



**Enabling the nation's future through  
discovery, learning and innovation**

The background of the slide is a blurred image of a computer keyboard, showing various keys in shades of blue and white. The text is overlaid on this background.

# NSF Draft Strategic Vision for Data, Data Analysis, and Visualization

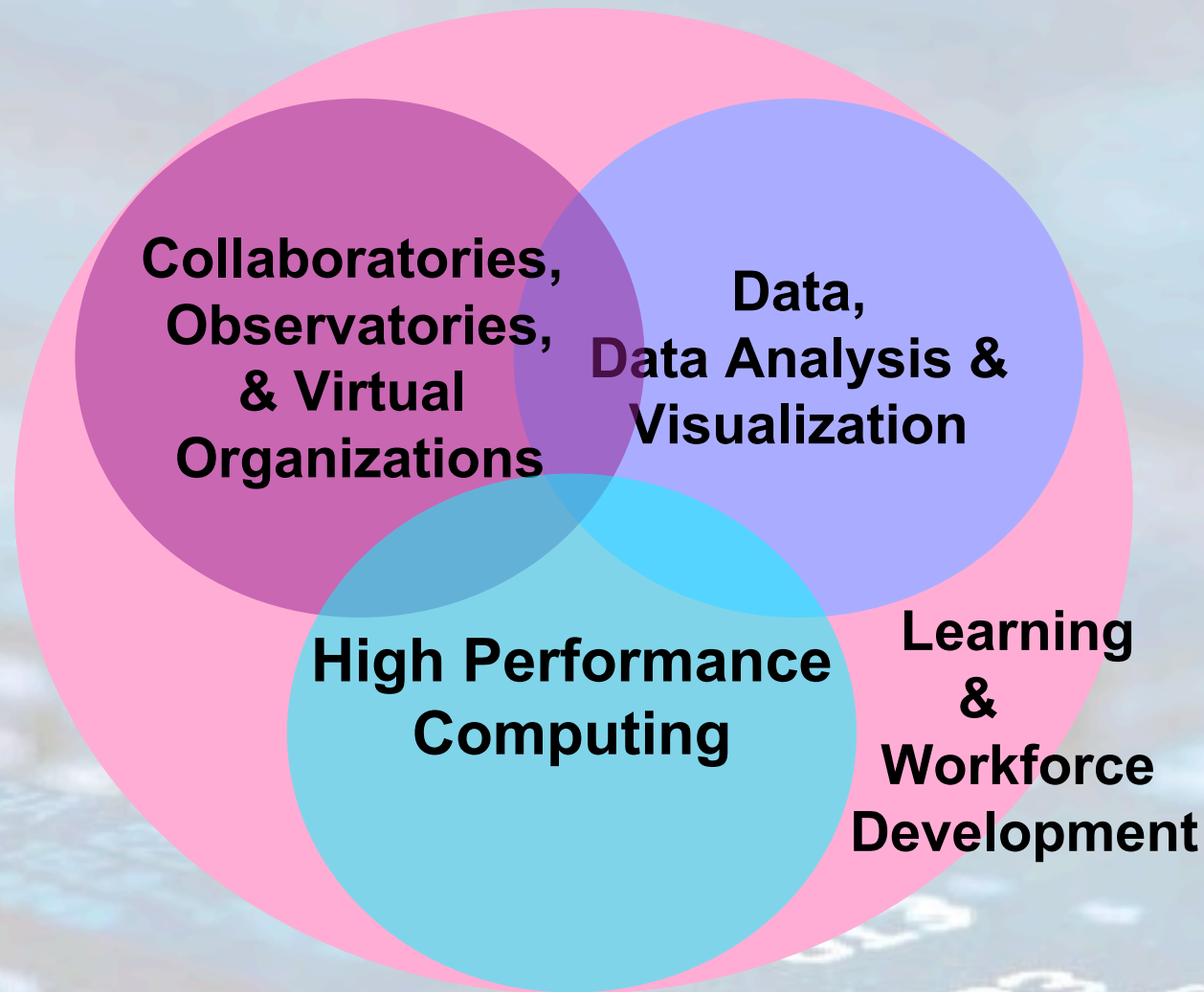
**Chris Greer**  
**National Science Foundation**

# Cyberinfrastructure ...

... is the organized aggregate of technologies that enable us to access and integrate today's information technology resources—data and storage, computation, communication, visualization, networking, scientific instruments, expertise—to facilitate science and engineering goals.

*- Fran Berman, Director, SDSC*

# ***Cyberinfrastructure Components***



“Sometime in the 2010s, if all goes well, the Large Synoptic Survey Telescope will start to bring a vision of the heavens to Earth.

Suspended between its vast mirrors will be a three-billion-pixel sensor array, which on a clear winter night will produce 30 terabytes of data. In less than a week this remarkable telescope will map the whole night sky ...

And then the next week it will do the same again - building up a database of billions of objects and millions of billions of bytes.”

*Nature 440, 383*

# BioLit: Tools for New Modes of Scientific Dissemination

## The Knowledge and Data Cycle

0. Full text of PLoS papers stored in a database

1. A link brings up figures from the paper

2. Clicking the paper figure retrieves data from the PDB which is analyzed

4. The composite view has links to pertinent literature text and back to the PDB

3. A composite view of journal and database content results



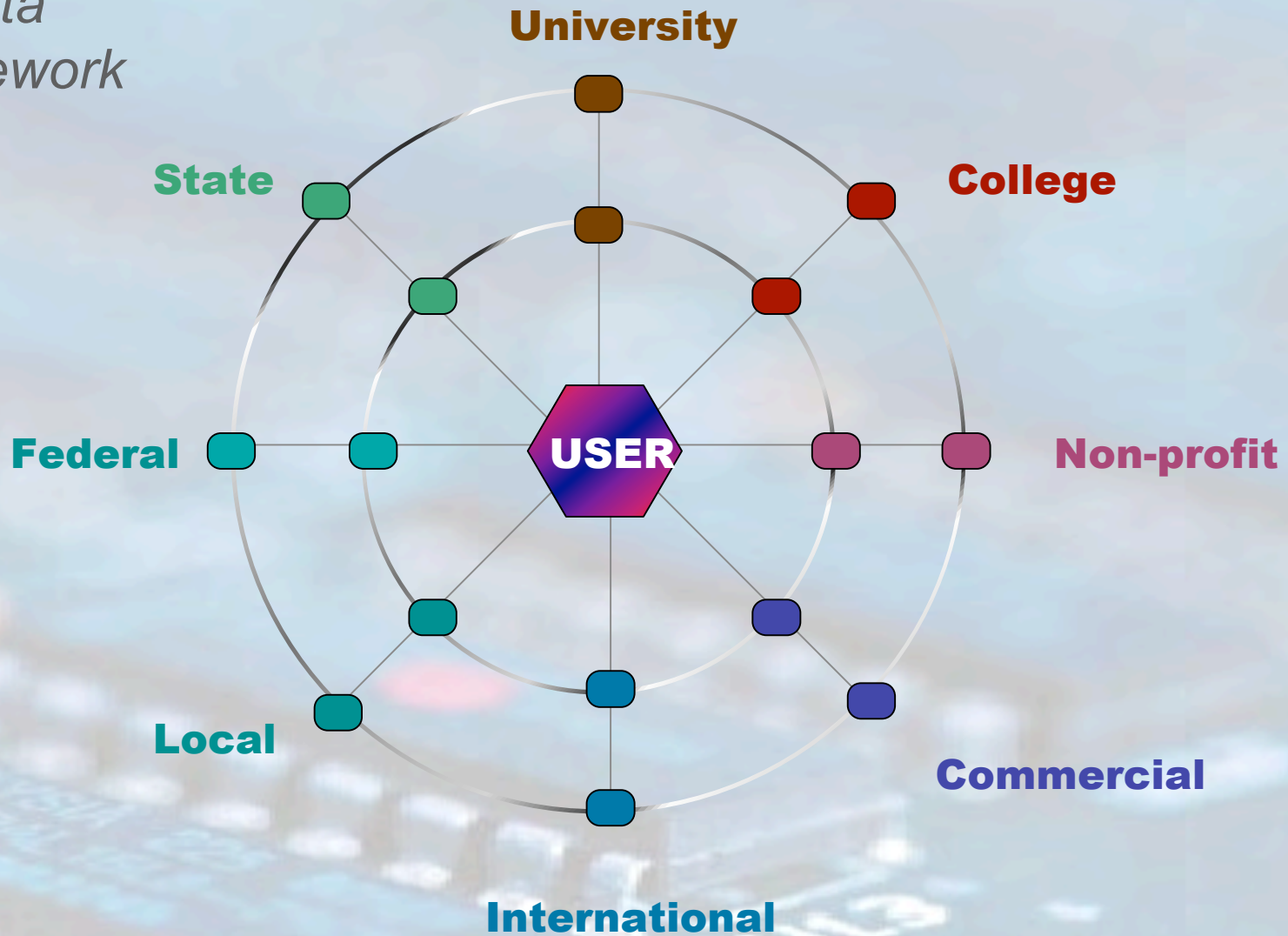
# Vision:

“... a vision in which science and engineering digital data are routinely deposited in 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.”

# Goals:

- To catalyze the development of a system of science and engineering data collections that is open, extensible and evolvable.
- To support development of a new generation of tools and services facilitating data acquisition, mining, integration, analysis, and visualization.

*National Digital  
Data  
Framework*



User-centric, Multilevel, Nimble, Sustainable, Reliable

# Principles:

- Data generated with NSF funding will be accessible and reliably preserved
- Research/education opportunities determine investment priorities
- Broad community engagement is necessary in reviewing and prioritizing data activities

# Principles (cont'd):

- Data is only useful if it can be found, understood, and analyzed
- Legitimate privacy, confidentiality, and intellectual property rights must be protected
- International, interagency, and public-private partnerships are essential

# Plan of Action

- Coherent Organizational Framework
- Flexible Technological Architecture
- Coherent Data Policies

## Plan of Action:

### A. Coherent Organizational Framework

- Diversity of approaches
- Communities of practice
- Community proxy roles of collections
- Dynamic and evolving system

## Plan of Action:

### B. Flexible Technological Architecture

- Layered capabilities
- Metadata
- Data analysis and visualization tools
- Promoting use and stability of standards

## Plan of Action:

### C. Coherent Data Policies

- Transparent policy frameworks
- Data management plans
- Interagency coordination
- International cooperation

# Summary

Strategic plan seeks to:

- Promote a change in culture
- Catalyze development of a national digital data framework
- Support new generations of tools, services, and capabilities

# NSF'S Cyberinfrastructure Vision for 21<sup>st</sup> Century Discovery

[http://www.nsf.gov/od/oci/ci\\_v5.pdf](http://www.nsf.gov/od/oci/ci_v5.pdf)

**Send comments to:  
ciinput@nsf.gov**