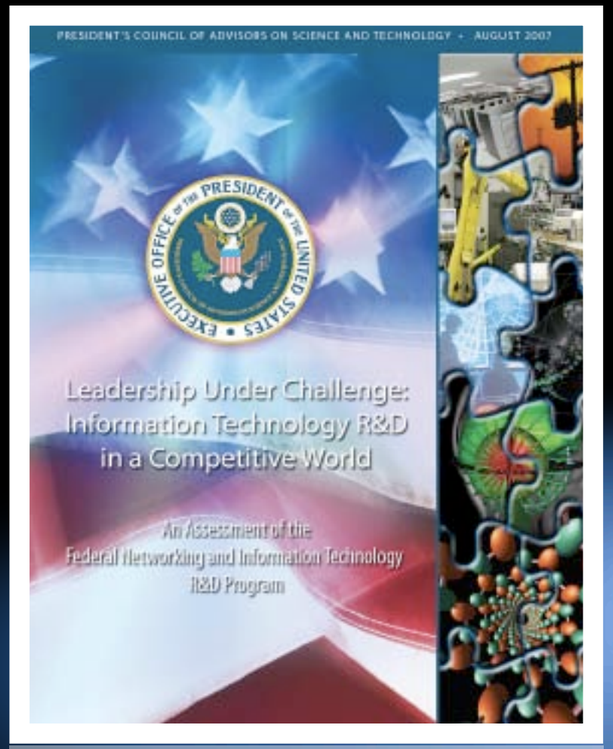


Lucy Nowell
lnowell@nsf.gov

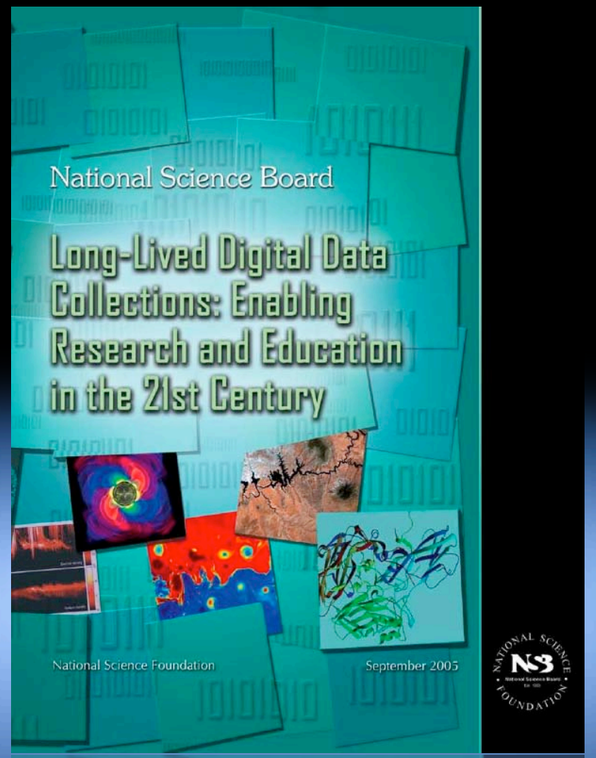


Lucy Nowell
lnowell@nsf.gov

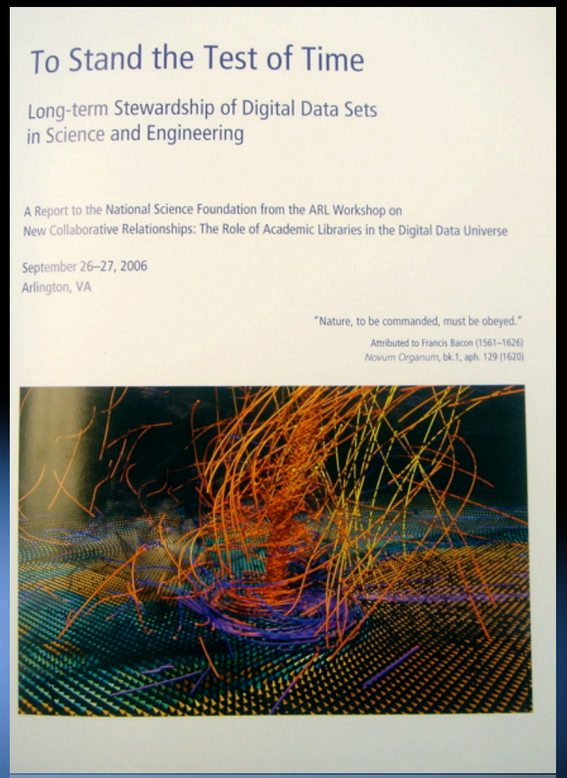
PCAST Recommendations on Digital Data



NSB Report: Long-Lived Digital Data Collections Enabling Research and Education in the 21st Century



NSF Supported Experts Study



Examples of Input Informing Data Activities

Lucy Nowell
lnowell@nsf.gov

Storage Networking Industry Association (SNIA) *100 Year Archive Requirements Survey Report*

"The results confirmed our view that there is a pending crisis in archiving…If we're going to be able to avoid this crisis we have to create long-term methods for not only preserving information, but also for making it available for analysis in the future."

Paperwork Reduction Act*: The purposes of this subchapter are to ….

(2) Ensure the greatest possible public benefit from and maximize the utility of information created, collected, maintained, used, shared, and disseminated by or for the federal government; ….

(7) Provide for the dissemination of public information on a timely basis, on equitable terms, and in a manner that promotes the utility of the information to the public and makes effective use of information technology

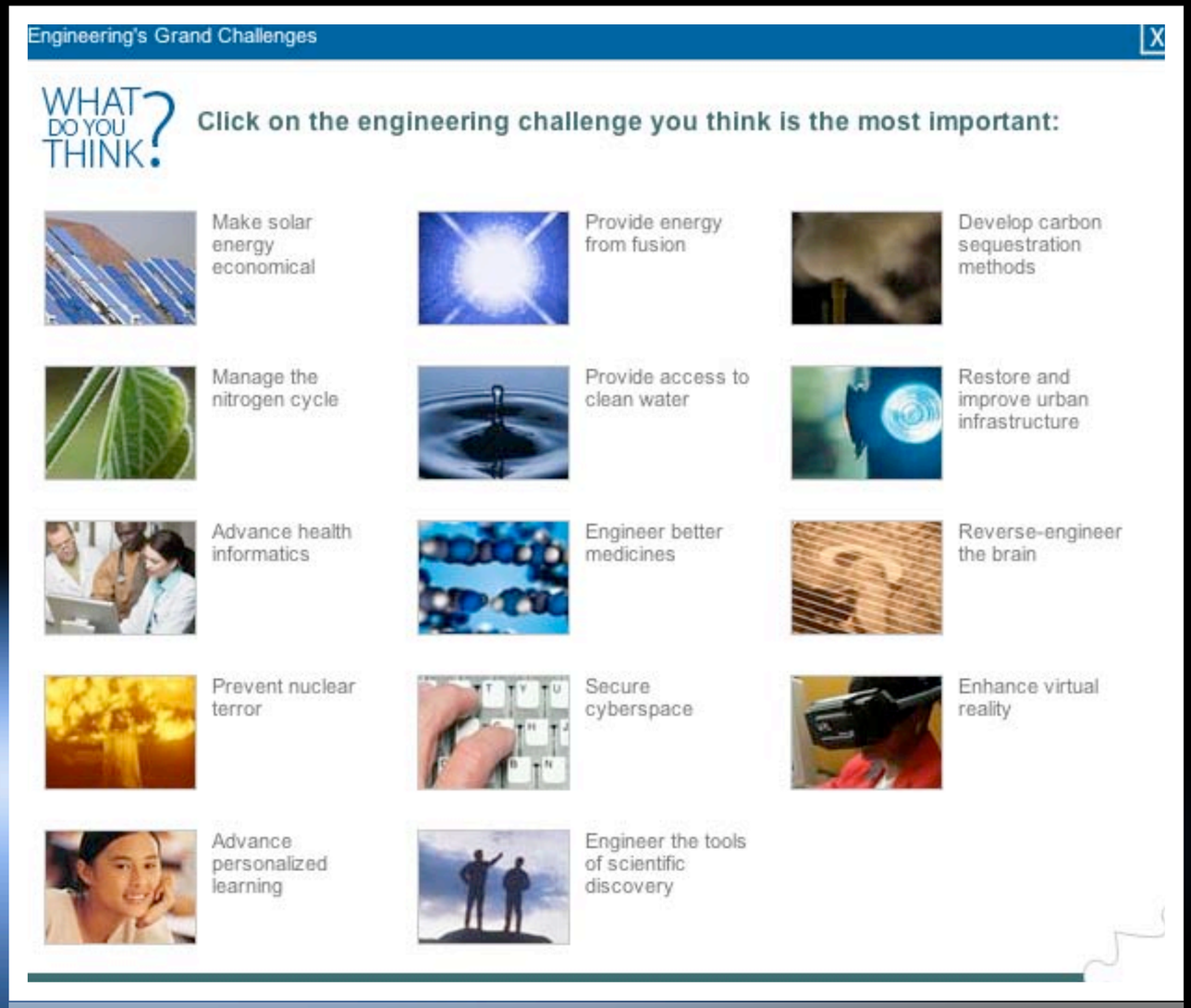
Office of Management and Budget (OMB) Circular A-130: Management of Federal Information Resources

Part 7. Basic Considerations and Assumptions:….

k. The open and efficient exchange of scientific and technical government information, subject to applicable national security controls and the proprietary rights of others, fosters excellence in scientific research and effective use of Federal research and development funds.

More Examples of Input Informing DataNet

Lucy Nowell
lnowell@nsf.gov



http://
www.engineeringchallenges.org/

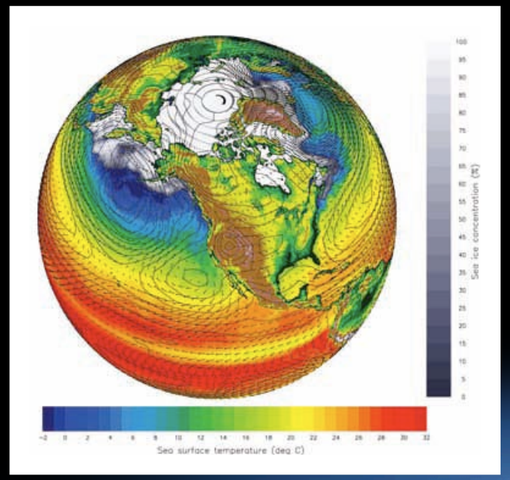
Grand Challenges

National Science Foundation
Where Discoveries Begin

Lucy Nowell
lnowell@nsf.gov

Office of
Cyberinfrastructure

http://www.nationalacademies.org/morenews/20080312.html



- **10 Questions Shaping 21st-century Earth Science**
- How did earth and other planets form?
- What happened during earth's "dark Age" (the first 500 million years)?
- How did life begin?
- How does earth's interior work, and how does it affect the surface?
- Why does earth have plate tectonics and continents?
- How are earth processes controlled by material properties?
- What causes climate to change – and how much can it change?
- How has life shaped earth? And how has earth shaped life?
- Can earthquakes, volcanic eruptions and their consequences be predicted?
- How do fluid flow and transport affect the human environment?

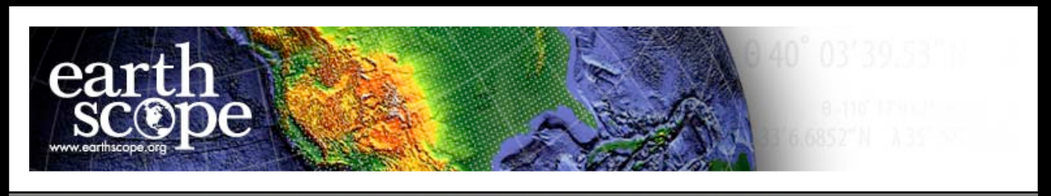
Grand Challenges

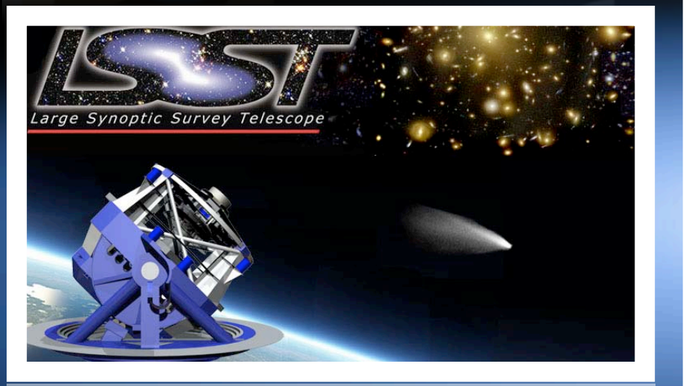
National Science Foundation
Where Discoveries Begin

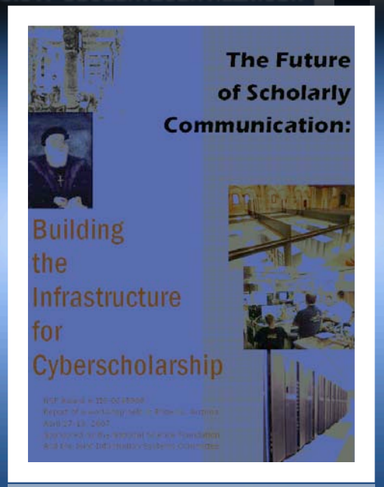
Lucy Nowell
lnowell@nsf.gov

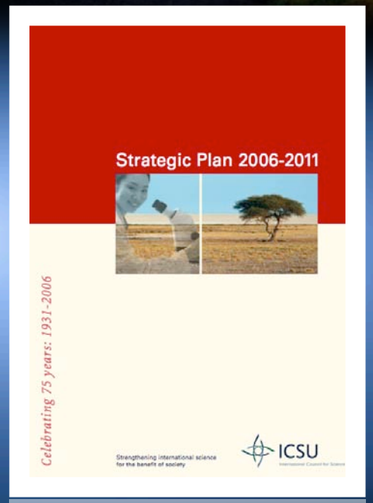
Office of Cyberinfrastructure

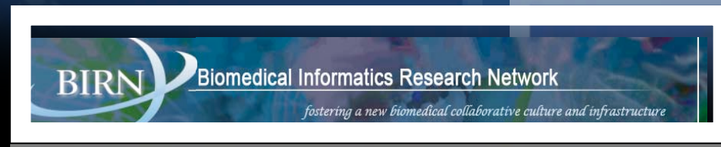
Examples of Science Drivers for OCI Data Activities

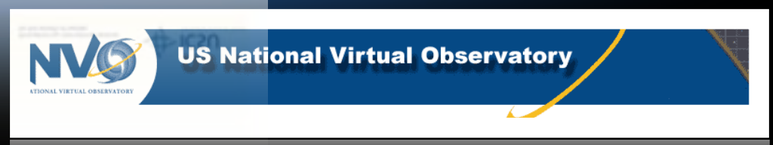










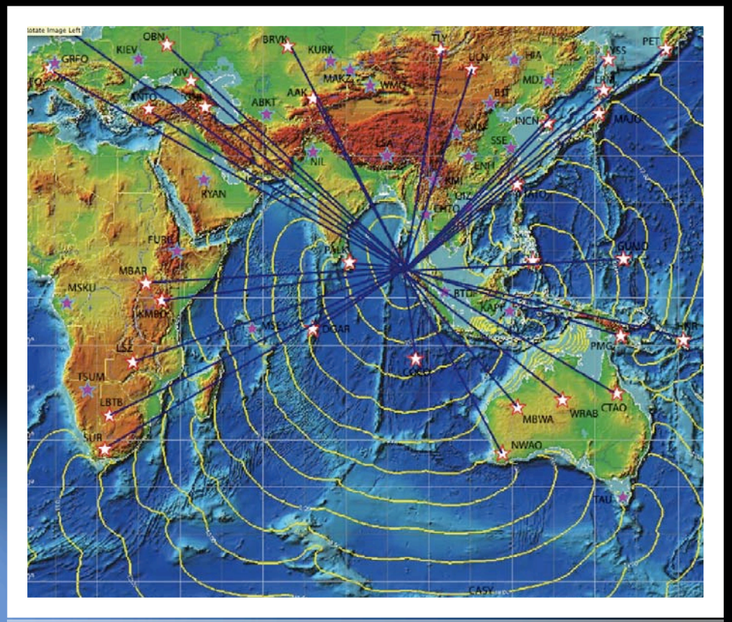


National Science Foundation
Where Discoveries Begin

Lucy Nowell
lnowell@nsf.gov

Office of Cyberinfrastructure

- Irreplaceable Data
- Replication of Results
- Longitudindal Science and the Impact of Human Activity
- Enabling Interdisciplinary Science & Engineering for Grand Challenges
- Broadening Participation



Why Preserve & Share Data?

Lucy Nowell
lnowell@nsf.gov

- **Irreplaceable Data**

“In 1964, the first electronic mail message was sent from either MIT, the Carnegie Institute, or Cambridge University. The message does not survive, however, and so there is no documentary record to determine which group sent the pathbreaking message.”

Report of the Task Force on Archiving of Digital Information, Commission on Preservation and Access and the Research Libraries Group

Why Preserve & Share Data?

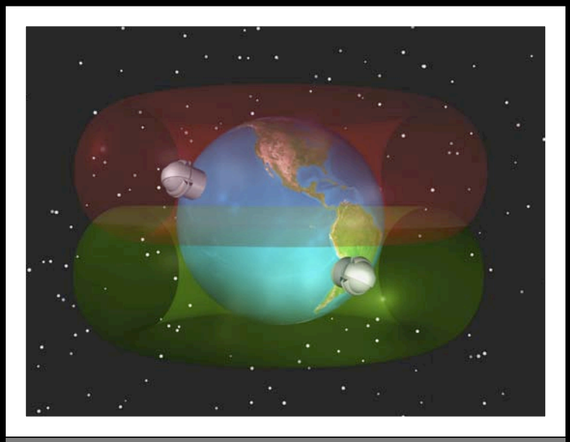
Lucy Nowell
lnowell@nsf.gov

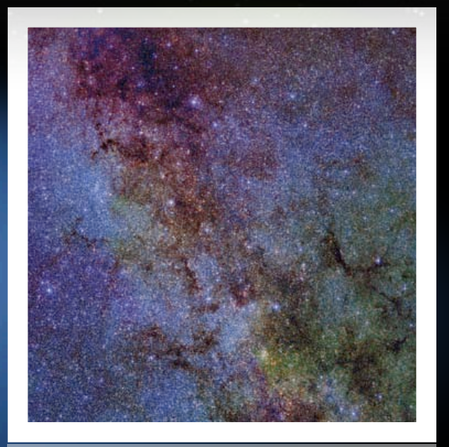
Irreplaceable Data

The Saga of the Lost Space Tapes

"Maybe somebody didn't have the wisdom to realize the original tapes might be valuable sometime in the future," she said. "Certainly, we can look back now and wonder why we didn't have better foresight about this." – Dolly Perkins, Deputy Director, Goddard Space Flight Center

NASA Is Stumped in Search for Videos of 1969 Moonwalk,
Marc Kaufman, Washington Post, January 31, 2007





- **WASHINGTON – NASA said today it was launching an official search for more than 13,000 original tapes of the historic Apollo moon missions.**

Seth Borenstein, Associated Press, August 15, 2006, 5:13 PM

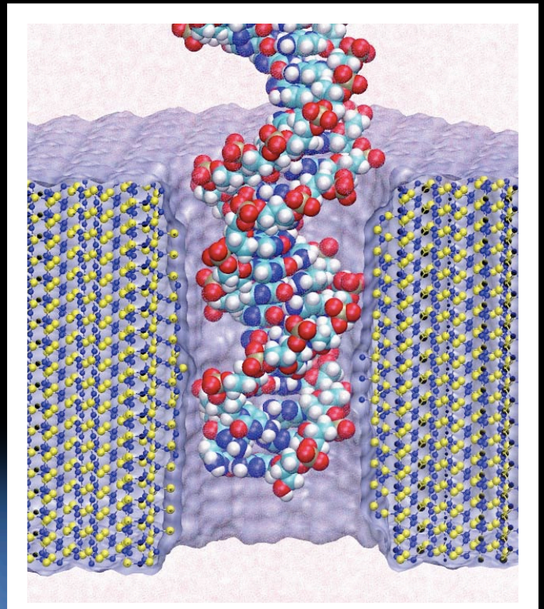
Why Preserve & Share Data?

Lucy Nowell
lnowell@nsf.gov

- Irreplaceable Data
- **Replication of Results**

"...the results of one scientist's experiment are not considered reliable until another scientist has replicated them. The reproducibility of results plays several different, crucial roles in science...[but] in many circumstances, considerations of time and money often make reproducibility impractical."

The Key Role of Replication in Science, Nancy S. Hall,
The Chronicle of Higher Education, November 10, 2000



Why Preserve & Share Data?

Lucy Nowell
lnowell@nsf.gov

- Irreplaceable Data
- Replication of Results
- **Data for Longitudinal Science**

With 43 years of hard-won water quality data, PI Stanley Grant observed, “When we plotted the data, our first reaction was ‘Wow! We can see the impact of the Clean Water Act!’ ”

Why Preserve & Share Data?

Lucy Nowell
lnowell@nsf.gov

August 1941

Courtesy of Mark Parsons of the National Snow & Ice Data Center (SNIDC)

A glacier melts and becomes a lake
August 2004

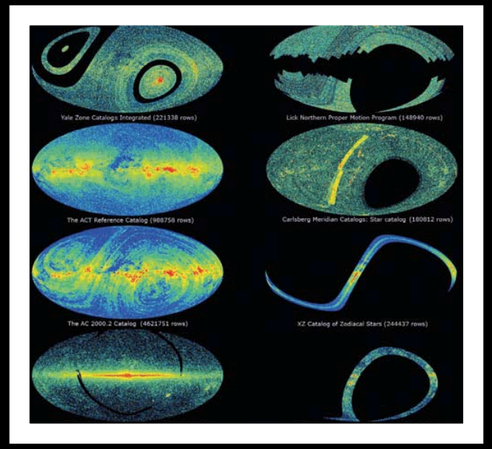
Why Preserve & Share Data?

National Science Foundation
Where Discoveries Begin

Lucy Nowell
lnowell@nsf.gov

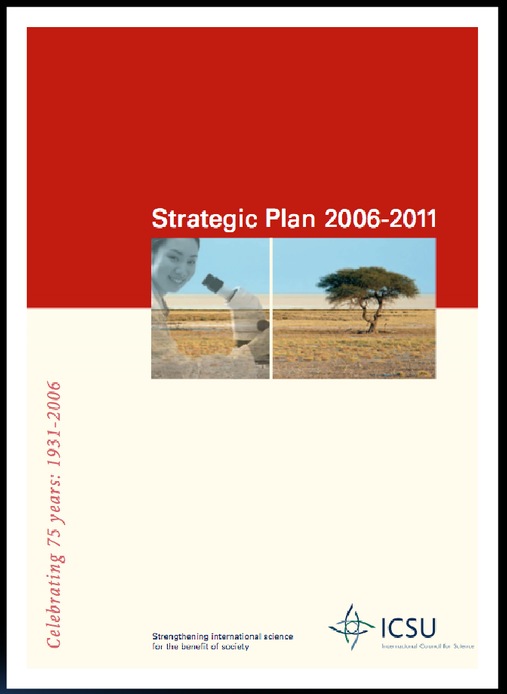
Office of Cyberinfrastructure

- Irreplaceable Data
- Replication of Results
- Data for Longitudinal Science



- **Interdisciplinary Science**

"The needs and opportunities for international interdisciplinary science are greater at the turn of the 21st century than at any previous period in history... Global research problems are invariably complex and require the collaboration of many disciplines as well as many countries."



ICSU
International Council for Science

Why Preserve & Share Data?

National Science Foundation
Where Discoveries Begin

Lucy Nowell
lnowell@nsf.gov

Office of Cyberinfrastructure

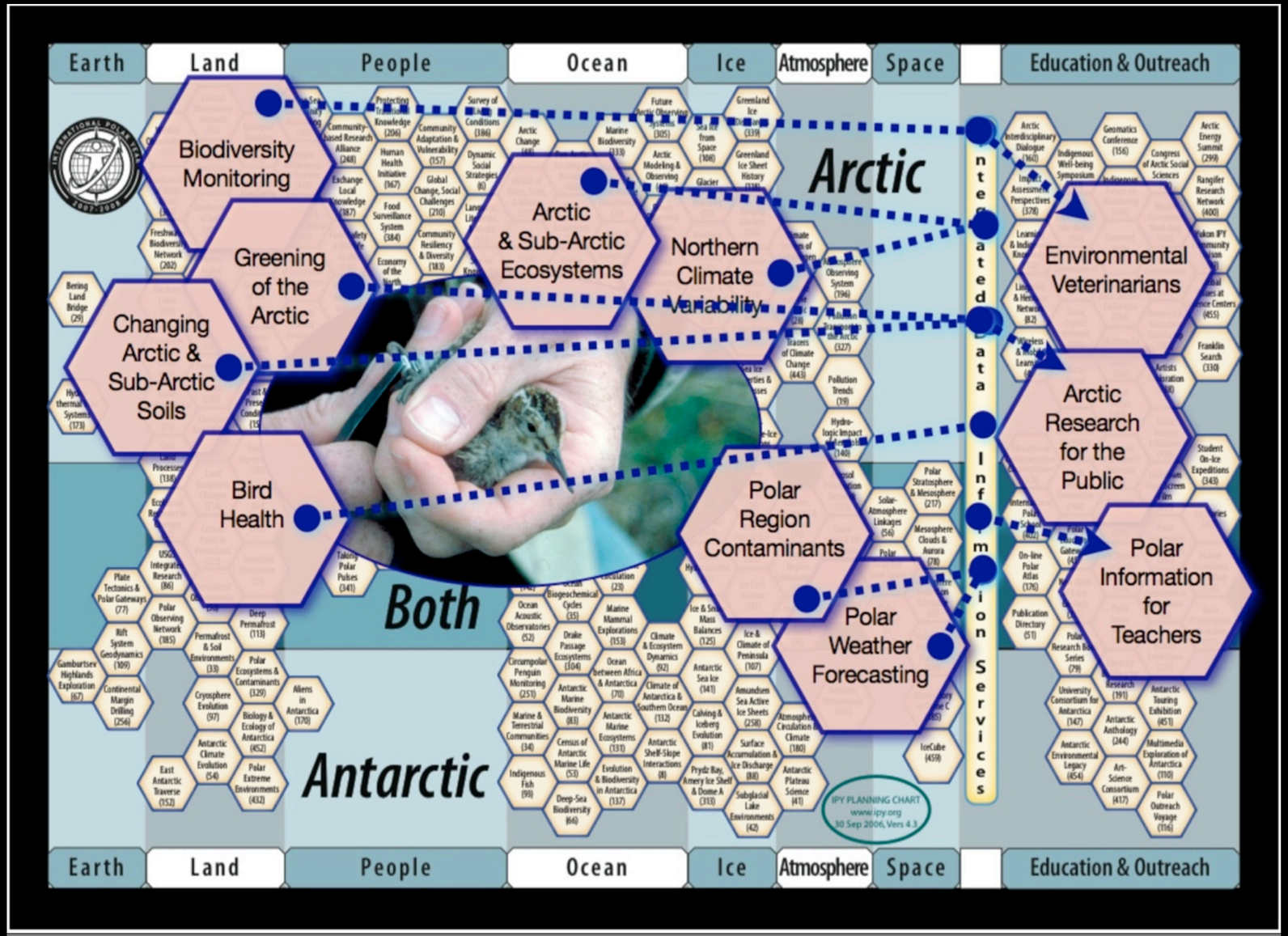
Irreplaceable Data

Replication of Results

Longitudinal Science

Interdisciplinary Science

Courtesy of Mark Parsons of the National Snow & Ice Data Center (SNIDC)



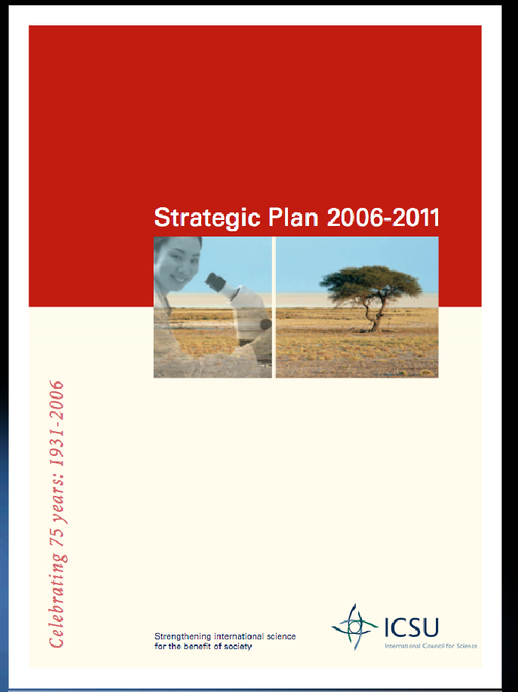
Lucy Nowell
lnowell@nsf.gov

- Irreplaceable Data
- Replication of Results
- Data for Longitudinal Science
- Interdisciplinary Science

ICSU
International Council for Science

- **Broadening Participation**

"The flow of scientific data and information is one of the most critical factors in promoting the participation of scientists...and ensuring the universality of science. As well as being of importance to science itself, publicly available scientific data are increasingly important for decision-making by governments and many sectors of society, from clinical practitioners to farmers."



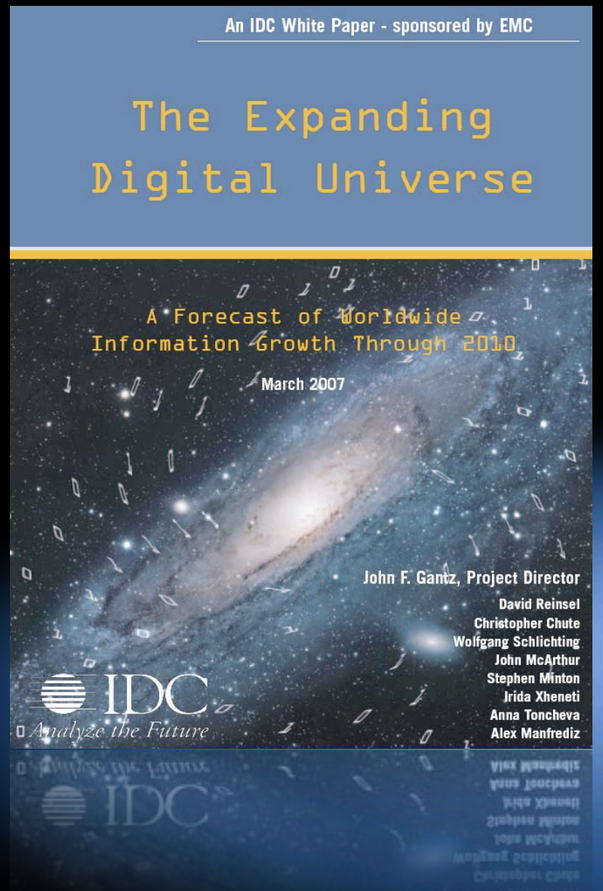
Why Preserve & Share Data?

National Science Foundation
Where Discoveries Begin

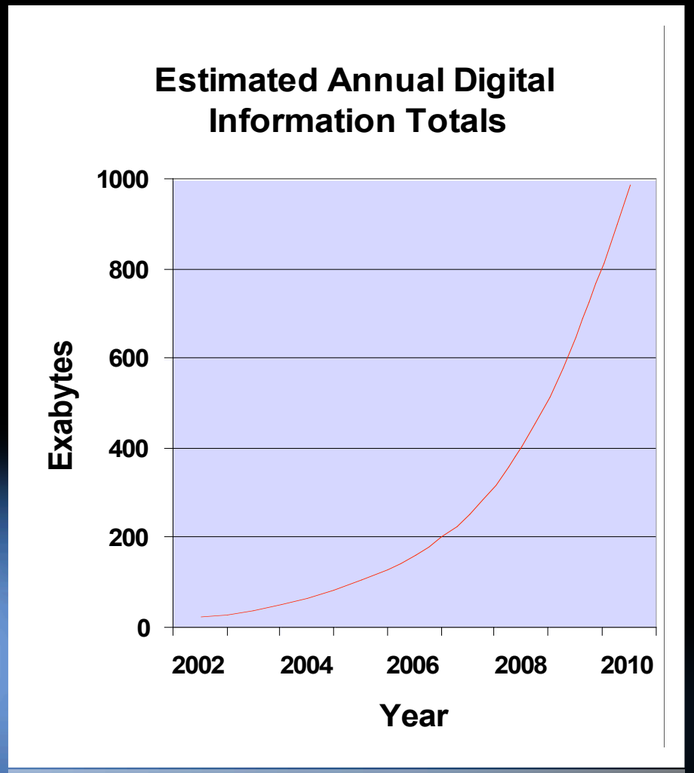
Lucy Nowell
lnowell@nsf.gov

Office of Cyberinfrastructure

IDC - Global Marketing Intelligence Company



"In 2007, the amount of information created will surpass, for the first time, the storage capacity available."



We Must Choose What to Keep

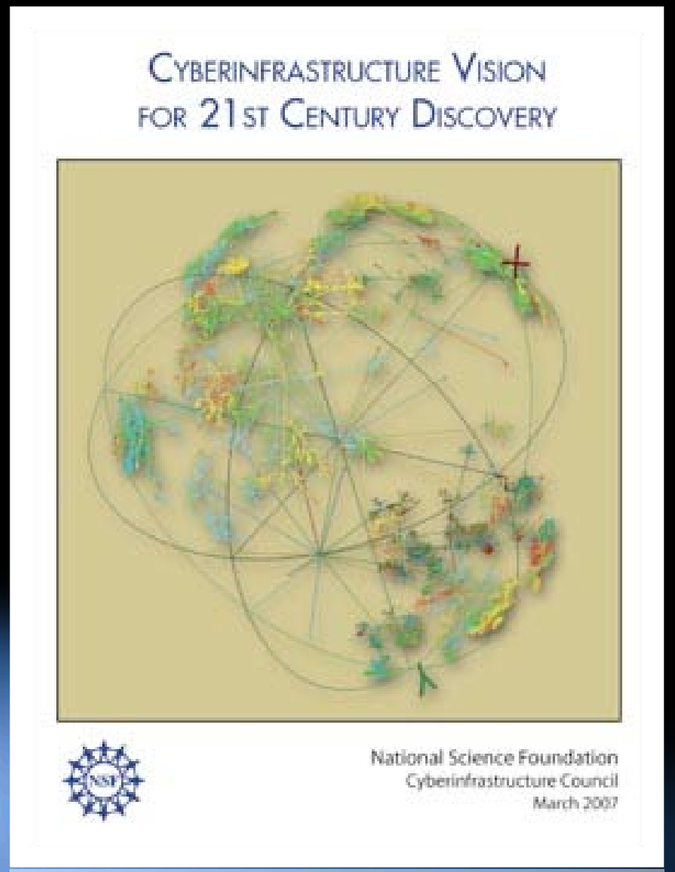
National Science Foundation
Where Discoveries Begin

Lucy Nowell
lnowell@nsf.gov

Office of Cyberinfrastructure

- “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.”

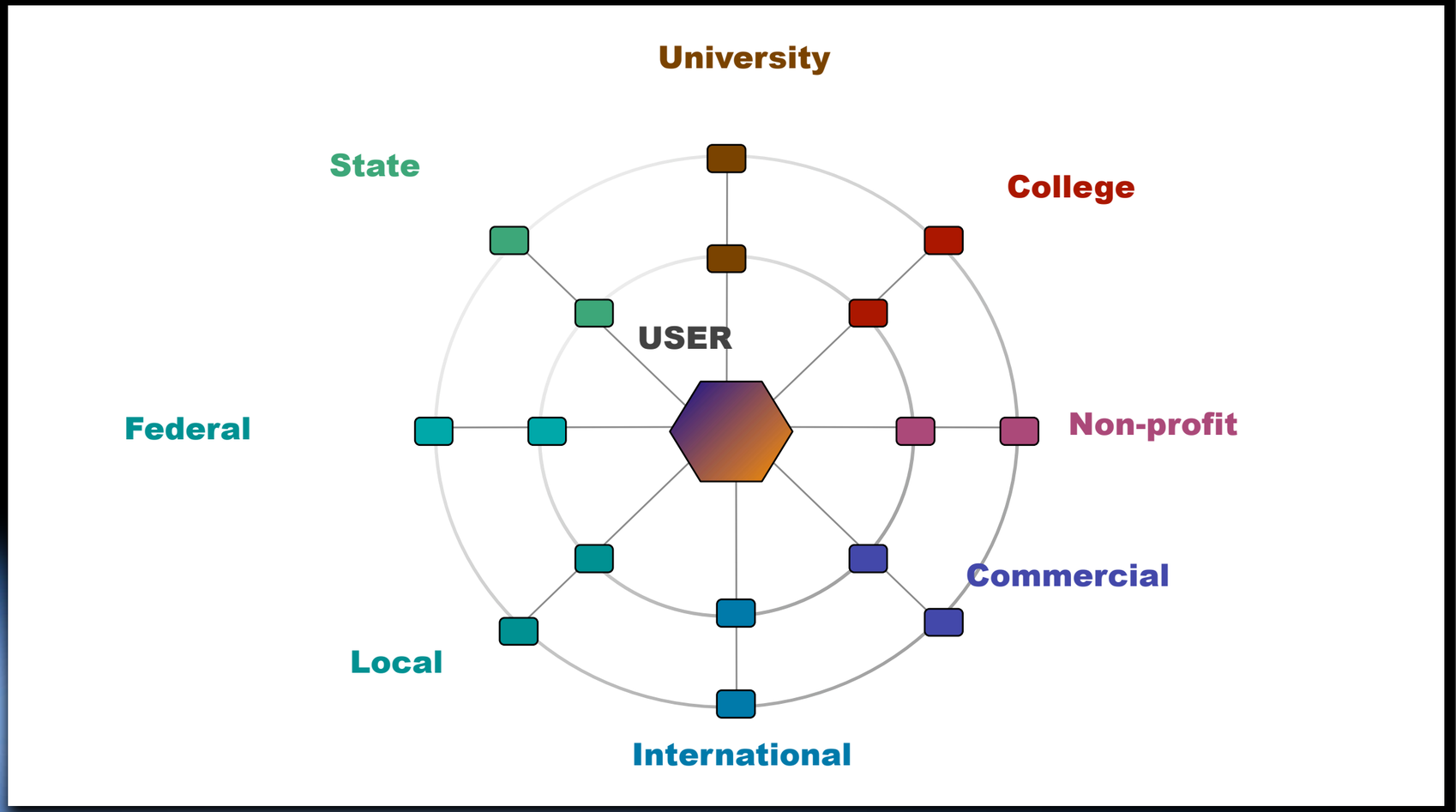
- “Scientific visualization, including not just static images but also animation and interaction, leads to better analysis and enhanced understanding.”



NSF Cyberinfrastructure Vision for 21st Century Discovery

Lucy Nowell
lnowell@nsf.gov

- User-centric
- Multi-Sector
- Open
- Extensible
- Evolvable
- Sustainable
- Nimble



Digital Data Preservation and Access Framework

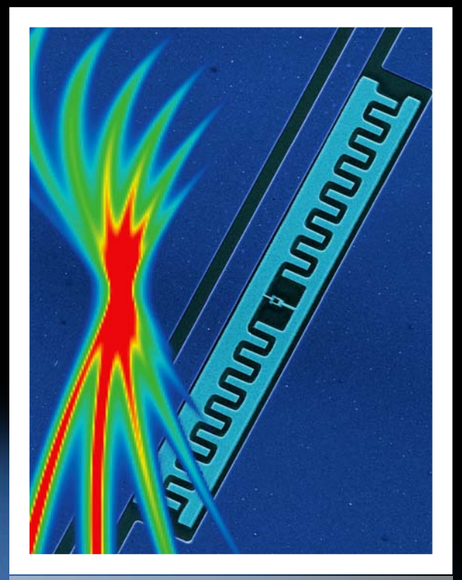
National Science Foundation
Where Discoveries Begin

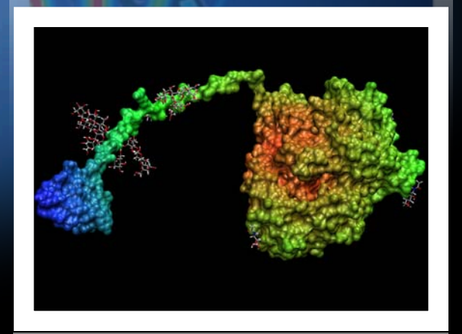
Lucy Nowell
lnowell@nsf.gov

Office of Cyberinfrastructure

DataNet Partners: Three Primary Goals

- 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.

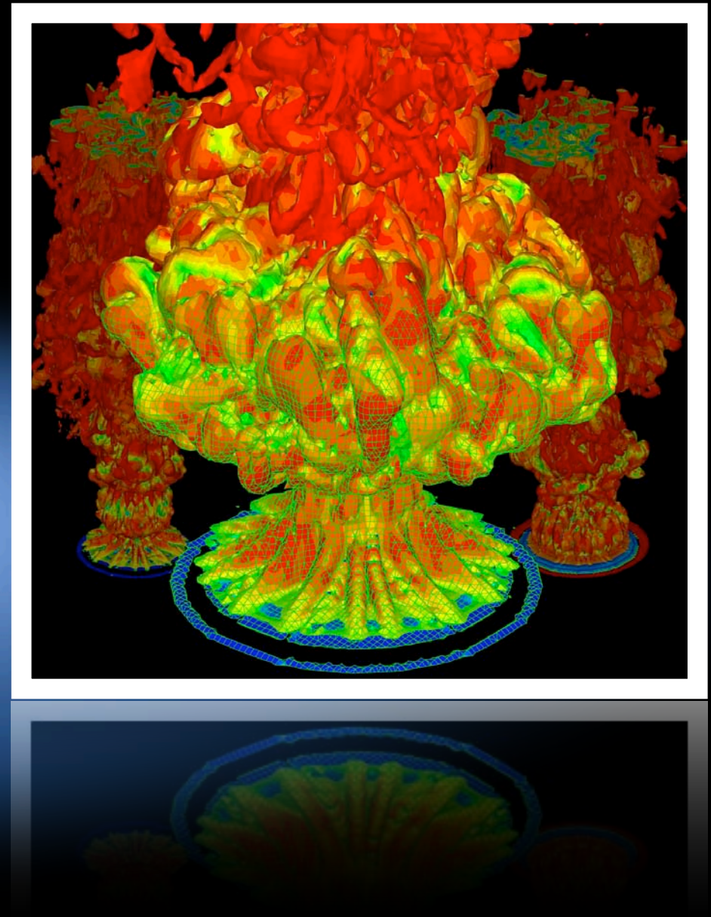




Lucy Nowell
lnowell@nsf.gov

DataNet Approach

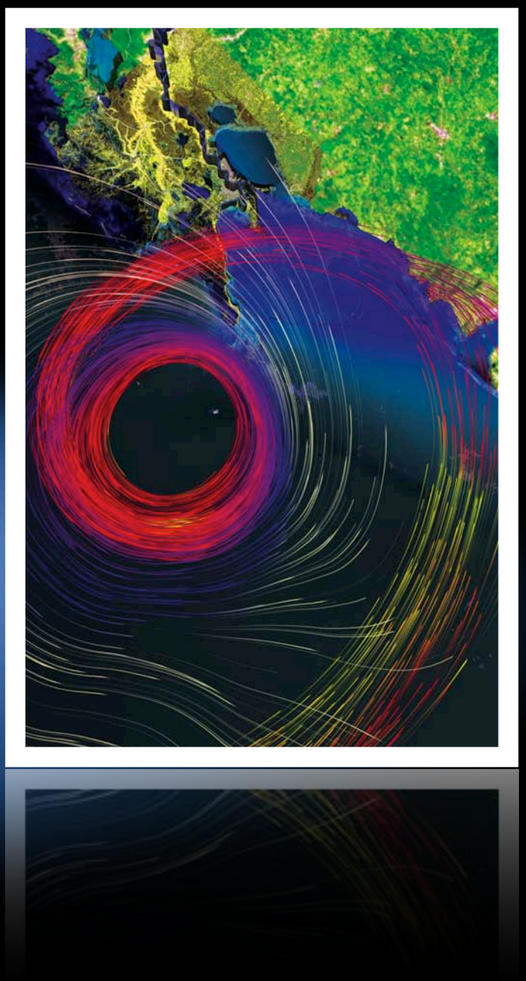
- 5 awards planned, 2 this year and 3 next year.
- Explore, demonstrate and understand diverse approaches to developing in sustainable ways with diverse scientific digital data content.
- Initial focus on several disciplinary areas, with active outreach to more communities and more disciplines



Lucy Nowell
lnowell@nsf.gov

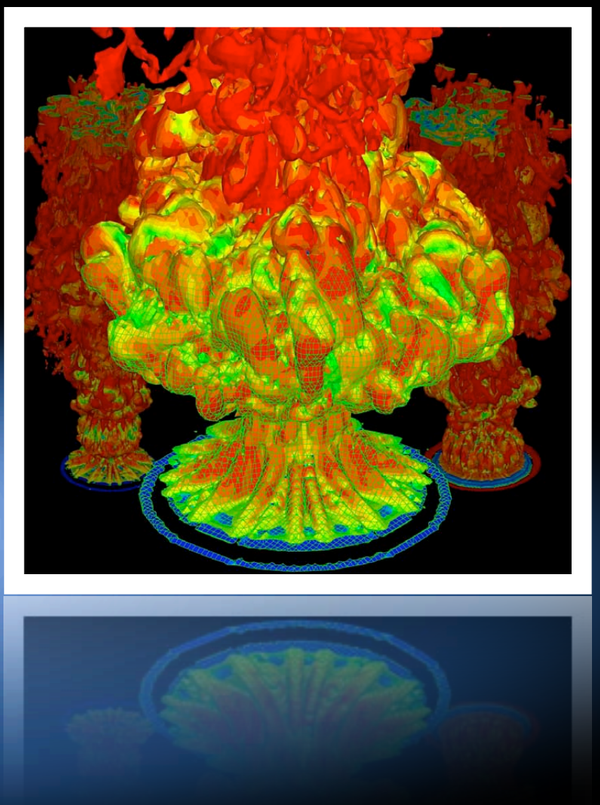
A Successful DataNet Partner Will...

- **Integrate** library and archival sciences, cyberinfrastructure, computer and information sciences, and domain science expertise.
- **Engage** at the frontiers of computer and information science and cyberinfrastructure with research and development to drive the leading edge forward.



Lucy Nowell
lnowell@nsf.gov

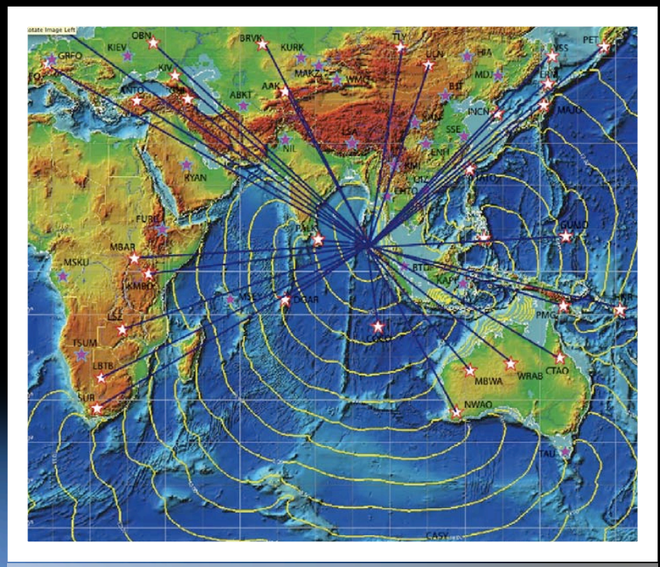
- **Provide** reliable digital preservation, access, integration, and analysis capabilities for science/ engineering data over decades-long timeline.
- Continuously **anticipate and adapt** to changes in technologies and in user needs and expectations.



Lucy Nowell
lnowell@nsf.gov

- Any information that can be stored in digital form and accessed electronically: numeric data, text, publications, sensor streams, video, audio, algorithms, software, models & simulations, images, etc.

- Caveats
  - Focus on data central to the scientific & engineering research & education mission of NSF.
  - Conversion to digital format is not supported by this program.

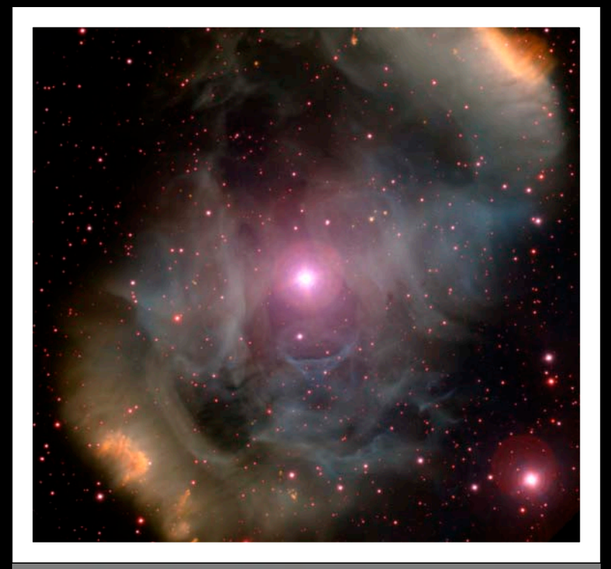


Digital Data

Lucy Nowell
lnowell@nsf.gov

DataNet Partner Responsibilities

- Provide for full data management life cycle
  - Data deposition/acquisition/ingest
  - Data curation & metadata management
  - Data protection, including privacy
  - Data discovery, access, use, & dissemination
  - Data interoperability, standard, & integration
  - Data evaluation, analysis, & visualization
- Engage in research central to DataNet responsibilities
- Education & training
- Community & user input assessment
- International engagement – collaborate & coordinate closely with preservation & access organizations to catalyze formation of a global data network
  - Foreign collaborators are expected to secure support from their own national sources.



Lucy Nowell
lnowell@nsf.gov

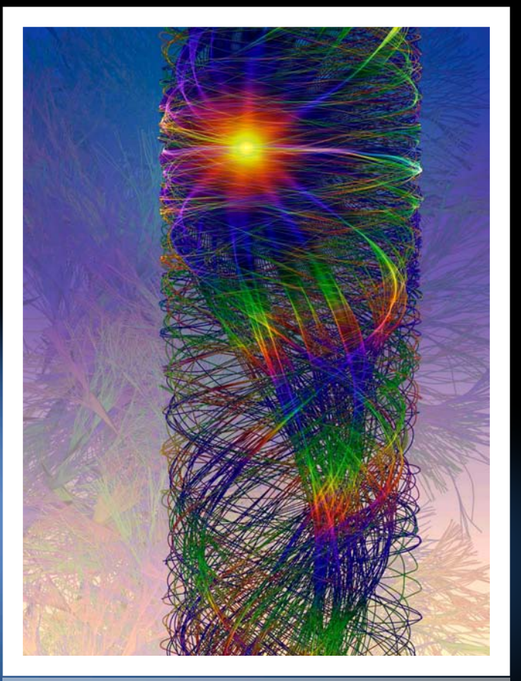
- **Science** must drive the proposal.
- Start with the **science drivers**: What data are needed? Where do they reside now? What is needed to integrate and analyze them that is missing now?
- Partnerships should be based on what needs to be accomplished to support the science.
- **Every** proposal must address **every DataNet requirement**.
- Every proposal must be of interest to **at least two NSF Directorates**.

Where to Begin?

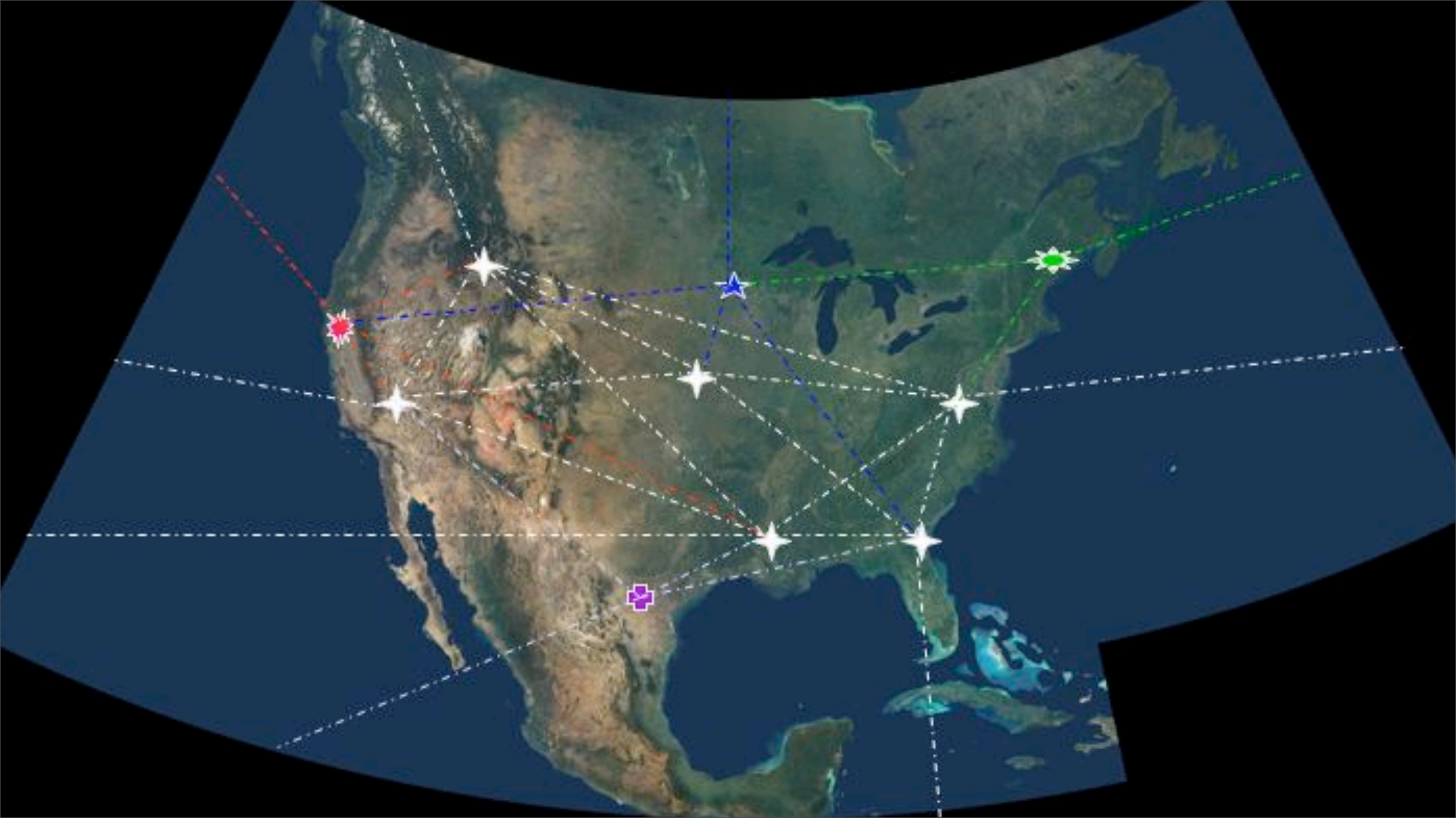
Lucy Nowell
lnowell@nsf.gov

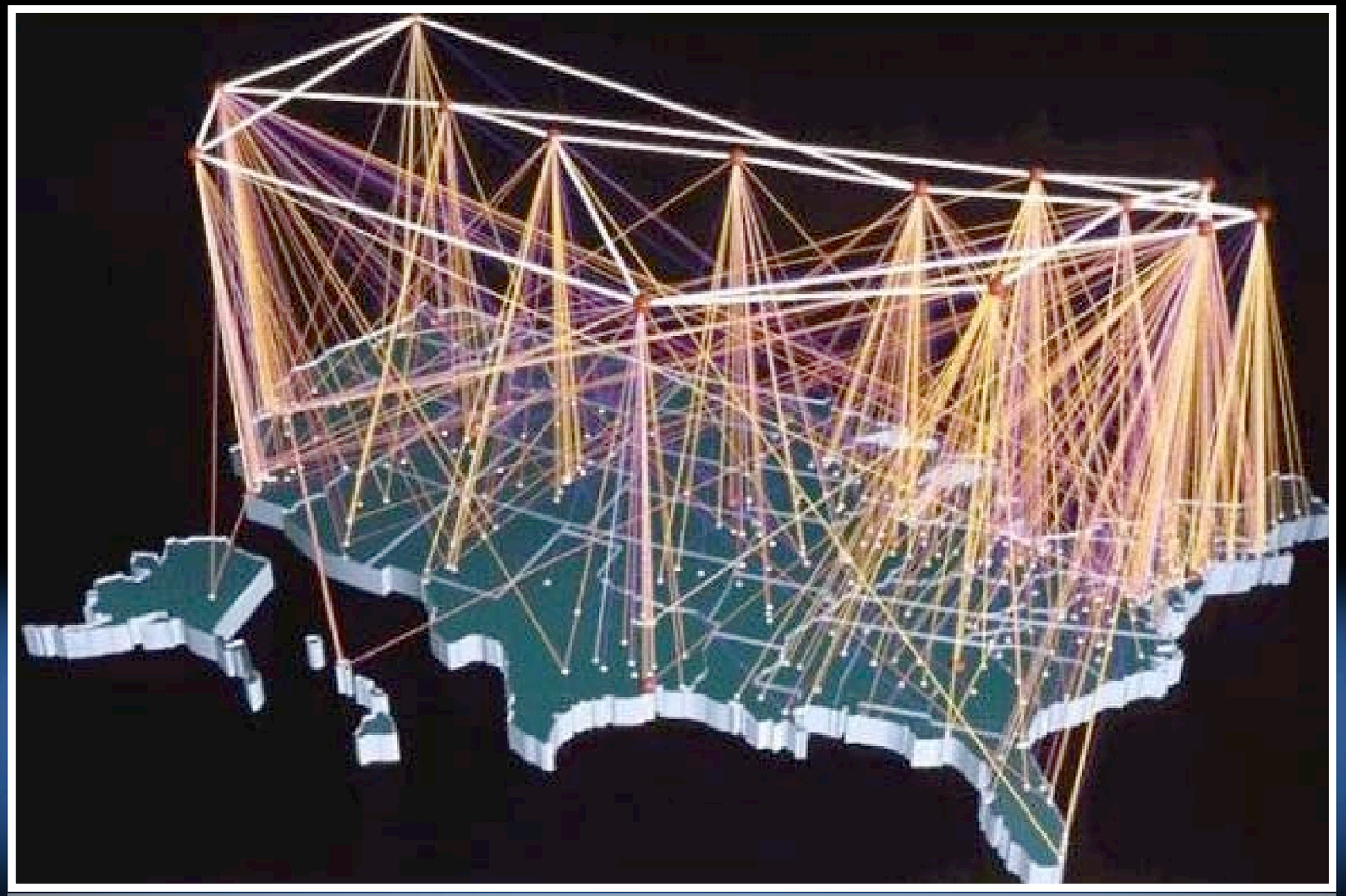
DataNet Status

- Two proposals undergoing continued review, with planned recommendation to the National Science Board for approval in December.
- FY09 DataNet Competition:
  - Preliminary Proposals due November 13, 2008
  - Invited Full Proposals due May 15, 2009
  - Site Visits planned for August 2009



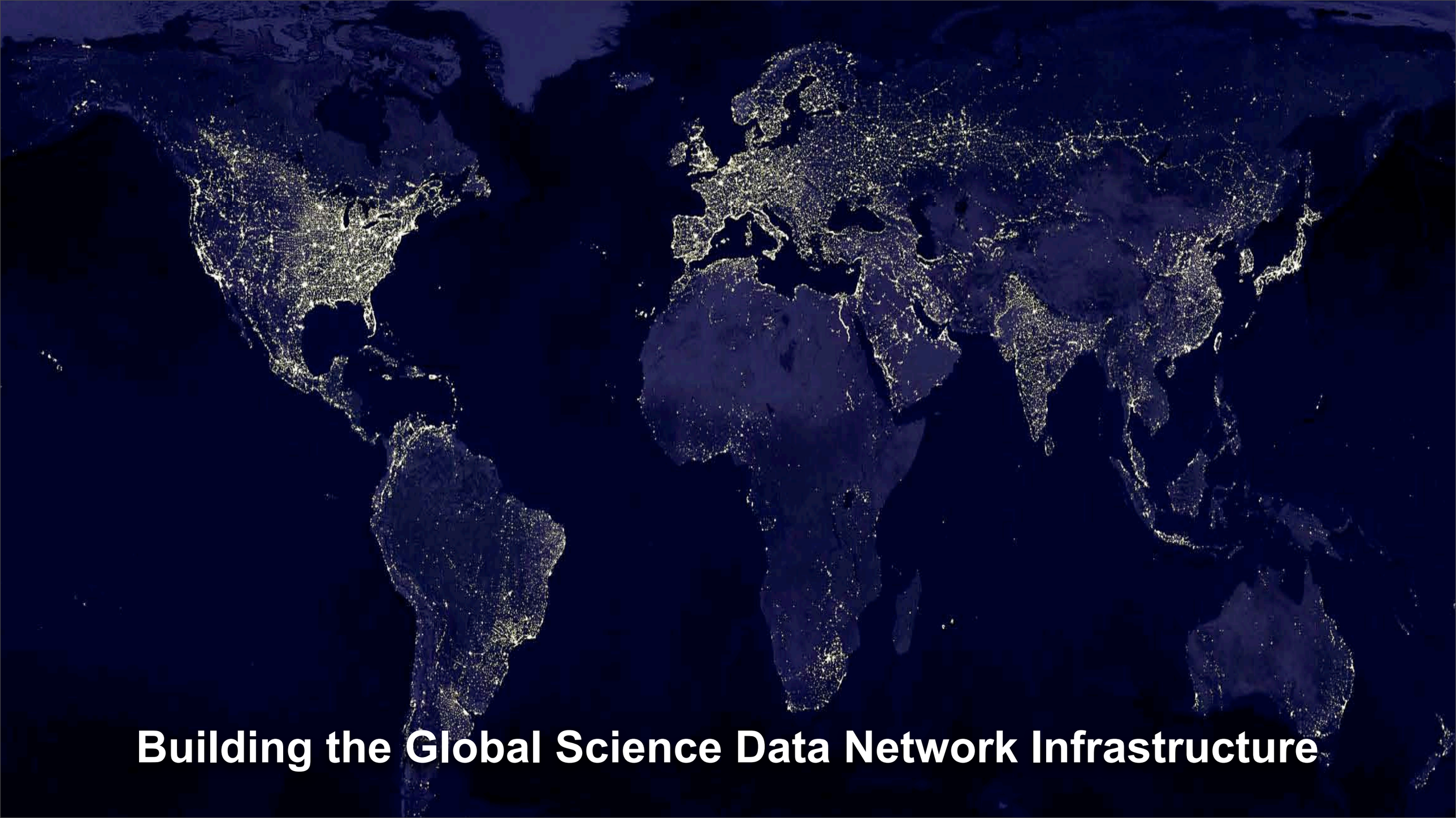
Lucy Nowell
lnowell@nsf.gov





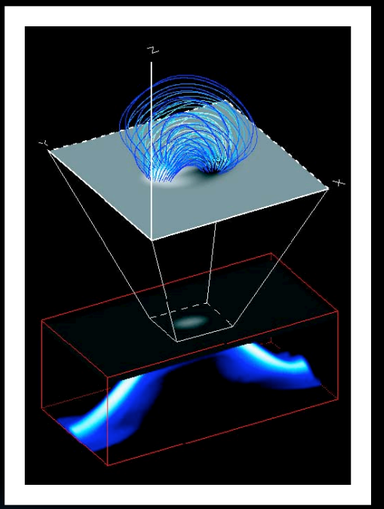
Goal: A DataNet Partners Network of Networks

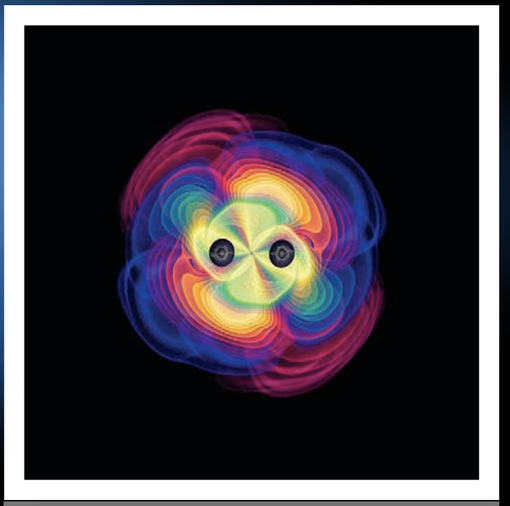
Lucy Nowell
lnowell@nsf.gov


Building the Global Science Data Network Infrastructure

Critical Needs & Opportunities

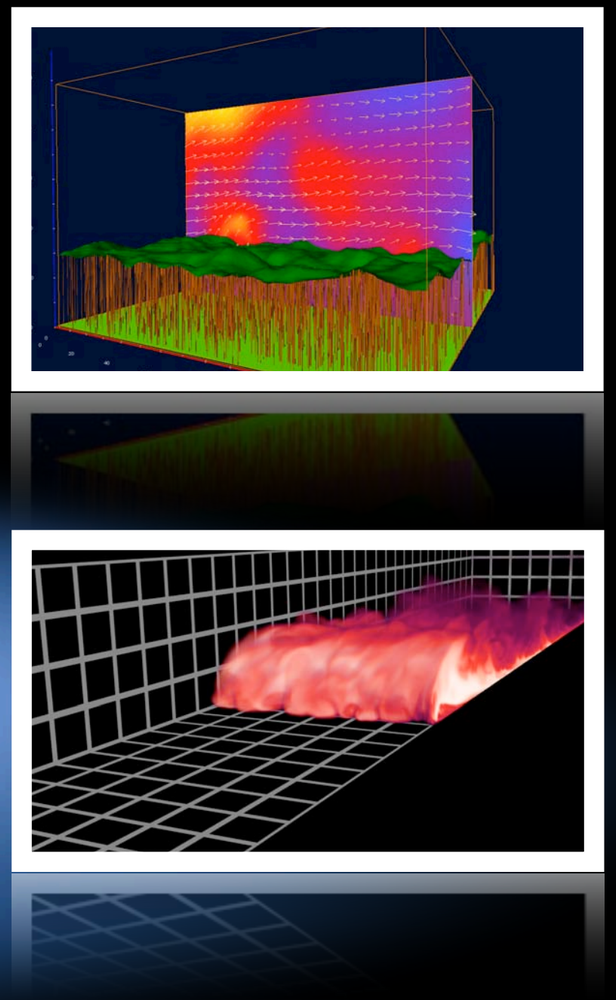
- Science-driven proposals for more DataNet Partners
- Extending/evolving collections & services of DataNet Partners
- Outreach
  - Changing the culture, practices & reward structures
  - Promote reuse & repurposing of data
- Support for interdisciplinary grand challenge research
- Interoperability across collections





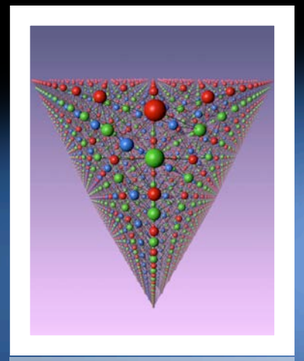
Lucy Nowell
lnowell@nsf.gov

- Integrating DataNet Partners with other CI activities
- Hardware & software architectures for data-intensive computing
- Information synthesis, data mining & analytic tools
- Broaden participation via data sharing/re-use
  - Geographically
  - K-12, Undergraduate, etc.
- Sustaining DataNet Partners for Phase 2
- Extend DataNet Partners beyond the first 5 awards



Lucy Nowell
lnowell@nsf.gov

“Without investments made by previous generations, we would not enjoy the seemingly invisible infrastructure that makes our modern lives possible. It goes without saying that if we don’t make similar investments now we will rob future generations of the quality of life they should enjoy.”



Princeton University Engineering School, Feb. 19, 2008, Greatest Technological Research Challenges of the 21st Century Identified by Expert Panel. *Science Daily*

http://www.sciencedaily.com/release/2008/02/080215151157.htm

Greatest Technological Challenges of the 21st Century

Lucy Nowell
lnowell@nsf.gov


Thank
you!