National Science Foundation EarthCube Layered Architecture
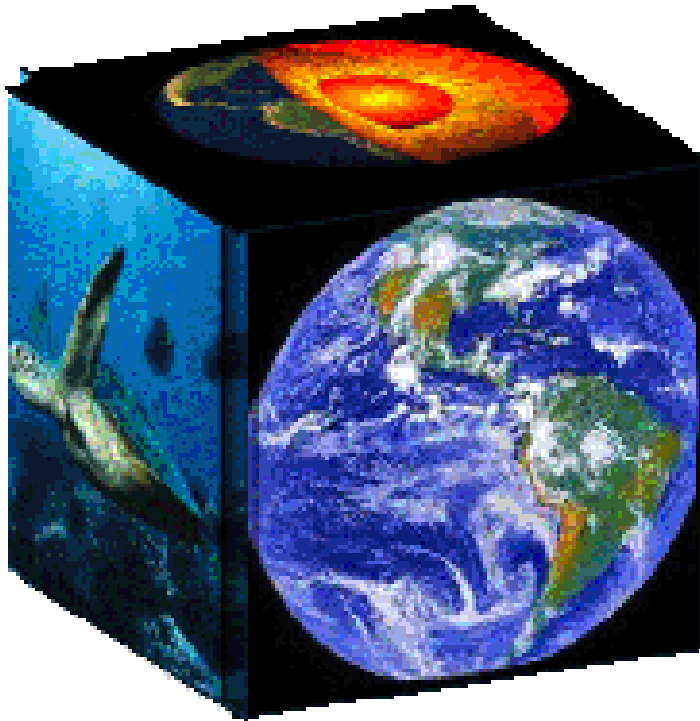
Chris MacDermaid

Colorado State University Cooperative Institute for Research in the Atmosphere (CIRA)
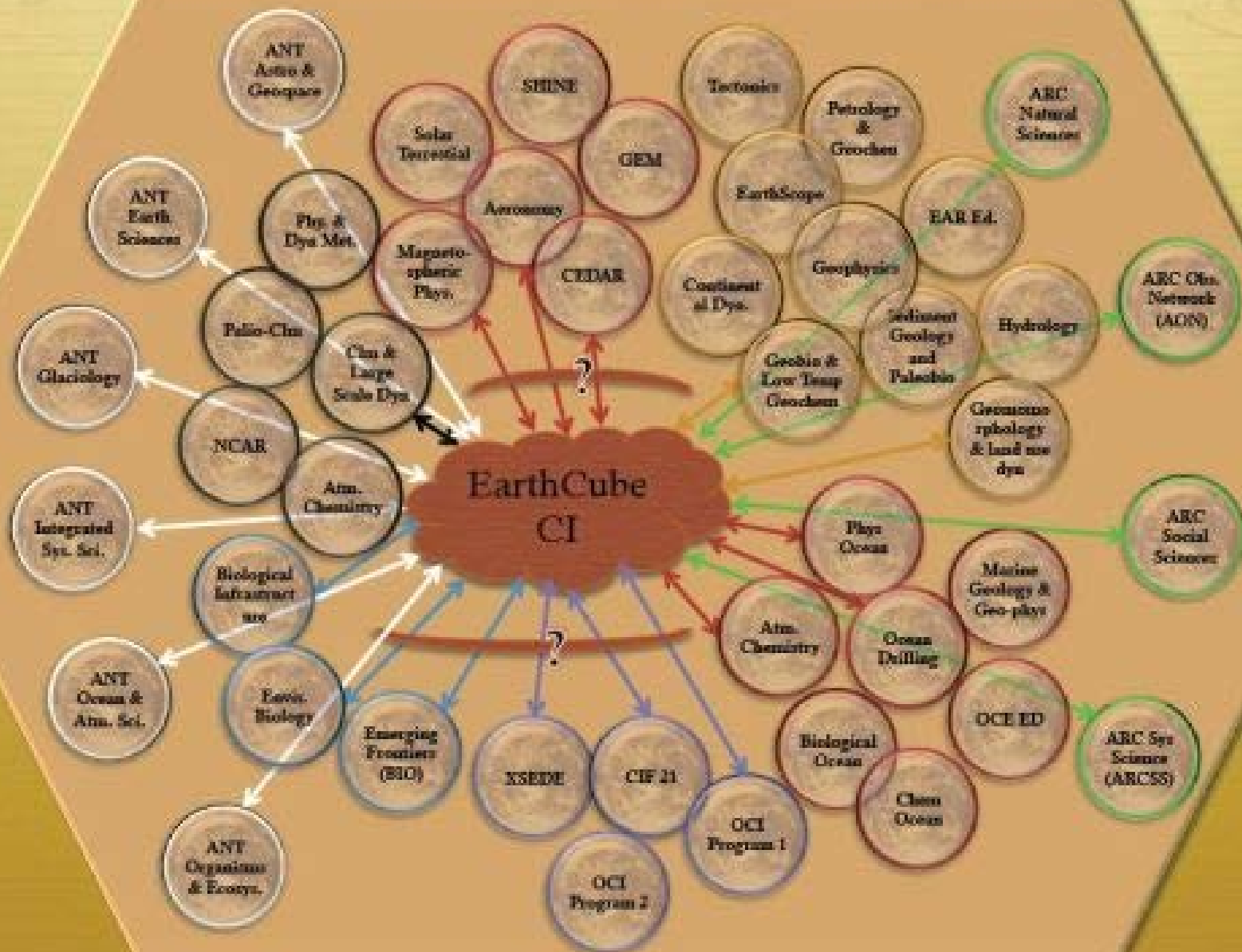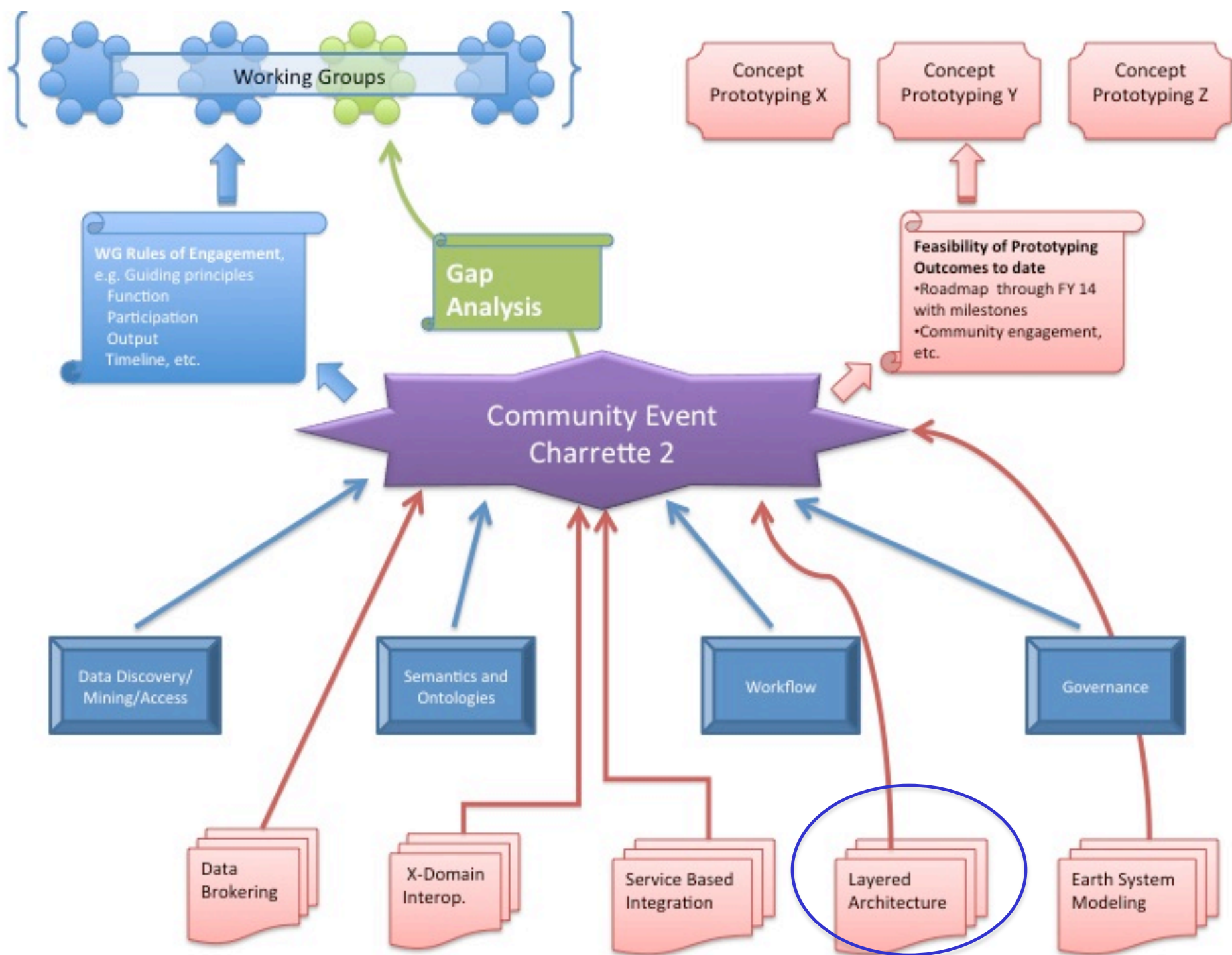
NOAA/OAR/ESRL/GSD

What is NSF EarthCube?

- NSF activity to create a data and knowledge management system for the 21st Century

- Goal is to transform the conduct of research by supporting the development of a community-guided cyberinfrastructure

- Funded by NFS Directorate of Geosciencies and the Office of Cyberinfrastructure

EarthCube Target Community

Layered Architecture Concept Award Team Members

Reagan Moore   (UNC-CH/DICE)
Ilkay Altintas   (UCSD)
David Arctur   (OGC)
Lawrence Band   (UNC-CH/IE)
Liping Di   (GMU)
Janet Fredericks   (WHOI)
Jeff Horsburgh   (Utah State University)
Yong Liu   (UIUC / NCSA)
Chris MacDermaid   (Colorado State Univ.)
Brian Miles   (UNC-CH/IE)
Michael Schoffner   (RENCI)
Antoine de Torcy   (UNC-CH/DICE)
Weiguo Han   (GMU)

Layered Architecture Goals

- Use Layered Architecture to explore interoperability between geoscience communities

- Develop Interoperibility testbed to identify:
  - The existing infrastructure components
  - The current types of interaction mechanisms
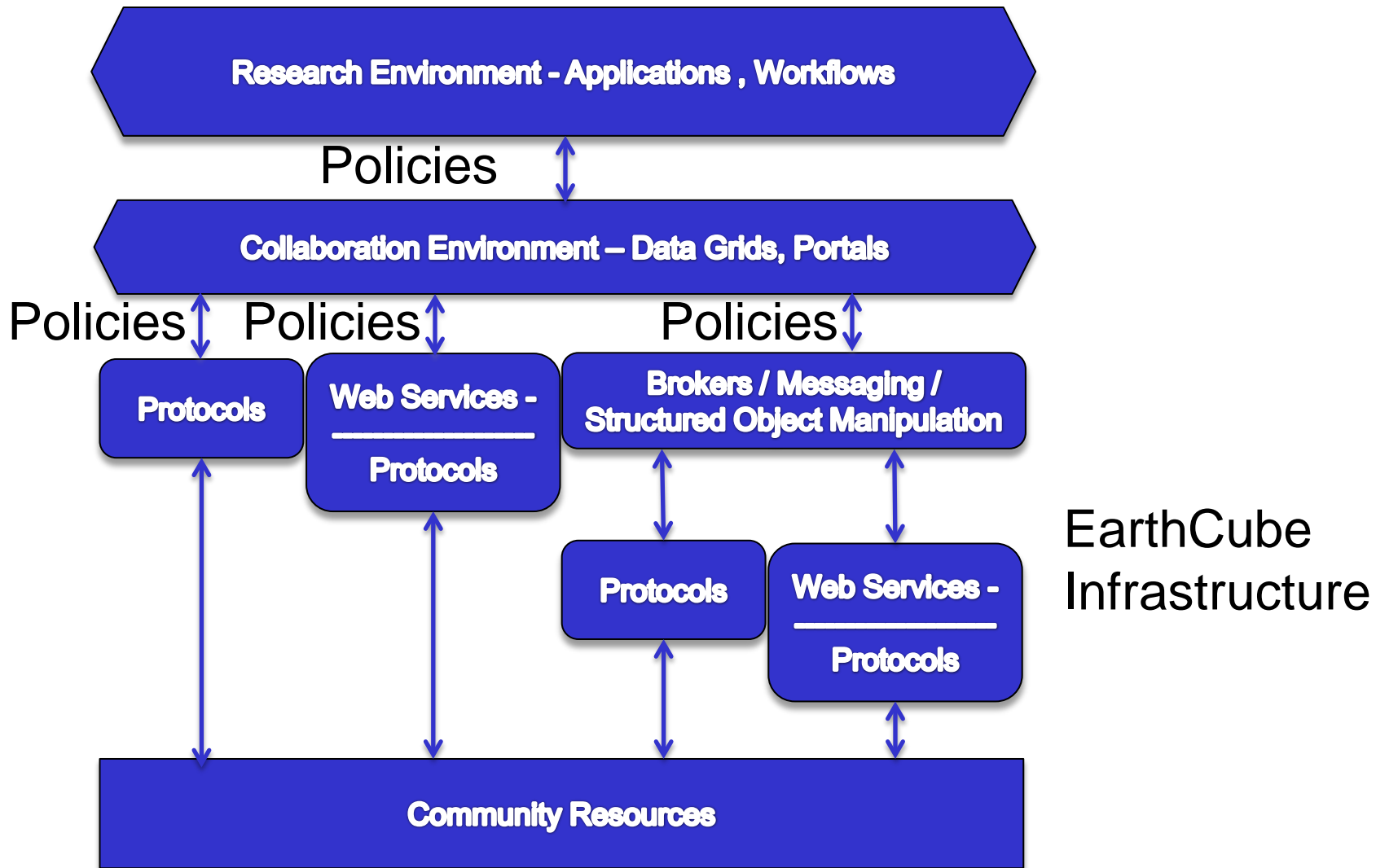 - The mechanisms that need to be implemented to improve interoperability

Layered Architecture Use Case

- Demonstrate reproducible science
  - Registration of data,
  - Storage of data
  - Sharing of data
  - And ability to re-execute workflows
- Automate data dissemination
- Integrate community resources

Two Main Layers

- Collaboration environment
  - Workspace supporting collaborative research
  - Formation of shared collection for sharing of files and workflow analyses
  - Data grids / scientific portals / Community models
- Community resources
  - Resources provided for use by discipline
  - Data repositories, information catalogs, ontologies, analysis services, presentation services

Loosely Coupled – Layered Architecture

Research Environment - Applications , Workflows

Policies

Collaboration Environment – Data Grids, Portals

Policies Policies Policies

Protocols

**Web Services -
_____
Protocols**

**Brokers / Messaging /
Structured Object Manipulation**

EarthCube
Infrastructure

Protocols

**Web Services -
_____
Protocols**

Community Resources
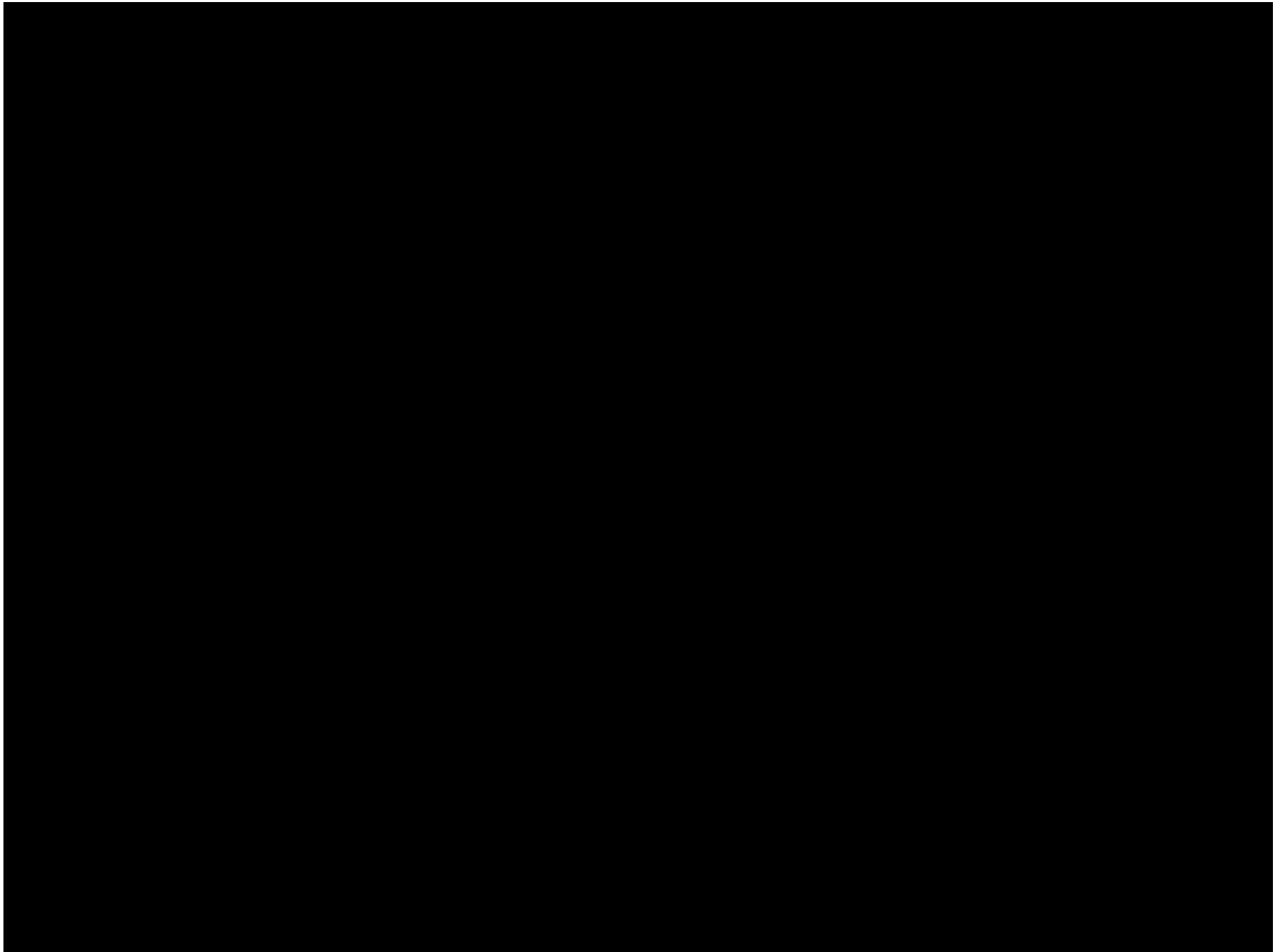
EarthCube Future

- EarthCube Test Enterprise Governance
    - Full Proposal Deadline Date: March 26, 2013

- EarthCube Research Coordination Networks
    - Full Proposal Deadline Date: March 26, 2013

- EarthCube Building Blocks
    - Full Proposal Deadline Date: May 22, 2013

- EarthCube Conceptual Designs
    - Full Proposal Deadline Date: May 22, 2013

https://www.nsf.gov/funding/pgm_summ.jsp?pims_id=504780

Questions

Chris.MacDermaid@colostate.edu

Backup

Goals

- Build a community consensus on infrastructure requirements to support research.
  - Community groups, concept awards
- Build upon existing community resources.
  - Use existing web services, brokers
- Enable reproducible science. Research results generated by one scientist should be repeatable by another.
  - Share workflows, re-execute workflows
- Enable both the long-tail of research by individual researchers and data-driven research.
  - Build shareable collections, move computation to data
- Support an extensible architecture, enabling new research initiatives to build upon results of prior research
  - Design interoperability mechanisms (virtualization of operations on name spaces)
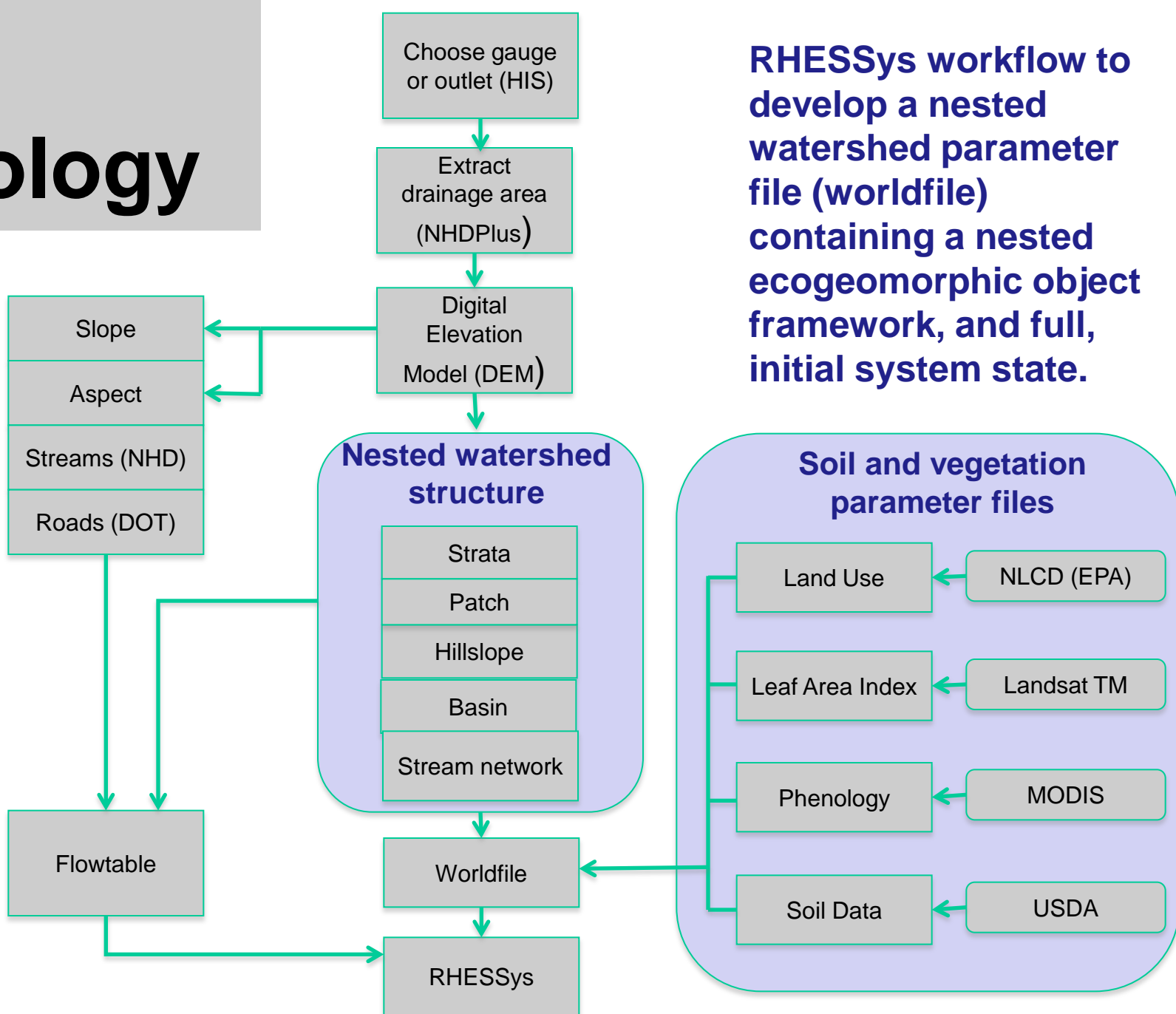
Community Resources

- Seek interoperability mechanisms to federate
  - Federal repositories
    - NOAA / USGS / EPA / NASA
  - Data management systems
    - XSEDE / DataNet / OSG / ESG / CLASS / DAAC
  - International resources
    - EUDAT

- And create Geo-science cyberinfrastructure
  - Develop list of interoperability mechanisms

Community Assertions

- Each community resource should publish assertions about the properties of the entities that can be retrieved
  - Example, all data are calibrated
  - Enforce with policy that verifies data were calibrated
- Each user group has expectations for the type of entities they will acquire
  - Example, data are projected to desired coordinate system
  - Enforce with a policy that applies the appropriate transformation

Eco-Hydrology

RHESSys workflow to develop a nested watershed parameter file (worldfile) containing a nested ecogeomorphic object framework, and full, initial system state.

Choose gauge or outlet (HIS)

Extract drainage area (NHDPlus)

Digital Elevation Model (DEM)

Slope

Aspect

Streams (NHD)

Roads (DOT)

Nested watershed structure

Strata

Patch

Hillslope

Basin

Stream network

Soil and vegetation parameter files

Land Use — NLCD (EPA)

Leaf Area Index — Landsat TM

Phenology — MODIS

Soil Data — USDA

Flowtable

Worldfile

RHESSys

Event-Driven Real-Time Drought Analysis/Prediction Workflow