Thoughts on cyberinfrastructure for the geosciences, emphasizing aspects complementary to high-performance computing

Dave Fulker for the Unidata Policy Committee
17-Apr-06
Motivation

- Convert cyberinfrastructure (CI) from concept to action, maximizing benefits for GEO
- Close gap between CI for supercomputing and CI in other (softer) dimensions
- Subtext: encourage inclusion of Unidata-like activities in plans/budgets for CI
Cyberinfrastructure Definitions

- **CI as underlying foundation (functionality)**
  - Well tested, dependable & adaptable

- **CI as a system of public works (common good)**
  - Transparent, standardized & low-cost or free

- **CI as a socio-technical environment (evolution)**
  - Emphasis upon ongoing, participatory design
    - "Deep & enduring changes are not technological but social & cultural in their core"
    - "Computers, different from passive technologies, can be extended in ways the designers did not foresee"
      - Gerhard Fischer (paraphrased)
Setting a Direction

A Proposed CyberGeo Statement of Purpose:
To establish a reliable, socio-technical environment that leverages creativity & learning in the geosciences.

Underlying assumption: NSF/GEO intends to foster & support cyberinfrastructure projects & programs—which I’ve dubbed “CyberGeo”—that benefit the geosciences.
Some Guiding Principles

- **Leveraged Activities**: some highly specialized (leading edge); others nearly universal (multi-disciplinary)

- **Dual Priorities**: technologies that A) serve large parts of GEO or B) enable advances otherwise unachievable

- **Evolving Boundaries**: experimental/discipline-specific concepts (via abstraction & engineering) ⇒ infrastructure

- **Central versus Distributed Activity**: an evolving balance—many functions that now are best performed in a large, central facility eventually will become well-matched to desktop or departmental systems
More Guiding Principles

- **Elevating Semantics**: general cyberinfrastructure trend ever-higher levels of meaning embedded in tools & data flows

- **Transcending the Disciplines**: common abstractions (IDV, LDM, NetCDF, CDM, GALEON, e.g.) enable & support interdisciplinary advances...

- **Standards/Transparency/Openness**: international standards/certifications; policies on accessibility & use...

- **Other Principles?**
Strategic Questions

- How might a broad concept of CyberGeo be converted to program activities that
  - Are practical & effective (in an NSF & GEO context)?
  - Yield the full promise of cyberinfrastructure?

- The answer proposed here:
  - Define “full promise” as 5 classes of activity that should be enhanced by cyberinfrastructure
  - Drawing on GEO history & promising experiments, lay out specific services & technologies that
    - Represent reasonable units of work
    - Cover all classes of CI-enhanced activity
Coverage Map

QuickTime™ and a TIFF (LZW) decompressor are needed to see this picture.
Classes of CI-Enhanced Activity

1. Earth-Systems Observation
   - Extending the measurements of the earth system

2. Earth-Systems Simulation
   - Enhancing models/theories of earth-system processes

3. Data Analysis & Synthesis
   - Quantitatively linking observation, theory & perception

4. Scholarly Communication
   - Exchanging/reusing/enriching the artifacts of science

5. Learning & Decision Making
   - Building scaffolds for cognition
End-to-End Use Cases

- Extreme events
- Multidisciplinary studies of water, volcanoes...
- Data repurposing
  - Data-access methods
    - Encapsulation
    - Web services...
  - Metadata
    - Higher semantics
    - Polymorphism...
Coverage Map

QuickTime™ and a TIFF (LZW) decompressor are needed to see this picture.
Some Questions for Discussion

- Is this complementary to other CI reports?
- Should it be less Unidata-centric and, if so, what are the best steps for getting there?
- Is the services/technologies granularity OK?
- Should services/technologies be sequenced or prioritized and, if so, by what means?
  - Individually or as a package?
  - In workshops or focus groups?
- Should this address programmatic issues?