Lessons for Educational Projects from Evaluation of Multi-Scale Quantitative Models

Louis J Gross, NIMBioS Director Emeritus
Professor of Ecology and Evolutionary Biology and Mathematics
University of Tennessee, Knoxville

Multi-Scale Evaluation in STEM Education

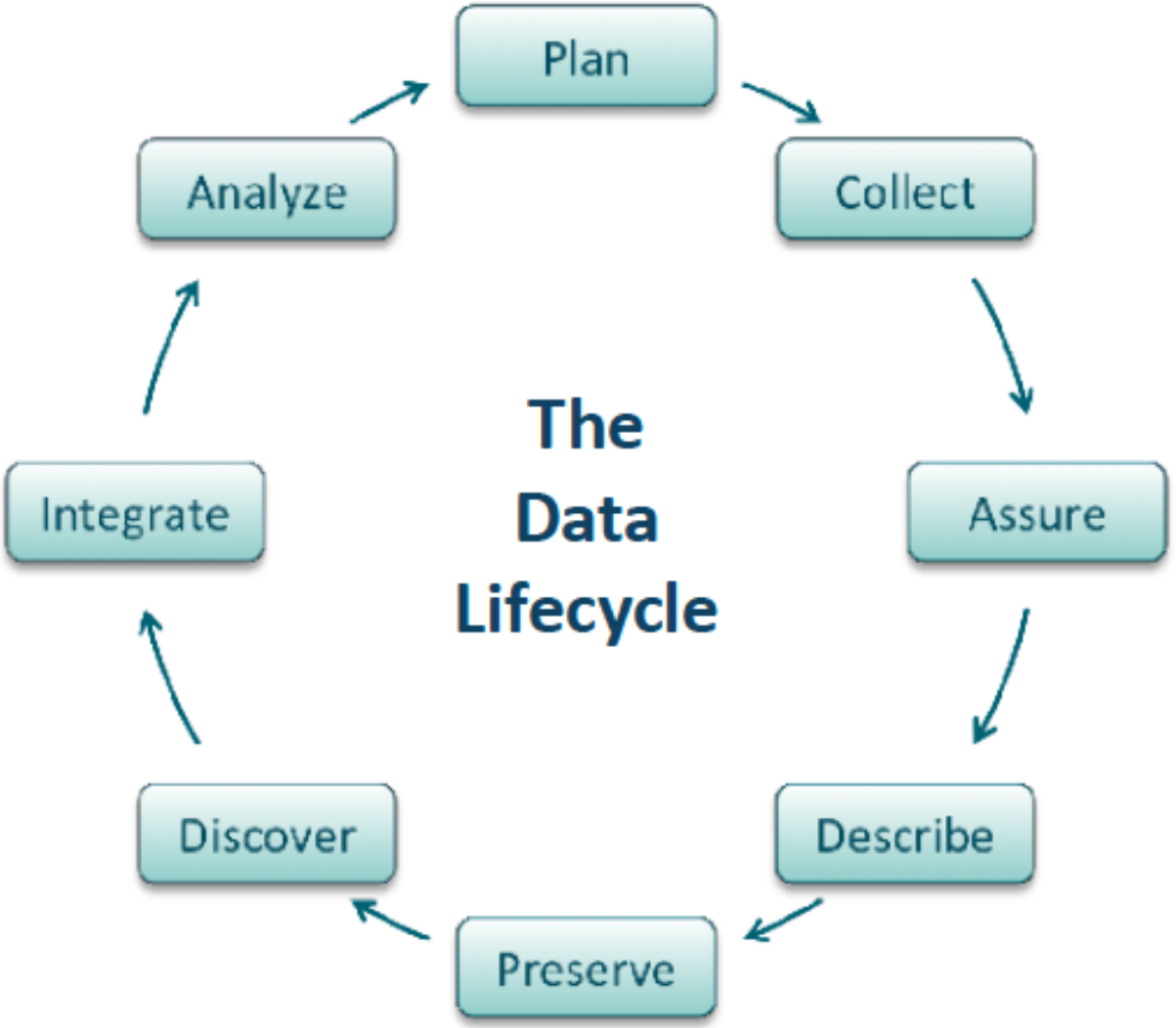
Inclusion across the Nation of Communities of Learners of Underrepresented Discoverers in Engineering and Science (NSF INCLUDES)

These STEM evaluation activities are supported by the National Science Foundation through award HRD-1650390 to the University of Tennessee, Knoxville.

The Evaluation Process

Cyclical Aspects of Data

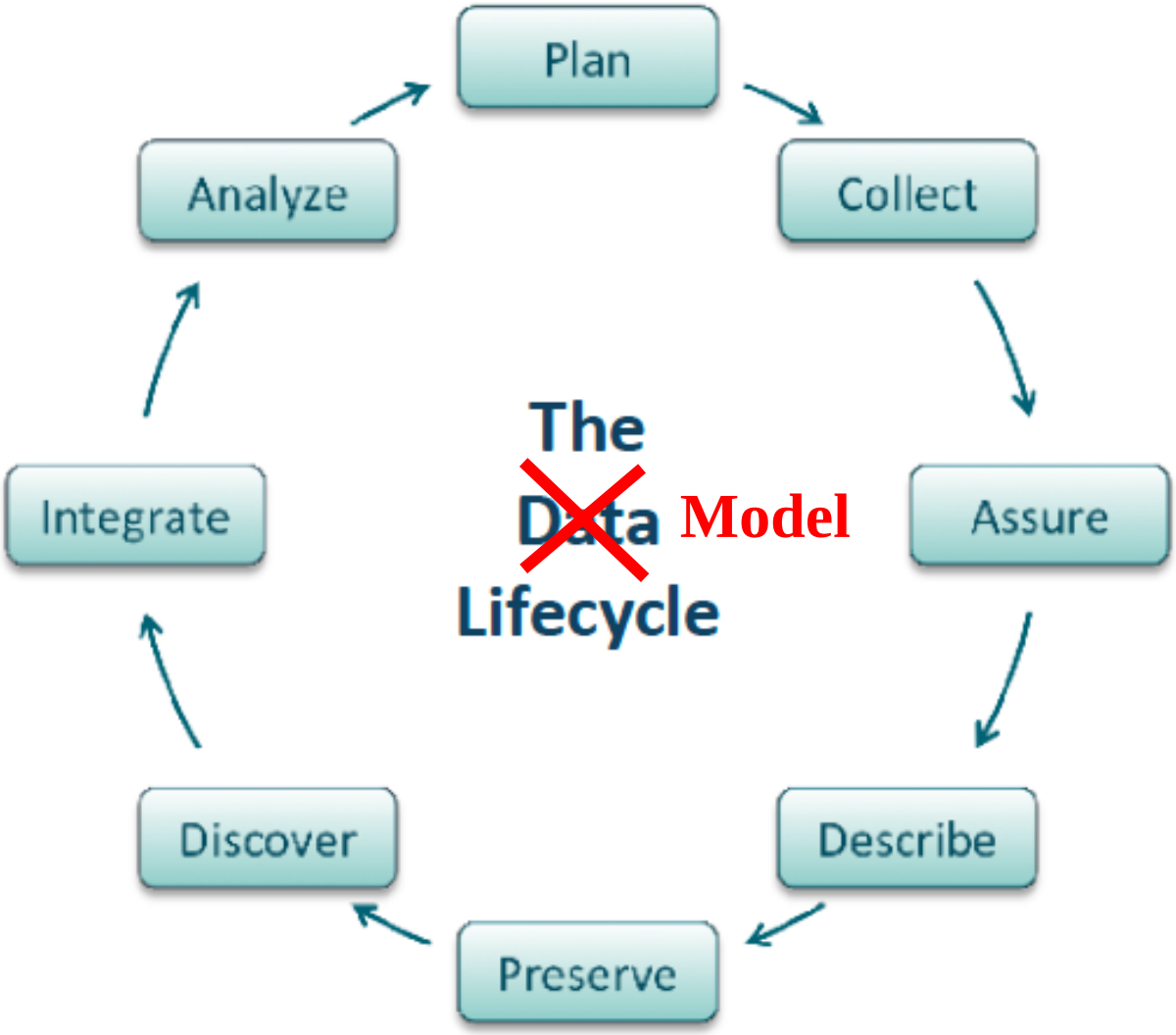
The stages through which well-managed data passes from project inception to conclusion.



From DataONE.org

Cyclical Aspects of ~~Data~~ Models

The stages through which well-managed data passes from project inception to conclusion.



From DataONE.org

Environmental Modeling

Species densities

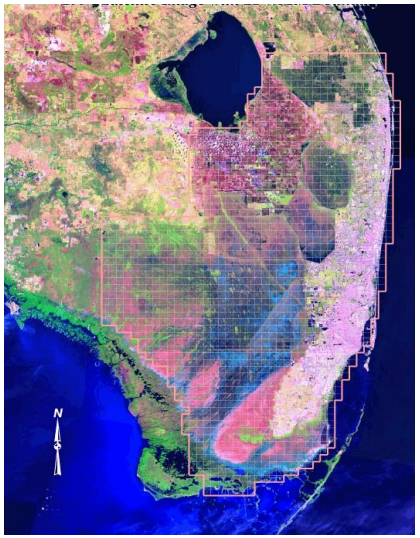
Animal telemetry

Physical conditions

Data sources

GIS map layers (Vegetation, hydrology, elevation),Weather, Roads, Species densities

Monitoring



Models

Statistical

Differential equations

Matrix

Agent-based

Management input

Harvest regulation

Water control

Reserve design

Evaluation/Analysis

Simulation

Visualization, corroboration, sensitivity, uncertainty

Matlab, C++, Distributed, Parallel

You make models all the time:

What decision do you make when faced with:

Your criteria may not be the same as others.

Objectives of Models

1. Suggest observations and experiments

2. Provide a framework to assemble bodies of facts/observations - standardize data collection

3. "Allows us to imagine and explore a wider range of worlds than ours, giving new perceptions and questions about how our world came to be as it is" F. Jacob - The Possible and the Actual, 1982

4. Clarifies hypotheses and chains of argument

5. Identifies key components in systems

6. Allow investigation while accounting for societal or ethical constraints

Objectives of Models

7. Allows simultaneous consideration of spatial and temporal change

8. Extrapolate to broad spatial or long temporal scales for which data can not easily be obtained

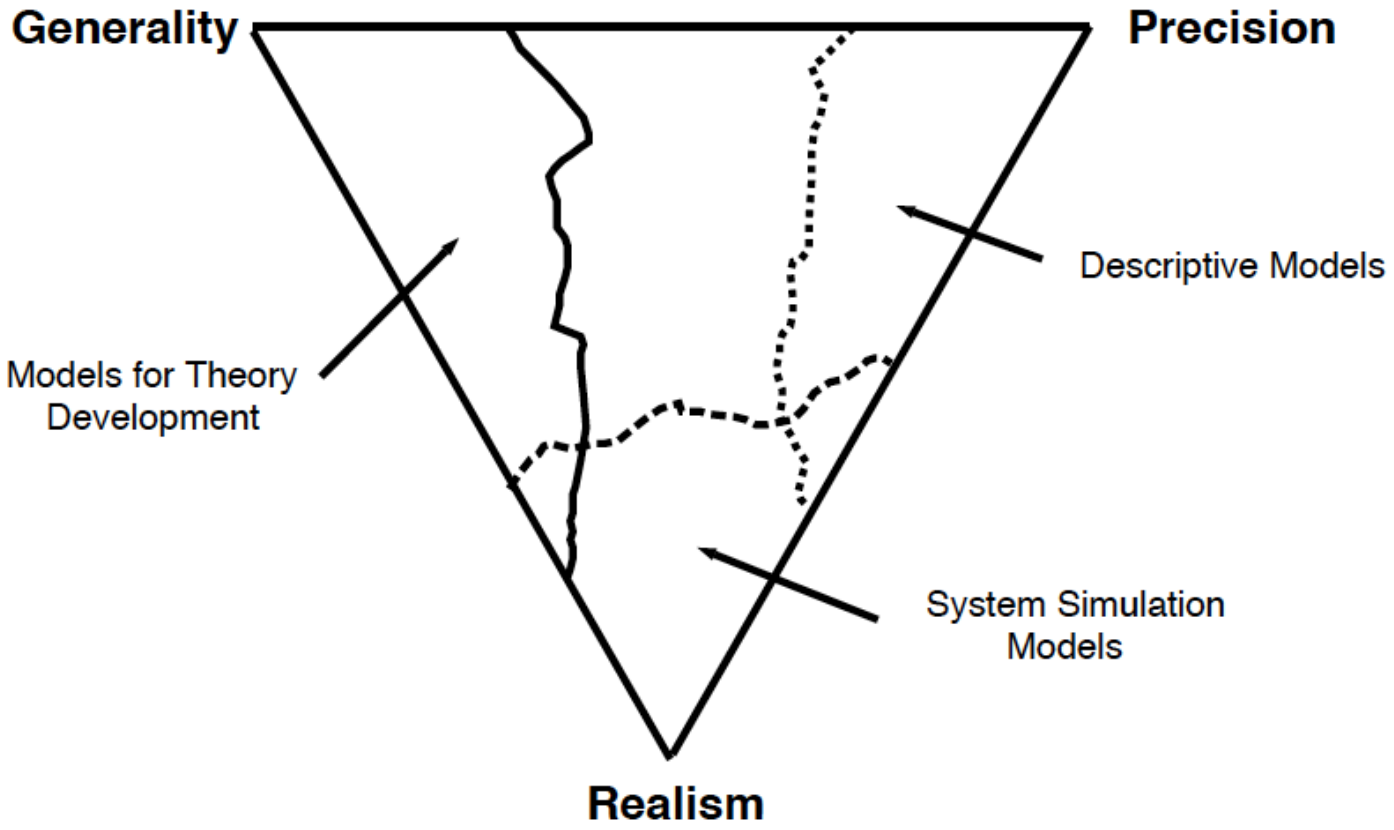
9. Prompts tentative and testable hypotheses

10. Serves as a guide to decision making in circumstances where action cannot wait for detailed studies

11. Provides an antidote to the helpless feeling that the world is too complex to understand in any generality - provides a means to get at general patterns and trends

12. To predict how a system will behave under different management, and control the system to meet some objective

Models and tradeoffs



No one model can do everything

Constraints on models

Data constraints: Available data may not be sufficient to specify appropriate functional forms, interrelationships, or parameters. May force aggregation of components. May not be sufficient to elaborate criteria for evaluation of model performance.

Effort constraints: Resource constraints may limit the amount of detail it is feasible to include. Limits on time modelers and collaborators may invest as well as pressure to produce results.

Computational constraints: Despite great enhancements in computational resources, there are many problems still not feasible to carry out computationally.

Other constraints: ethical or other societal considerations.

Model evaluation – some terminology

Verification - model behaves as intended, i.e. equations correctly represent assumptions; equations are self-consistent and dimensionally correct. Analysis is correct. Coding is correct - there are no bugs.

Calibration - use of data to determine parameters so the model "agrees" with data. This is specific to a given criteria for accuracy. Some call this Tuning or Curve-fitting.

Corroboration - model is in agreement with a set of data independent from that used to construct and calibrate it.

Validation - model is in agreement with real system it represents with respect to the specific purposes for which it was constructed. Thus there is an implied notion of accuracy and domain of applicability.

Evaluation (testing) - appropriateness to objectives; utility; plausibility; elegance; simplicity; flexibility.

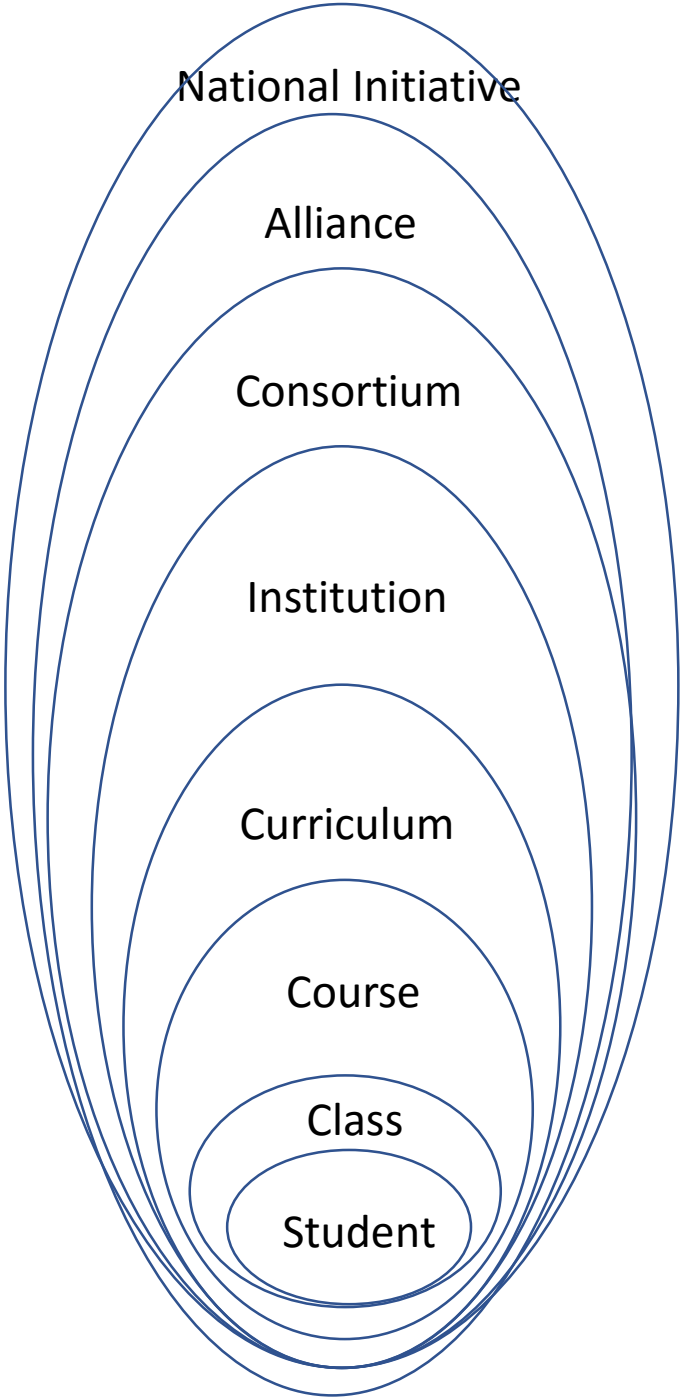
Models and evaluation

- Given the many objectives for models, we should expect just as many criteria for evaluating whether a model is useful
- Before developing a model in any detail, criteria should be established for evaluating its use
- Evaluation should account for constraints of Data, Effort and Resources, Computation
- Include evaluation of alternative approaches based on constraints to assess most appropriate methods, decide level of detail, scale, and what to ignore.

In general there is relatively little attention paid to evaluation of models. Why?

1. It's difficult and requires potentially different skill sets from those constructing and using models.
2. Science is very much a human enterprise and it is natural that once one has devoted considerable effort to developing a particular model, it is difficult to critique yourself.
3. Modern settings with a great amount of team effort to develop models or experimental protocols can constrain individuals who do not wish to be an outcast in a lab.

Gross, L. J. 2013. Selective ignorance and multiple scales in biology: deciding on criteria for model utility. *Biological Theory* **8**:74-79


National Initiative
Alliance
Consortium
Institution
Curriculum
Course
Class
Student

ATLSS STRUCTURE
Across Trophic Level System Simulation

Model Type

- **Individual-Based Models**
  - Cape Sable Seaside Sparrow
  - Snail Kite
  - Wading Birds
  - White-tailed Deer
  - Florida Panther
  - Radio-telemetry Tracking Tools
- **Age/Size Structured Models**
  - Fish Functional Groups
  - Alligators
  - Reptiles and Amphibians
- **Linked Cell Models**
- **Process Models**
  - Lower Trophic Level Components
  - Vegetation
- **Spatially-Explicit Species Index Models**
  - Cape Sable Seaside Sparrow
  - Snail Kite
  - Long-legged Wading Birds
  - Short-legged Wading Birds
  - White-tailed Deer
  - Alligators
- **Abiotic Conditions Models**
  - High Resolution Topography
  - High Resolution Hydrology
  - Disturbance

Lessons from ATLSS - Interacting with Stakeholders:

Constantly communicate with stakeholders.

Regularly explain the objectives of your modeling effort, as well as the limitations, to stakeholders. Be prepared to do this over and over for the same people, and do not get frustrated when they forget what you are doing and why.

Be prepared to regularly defend the scientific validity of your approach.

Lessons from ATLSS - Interacting with Stakeholders:

Don't limit your approach because one stakeholder/funding agency wants you to.

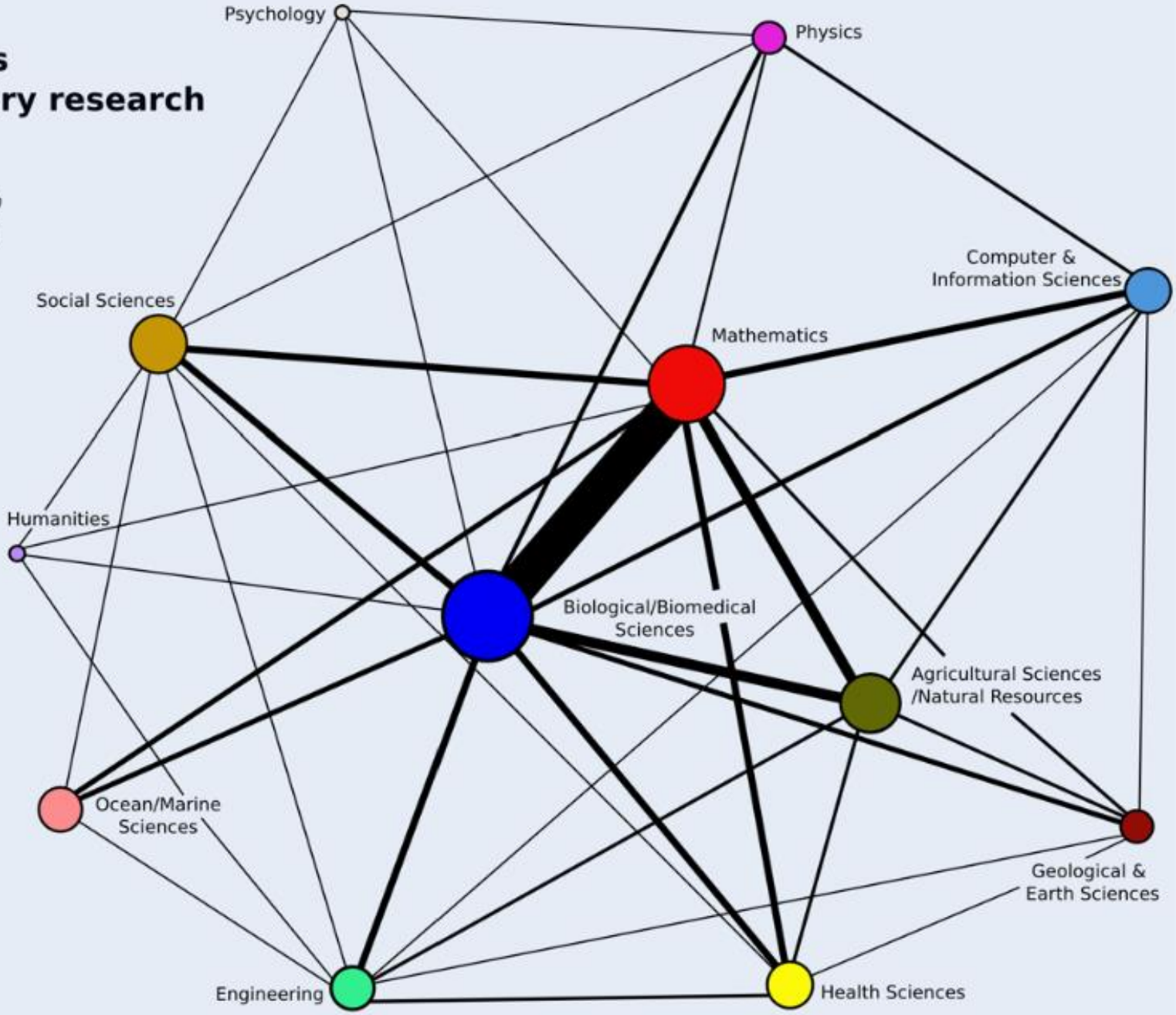
Be prepared for criticism based upon non-scientific criteria, including personal attacks.

Ignore any of the stakeholders at your peril.

NIMBioS fosters cross-disciplinary research

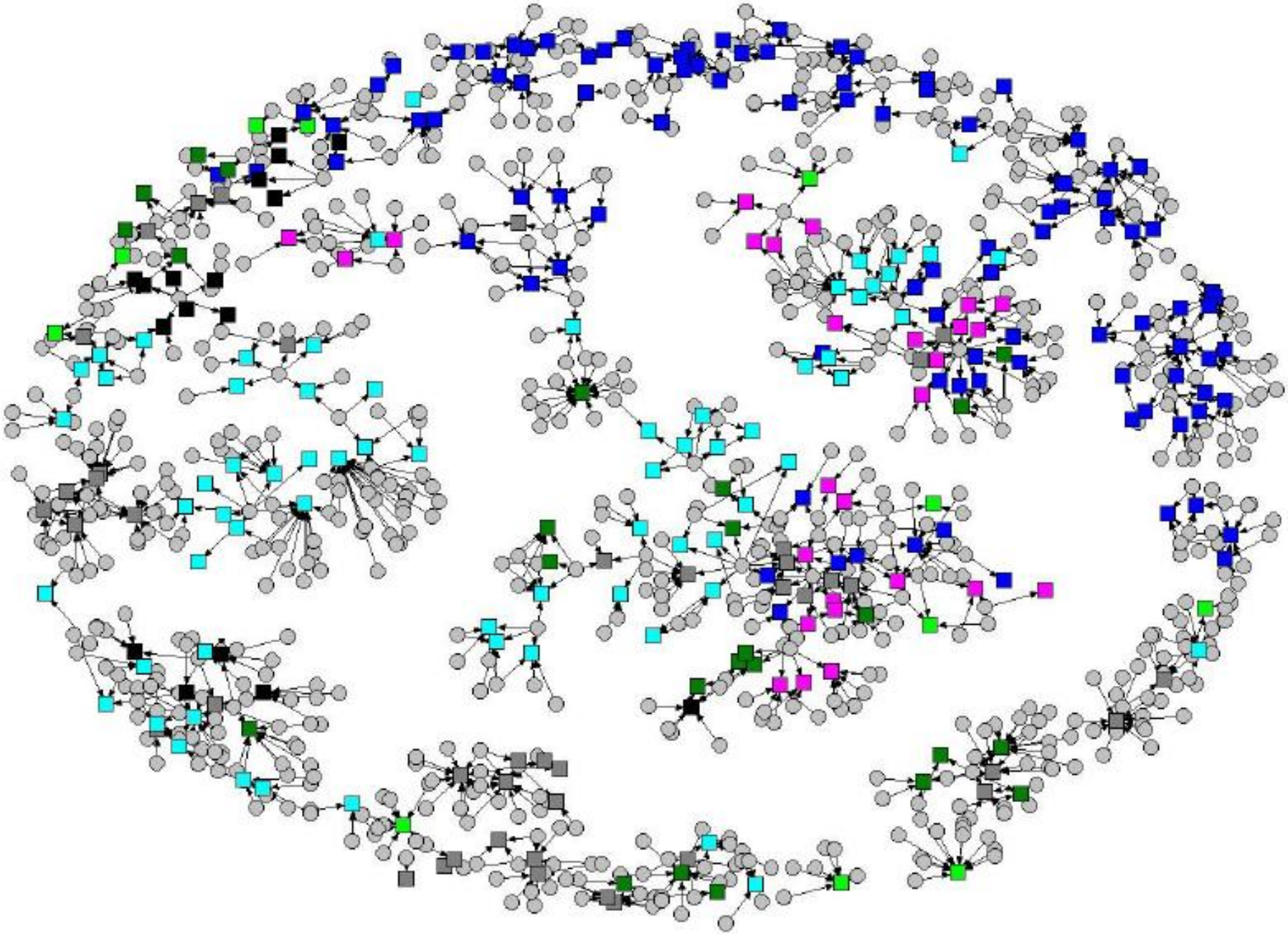
Node size
number of Working Group participants in a given research area, where the node radius is the log number of participants

Line size
number of collaborations between research areas within Working Groups



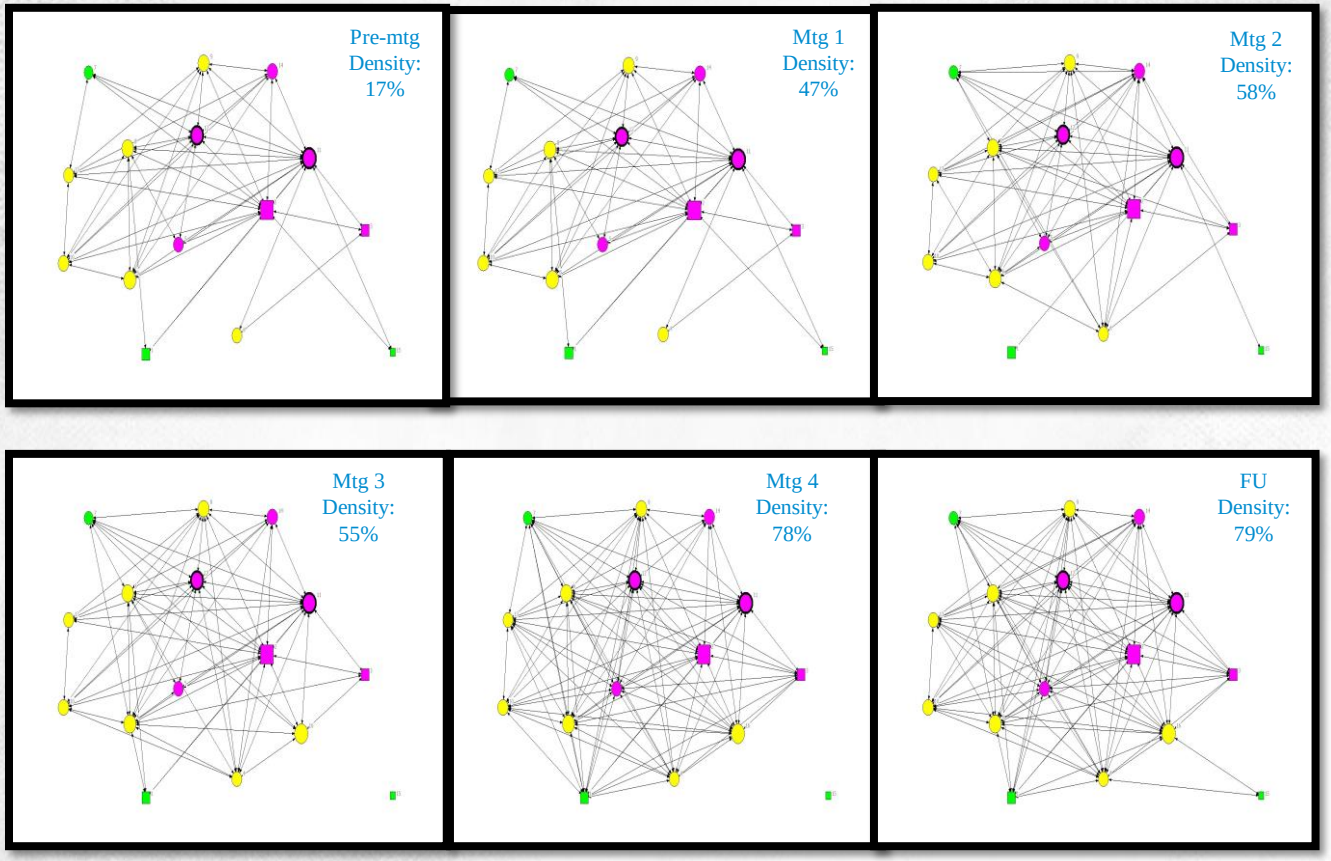
Working Groups focus on major scientific questions at the interface between biology and mathematics that require insights from diverse researchers who meet several times over a two-year period.

Figure 38. Participant paper collaboration for all NIMBioS events



Network Evolution

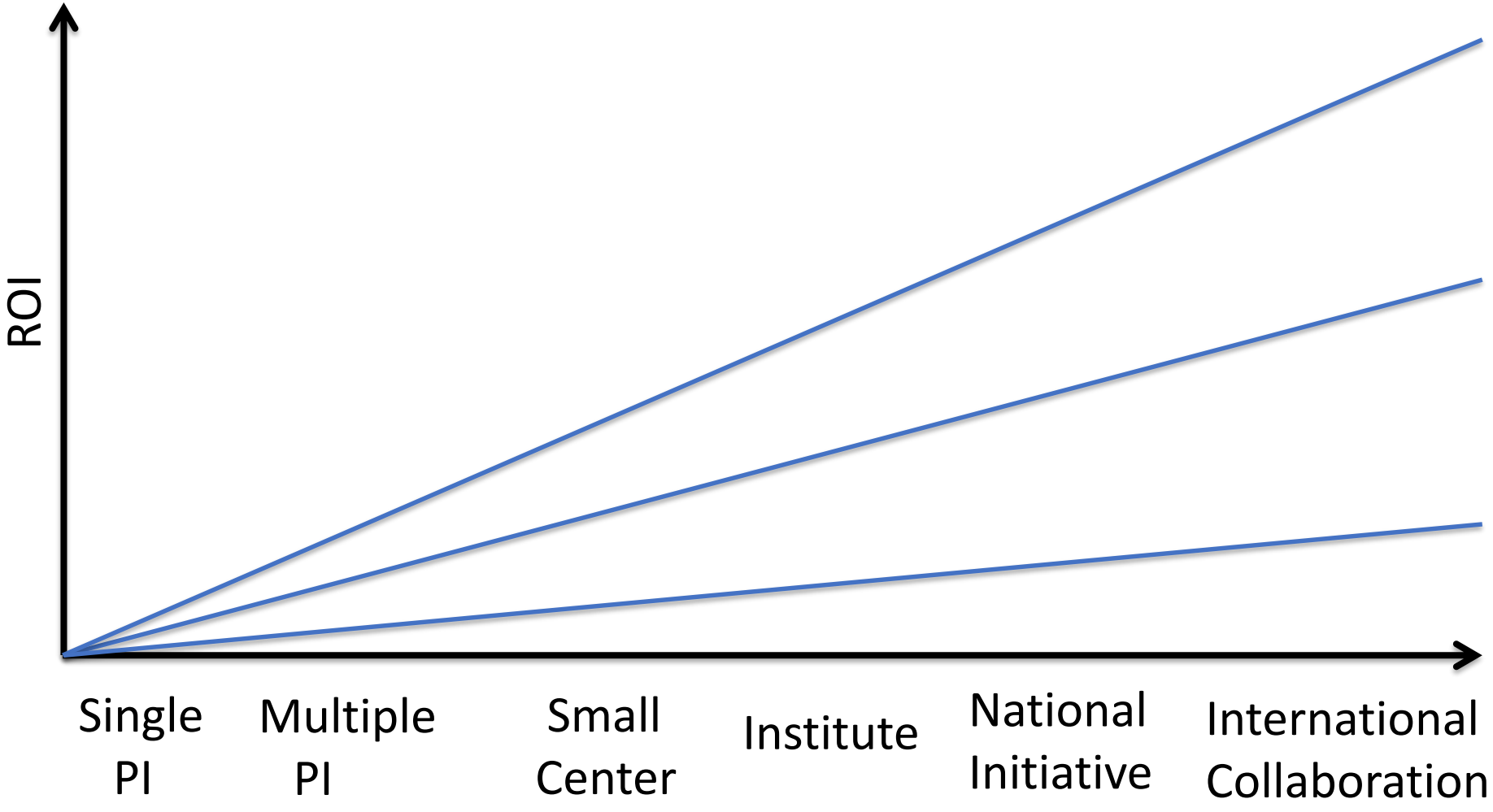
What are the patterns of change in the connections of working group participants over time?



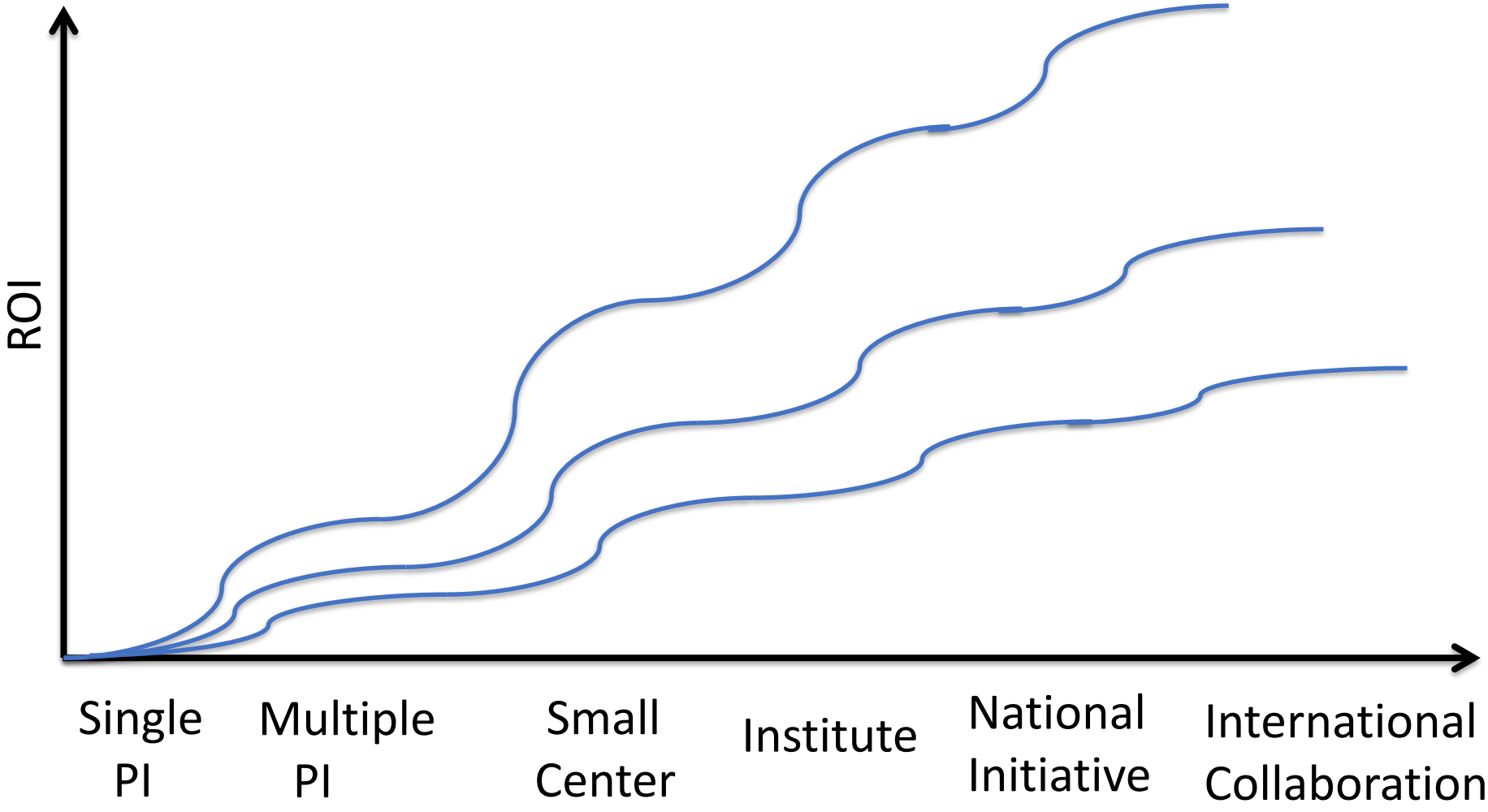
Density: # of ties in network expressed as a proportion of the maximum possible ties

One Working Group n = 15

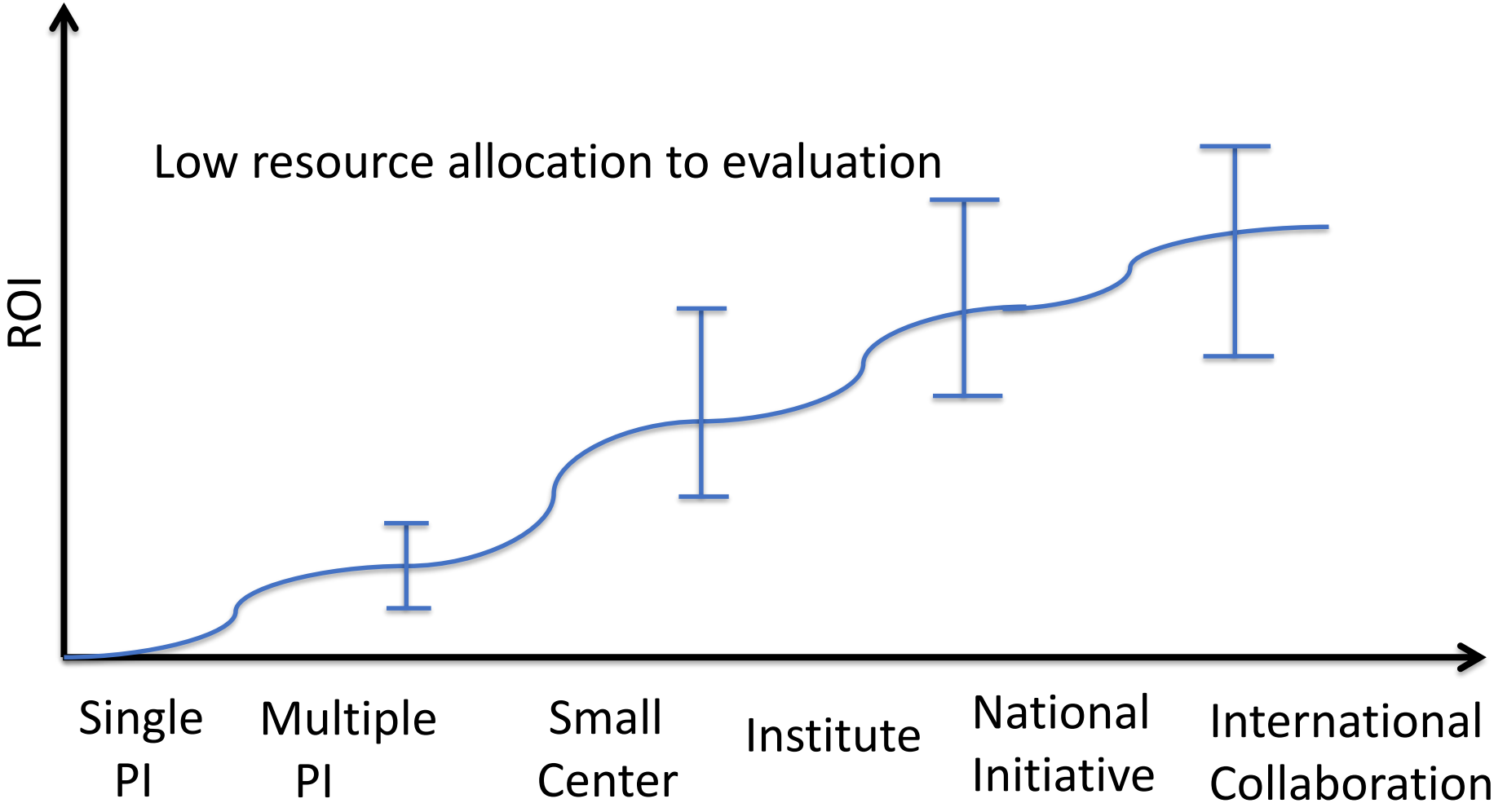
Scale, Infrastructure, ROI and Evaluation



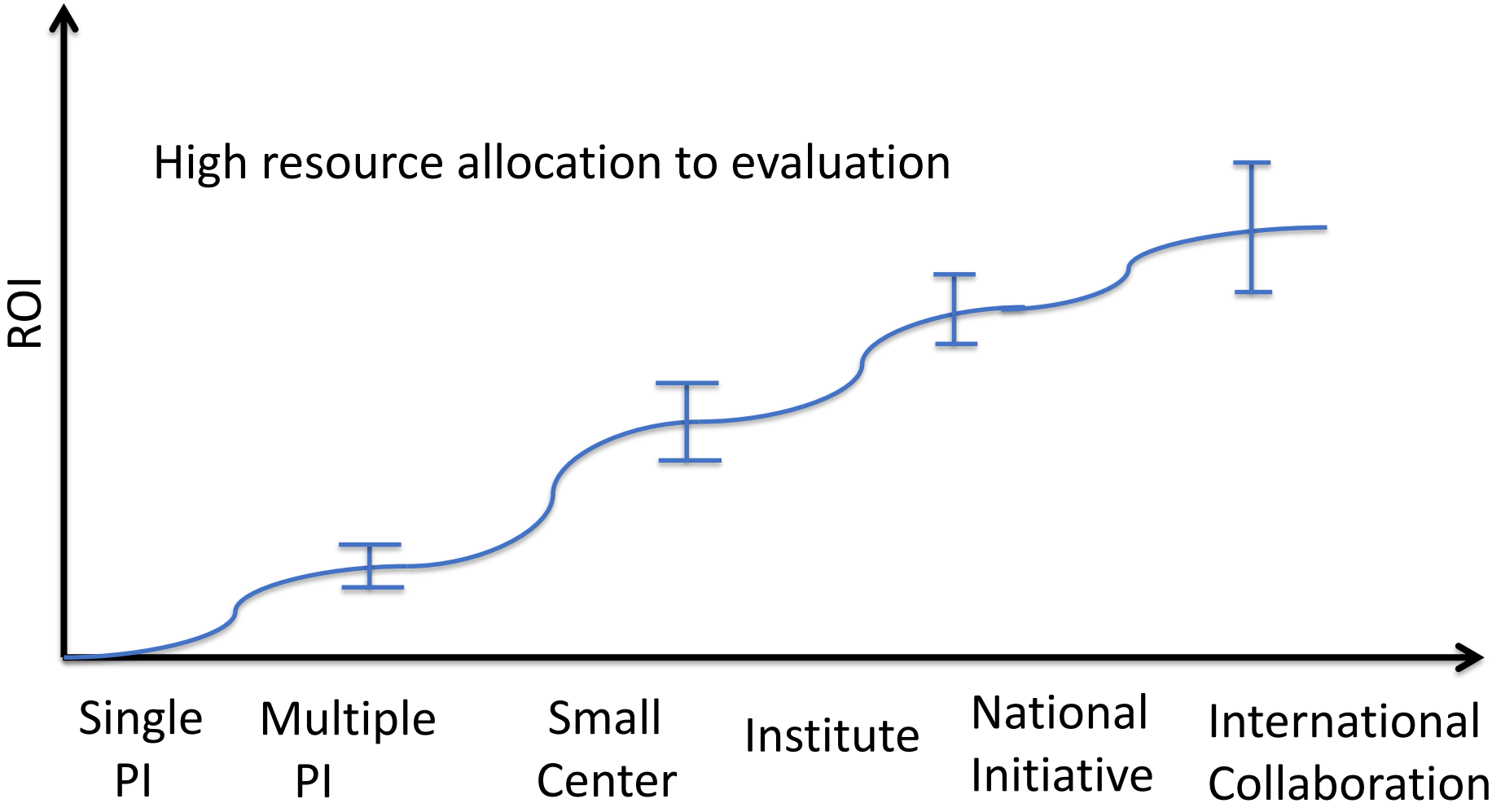
Scale, Infrastructure, ROI and Evaluation



Scale, Infrastructure, ROI and Evaluation



Scale, Infrastructure, ROI and Evaluation



Scale, ROI and Evaluation

- Agencies (independently or collaboratively) can potentially invest in collaborative impact efforts at diverse scales
- Multiple impact metrics might be considered, possibly dependent upon the scale
- Determining portfolio allocations across scales will depend upon how impact metrics vary with scale
- Evaluation efforts to assess impacts might best be built in from the start
- Potential for adaptive resource allocation requires both evaluation for impact metrics and policies for modification
- These issues are particularly pertinent for INCLUDES due to its multi-scale nature, need for synthesis within and across time/space/social scales and convergence of diverse disciplines

How we are evaluating the activities for this project?

Webinar

Post survey of participant experience/learning

Tutorial

In person participant pre- and post-survey

Distance participant post-survey

Conference

In person participant pre- and post-survey

Formative evaluation session, end of Conference day 1 to determine:

what participants felt were the most useful aspects of the day's activities

what could be improved upon

what they would like to learn more about on day 2

Distance participant post-survey

NIMBioS
National Institute for Mathematical and Biological Synthesis

NISER
NATIONAL INSTITUTE FOR STEM EVALUATION AND RESEARCH