THE CHALLENGE OF MEASURING PERFORMANCE

(FROM THE NSF POINT OF VIEW)

PAMELA O'NEIL
DEPUTY PERFORMANCE IMPROVEMENT OFFICER
NATIONAL SCIENCE FOUNDATION

THE FEDERAL CONTEXT

M-12-14 from Jeff Zients to the Heads of Agencies – May 18, 2012

"Since taking office, the President has emphasized the need to use evidence and rigorous evaluation in budget, management, and policy decisions to make government work effectively. … Where evidence is strong, we should act on it. Where evidence is suggestive, we should consider it. Where evidence is weak, we should build the knowledge to support better decisions in the future. …

3. Infusing evidence into grant-making. Grant-making agencies should demonstrate that, between FY 2013 and FY 2014, they are increasing the use of evidence …."

GPRA MODERNIZATION ACT

GOVERNMENT PERFORMANCE AND RESULTS ACT OF 2010

"An Act To require quarterly performance assessments of Government programs for purposes of assessing agency performance and improvement, and to establish agency performance improvement officers and the Performance Improvement Council."

- Established performance improvement roles and responsibilities (PIO)
- Emphasis on goal setting
- Frequent data-driven management reviews
- Agency performance information reported on performance.gov website

TAKE HOME MESSAGE:

The importance of using evidence for decision making is continually articulated and emphasized in government.

A strong message from the top that impact is tied to $.

OVERVIEW OF NSF ASSESSMENT MECHANISMS

Proposal	Project	Program	Foundation	Science
Ad hoc Review	Annual Reports	Monitoring	Committees of Visitors	Science of Science Policy
Panel Review	Site visits	Formative evaluation	Performance and Priority Goals	
Site visits	Final Reports	Formal impact evaluation	Merit Review Report	
Director's Review Board	External formative or summative evaluation			
National Science Board				

OVERVIEW OF NSF ASSESSMENT MECHANISMS

Proposal	Project	Program	Foundation	Science
Ad hoc Review	Annual Reports	Monitoring	Committees of Visitors	Science of Science Policy
Panel Review	Site visits	Formative evaluation	Performance and Priority Goals	
Site visits	Final Reports	Formal impact evaluation	Merit Review Report	
Director's Review Board	External formative or summative evaluation			
National Science Board				

Me

NSF Program Officers

PIs/ Project Staff

Points of View

PERFORMANCE MANAGEMENT

(A PROCESS BY WHICH ORGANIZATIONS ALIGN THEIR RESOURCES, SYSTEMS AND EMPLOYEES TO STRATEGIC OBJECTIVES AND PRIORITIES.)

What limits the success of Performance Measurement?

1) *The clear articulation of goals.*
2) *Alignment of goals, activities and indicators.*

Strategic Plan (Articulates and prioritizes goals)

Includes: Strategy for measuring progress toward goals

(Performance Plan, Evaluation Plan, etc.)

Strategic Monitoring	Formative Evaluation		Summative Evaluation	
Collect data to help you establish a baseline and measure progress	Study your progress to inform next steps.		Study what you do to document that you have made a difference	
Audience = your leadership: Director, Senior Staff, Advisory Board	Audience = Director, Senior Staff, Advisory Board	Audience = NSF Program Officers, Site visit members, etc.	Audience = NSF program officers	Audience = All of the above + congress and the world
Measures and documents your progress	Improvement of Programs	Evidence- Based programs	Additional funding for evidence-based programs	Renewal funding from NSF	More funding for Science in the Federal Budget

HOW TO CREATE ALIGNMENT BETWEEN WHAT YOU DO AND WHAT YOU MEASURE?

Logic model (Theory of Action)

is a systematic and visual way to present and share your understanding of the relationships among

- the resources you have to operate your program (inputs), the
- activities you plan, and
- the results you hope to achieve.

LOGIC MODEL

Inputs

- Resources that are needed to operate your program.

Activities

- If you have access to resources, the resources can be used to accomplish planned activities.

Outputs

- If you accomplish your planned activities, then you will hopefully deliver the amount of product and/or service that you intended.

Outcomes

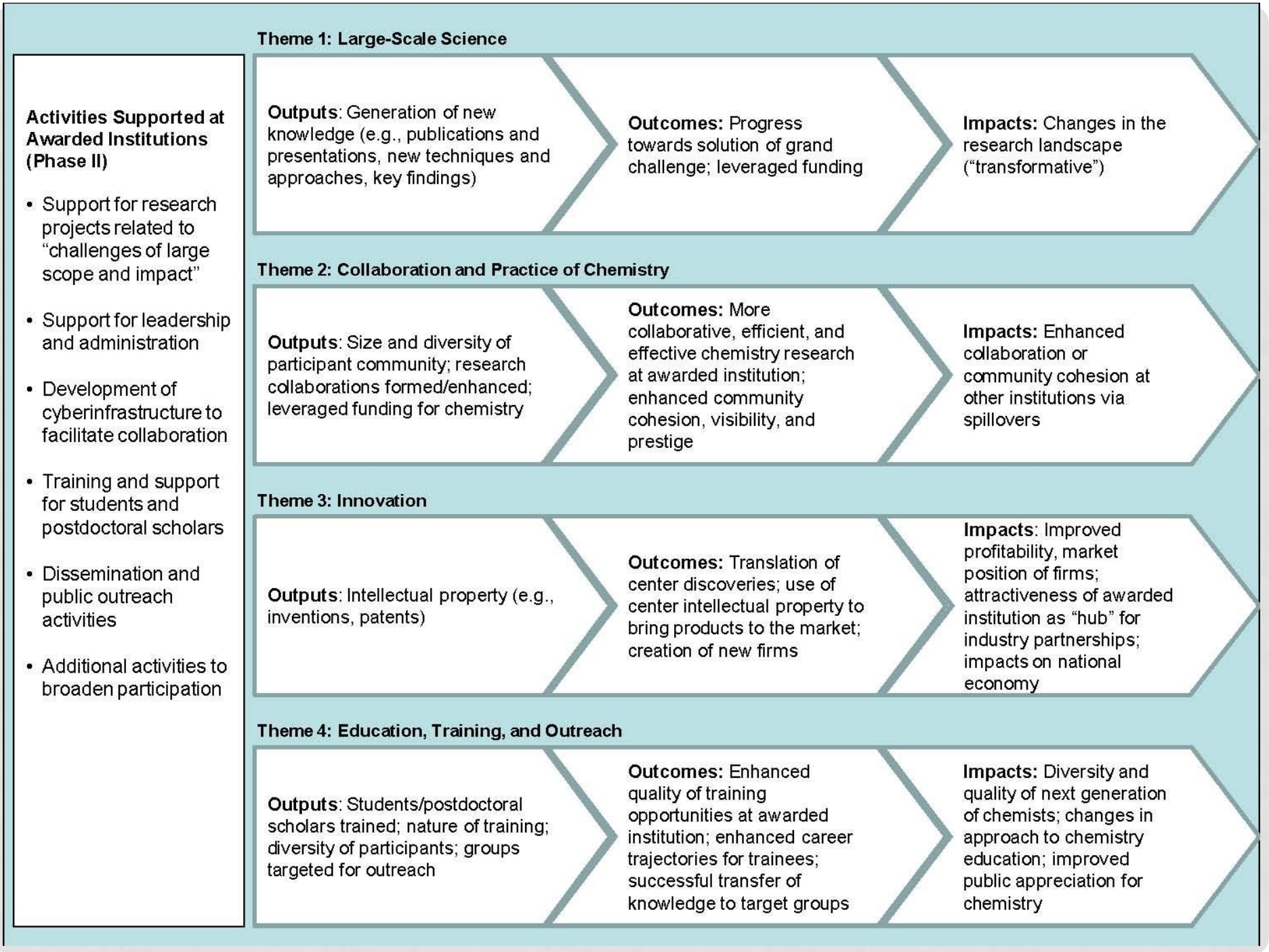
- If you accomplish your planned activities to the extent you intended, then your participants will benefit in certain ways.

Impact

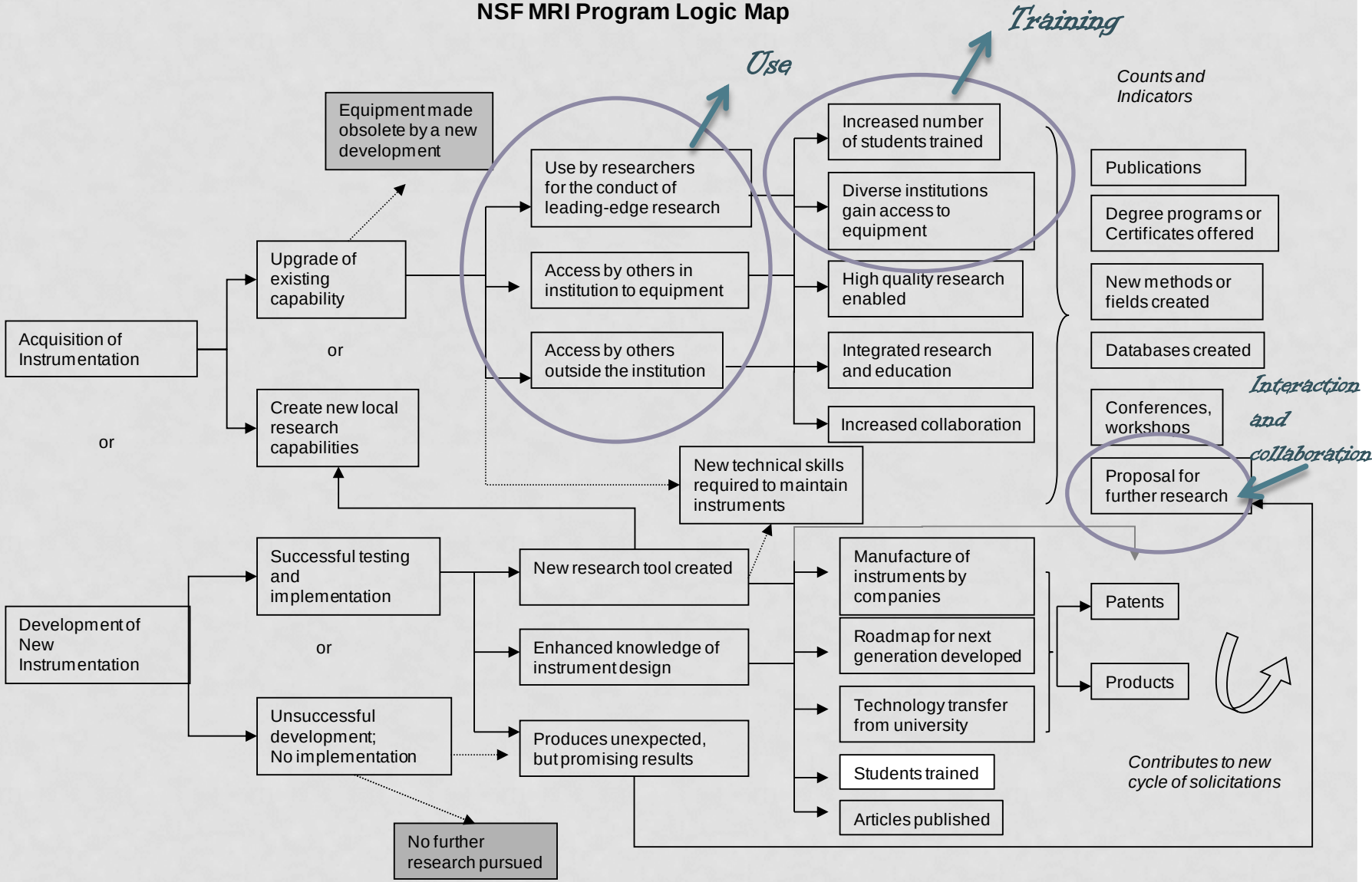
- If these benefits to participants are achieved, then certain changes might be expected to occur.

Your Planned Work

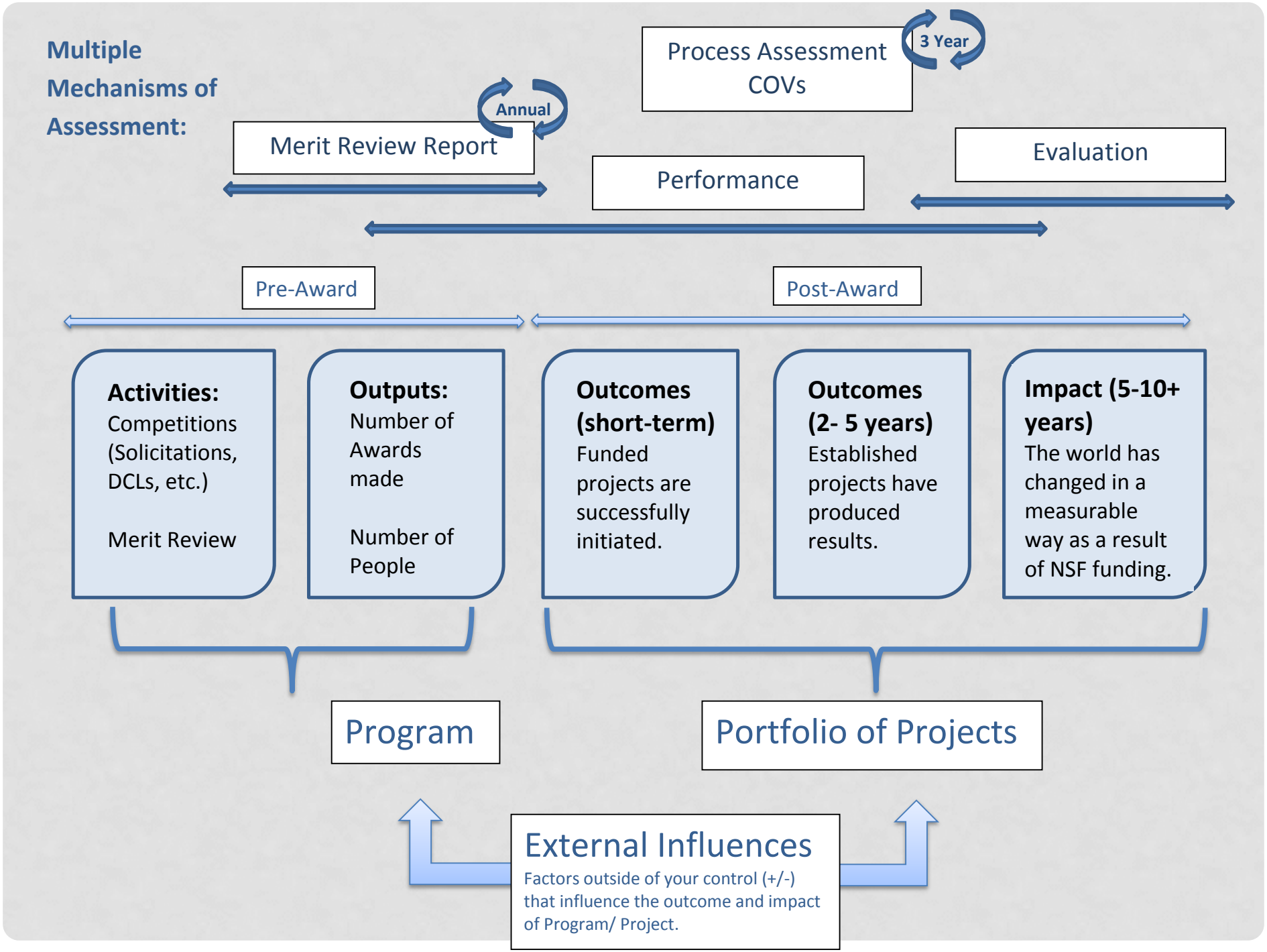
Your Intended Results


Activities Supported at Awarded Institutions (Phase II)
• Support for research projects related to "challenges of large scope and impact"
• Support for leadership and administration
• Development of cyberinfrastructure to facilitate collaboration
• Training and support for students and postdoctoral scholars
• Dissemination and public outreach activities
• Additional activities to broaden participation
Theme 1: Large-Scale Science
Outputs: Generation of new knowledge (e.g., publications and presentations, new techniques and approaches, key findings)
Outcomes: Progress towards solution of grand challenge; leveraged funding
Impacts: Changes in the research landscape ("transformative")
Theme 2: Collaboration and Practice of Chemistry
Outputs: Size and diversity of participant community; research collaborations formed/enhanced; leveraged funding for chemistry
Outcomes: More collaborative, efficient, and effective chemistry research at awarded institution; enhanced community cohesion, visibility, and prestige
Impacts: Enhanced collaboration or community cohesion at other institutions via spillovers
Theme 3: Innovation
Outputs: Intellectual property (e.g., inventions, patents)
Outcomes: Translation of center discoveries; use of center intellectual property to bring products to the market; creation of new firms
Impacts: Improved profitability, market position of firms; attractiveness of awarded institution as "hub" for industry partnerships; impacts on national economy
Theme 4: Education, Training, and Outreach
Outputs: Students/postdoctoral scholars trained; nature of training; diversity of participants; groups targeted for outreach
Outcomes: Enhanced quality of training opportunities at awarded institution; enhanced career trajectories for trainees; successful transfer of knowledge to target groups
Impacts: Diversity and quality of next generation of chemists; changes in approach to chemistry education; improved public appreciation for chemistry

NSF MRI Program Logic Map



NSF Activities (Inputs) **Actions** **Results** **Outputs** **Outcomes**


Multiple Mechanisms of Assessment:
Process Assessment COVs
3 Year
Annual
Merit Review Report
Performance
Evaluation
Pre-Award
Post-Award
Activities: Competitions (Solicitations, DCLs, etc.)
Merit Review
Outputs: Number of Awards made
Number of People
Outcomes (short-term) Funded projects are successfully initiated.
Outcomes (2- 5 years) Established projects have produced results.
Impact (5-10+ years) The world has changed in a measurable way as a result of NSF funding.
Program
Portfolio of Projects
External Influences
Factors outside of your control (+/-) that influence the outcome and impact of Program/ Project.

PROGRAM EVALUATION

is the systematic collection of information about the activities, characteristics, and outcomes of programs to make judgments about the program, improve program effectiveness, and/or inform decisions about future programming.

Patton, 2011

A "CONTINUUM" OF PROGRAM EVALUATION PROCESSES:

(1) Set up baseline information,
(2) Develop clear current program goals and program theory of action ***(Logic Models)***,
(3) Finalize meaningful and useful program measurement outcomes,
(4) (4) Put into place an appropriate program performance management system,
(5) Accumulate sufficient data and information from the performance management system, and
(6) Use evidence and data to implement target program improvements.
(6) If possible, measure impact using quasi-experimental design

CHALLENGES OF MEASURING IMPACT
(CAN THE CHANGE BE ATTRIBUTED TO THE INTERVENTION?)

- Measurement of impact requires a counterfactual analysis (a comparison between what actually happened and what would have happened in the absence of the intervention.
- In science it is often difficult or impossible to find the right counterfactual.
- There is no specific time interval where we can be sure that any or all impact has occurred.

Quasi-Experimental methods (matching, differencing, etc.) comparing program participants with non-participants based on observed selection characteristics.

QUASI-EXPERIMENTAL METHODS
(MATCHING, DIFFERENCING, ETC.)

- Matching involves comparing program participants with non-participants based on observed selection characteristics.
- Differencing, Difference-in-differences or double differences, which uses data collected at baseline and end-line for intervention and comparison groups, can be used to account for selection bias.

Knowing what type of quasi-experimental design you will use in the beginning will make your monitoring data more useful (i.e. you might collect baseline data on non-participants)

ADVICE FROM NSF?

- Make a plan that is a living document.
- Start at the beginning.
- Align metrics with goals.
- Use a logic model for alignment.
- Choose to measure things that are meaningful to you.
- Choose quality over quantity.
- Have a few key metrics that track critical success factors.
- You can't possibly measure the true value of science. Don't let that discourage you. If you are successful at measuring a fraction of your accomplishments it will be enough.

Measuring what you accomplish helps you describe what you do for a variety of audiences including: NSF Staff, site visitors, panelists reviewing your renewal, the public, congress, OMB. ***The most important user of your data should be you!***