The Situation Today

Earth Science **Stuff** is (still) hard to use...

- data
- science tools / svcs
- analysis results
- knowledge about
  - data
  - tools
  - analysis methods

- find
- share
- reuse
- put together
  - data + data
  - data + tool
  - tool + tool
  - desktop + online svc
Proposed: Convergent Evolution to an Earth Science Collaboratory (ESC)
Cyberinfrastructure Services

*used by all other components*

- **Security**
  - authentication
  - authorization
  - code audit/padded cell
  - integrity checking
- **Social**
  - tagging
  - sharing
  - discussions
  - groups
  - reputation
- **Cloud**
  - elastic provisioned storage and computing
- **Discovery**
  - data, tools, workflows, experiments
  - search by keyword, variable, time, author
- **Information Mgmt**
  - provenance
  - identifiers
  - archive
- **Semantic Web**
  - data ontology
  - tools ontology
Why now?

• Because we can do it (finally)!
  – Advances in standards acceptance and implementation (e.g., OPeNDAP, autoconf)
  – A consistent, coherent, loosely coupled architecture encapsulates complexity and maximizes flexibility
  – Social networking has reached the mainstream
  – Key lessons can be learned from prior efforts
  – Cloud technology helps with provisioning: resources, tools, workflows
    • Suggestion: entice Nebula with a Grand Challenge to get their help

• The need is growing
  – Interest in working with multiple datasets is growing
  – Calls for transparency and reproducibility are growing
Prior Art

- Talkoot, myexperiment.org – workflow sharing, virtual notebooks
- Earth System Grid – provisioned tools, format standards/checkers
- Land Information System – OPeNDAP as access infrastructure
- Earth Science Modeling Framework – programmatic approach to integration
- Giovanni, LAS – community services/tools
- Nebula – cloud provisioning
- RAMADDA – management of diverse information objects
- NASA Earth Exchange – collaborative framework for NASA Earth Science projects
- HUBzero, Zooniverse – science collaboration frameworks
- EOSDIS – Federated data centers, federated discovery
How to move forward?

• Programmatic element
  – RFC to community on feasibility, challenges, approach
  – Followed by RFPs for component and integration

• Prototype element
  – Narrow end-to-end prototypes
  – Followed by refactoring, broadening and convergence