

Evaluation and Performance Measurement for Research and Technology Development Programs

SKEP ERA-NET Workshop

April 18 - 20, 2007

Aland Islands, Finland

Dr. Gretchen Jordan

Sandia National Laboratories, USA

Gbjorda@sandia.gov, 505-844-9075

Some of the work presented here was completed for the U.S. Department of Energy (DOE). Sandia National Laboratories is operated by Sandia Corporation, a subsidiary of Lockheed Martin Corporation, for the DOE under Contract DE-AC04-94AL8500. **Opinions expressed are solely those of the author.**

Outline

- Definitions
- Requirements and challenges
- When and why we evaluate
- Managing program evaluations
- Key evaluation questions - an example

Program Evaluation and Performance Measurement both provide performance information

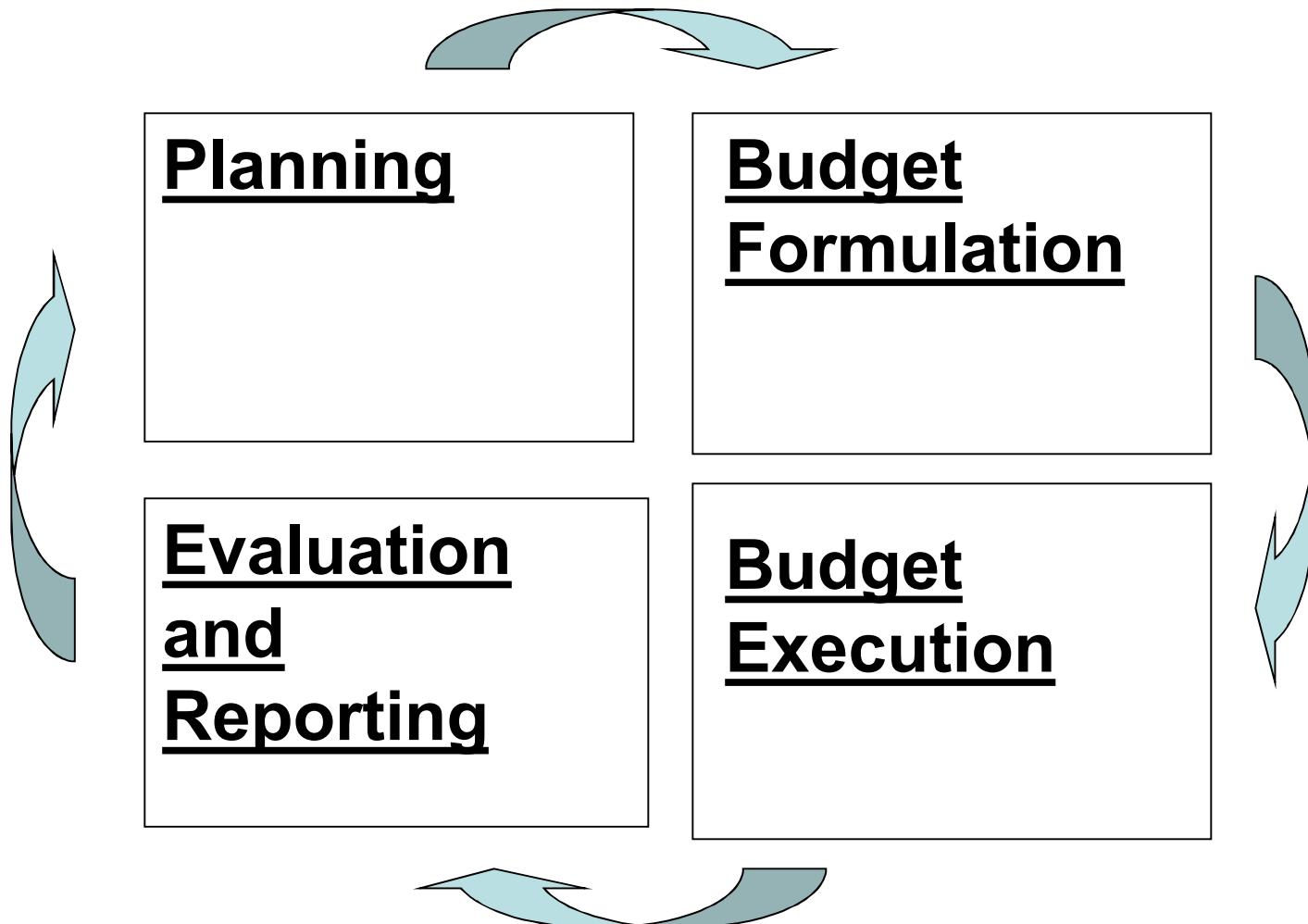
Program evaluation, as used in the Government Performance and Results Act (GPRA) is an assessment, through objective measurement and statistical analysis, of the manner and extent to which federal programs achieve intended objectives.

Performance measurement, which is complementary, is the ongoing monitoring and reporting of program accomplishments, particularly towards pre-established goals.

GAO/GGD-98-53

A program is an intentional allocation of resources in support of specific strategies or activities to produce defined products and/or services for a defined client group so that they will change their behavior which will lead to reducing strategic problems necessary to achieve the program's mission.

Performance Based Management Uses Performance Information in All Areas of Management



Differences

Measurement/Metrics	Evaluation
Quick	Slow
Continuous	Periodic
Inexpensive	Costly
Primarily Outputs and Outcomes	Processes, Outputs, Outcomes, Impacts
Answers “What?” (data)	Answers “Why?” (causality)

Requirements and Challenges

Objectives common in OECD performance management frameworks

- Continuous improvement in performance - to improve internal functioning
- Improve mechanisms to distribute and clarify responsibilities and control – to define new ways of interacting with partners in the societal network
- Realize savings by shrinking activities and budgets and increasing efficiency gains - to deal with need for fiscal restraint (or reduction)

Source: In Search of Results: Performance Management Practices,
Organization For Economic Co-operation and Development, 1997.

Current general trends require a more comprehensive performance management, which means a change to . . .

- General strategic review rather than ex ante control
- A center which devolves operational responsibility and holds those in charge accountable
- An autonomous periphery which accepts responsibility and is accountable
- A contract-oriented and results-focused steering of activities, rather than steering focused on inputs and administrative law

Requirements: in the U.S. and around the world, legislation like the Government Performance and Results Act (GPRA)

Strategic Plan (Agency Level)

- Mission statement, Goals and objectives
- How (resources, etc.) goals will be achieved
- External factors affecting goal achievement
- Program evaluations that will be used to set and revise goals

Annual Performance Plans (Budget Level)

- Linked to Agency Strategic Plan
- Measurable performance goals
- Performance indicators for output and outcomes
- Basis for comparison (baseline) and Means to verify/validate measured values

Performance Report (Budget Level)

- Performance indicators and actual performance for FY
- Explanation of why goals were not met
- Plans for achieving goals not yet met, which can never be met
- Summary of program evaluations completed

White House Office of Management and Budget Requirements: Program Assessment Rating Tool

OMB PART for Applied R&D Programs

Program Purpose & Design	Strategic Planning	Program Management	Program Results/ Accountability
<p>1.1 Purpose clear?</p> <p>1.2 Address a specific problem, interest, or need?</p> <p>1.3 Not duplicative of other Federal, state, local or private efforts?</p> <p>1.4 Design free of major flaws?</p> <p>1.5 Effectively targeted – resources reach intended beneficiaries and/or address purpose directly?</p>	<p>2.1 Meaningful long-term performance measures?</p> <p>2.2 Targets & timeframes for long-term measures?</p> <p>2.3 Annual performance measures?</p> <p>2.4 Baselines and targets for annual measures?</p> <p>2.5 Partners work toward long-term goals?</p> <p>2.6 Independent evaluations?</p> <p>2.7 Budget requests tied to annual and long-term goals?</p> <p>2.8 Correcting strategic planning deficiencies?</p> <p>2.RD1 Compare program benefits to similar efforts?</p> <p>2.RD2 Prioritization process for budget and funding decisions?</p>	<p>3.1 Regular collection of performance information to manage program?</p> <p>3.2 Managers and partners held accountable?</p> <p>3.3 Funds obligated timely and spent for intended purpose?</p> <p>3.4 Procedures to measure & achieve efficiencies & cost effectiveness?</p> <p>3.5 Collaborate and coordinate with related programs?</p> <p>3.6 Strong financial management practices?</p> <p>3.7 Addressing management deficiencies?</p> <p>3.RD1 Allocate funds and use management processes that maintain program quality?</p>	<p>4.1 Demonstrated progress towards long-term performance goals?</p> <p>4.2 Achieve annual performance goals?</p> <p>4.3 Improved efficiencies or cost effectiveness towards program goals?</p> <p>4.4 Compare favorably to similar government or private efforts?</p> <p>4.5 Independent evaluations indicate that program is effective and achieving results?</p>

Will requirements be effective? Criteria to judge approaches for evaluation or performance indicators

1. Is there a plausible theory of how it is intended to contribute to program improvement (that is, the program meeting its customers' needs)
2. Is there detailed guidance for others to implement the approach?
3. Is there evidence that it can and often does contribute to program improvement?

Patricia Rogers, Royal Melbourne Institute of Technology

http://www.city.grande-prairie.ab.ca/perfm_a.htm (no longer active)

Rogers: Two of the ways performance indicators are thought to improve programs are flawed.

- It is plausible that indicators make program managers more informed
 - a. supplement considerable knowledge of program
 - b. are the source of questions, not answers
- Making public sector more accountable or Informing resource allocation depend on treating indicators as actual measures of performance – which they are not.
 - a. must be publicly reported and comprehensible to people outside the program
 - b. tendency is to simplify if objectives are complex, conflicting, changing, or necessarily opportunistic

The demand for research performance assessment has put pressure for change on both evaluators and the bench scientist.

Evaluators

- Increased emphasis on indicators, not on program evaluation
- Getting evidence of current quality, relevance, and performance is difficult
- No new evaluation methods, little guidance on indicators
- No research on efficacy of performance indicators
- Lack of skills, no network

Scientists, Science managers

- Goal displacement when indicators are too simple
- Unintended outcomes are ignored
- Rigid use of indicators means can't respond to changes
- Use of too narrow a set of indicators means choose inferior projects/contractors

Research Assessment in U.S.

Strengths Project-level review by peers Bibliometrics data for some fields Retrospective case studies A culture of evaluative inquiry/experimentation	Weaknesses No agreed upon way to assess portfolios Small, splintered R&D evaluation community Few new methods; existing methods often expensive Data issues, including attribution
Opportunities Requirements are centralizing, emphasize evaluation Requirements have reached level of labs and bench New computing power WREN	Threats Tension between control mentality and nature of scientific work Performance indicators selected/used out of context Reliance on linear model

Several Strategies for Building On Strengths and Guarding Against Threats -1

- Participate in building processes and methods that are
 - Responsive to both short term and longer term values
 - Cost-effective and Useful
 - Strategically timed
 - Integrated
 - Accurate
 - Engaging and Participatory
 - Practical and Helpful
- Increase training and networking for evaluators
- Develop less intrusive evaluation methods using new computing power, especially at the portfolio level

Several Strategies for Building On Strengths and Guarding Against Threats -2

- Select indicators that do no harm and are targets along the way to desired outcomes
- Integrate performance monitoring and evaluation (including peer review)
- Assess effects of performance management on the research environment and on progress
- Agree on an innovation system framework so that comparable data can be collected and disseminated
- Get scientists involved in defining outcomes that can be generalized across fields/areas of research

When and Why We Evaluate

Five forms of evaluation

- Proactive
- Clarificative
- Interactive
- Monitoring
- Impact

Owen with Rogers, Program Evaluation:
Forms and Approaches, 1999

Proactive evaluation (formative, prospective)

- When: before the program begins
- Orientation: Synthesis, to assist with planning decisions about what type of program that is needed
- Major focus: Program context
- Typical issues: Is there a need? What do we/others know about the problems to be addressed? Best practices?
- Major approaches:
 - Needs assessment or analysis
 - Research/literature review
 - Review of best practice, creation of benchmarks
- Assembly of evidence: Review of documents/databases, site visits, other interactive methods such as focus groups, delphi technique

Clarificative evaluation

- When: during program development
- Orientation: Clarification
- Major focus: All elements
- Typical issues: What are intended outcomes and how is program designed to achieve them? Underlying rationale? Plausible? Elements to be modified? What to assess?
- Major approaches:
 - Evaluability assessment
 - Logic/theory development
 - Accreditation
- Assembly of evidence: Combination of document analysis, interview and observation

Interactive evaluation

- When: during program development
- Orientation: Improvement
- Major focus: Delivery
- Typical issues: What is program trying to achieve? Is delivery working, consistent with plan? How could program or organization be changed to be more effective?
- Major approaches:
 - Responsive
 - Action research
 - Quality review
 - Developmental
 - Empowerment
- Assembly of evidence: Relies on intensive onsite studies, including observation. May involve providers and program participants

Monitoring evaluation

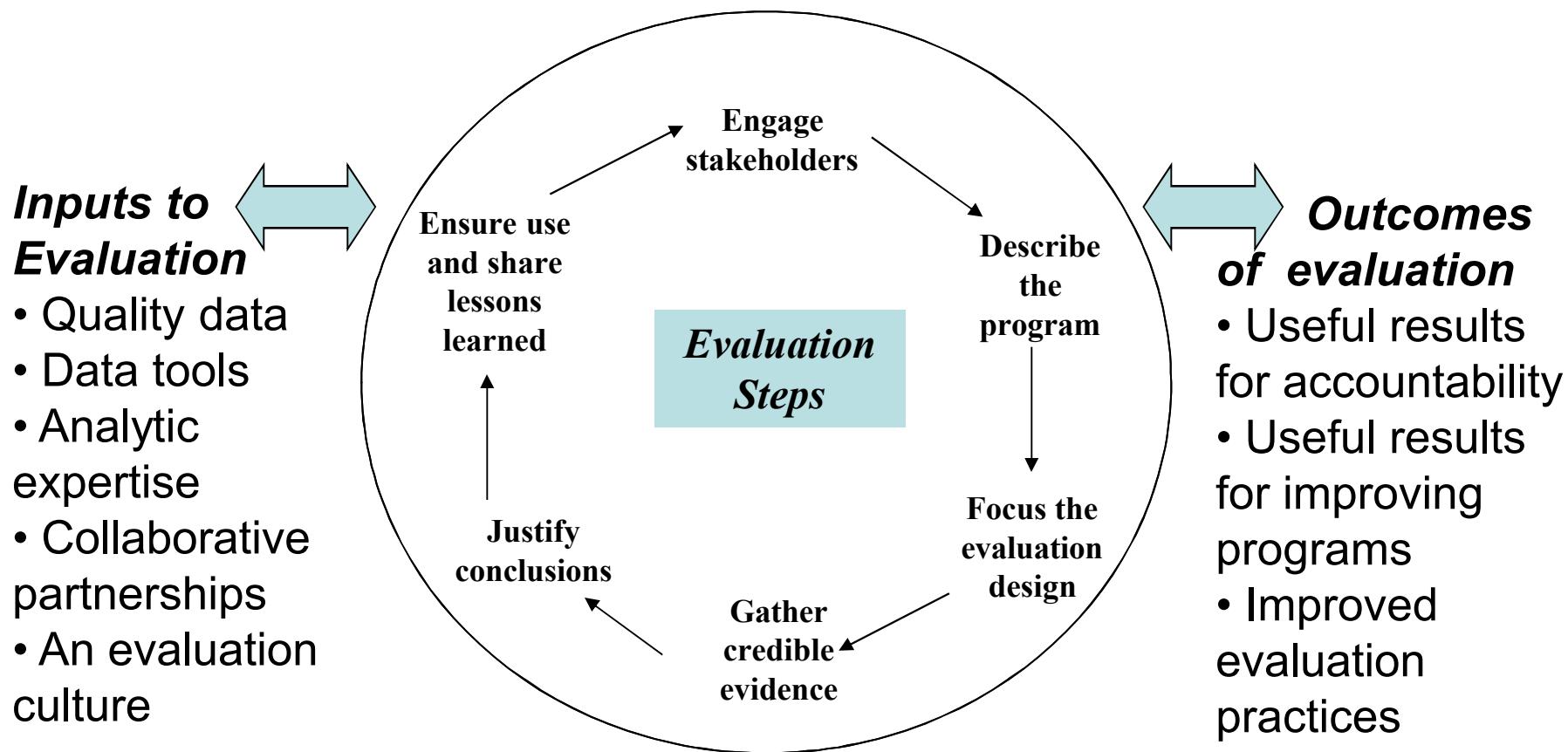
- When: once program has settled
- Orientation: Justification/fine tuning
- Major focus: Delivery and outcomes
- Typical issues: Is program reaching target population? Is implementation meeting benchmarks? Differences across sites, time? How/what can be changed to be more efficient, effective?
- Major approaches:
 - Component analysis
 - Devolved performance assessment
 - Systems analysis
- Assembly of evidence: Requires availability of management information systems, meaningful use of indicators and other performance information

Impact evaluation (summative, retrospective)

- When: after a program has settled
- Orientation: Justification/accountability
- Major focus: Delivery/ outcomes
- Typical issues: Program implemented as planned?
Stated goals achieved? Needs served? Can you
attribute goal achievement to program? Unintended
outcomes? Cost effective?
- Major approaches:
 - Objectives based -- Goal free evaluation
 - Process-outcome studies -- Performance audit
- Assembly of evidence: Preordinate research designs
perhaps with control groups, tests and other
quantitative data; for all outcomes, more exploratory
and qualitative

Managing program evaluation

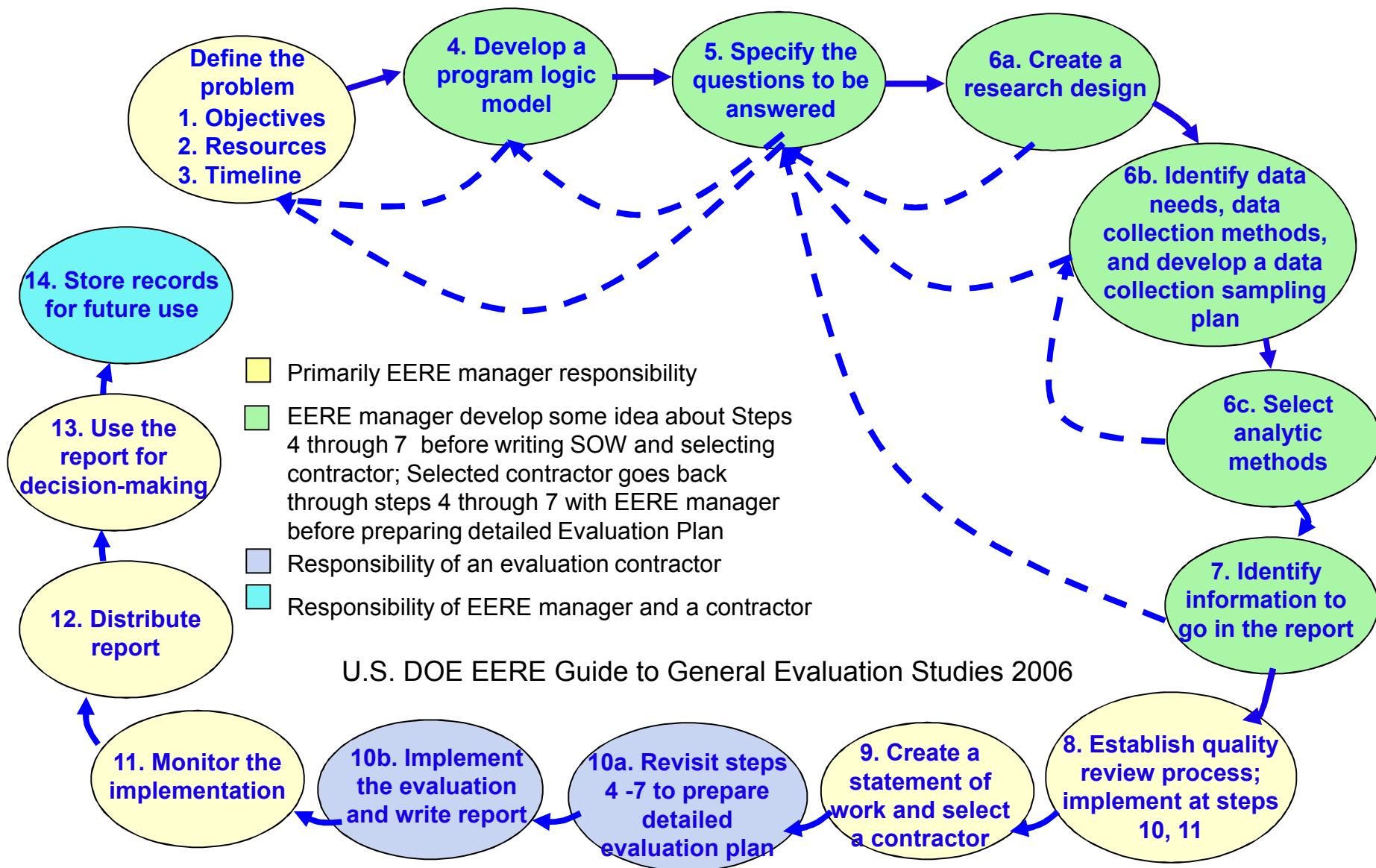
A Framework for the Program Evaluation Process



GAO-03-454, modified

Center for Disease Control and
Prevention (CDC)]

Overview of Steps for Managing an Evaluation



Fostering Organizational Commitment and Capacity for Data Quality

- Communicate support for quality data
- Review organizational capacities and procedures and use these
- Facilitate agency-wide coordination and cooperation
- Assign clear responsibilities
- Adopt mechanisms that encourage objectivity in collecting and managing data
- Provide responsible staff with training and guidance for needed skills and knowledge

Performance Plans: Selected Approaches for Verification and Validation of Agency Performance Information", GAO 1999 GGD-99-139

Uphold Evaluation Standards

- **Utility** - Serve the information needs of users
- **Feasibility** - Be practical, realistic, diplomatic, frugal
- **Propriety** - Behave legally, ethically, with respect
- **Accuracy** - Reveal and convey accurate information

American Evaluation Association
<http://www.eval.org/Publications/GuidingPrinciples.asp>

Choose indicators carefully. Various levels of the organization each need a small set. Each indicator in the set will

- **Link to desired outcomes.** And at least one should link to goals in the organizational/reporting hierarchy.
- **Communicate well.** Simple to report and understand; help the public understand how the program is doing.
- **Have benefits greater than costs.** Be sure benefits of measuring it are greater than the costs
- **Drives performance the right way,** or perverse effects are offset by another indicator in the set.

A Balanced Set tells a brief, convincing performance story and drives performance the right way by measuring the strategies and by covering all aspects of the program logic and of stakeholder information needs.

Weigh Pros and Cons of Various Methods

Methods	Pro	Con
Bibliometric analysis	Quantitative; useful on aggregate basis to evaluate quality for some programs and fields	Measures only quantity; not useful across all programs & fields; comparisons across fields or countries difficult; can be artificially influenced
Economic rate of return	Quantitative; shows economic benefits of research	Measures only financial not social benefits (such as health-quality improvements); time separating research from economic benefit is often long; not useful across all programs and fields
Peer review	Well-understood method; provides evaluation of quality of research; already an existing part of most Federal-agency programs in evaluating the quality of research projects	Focuses primarily on research quality; other elements are secondary; evaluation usually of research projects, not programs; great variance across agencies; concerns regarding use of “old boy network”; results depend on involvement of high-quality people in process
Case studies	Provides understanding of effects of institutional, organizational, and technical factors influencing research process so process can be improved; illustrates all types of benefits of research process	Happenstance cases not comparable across programs; focus on cases that might involve many programs or fields making it difficult to assess Federal-program benefit
Retrospective analysis	Useful for identifying linkages between Federal programs and innovations over long intervals	Not useful as a short-term evaluation tool because of long interval between research and practical outcomes
Benchmarking	Provides a tool for comparison across programs and countries	Focused on fields, not Federal research programs

Source: COSEPUP, on GPRA and Basic Research

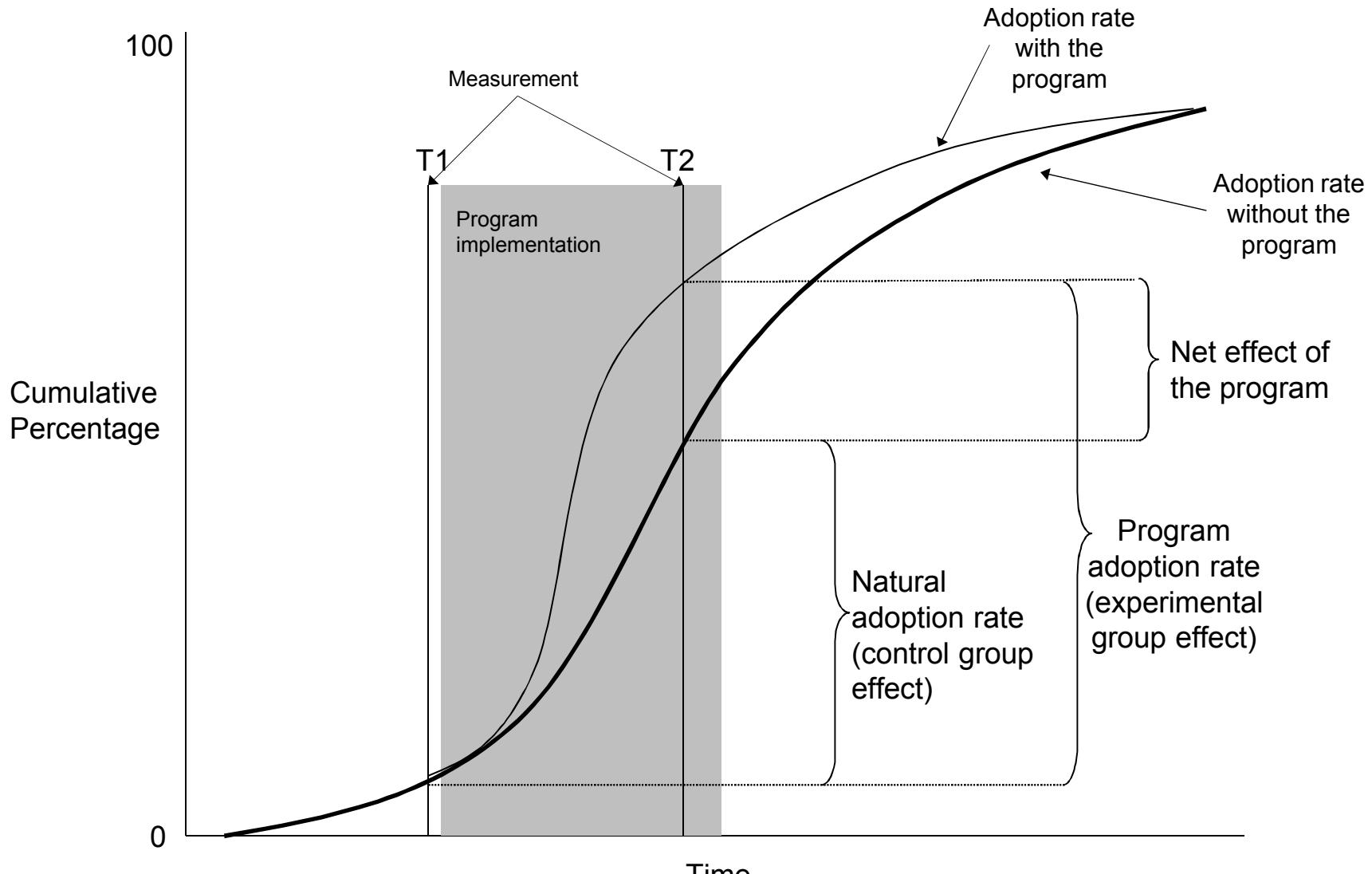
U.S. DOE EERE PEER REVIEW GUIDE (2004)

Sets Minimum Requirements

- **Frequency of Review** -- All EERE programs and projects will be reviewed by qualified and objective peers at least every 2 years, or more frequently if needed. This includes all TD research, development, demonstration, and deployment programs, and all BA analysis and business management programs.
- **Timely Preparation** – Preparation done well in advance
- **Core Evaluation Criteria** -- Unambiguous evaluation criteria will be defined: technical quality, productivity, and accomplishments; relevance to EERE and programmatic goals; management; and impact.
- **Reviewers** -- There will be a minimum of 3 reviewers for each discrete program element, each competent and objective, selected by a credible process.
- **Plan for Collecting Reviewer Data** – plan ahead for how review inputs documented, analyzed, reported; sufficient time onsite for rigorous Q&A periods
- **Producing the Peer Review Report** --The summary report should reflect the full range of reviewer comments with high fidelity. The report should also include all individual inputs from the reviewers.
- **Program Manager /Office Director Review** -- Before the report is finalized, the program manager/office director is to have the opportunity to provide written responses to peer reviewer findings.
- **Final Peer Review Report and Response** – final report with manager's response.
- **Peer Review Record and Complete Ex-post Evaluation** – establish peer review record & evaluate the peer review process

Flexibility is provided within these requirements

Experimental Designs with Control Groups Are Desirable for Some Impact Evaluations



Source: John H Reed, Innovologie Inc

Standard experimental design

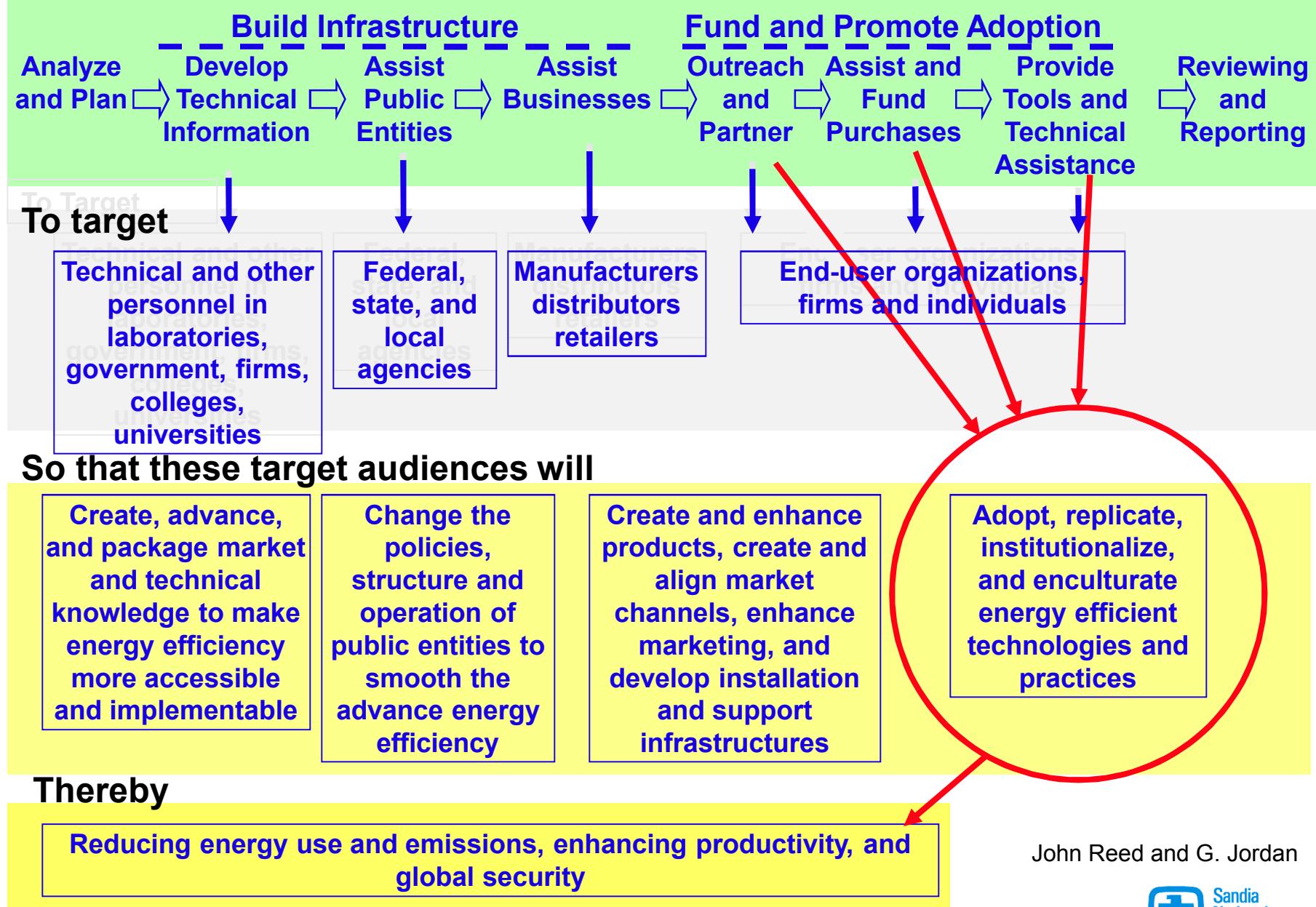
Continuously improve your indicators and your performance management system.

PROGRESS IN MEASURING PERFORMANCE						
Coverage	Drivers	Focus	MATURE	Structure	Ownership	Usage
Results, processes, & capabilities	Internal management processes	Strategic AND operational goals	MATURE	Integrated performance model	Led by senior management	Guide decisions, improve effectiveness
Budgets, variances	External demands	Internal activities	INITIAL	Ad hoc, no coherence	Primarily a staff activity	Control inputs & spending

Source: Bob Frost, Measurement International

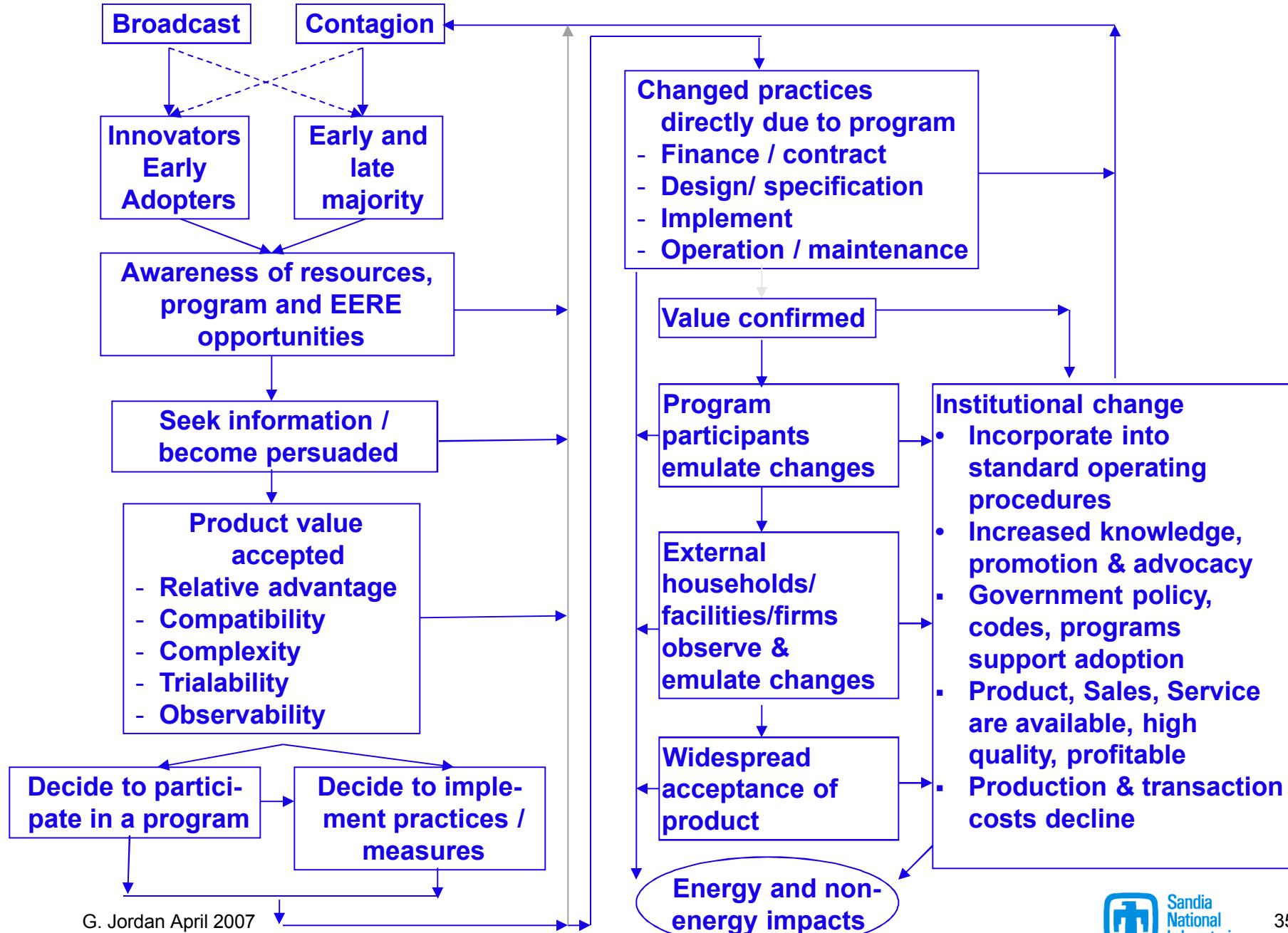
Key Questions - An example for evaluation of diffusion of a technology, idea or practice

The Office of Energy Efficiency and Renewable Energy of the US Department of Energy Undertakes Activities to



The Diffusion of Innovation and the Important Questions

John Reed and G. Jordan



The Diffusion of Innovation and the Important Questions

Is the word spreading through program channels and to whom is it being spread?

In response to program activities are target audiences aware of the program and the technologies and practices being promoted by the program?

In response to program activities, are the target audiences seeking more information and becoming persuaded?

Do the program's products, and the technologies and practices the program is trying to sell, have the "right stuff"?

Are the target audiences deciding to implement because of the program?

Are target audiences implementing their decisions to adopt and can adoption be traced to the program?

When target audiences adopt do they value what they have done?

Do target audiences repeat the behavior in the same or different ways?

Do others observe the changes to behavior and decide to investigate or try the behavior?

Do the changes to behavior become ingrained or do the target audiences revert to earlier behaviors?

Thank you for your attention.

Gretchen B Jordan
gbjorda@sandia.gov