Cyberinfrastructure Planning within the LTER Network Planning Grant Context

Barbara Benson
James Brunt
Peter McCartney
John Porter
John Vande Castle

LTER Network Planning Proposal

- New initiatives in long-term thematic, regional, and network-scale science
- Increasing the capabilities of scientists/sites (e.g., cyberinfrastructure, technical expertise) to perform research and education for the new environmental challenges

“The LTER Network will identify and pursue the appropriate strategy to accelerate its transition from an association of sites driven by local goals and resources to a fully functional Network driven by regional and national research priorities and shared resources.”

Goals: Cyberinfrastructure (CI) Planning

- strengthen the LTER planning effort through a broadly based consideration of needed cyberinfrastructure
- engage computer and information scientists to address the new integrative challenges presented by the expanding spatial, temporal and interdisciplinary scope of LTER network science
- provide cross-fertilization between LTER CI planning and that of other concurrent efforts within and beyond the ecological science community.

Cyberinfrastructure encompasses the computing power, storage capacity, networking capability … and specialized software and hardware environments … also includes people and organizations that operate and maintain equipment, develop and support software, create standards and best practices, and provide other key services like security and user help-desk support

Tasks

- Assess the CI capabilities of LTER and non-LTER resources
- Identify areas where CI improvement is required to support new network science
- Develop an information system design that scales to network level
- Develop an implementation strategy

Ecological Approach

1. Observational:

- Capture gradients and spatiotemporal variation: human-dominated, climatic, N-loading, etc.
- Measure variables above in consistent, coordinated manner over long-term.
- Inclusion of sites within and outside of LTER network.

Ecological Approach

2. Experimental

A. Manipulations:

- press driver * pulse driver * biotic structure

  Ex: N deposition * fire/drought/storm * dominant taxa

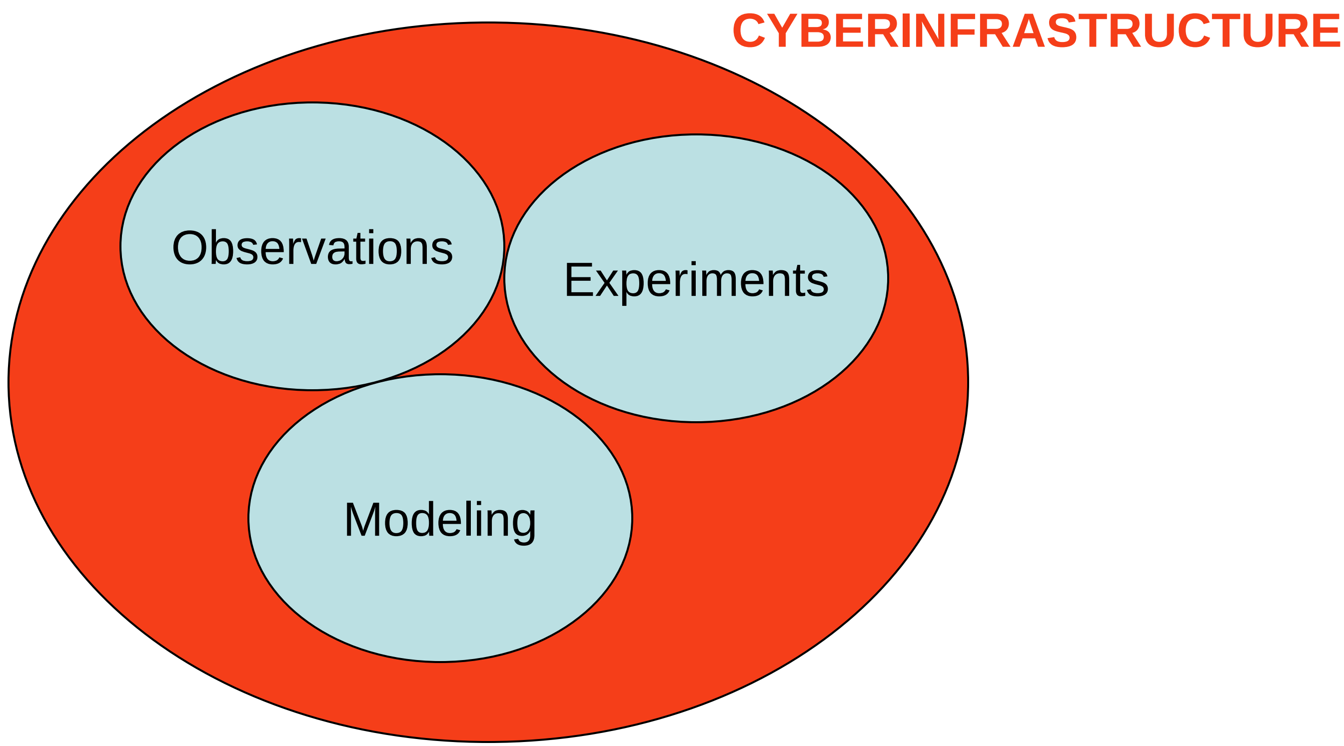
B. Measurements:

- coordinated & comparable response variables
- scale-independent measures of richness & dominance in each trophic level
- some measure of connectivity among trophic levels (stable isotope analyses)
- rates of primary & secondary production / community metabolism
- system efficiency (retention & export of C, N, P)

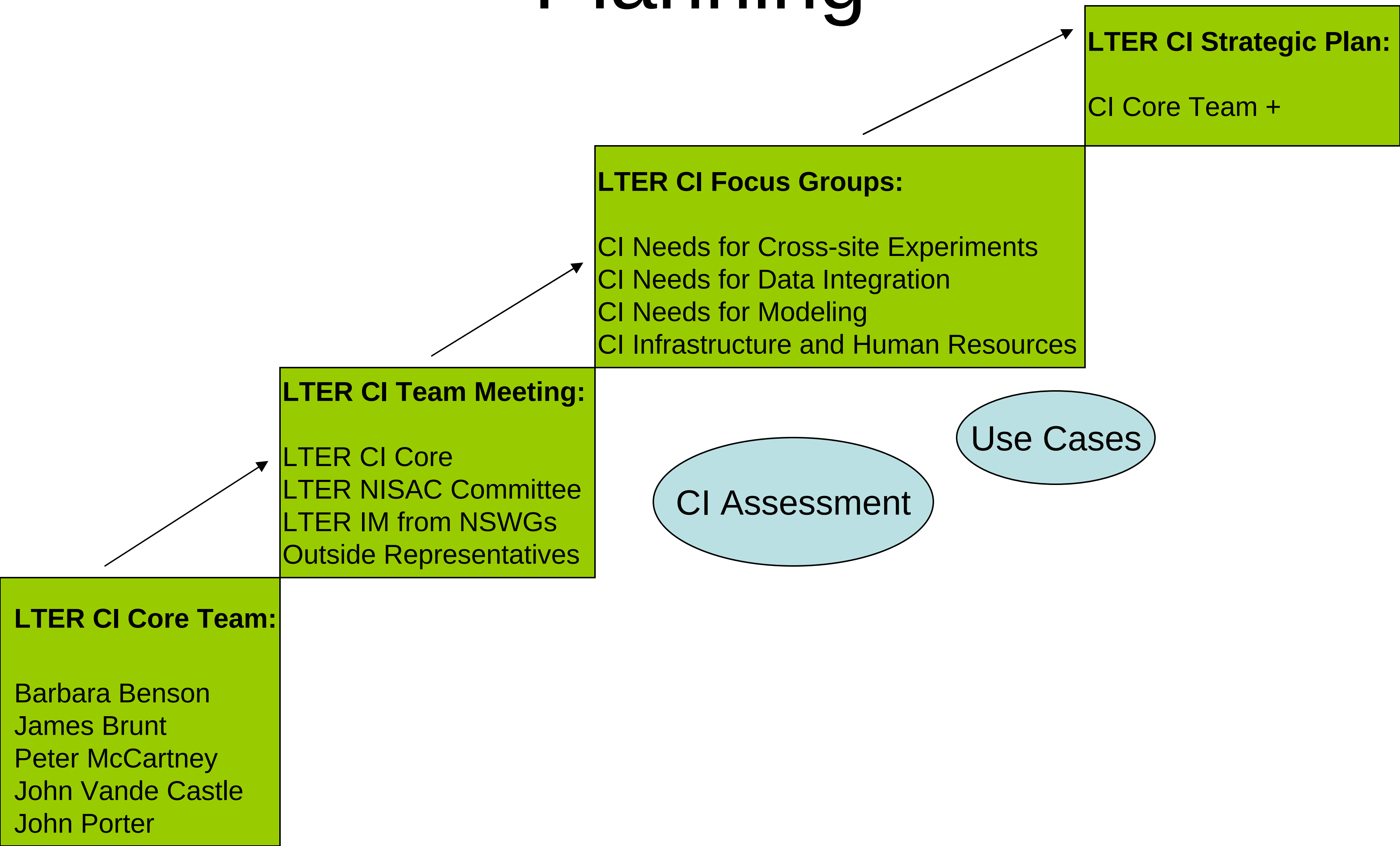
Ecological Approach

3. Modeling

- Simulation models
- Conceptual models
- Forecasting/scenario models
- Economic models
- Human demographics and land use change models


CYBERINFRASTRUCTURE
Observations
Experiments
Modeling

LTER Cyberinfrastructure Planning



LTER Cyberinfrastructure Assessment Survey
(August 2005)

•There is a **very** wide range of capabilities among LTER sites.

•Most sites maintain IM support with 1 FTE, range is 0.5 to 3 from LTER funding.
- 17 sites obtain external support for IM ranging from 0.25 to 2.5 FTE.

•Implementation of structured metadata (EML) generally strong
- 25% of sites have little completed

•LTER site data span a wide variety of forms, from streaming sensor data and automated shipboard systems to manual recordings of meteorological data.

•Approximately half of the LTER sites
- have internet connections 10mb/s or less
- have field wireless connectivity

•All LTER sites use phone teleconferencing, although only half have video capabilities.

•For a large increase in wireless site data collection, sites say they would need updates to site wireless installation, more storage capacity and more personnel.

•Site and LNO surveys available at “http://lternet.edu/technology”

LTER Affiliated Groups in LTER CI Planning

- CI-Core: Barbara Benson, James Brunt, Peter McCartney, John Porter and John Vande Castle

- LTER Information Managers: Corinna Gries, Kristin Vanderbilt, Karen Baker, Ken Ramsey, Jonathan Walsh, Don Henshaw, Barbara Benson, John Porter

- LTER Network Information System Advisory Committee (NISAC): Barbara Benson, Emery Boose, James Brunt, Stuart Gage, Mark Harmon, Don Henshaw, Tim Kratz, Peter McCartney, William Michener, Debra Peters, Robin Ross, Mark Servilla, John Vande Castle, Robert Waide

- LTER Principal Investigators/ Site Personnel: Stuart Gage, Mark Harmon, Tim Kratz, Debra Peters, Robin Ross, Paul Hanson, Hank Shugart

Associated Groups in LTER CI Planning

- **Chaitan Baru (GEON/SDSC, NEON)**
- **Kai Lin (GEON)**
- **Bryan Beecher (UM/ICPSR)**
- **Mark Schildhaur (NCEAS)**
- **Chris Jones (PISCO)**
- **Mandy Lane, Herbert Schentz (ALTER-Net)**
- **Bob Cook, Tim Rhyne (ORNL/NASA)**
- **Peter Cornillon, Nathan Potter (OPeNDAP/OGC)**
- **Mark Stromberg (OBFS)**
- **David Maidment (CUAHSI)**
- **Mike Freeman (NBII/NCSA)**
- **Gordon Bonan (NCAR)**
- **George Hurtt (UNH EOS)**
- **Peter Franks (SIO/CCE LTER)**
- **Jennifer Eakins (SIO IGPP, RoadNet)**
- **Michael Piasecki (Drexel OWL, CUAHSI)**
- **Patrick Mulholland (ORNL/ESD – LINX)**
- **Michael Hamilton (CENS/James Reserve, NEON)**
- **Shawn Bowers (DAKS/UCSD)**

Products

- Strategic plan for cyberinfrastructure
  - Addresses vision, design, and implementation
- Cyberinfrastructure initiatives
  - LTER IM activities – metadata, LTERGrid
  - NSF proposals – SEI&II, BDI, CyberTools
  - Integrated CI in LTER Network Science Proposal

LTER Network Cyberinfrastructure

- Strengths
  - Online data & metadata
  - Standards for data release, metadata documentation, evaluation
  - Tool development (Site, LNO, and partners)
- Challenges
  - Distribution of responsibilities and resources
  - Diversity among sites
    - standards for data formats & protocols
    - tracking and notification
    - Completeness of data and metadata

CI Focus Groups

- Infrastructure for data integration
- Support for integrated modeling
- Framework for cross-site/network experiments
- Systems architecture & human capacity

Cyberinfrastructure for Cross-site/Network Experiments

Barbara Benson, James Brunt, Jennifer Eakins, Mike Freemon, Paul Hanson, Chris Jones, David Maidment, Pat Mulholland, Mark Servilla, John Vande Castle

Cross-site/Network Experiment CI

- **Organizational components**
  - Decision making process and policies, governance
  - Authorship issues
- **Procedural components**
  - Experimental design
  - Protocol development
  - Data acquisition
  - Data management
  - Analysis
  - Archiving

Cross-site/Network Experiment CI

- Cyberinfrastructure Components
  - Human resources
  - Computing and information resources
    - Acquisition
    - Storage
    - Processing (qa/qc, derived data)
    - Retrieval and analysis
    - Communication (connectivity, security, collaboration)

Framework for Cross-site Experiments

- Tools and expertise based on centralized access architecture
- Centralized: personnel to provide design and development support
  - customizable data entry software
  - designing and curating databases
  - Tools for data quality screening and data query
  - Sophisticated environment for analysis and visualization
- Resources at research site and research project level to enact technological solutions

Cyberinfrastructure for Data Integration

James Brunt, Shawn Bowers, Kai Lin, Tim Rhyne, Herbert Schentz, Mark Schildhauer, Mark Servilla, John Vande Castle, and Barbara Benson

Challenges of Cyberinfrastructure for Data Integration

- Differences in:
  - file formats
  - data models - table layout & data types
  - units, dynamic resolution (precision), temporal and spatial resolution and extents or coverage
  - spatial, temporal and thematic scales
  - semantic classifications and field definitions
  - QA flags, including missing value
  - in structural heterogeneity
- Auditability (documentation of assumptions, decisions and lineage)

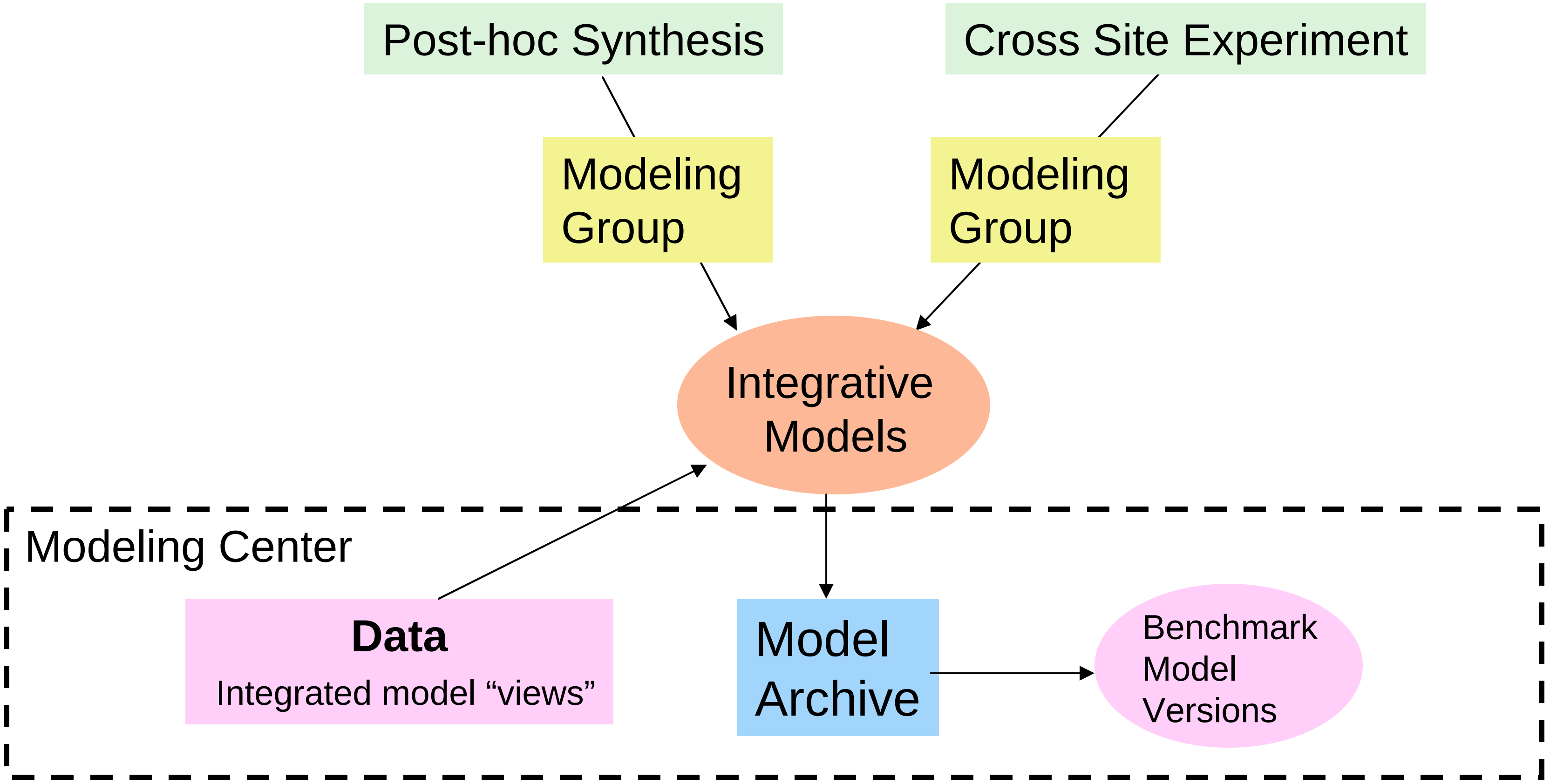
Summary and Recommendations

- Experimental data where the experiment is designed a priori will benefit from working from a global schema approach.
- Post-collection data integration efforts where an ongoing value-added data product is expected should be federated if feasible.
  - one-time value-added data products could use manual data processing techniques.
- For all data holdings, structural and ontological metadata should be continued to be defined and developed to make it possible to do semi-automated data integration for ad hoc analysis.
- Tools for registration and integration of existing databases should be made available

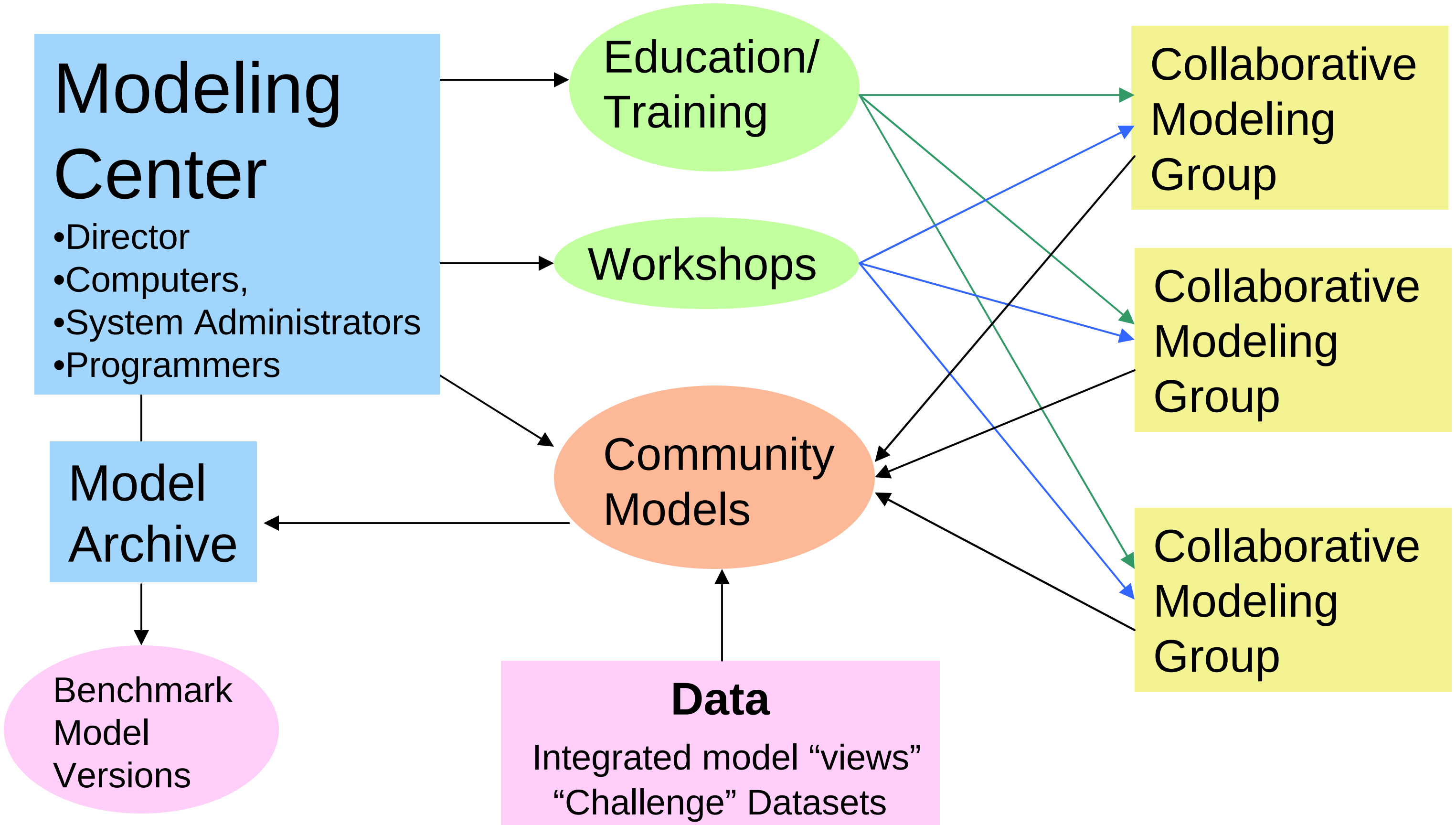
Cyberinfrastructure for Ecological Modeling

Gordon Bonan, Robert Cook, Peter Franks, George Hurtt, John Porter, Enrique Reyes, Hank Shugart, John Vande Castle

Project-based Modeling



Community Modeling Cyberinfrastructure



Cyberinfrastructure: Organization and Human Resources

Peter McCartney, Barbara Benson, James Brunt, John Vande Castle, Mark Schildauer, Nathan Potter, Mark Stromberg, Jennifer Eakins, Michael Piasecki, Bryan Beecher, Michael Hamilton, Karen Baker, Corinna Gries, Don Henshaw, John Porter, Mark Servilla

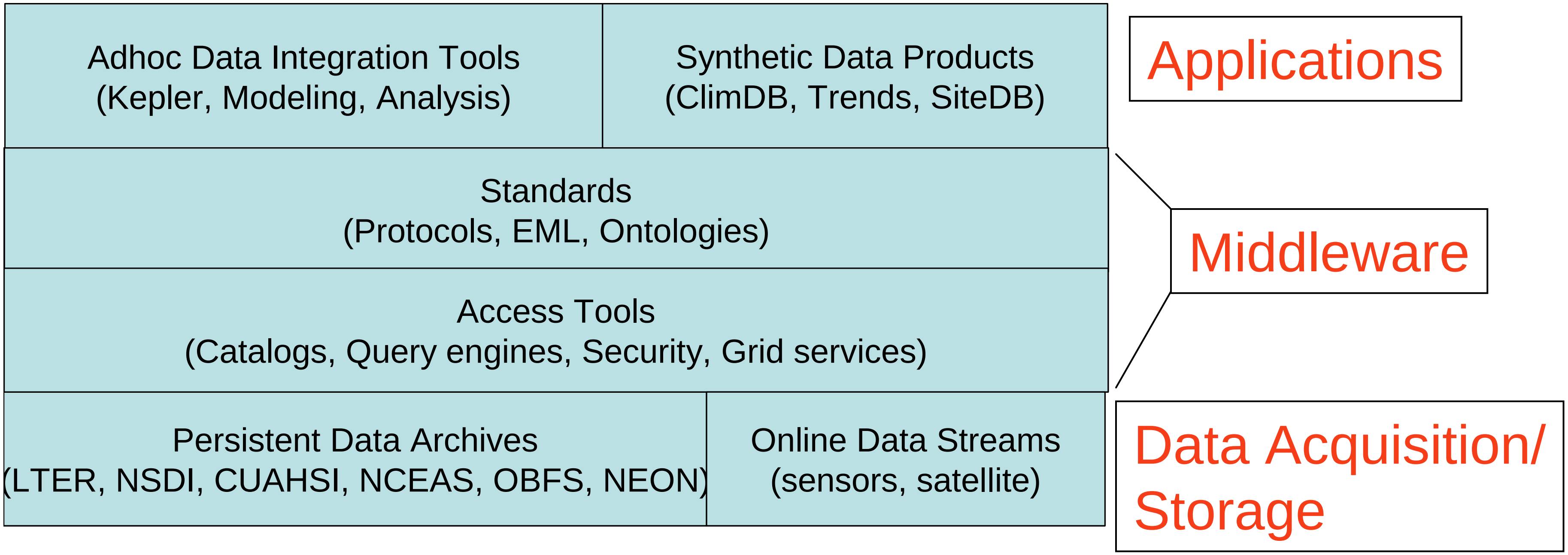
Organizations Requiring Human Resources

- Study Sites
  - Source of long-term data
- Multi-site Projects
  - short-term perspective, specialized CI (synthesis projects or cross site experiments)
- Multi-site Networks
  - long-term perspective with generalized CI (LTER Network)
- Centers
  - Long-term perspective, specialized tasks, software development
- Open-source Communities
  - individuals and organizations
- Academic Programs
  - Formal engineering and computer science training for ecoinformaticians

Positions

- ecological knowledge engineers
- system administrators
- software developers
- data processors
- program managers
- training coordinators
- Information managers
- information scientists
- data archivists
- web developers
- technicians

Tiered Infrastructure



Timetable

- Mar 2006 CI Core + NISAC Chair
  - Integrate Focus Group whitepapers and needs assessments into final CI strategic plan
- May 2006 CI Team + NISAC
  - Review CI plan with NISAC and finalize for STF
- Jun 2006 CI Core + STF
  - Merge CI strategic plan into overall network plan
- Sept 2006 LTER All Scientists Meeting with partners
  - Roll out of network science and CI plan; implementation workshops