Designing and Evaluating Innovative Research Programs: U.S. Examples

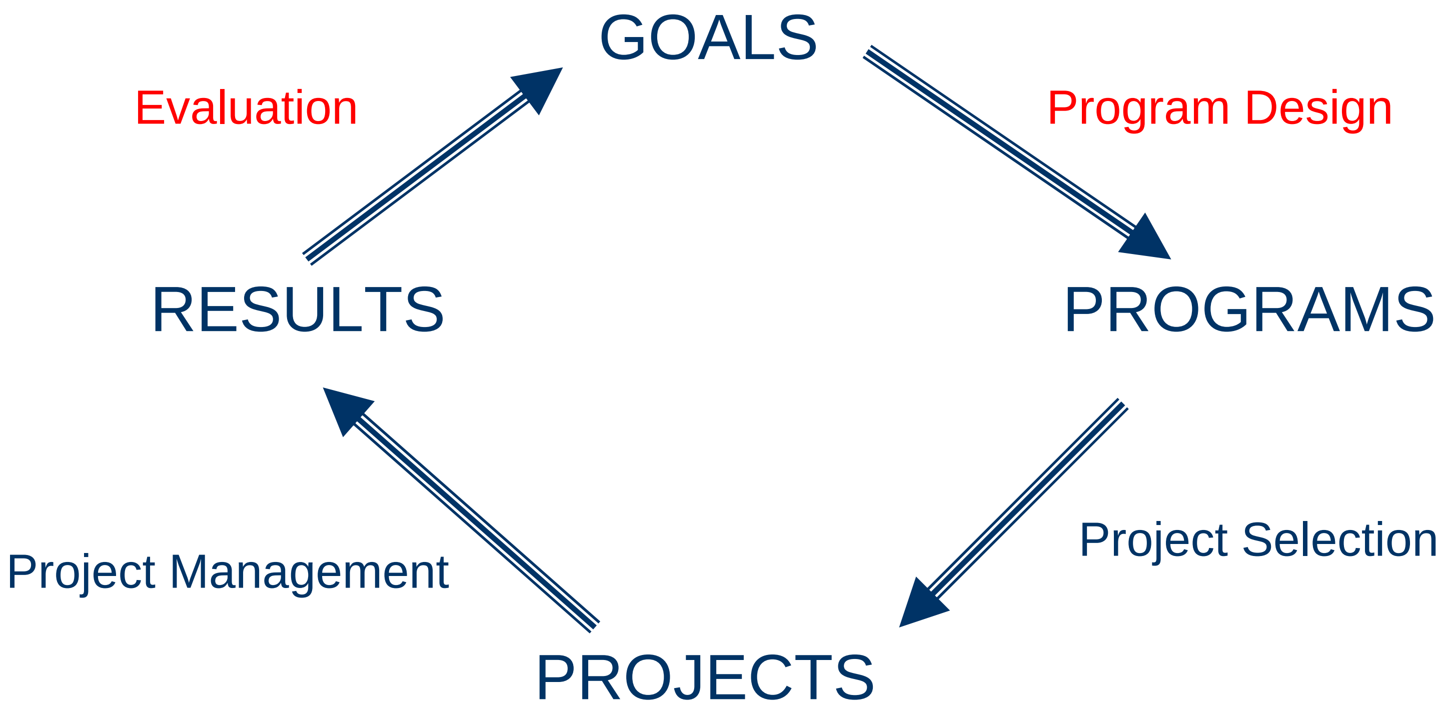
Susan E. Cozzens

Technology Policy and Assessment Center

School of Public Policy

Georgia Institute of Technology

Results-oriented management



In reality…

- Evaluation information is not often available at the time of program design.
  - New programs
  - Programs that have not been evaluated
- Past effectiveness is only one consideration in designing a program.
  - Goals may change.
  - Circumstances may change.
  - Constituencies may be entrenched.

Outline of talk

- What relationships between evaluation and design can we see in U.S. research programs?
- Part One: National Science Foundation
  - General information on NSF program design and evaluation
  - Examples
    - STCs – Science and Technology Centers
    - IGERT – Integrative Graduate Education and Research Training grants
- Part Two: Government laboratories
    - NIOSH: logic models and matrix management
    - ARS: stakeholder input and evaluation
    - NRL: competition for ideas and resources

National Science Foundation

Goals

Programs

Evaluation

NSF mission

- The National Science Foundation Act of 1950 (Public Law 81-507) set forth NSF's mission and purpose:
- *To promote the progress of science; to advance the national health, prosperity, and welfare; to secure the national defense....*

Specific purposes

- The Act authorized and directed NSF to initiate and support:
  - basic scientific research and research fundamental to the engineering process,
  - programs to strengthen scientific and engineering research potential,
  - science and engineering education programs at all levels and in all the various fields of science and engineering,
  - programs that provide a source of information for policy formulation,
  - and other activities to promote these ends.

Amended over the years

- 1968: authority to support applied research.
- 1980: authority to support activities to improve the participation of women and minorities in science and engineering
- 1986: engineering given equal status to science
- But not a basic redesign.

Evolving Goals -- 1995

- Enable the United States to uphold a position of world leadership in all aspects of science, mathematics, and engineering.
- Promote the discovery, integration, dissemination, and employment of new knowledge in service to society.
- Achieve excellence in U.S. science, mathematics, engineering, and technology education at all levels.

Evolving Goals: 1997

- NSF expects the following as outcomes from its investments, taken in the aggregate and observed over time.
  - Discoveries at and across the frontier of science and engineering;
  - Connections between discoveries and their use in service to society;
  - A diverse, globally-oriented workforce of scientists and engineers;
  - Improved achievement in mathematics and science skills needed by all Americans; and
  - Timely and relevant information on the national and international science and engineering enterprise.

Evolving Goals -- 2003

- **People Goal** – A diverse, competitive, and globally-engaged U.S. workforce of scientists, engineers, technologists and well-prepared citizens
- **Ideas Goal** - Discovery across the frontier of science and engineering, connected to learning, innovation and service to society
- **Tools Goal** – Broadly accessible, state-of-the-art S&E facilities, tools and other infrastructure that enable discovery, learning and innovation

Evolving Goals: Current

- Discovery
  - Foster research that will advance the frontiers of knowledge, emphasizing areas of greatest opportunity and potential benefit and establishing the nation as a global leader in fundamental and transformational science and engineering
- Learning
  - Cultivate a world-class, broadly inclusive science and engineering workforce, and expand the scientific literacy of all citizens
- Research Infrastructure
  - Build the nation's research capability through critical investments in advanced instrumentation, facilities, cyber-infrastructure, and experimental tools
- Stewardship
  - Support excellence in science and engineering education through a capable and responsive organization

Bottom line

- The specifics of goals change
- But core goals stay the same

NSF Program Portfolio, FY 2010

- Discovery: $5.7 billion
  - Mostly standard grants
- Learning: $857 million
  - A wide variety of programs
- Infrastructure: $117 million
  - Several specialized programs
- Stewardship: $318 million

Elements of program design

- Reflected in document called program solicitation
- Scope and purpose of program
- What should projects consist of?
  - Collaboration requirements and permissions
  - Required elements
- Selection process
  - Eligibility -- How many per institution and investigator?
  - Pre-proposals?
- Review and reporting requirements

Project cycle

- For standard projects
  - Selection through peer review (ad hoc and panel)
  - Management fully delegated
  - Results
    - Annual report
    - Final project reports
- For larger grants, a much more intensive process.
  - E.g., for centers, annual and mid-term reviews

Ex Post Program Evaluation

- Special programs have received special attention for decades.
- Standard programs examined by Committees of Visitors – mostly focused on project selection process.
- Agency reports on performance through Advisory Committee on GPRA (Government Performance and Results Act)

ACGPA 2009 Report

- Evaluating against strategic plan for 2006-2011
- Goals of Discovery, Learning, Research Infrastructure
- Provided with highlights of program accomplishments in these three areas
- YES or NO ("successfully met performance objectives")

ACGPA on "future assessment"

RECOMMENDATIONS

1. *Consider an assessment framework that uses multiple measures and methods, applied over various time scales. Use both quantitative and qualitative evidence, including highlights.*
2. *Emphasize the dynamic relationships among strategic goals and outcomes.*
3. *Use performance assessment as an opportunity and means to document the strategic value of NSF's science investments to the nation and the public.*
4. *Engage the scientific community as a partner in performance assessment.*
5. *Build assessment into the organizational and programmatic infrastructure of NSF.*

NSF Centers Programs

- Lots of evaluation
- New designs in new programs

Centers programs

- Centers goals
  - Used for many purposes over the years
  - 1990s used as instrument of institutional change
- Centers program designs
  - Generally experimental at least at first
- Centers evaluations started early and happen often.
  - They have frequently pioneered methods.
  - Often high profile, many stakeholder groups
  - Combined from early stages with monitoring
  - Used more for justification than for program design

Materials Research Centers

- First centers program at NSF.
  - Started at another agency.
  - Focused on provision of instruments.
- Evaluated in 1978.
  - Explicitly comparative – were centers as effective as individual project grants?
    - Used 15 top universities in project grants as comparison
  - Early use of “converging indicators” – both bibliometric and reputational.
    - Peer review of papers; plus citation analysis.
  - Overhead comparisons

MRC Evaluation Findings

- Comparison with project grants
  - Centers much more efficient in overhead costs.
  - Research at MRLs were much more experimental rather than theoretical.
  - Highest citations to investigators with individual NSF grants.
  - MRLs had more achievements in the top 15%.

Diffusion of the Center Concept

- University-Industry Cooperative Research Centers, started late 1970s.
- Engineering Research Centers, started mid-1980s.
- Minority institution centers: RIMI (1982-95), then CREST
- Science and Technology Centers, started 1987.

Evaluation of IUCRCs

- Program design: small, focused on collaboration
- Center evaluator built in from the beginning
- Monitoring also built in from the beginning
- Provides baseline for comparison with other programs
- Tracks best practices
- Collects examples of breakthroughs

Engineering Research Centers

- Goal: Change the culture of engineering, increase competitiveness of U.S. firms
- Program design
  - High levels of interaction with industry
  - Integrate engineering knowledge
    - Interdisciplinary, team-based collaborative research
  - Contribute to effectiveness of all levels of engineering education

ERC Evaluations

- ERCs evaluated by blue-ribbon Academy panel in late 1980s; positive findings.
- Ten-year out studies
  - 1994 ERC-industry Interaction Study
    - More is better
  - 1994 ERC Graduate Effectiveness Study
    - Culture had spread; hard to find differences
- Recent pilot studies on regional impact

Science & Technology Centers (STCs)

- Started 1989
- Goals
  - Exploit special opportunities
    - Complexity, scale, duration, facilities
  - Educate students to be aware of applications
  - Provide stable, long term funding
- Structure similar to ERCs
- Two competitions led to 25 centers by early 1990s

Center Evaluation Workshop

- NSF, 1992, considered the several programs, with special attention to ERCs and STCs
- Three dimensions
  - Research
  - Technology Transfer
  - Education
  - Plus institutional impacts
- Systematic data collection good for activities, collaborations, short on “value added” and outcomes.

STC Evaluation(s) 1996

- Large contract awarded to Abt Associates for the GPRA pilot project evaluation.
  - Pioneered "alternative format" GPRA goals.
  - Interviews, historical review, analysis of secondary data, bibliometric and patent data, surveys of related populations.
- Panel review of data done at the National Research Council.
- Congress asks NAPA to do a third, which focuses on NSF management.

STC 1996 Results

- Results:
  - The program has been successful and should be continued.
  - The central office should coordinate better with the directorates.
  - Program had increased emphasis on multi-disciplinarity and K-12 education – this could be moderated
- Program redesigned by NSF leadership, based on their vision of the future.
  - Required collaboration across institutions.
  - Now called “integrative partnerships”

Current STC Design

- Three elements maintained
  - Research (“ambitious research vision or theme”)
  - Education and human resources
  - Knowledge transfer
- Must be multi-institutional
  - Partners can be industry, government laboratories, or others
  - Reflected in External Advisory Board
- Diversity
  - Includes “under-represented groups”
- Leadership and management is a key element.
- Participate in national network

STC Award Conditions

- **Special Award Conditions:** STC awards are made in the form of Cooperative Agreements. The STC Cooperative Agreements will have an extensive section of Special Conditions relating to the period of performance, statement of work, awardee responsibilities, NSF responsibilities, joint NSF-awardee responsibilities, funding and funding schedule, reporting requirements, key personnel, and other conditions. NSF has responsibility for providing general oversight and monitoring of STCs to help assure effective performance and administration, as well as facilitating any coordination among the STCs as necessary to further the objectives of the STC program. Prior to finalizing the Cooperative Agreement, a retreat of the Center's key personnel to address strategic planning of the STC will be required.

2007 MRSEC Evaluation

- Goals
  - Assess performance and impacts
  - Consider future directions
- National Academies committee
  - Used questionnaires, telephone interviews, site visits
- Still had a hard time attributing achievements to the Centers specifically.

2007 MRSEC recommendations

- Restructure program to allow more efficient use and leveraging of resources
- Invest in both centers of excellence and stand alone teams of researchers
  - Materials Centers of Excellence (MCEs)
  - Materials Research Groups (MRGs)
- MRSECs should operate as a national network.
- Remarkable: Almost no mention of nanotechnology

IGERT (Integrative Graduate Education and Research Training Program)

- Background
  - Tension between research and education
  - COSEPUP report on reshaping graduate education (1995)
    - Flexibility, versatility in careers
    - Interdisciplinarity
    - Control time to degree
    - Attract more women and minorities
- Four major studies followed; similar conclusions
  - Include ethics and values
  - Include global exposure

Program Design 1998

- Goals:
  - The program is intended to catalyze a cultural change in graduate education, for students, faculty, and institutions,
  - by establishing innovative models for graduate education and training
  - in a fertile environment for collaborative research that transcends traditional disciplinary boundaries.
  - It is also intended to contribute to a world-class, broadly inclusive, and globally engaged science and engineering workforce.
- Institutions receive $3 million over five years
- Mostly spent on student support, curriculum development

Awards Made

- In first eight years
  - 125 programs supported
  - 2900 graduate students trained
  - $300 million in committed funds
- Examples
  - Health assistive smart environments
  - Resilience of ecological and social systems
  - Motor control and movement
  - Materials for a sustainable future

Evaluation in 2006

- Surveys and interviews
  - Both IGERT and comparison group
- Results
  - Impacts on recruitment: Modest; achieving national averages
  - Impacts on faculty: Positive reports
  - Impacts on students: Education is distinctive.
  - Institutionalization: Evidence of change in policies

Current IGERT Program Design

- Comprehensive, interdisciplinary theme
- Integration of research with graduate education
  - Fosters understanding of global nature and context of theme
  - State of the art equipment and methods
  - Career development opportunities
- Program strategy for recruiting under-represented groups
- Strategy for assessing project results
- Plan for disseminating experience
- Institutional commitment

Summary

- NSF's basic programs follow a standard design that has not changed much.
- NSF sometimes introduces new programs with experimental designs.
  - These are more likely to receive in-depth evaluations.
- Once a program is moving, changes in design are more likely to happen within projects than at program level.

Where does NSF innovate?

- Goals: Wording changes but basic functions are maintained.
- Portfolio of programs: Innovation takes the form of new programs with new designs.
- Programs: Considerable momentum in design; elements may be strengthened.
- Projects: Innovation strongly encouraged; turnover enforces this.

Laboratory Programs

Design and Evaluation

U.S. Federal Laboratories

- $25 million spent intramurally
  - Half of the intramural is Department of Defense
  - Examples: Army Research Laboratory, Naval Research Laboratory
- An additional $10 billion in FFRDCs (federally funded research and development centers) mostly under Department of Energy
  - Examples: Lawrence Berkeley Laboratory, Oak Ridge National Laboratory, Brookhaven National Laboratory
- Mostly “mission oriented” but also basic energy sciences

Issues in the lab systems

- Clarify mission
- Let function determine size
- Develop independent personnel system
- Link to private industry (CRADAs)
- Work on technology transfer
- Improve management systems
- Develop core competencies

Program design?

- Laboratory goals are complicated by multiple sponsors.
- Priority setting may be externally driven.
- Individuals and laboratories have their own trajectories.
- Lines of activity are very hard to redirect or shut down.
- But the collection needs to be coherent enough to explain to core funders.

National Institutes of Health

- External Boards of Scientific Counselors do program and personnel assessment
- Evaluation focuses at individual level, with some attention to laboratory
- Indicators used in personnel review are like university labs

Department of Energy

- Assessment takes place in annual budget cycle
- Projects peer-reviewed prospectively, and sometimes retrospectively
- Program review processes at labs
- Technical targets serve as annual performance goals

National Institute of Standards & Technology

- Program review done externally by NRC
- Strong management involvement
- Case studies of economic impact at project level
- Some technical benchmarking
- Qualitative review has limits in external communication

Army Research Laboratory

- Stakeholder involvement through a special board
- Technical advice through contract with NRC
- Customer feedback through project-level survey
- Metrics used sparingly, privately

Environmental Protection Agency

- Program structure keyed to strategic plan, reflected in budget
- Criteria focus on impact
  - Will it make a big difference in risk assessments?
  - Will it improve risk management by an order of magnitude?
- Strong peer review of rulings and reports

Naval Research Laboratory

- 80 percent of funding comes through projects supported by external clients
- 20 percent comes through Base Program funded by Office of Naval Research
- Strong competitive internal process
- Good criteria for personnel decisions
- Strong management integration of client needs with long-term investments

NRL basics

- $800 million budget
- Pays personnel, overhead, partners
- 3000 people
  - including 1500 scientists and engineers
  - 1500 additional at partner organizations
- Strongly project based system
- Basic research through testing and development

Base funding

- About 20% comes from Office of Naval Research
- Laboratory’s Research Advisory Committee allocates these funds.
  - Includes “focus area coordinators”
- They propose a program to ONR based on what they need to maintain capabilities.

Project competition

- Each unit receives turnover target.
- Criteria for selection:
  - S&T merit
  - Performer credentials
  - Facilities available
  - Navy relevance
  - Transition potential
- Units can make room by not continuing something.
- Projects go through informal retrospective program review.

Customer supported program

- 80 percent of lab work is chosen by customers
  - Navy systems commands
  - Army, Air Force, Defense Advanced Research Projects Agency (DARPA), NASA, Department of Energy
- Working relationship plays critical role.

Personnel assessment

- Standard government processes
- Tracks publications, citations, patents, CRADAs (cooperative research and development agreements)
- Project based system gives clear criterion for unproductive employees.
- About 3% leave each year under these criteria.

Strengths and weaknesses

- Strengths
  - Flexibility, responsiveness to change
  - Pluralism – creates funding stability
- Weaknesses
  - Frenetic marketing
  - Taking in laundry
- Overall, very strong in comparison with other federal laboratories.

Agricultural Research Service

- Size
  - $745 million
  - 1100 projects
  - 22 national programs
- Similarities to NIOSH
  - traditional laboratory tasks
  - physically scattered facilities (100 locations)
- Stimulus for new structure
  - NAS study in 1998
  - GPRA implementation also 1998

Process

- Strategic plan goals set under GPRA
- Research goals reflected Department goals when this was first done in the late 1990s.
  - No longer the case now, with new Department goals.
- Every project aligned to a strategic goal, which formed the basis for evaluation.
- National program leaders appointed.

Project assessment

- New Office of Scientific Quality Reviews
- All projects go through external reviews at the same time.
  - Reviewers are PhD scientists from outside ARS, both government and non-government.
- Projects can be revised if they are not acceptable on the first review.

Assessment in the Research Process

- Between program reviews, managers maintain progress towards goals.
  - Area offices review
  - Submit report on accomplishments and impacts
- Individual scientists also subject to rigorous review processes.
  - Process has great credibility in the agency.
- Retrospective external review being considered.

Communication

- The new structure communicates value well outside the organization.
- It also communicates organizational goals to the staff, who now know where they fit.
- Annual reports from national programs are available on the web site.
- Congressional staff are noticing, paying more attention to value for the public.

ARS summary

- Insuring, measuring, and demonstrating all rest heavily on strategic plan
- New national program structure, keyed to plan
- New program review system, looking at all projects in a program at once
- Improved internal and external communication

NIOSH

- National Institute of Occupational Safety and Health
- $424 million
- Scattered laboratories
- Very complicated stakeholder groups
  - Unions
  - Employers
  - Manufacturers
  - Congressional delegations

Strategic planning process

- National Occupational Research Agenda (NORA)
  - Stakeholder driven process setting long term priorities
  - Unveiled in 1996
  - Town Hall meetings 2005-6
- Priorities set by
  - The numbers of workers at risk for a particular injury or illness
  - The seriousness of the hazard or issue
  - The probability that new information and approaches will make a difference
- Now organized on a sector basis – sector agendas coming out in late 2008 - mid 2009.

Laboratory structures

- Laboratories scattered
  - Washington, Atlanta, Cincinnati OH, Morgantown WV, Pittsburgh PA, Spokane WA, Anchorage AK, Lakewood CO
- No simple correspondence with NORA categories
- Firewall between intramural and extramural research

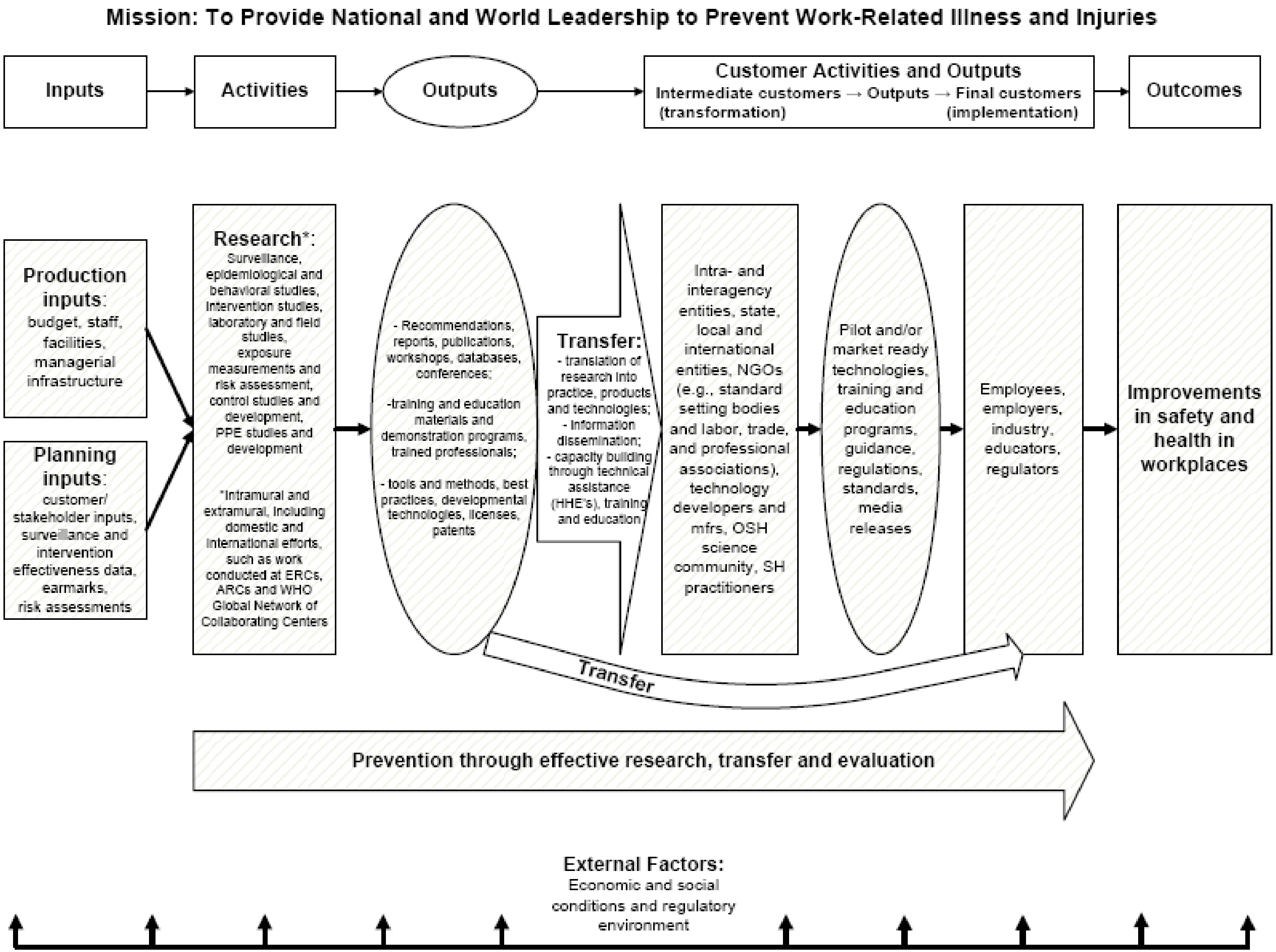
Commissioned IOM evaluations

- Pressure from OMB
- Framework Committee established
  - Developed detailed guidance for evaluation committees
  - Evidence packages evolved as the work went forward
- Worked with outside consultant on logic model

Review on logic models

- Linear representation of program logic
  - Inputs, activities, outputs, outcomes, impacts
- Becomes a set of “If.. Then… “ statements about the program
- Forms the basis for choosing performance indicators

Figure 1. The NIOSH operational plan presented as a logic model



Committees use the logic model

- Had specific versions for each “program”
- Could sort into the categories, with some disagreement
- Intermediate outcomes were the most influential in review results because…
- Outcome and impact data were largely not available (surveillance)

NIOSH use of logic models

- Groupings of projects that made sense did not follow traditional lines
- Groupings also stretched across physical laboratories
- Changing leadership and strategic plans made “program” structure into a paper exercise
- Very hard for evaluation to line up people, projects, and goals
- Therefore: evaluation may not have helped program design very much.

Summary of Laboratory Systems

Indicators of success

Excellence in Govt. Science

- Contributes to mission
  - produces outcomes for the public
  - objectives set in strategic plan
- Sound science
  - passes muster with peers
  - improving over time
  - higher standards for regulatory products
- Keeps customers happy

Excellence in science management

- Negotiate strategic goals
- Serious use of external technical advice
- Using both of these for budget and program decisions
- Tough, credible personnel evaluation
- Sensible application of metrics
- Incorporate communication

Characteristics of Excellence: Project Selection

- Users identified clearly, consulted in planning
- High need, high potential impact
- High quality of research
  - sound methods and research plan
  - good team, track record of success
  - necessary infrastructure
- Clear dissemination plan
- Fits into balanced portfolio?
- Review process should indicate priority

Characteristics of Excellence: Scientific Inquiry

- Project monitoring
  - Are the milestones and objectives being met?
  - Is the need still there?
  - Uh-oh, OK, or wow
- Program review
  - Balancing flexibility and continuity?
  - Responding to changes in external environment?
  - Responding to changes in science?

Characteristics of Excellence: Results

- Project level
  - Did the project achieve its planned result?
  - Are the results being used?
  - Did it contribute to a body of knowledge?
- Program level
  - Project level questions apply again.
  - Has the set of projects contributed to the strategic goal of the program and agency?

Characteristics of Excellence: Impact

- Impacts (long-term outcomes for the public) should be defined clearly at agency level.
- Sometimes, research programs will clearly contribute to achieving those impacts.
- Often, the links will be hard to trace.
- “Best judgments” should be informed by clear objectives and solid information.

Characteristics of Excellence: Communication...

- … of the science itself
  - delivered in usable form
  - accessible, timely
  - geared to several levels
- … of the excellence of the science
  - Success stories and awards are widely used.
  - Describe management processes clearly.
  - Users attest best to quality and impact.

Overall Summary

Design and Evaluation

- The theoretical relationship among goals, designs, and evaluation seldom appear in the reality of research programs.
- Program design is an eclectic art, rather than an applied science.
  - Responding to changes in the environment
  - Designing processes and relationships
  - Leaving room for creative project adaptation
  - Finding places for both ex ante and ex post evaluation