Strengthening Federal Capacity to Conduct Evaluations to Inform Future R&D Program Planning

Daniel Basco

This document was submitted as a dissertation in May 2018 in partial fulfillment of the requirements of the doctoral degree in public policy analysis at the Pardee RAND Graduate School. The faculty committee that supervised and approved the dissertation consisted of Dave Baiocchi (Chair), Matthew Lewis, and Kei Koizumi.
Abstract

This dissertation identifies strategies for strengthening the Federal Government’s capacity to evaluate its research and development (R&D) programs using insights identified from philanthropic foundations. While there is broad recognition that Federal investments in R&D have resulted in breakthrough technologies and scientific achievements, there is little evidence to support whether the Federal Government’s vast investments are being made wisely. Few systematic evaluations have been conducted to assess whether the structure of these investments and the requirements put on federally funded scientists bolster or discourage impactful work. Federal science agencies and departments must develop better evaluation capacity to conduct the evaluations necessary to inform the design and structure of R&D programs to spur greater, faster, and cheaper innovation.

Strategies for strengthening Federal R&D evaluation capacity were identified through a two-stage approach. In Stage 1, a range of R&D evaluation approaches were identified through interviews with science program officers, evaluation personnel, and leaders within philanthropic foundations that fund R&D in the United States. In Stage 2, a diverse group of Federal science policy, planning, and evaluation personnel participated in discussions and activities to identify which foundation approaches could benefit the Federal Government and how they would need to be adapted to fit a Federal context. This process resulted in a series of strategies to improve Federal science agencies’ capacities to conduct R&D evaluations that can inform future program planning and strategy.

This dissertation finds that the Federal Government can strengthen its R&D evaluation capacity using four strategies adapted from philanthropic foundations. The Federal Government could hire additional evaluation-trained personnel within science agencies to provide the expertise needed to improve existing evaluations and develop new evaluation activities that can inform future program planning. Congress and the White House could explore how to centralize oversight and support for R&D evaluations to provide consistent advice to science agencies on how to strengthen evaluation capacity. The Federal Government could also invest in additional evaluation activities to provide the resources needed to conduct large-scale evaluations that are designed to inform future program planning. Finally, the White House and science agencies could improve the design of and resources for R&D evaluations to improve their effectiveness.
Executive Summary

This dissertation identifies strategies for strengthening the Federal Government’s capacity to evaluate its research and development (R&D) programs using insights identified from philanthropic foundations. While there is broad recognition that Federal investments in R&D have resulted in breakthrough technologies and scientific achievements, there is little evidence to support whether the Federal Government’s vast investments are being made wisely. Few systematic evaluations have been conducted to assess whether the structure of these investments and the requirements put on federally funded scientists bolster or discourage impactful work. If the Federal Government could better assess the impacts of its previous investments, it could make more informed decisions on how to structure future programs to increase the likelihood that R&D positively impacts academia, government, the economy, and society. Federal science agencies and departments must develop better evaluation capacity to conduct the evaluations necessary to inform the design and structure of R&D programs to spur greater, faster, and cheaper innovation.

Strategies for strengthening Federal R&D evaluation capacity were identified through a two-stage approach. In Stage 1, a range of R&D evaluation approaches were identified through interviews with science program officers, evaluation personnel, and leaders within philanthropic foundations that fund R&D in the United States. In Stage 2, a diverse group of Federal science policy, planning, and evaluation personnel participated in discussions and activities to identify which foundation approaches could benefit the Federal Government and how they would need to be adapted to fit a Federal context. This process resulted in a series of strategies to improve Federal science agencies’ capacities to conduct R&D evaluations that can inform future program planning and strategy.

This dissertation finds that the Federal Government can strengthen its R&D evaluation capacity using four strategies adapted from philanthropic foundations:

1. **Hire Evaluation-Trained Personnel**— Above all other components and characteristics, the foundations that were interviewed have utilized personnel to increase their capacity to conduct R&D evaluations that inform strategic decisions. Similarly, the Federal Government could hire additional evaluation-trained personnel within science agencies to provide the expertise needed to improve existing evaluations and develop new evaluation activities that can inform future program planning.
2. **Explore Centralizing Oversight of and Centralize Support for R&D Evaluation**—Foundation presidents have been the activating energy behind increased efforts to strengthen the capacity for R&D evaluation within foundations. Several foundations that recently created evaluation units credited new foundation presidents with initiating and supporting these efforts. Congress and the White House could explore how to centralize oversight and support for R&D evaluations to provide consistent advice to science agencies on how to strengthen evaluation capacity.

3. **Invest in Evaluation Activities**—Several foundations that were interviewed indicated that they spend between 1% and 10% of their total operating budgets on evaluation activities each year. Federal spending on R&D evaluations is not possible to track using existing data but a survey of Federal science leaders conducted as part of this dissertation indicates low estimates of spending on evaluation activities. The Federal Government could invest in additional evaluation activities to provide the resources needed to conduct large-scale evaluations that are designed to inform future program planning.

4. **Improve Evaluation Designs and Resources**—Foundations are also developing new evaluations designs to better inform strategic decisions at the portfolio level such as whether to increase or decrease funding for a program, to identify how to leverage one program to support another, or explore how to change individual program requirements to improve the likelihood of program success. The White House and science agencies could improve the design of and resources for R&D evaluations to improve their effectiveness.

The Federal Government can increase its capacity to conduct evaluations that inform future programming by adapting foundations’ approaches of investing in governance structures, personnel, funds for evaluation, and evaluation design advancements. The recent Foundations for Evidence-Based Policymaking Act indicates that Congress is interested in encouraging reform. However, the bill does not currently address the unique needs of science agencies. The Federal Government should capitalize on current interests in developing evidence-based policy to enact a series of science policy focused reforms to strengthen its capacity to conduct evaluations that can inform the structure and design of future R&D programs.
# Table of Contents

Abstract ........................................................................................................................................................................................................................................... iii

Executive Summary ...................................................................................................................................................................................................................... iv

Figures ........................................................................................................................................................................................................................................ xii

Tables .................................................................................................................................................................................................................................... xiii

Acknowledgments ............................................................................................................................................................................................................... xv

Abbreviations .................................................................................................................................................................................................................. xvii

Introduction ...................................................................................................................................................................................................................... 1

Motivations .................................................................................................................................................................................................................. 2

*Limited Federal Efforts to Evaluate R&D to Inform Strategy* ........................................................................................................................................................................... 2

*Difficulty in Evaluating R&D Outcomes and Impacts* .................................................................................................................................................................................. 3

*Potential Learning from Foundations and Firms* ................................................................................................................................................................................ 4

Research Questions ........................................................................................................................................................................................................... 4

Theoretical Framework ........................................................................................................................................................................................................ 5

Approach and Methods ..................................................................................................................................................................................................... 7

*Stage 1* ............................................................................................................................................................................................................. 8

*Stage 2* ........................................................................................................................................................................................................... 14

Overview of Dissertation ........................................................................................................................................................................................................ 18

Chapter 1: Literature Review .................................................................................................................................................................................................. 19

Overview of Federal R&D Enterprise .............................................................................................................................................................. 20

Federal R&D Evaluation Capacity and Approaches ...................................................................................................................................................................... 22

*Federal-Wide Efforts to Improve Government Programs* .................................................................................................................................................................. 22

*Efforts to Improve R&D Evaluation Capacity and Practices* ................................................................................................................................................................ 24

*Current Approach to Federal R&D Evaluation* ................................................................................................................................................................. 26

Philanthropic Foundation R&D Landscape and Evaluation Practices ......................................................................................................................................................... 27

Private Sector R&D Landscape and Evaluation Practices ......................................................................................................................................................... 29
Gaps in Existing Literature ............................................................................................................. 32

Chapter 2: Foundation and Firm Findings ..................................................................................... 34

Insights on R&D Evaluation from Foundations ............................................................................. 36
  Use of Evaluation Personnel ........................................................................................................ 36
  Leadership Influence on Evaluation Capacity .............................................................................. 41
  Variety of Evaluation Activities Serve Several Purposes .............................................................. 43
  Growing Emphasis on Learning .................................................................................................. 51
  Funding for Evaluations ............................................................................................................ 52

Limitations of the Socio-Technical Framework for Describing Foundations .................................. 53
  Lack of Data Collection Tools .................................................................................................. 54
  Limited Policy Underpinnings .................................................................................................. 56

Limited Insights on R&D Evaluation from Firms ........................................................................... 57

Chapter 3: Workshop Findings .................................................................................................. 59

Potential Evaluation Research Questions ....................................................................................... 61
Cluster Evaluation Findings .......................................................................................................... 64
Financial Resource Findings ......................................................................................................... 66
Personnel Findings ........................................................................................................................ 67
  Centralized Evaluation Service .................................................................................................. 67
  Internal Evaluation “Consultants” ............................................................................................... 68
  Program Officers as Evaluators .................................................................................................. 69
  Hybrid Approach Combining Centralized Evaluation Service and Internal Evaluation Consultants.. 69
Culture for Evaluation .................................................................................................................. 70
Governance and Leadership ........................................................................................................... 71
Connections to the Commission on Evidence-Based Policymaking ............................................. 72

Chapter 4: Policy Implications and Conclusions ......................................................................... 74

Hire Evaluation-Trained Personnel ............................................................................................... 77
  1. Develop Centralized Evaluation Services with Evaluation-Trained Staff .................................. 81
  2. Develop Evaluation Capacity at Lower Levels .......................................................................... 84
  3. Increase the Number of FTEs for Evaluation Experts .............................................................. 85
  4. Designate Learning Officers .................................................................................................... 85

Explore Centralizing Oversight of and Centralize Support for R&D Evaluation ............................. 86
  1. Explore How to Concentrate Oversight to Standardize and Maximize Advice to Science Agencies.. 89
  2. Develop a Dedicated NSTC Subcommittee to Support R&D Evaluation Activities .................. 90
Invest in Evaluation Activities ...................................................................................................... 91
Figures

Figure 1. Socio-Technical System Design ................................................................. 6
Figure 2. Adaptation of Socio-Technical Theory Framework .............................. 7
Figure 3. Overall Approach to Dissertation ......................................................... 8
Figure 4. Foundation Selection Process ............................................................... 10
Figure 5. Variations in R&D Program Design ....................................................... 21
Figure 6. Best Practices for Industry R&D ............................................................ 31
Figure 7. Adaptation of Socio-Technical Theory Framework ......................... 35
Figure 8. Adapted Theoretical Framework ........................................................... 54
Figure 9. Revised Theoretical Model ................................................................. 56
Figure 10. Analytical Process for Developing Recommendations ..................... 75
Tables

Table 1. Comparing Personnel Models ................................................................. 41
Table 2. Intersection Between Purposes and Types of Evaluation ....................... 46
I am indebted to numerous individuals and organizations who helped me throughout the dissertation process.

First and foremost, thank you to the numerous anonymous individuals within philanthropic foundations, private firms, and the Federal Government who generously provided their time and insights. This dissertation is built on a rich set of qualitative data that was collected from numerous experts who share an interest in ensuring that R&D positively impacts academia, government, the economy, and society.

Thank you to my committee, Dave Baiocchi, Matt Lewis, and Kei Koizumi for their invaluable help throughout the dissertation process. Their advice, nudges, and healthy skepticism pushed me to be a better researcher and improved the quality of this dissertation.

Thank you to all the Pardee RAND Graduate School students, faculty, and administrators who encouraged and guided me throughout the program. In particular, thank you to Jesse Lastunen and Christina Steiner for taking notes and editing my dissertation.

Thank you for the generous funding provided by RAND’s National Defense Research Institute, the Anne and James Rothenberg Dissertation Award, and RAND’s Science, Technology and Policy Program. This work would not have been possible without their support.

Thank you to my parents, who have inspired and supported me from the beginning. Their interminable perseverance and their devotion to helping others led me to this field and encouraged me to add another Dr. Basco to the family.

Finally, thank you to Jon for being my anchor, my cheerleader, and my proofreader. We made it to the finish line and I can’t wait for the adventures that will follow!
## Abbreviations

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARPA-E</td>
<td>Advanced Research Projects Agency-Energy</td>
</tr>
<tr>
<td>BRAIN</td>
<td>Brain Research through Advancing Innovative Neurotechnologies Initiative</td>
</tr>
<tr>
<td>CDC</td>
<td>Centers for Disease Control and Prevention</td>
</tr>
<tr>
<td>CEP</td>
<td>Commission on Evidence-Based Policymaking</td>
</tr>
<tr>
<td>DHS</td>
<td>Department of Homeland Security</td>
</tr>
<tr>
<td>DHS S&amp;T</td>
<td>Department of Homeland Security Science and Technology</td>
</tr>
<tr>
<td>DOD</td>
<td>Department of Defense</td>
</tr>
<tr>
<td>DOE</td>
<td>Department of Energy</td>
</tr>
<tr>
<td>EAC</td>
<td>NSF’s Evaluation and Assessment Capability office</td>
</tr>
<tr>
<td>EPSCoR</td>
<td>NSF’s Experimental Program to Stimulate Competitive Research</td>
</tr>
<tr>
<td>FDO</td>
<td>Foundation Directory Online</td>
</tr>
<tr>
<td>FTE</td>
<td>Full-Time Equivalent</td>
</tr>
<tr>
<td>GAO</td>
<td>Government Accountability Office</td>
</tr>
<tr>
<td>GPRA</td>
<td>Government Performance and Results Act</td>
</tr>
<tr>
<td>GPRA Mod</td>
<td>GPRA Modernization Act of 2010</td>
</tr>
<tr>
<td>GPRAMA</td>
<td>GPRA Modernization Act of 2010</td>
</tr>
<tr>
<td>GS</td>
<td>General Service or General Schedule</td>
</tr>
<tr>
<td>HHS</td>
<td>Department of Health and Human Services</td>
</tr>
<tr>
<td>IP</td>
<td>Intellectual Property</td>
</tr>
<tr>
<td>IRI</td>
<td>Industrial Research Institute, now Innovation Research Interchange</td>
</tr>
<tr>
<td>MCC</td>
<td>Millennium Challenge Corporation</td>
</tr>
<tr>
<td>MIS</td>
<td>Management Information System</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>MOU</td>
<td>Memorandum of Understanding</td>
</tr>
<tr>
<td>NASA</td>
<td>National Aeronautics and Space Administration</td>
</tr>
<tr>
<td>NDPA</td>
<td>NIH Director’s Pioneer Award program</td>
</tr>
<tr>
<td>NIH</td>
<td>National Institutes of Health</td>
</tr>
<tr>
<td>NIST</td>
<td>National Institute of Standards and Technology</td>
</tr>
<tr>
<td>NOAA</td>
<td>National Oceanic and Atmospheric Administration</td>
</tr>
<tr>
<td>NRL</td>
<td>Naval Research Laboratory</td>
</tr>
<tr>
<td>NSF</td>
<td>National Science Foundation</td>
</tr>
<tr>
<td>NSTC</td>
<td>National Science and Technology Council</td>
</tr>
<tr>
<td>OECD</td>
<td>Organization for Economic Cooperation and Development</td>
</tr>
<tr>
<td>OMB</td>
<td>Office of Management and Budget</td>
</tr>
<tr>
<td>ONR</td>
<td>Office of Naval Research</td>
</tr>
<tr>
<td>OSTP</td>
<td>Office of Science and Technology Policy</td>
</tr>
<tr>
<td>PMS</td>
<td>Performance Management System</td>
</tr>
<tr>
<td>PO</td>
<td>Program Officer</td>
</tr>
<tr>
<td>R&amp;D</td>
<td>Research and Development</td>
</tr>
<tr>
<td>RFP</td>
<td>Request for Proposal</td>
</tr>
<tr>
<td>RPPR</td>
<td>Research Performance Progress Report</td>
</tr>
<tr>
<td>RWJF</td>
<td>Robert Wood Johnson Foundation</td>
</tr>
<tr>
<td>SBIR</td>
<td>Small Business Innovation Research program</td>
</tr>
<tr>
<td>SciSIP</td>
<td>NSF’s Science of Science and Innovation Policy program</td>
</tr>
<tr>
<td>STEM</td>
<td>Science, Technology, Engineering, and Mathematics</td>
</tr>
<tr>
<td>STS</td>
<td>Socio-Technical System</td>
</tr>
<tr>
<td>USAID</td>
<td>United States Agency for International Development</td>
</tr>
<tr>
<td>USGS</td>
<td>U.S. National Geological Survey</td>
</tr>
<tr>
<td>VA</td>
<td>Department of Veterans Affairs</td>
</tr>
</tbody>
</table>
Introduction

While there is broad recognition that Federal investments in research and development (R&D) have resulted in breakthrough technologies and scientific achievements (Sargent et al. 2016), there is little evidence to support whether the Federal Government’s vast investments are being made wisely (Fealing et al. 2011). Few systematic evaluations have been conducted to assess whether the structure of these investments and the requirements put on federally funded scientists bolster or discourage impactful work (Heinze 2008, Stephan 2015). The Federal Government cannot answer simple portfolio analysis questions such as: Can we change R&D program requirements or funding structures to encourage faster, cheaper, or greater innovation? How will changing from a center funding model to an individual investigator funding model alter the outcomes and impacts from R&D? In order to answer these questions, Federal science agencies and departments must design, implement, and utilize complex evaluations that examine the outcomes and impacts of R&D to inform future program planning and strategy. This dissertation aims to identify approaches from philanthropic foundations that can be adapted to strengthen the Federal Government’s capacity to conduct these types of evaluations.

In order to identify methods for strengthening R&D evaluation capacity, I took a two-stage approach. In Stage 1, I identified a range of approaches and practices for conducting R&D evaluations within philanthropic foundations. Rather than identify best practices, this exploration aimed to identify a myriad of ways that foundations use personnel, organizational structures, data collection tools, evaluation designs, and foundation policies to strengthen their ability to conduct evaluations that inform strategy. In Stage 2, I brought together senior Federal science policy leaders to explore how, and if, these approaches and practices could be adopted by the Federal Government. Through these two stages, I aimed to identify methods that the Federal Government could undertake to build the capacity to collect and analyze the data necessary to inform policymakers at multiple levels on ways to alter R&D programs to encourage faster, cheaper, and greater innovation.

The following introduction describes the motivations for this dissertation, the research questions it aims to answer, the theoretical framework that grounds the approach, and the approach and methods used to answer the research questions.

---

1 Surveys indicate that, over the last several years, philanthropic foundations have increased the number of personnel and amount of funding devoted to evaluation (Center for Effective Philanthropy and Center for Evaluation Innovation 2016). This increase in evaluation focus could signal experimentation and new approaches to evaluation that could provide valuable lessons for the Federal Government. Evaluation practices within private firms were also investigated but ultimately dropped due to lack of focus on evaluations and limited sample size.
Motivations

This dissertation is motivated by three inter-related factors: limited Federal efforts to evaluate its R&D to inform strategy development; the difficulties in systematically evaluating the outcomes and impacts of R&D; and the potential to learn from philanthropic foundations and private firms\(^2\) efforts to evaluate their own R&D investments. This section briefly describes these motivations to frame the research questions for this dissertation. A full description of current evaluation approaches by the Federal Government, foundations, and firms is provided in Chapter 1.

Limited Federal Efforts to Evaluate R&D to Inform Strategy

In 2005, President Bush’s science advisor, John Marburger III, proclaimed in a ground-breaking speech to the science policy community, “I worry constantly that our tools for making wise decisions… are not yet sharp enough to manage the complexity of our evolving relationship with the awakening globe” (Marburger 2005). Marburger saw that expert opinion drove science policy decisions and a lack of data inhibited robust, scientific inquiry on how to improve science investments (Fealing et al. 2011). By 2008, Heinze notes that there was common agreement that “how research money is packaged, earmarked, distributed and finally spent within and across research institutions influences the working conditions under which research is carried out in laboratories”. So, funding requirements and structures shape how scientific research is conducted. Yet, as of 2017 when this dissertation was developed, few studies have been conducted to understand how R&D program variations influence innovation.

In part, few studies have been conducted due to a lack of evaluation capacity within the Federal Government. A survey I conducted with eighteen senior science policy leaders as part of this dissertation found that few funds and personnel are dedicated to R&D evaluation by the Federal Government. While spending on R&D evaluation activities is not tracked within the Federal Government, survey respondents estimate that between $1 and $30 million is spent per agency per year. Comparatively, the Federal Government spent approximately $148 billion in fiscal year 2016 alone on R&D (Hourihan and Parkes 2016b). Therefore, less than 0.1% of R&D funds were spent assessing the outcomes of these investments and how they can be improved. Similarly, few full-time staff focus exclusively on developing and overseeing R&D evaluations. Most survey estimates indicate that fewer than five full-time equivalents (FTEs) work on evaluations per agency. A few respondents even indicated that their agencies have no FTEs dedicated to evaluation. This lack of capacity inhibits agencies from designing and implementing the complex evaluations needed to inform strategy.

\(^2\) For this dissertation, private firms were U.S.-based, for-profit businesses that conduct R&D.
In addition, few agencies have created mechanisms to use existing evaluations for broader learning and future program planning. Survey respondents indicated that R&D evaluations that are conducted are often program-specific and are not designed or used to inform broader agency strategy. In addition, there are limited mechanisms for sharing the results of R&D evaluations with other programs so they can learn from prior efforts. This dissertation aims to identify approaches for strengthening the capacity for the Federal Government to conduct and use evaluations for program planning purposes.

**Difficulty in Evaluating R&D Outcomes and Impacts**

R&D has been notoriously difficult to systematically evaluate to determine the outcomes and impacts that result from funded projects (Fealing et al. 2011). In the workshop that is described later in this dissertation, senior science policy, planning, and evaluation leaders were asked to briefly share evaluation challenges they have experienced when trying to assess their own portfolios. Participants identified several issues that are also validated in the literature, including:

- **Identifying and Capturing Data**
  - Identifying the correct output measures, particularly for basic research
  - Capturing innovation-related data
  - Digitizing old, physical documents to create searchable, quantitative data for evaluations
  - Measuring improvements that resulted from R&D investments
  - Time-lags to measurable outcomes and impacts
  - The non-linear path of research and innovation
  - Attribution concerns when agencies are not the only funders in a scientific area

- **Leadership and Culture**
  - Agency culture that drives evaluations to be used primarily for accountability purposes
  - Shifting agency mindsets to focus on forward-looking evaluations rather than focusing on accountability only
  - Leadership indifferences to the value and results of evaluations
  - Multi-level approvals needed to conduct evaluations

- **Using Evaluation Results**
  - Balancing the mix of projects in a portfolio based on evaluation results
  - Misusing evaluation metrics to prescribe future grant requirements that stifle innovation

These challenges complicate science agencies’ abilities to conduct systematic evaluations of outcomes and impacts that can inform future program planning and strategy. However, there are non-Federal actors who also fund R&D and have dedicated significant resources to developing evaluation approaches to overcome some of these issues. Their related work could provide valuable insights for the Federal Government. This dissertation aims to identify approaches from these organizations and explore how they could be adapted to help the Federal Government overcome some of these issues.
Potential Learning from Foundations and Firms

Philanthropic foundations fund similar types of research as the Federal Government and similarly began to face external pressure in the early 1990’s to better evaluate the impacts of their efforts, although from different sources (Behrens and Kelly 2008, Micheli and Kennerley 2005). In the Federal Government, these pressures led to legislation that aimed to standardize performance measurement across agencies. Foundations did not face a similar call for standardization and therefore foundations began experimenting with a variety of approaches to better evaluate the results of their investments (Behrens and Kelly 2008). So, foundations and government faced similar pressures in the early 1990’s to document their impacts but developed different approaches to evaluation. The diversity of foundation approaches provides a potential opportunity for government to learn from foundations’ experimentation.

In addition, firms may offer potential lessons for the Federal Government. Common assumptions about the differences between private and public organizations may lead researchers to question the value of examining private sector practices to inform public sector planning. However, decades of research highlight that there are many commonalities between private firms and government organizations (Rainey and Bozeman 2000). With regards to R&D in particular, there are a variety of overlaps in processes and functions. In addition, the breadth of private firms that have experimented with different ways to design and implement evaluation systems since the 1950’s provides a rich set of experiences to learn from.

Firms and foundations offer potential lessons for the Federal Government if their approaches can be documented and if these approaches can be translated by Federal stakeholders. Insights from these organizations would offer a new way to examine a pervasive issue.

Research Questions

This dissertation aims to identify approaches to R&D evaluation from philanthropic foundations and explore how they can be applied to the Federal Government to strengthen their capacity to conduct evaluations that inform future program planning. The following primary and secondary research questions will be investigated in achieving this objective:

Primary Research Question:
1. What characteristics and components from philanthropic foundations’ approaches to R&D evaluation can be adapted to strengthen the capacity of Federal Government to conduct evaluations that inform future R&D program planning?

Secondary Research Questions:
1. How do philanthropic foundations plan and develop their R&D programs?
2. How do philanthropic foundations evaluate the outcomes and impacts of their R&D portfolios?
3. How do foundations use evaluations of their past R&D projects to inform future planning of R&D investments?

Additional research questions were originally proposed to identify approaches from private sector firms for evaluating the outcomes and impacts of R&D investments. However, these research questions were ultimately dropped due to limited use of retrospective evaluations to examine the outcomes and impacts of their R&D among the firms interviewed and significant difficulties in recruiting firms to participate in interviews. Literature review findings on firms are presented in Chapter 1. Limited findings from firms that were interviewed are presented in Chapter 2. Methodological challenges in recruitment are described in Chapter 4 to note potential areas for further research and possible ways to avoid similar challenges. However, the majority of the following dissertation focuses on foundation efforts and how they could apply to the Federal Government.

**Theoretical Framework**

In order to capture comprehensive, systematic information on R&D evaluation approaches within foundations, it is important to identify a theoretical framework that outlines the various components and characteristics that define evaluation capacity. Initially, I explored the literature on foundations and firms to try to identify a framework or theory describing R&D evaluation approaches. Unfortunately, there does not appear to be specific models for R&D evaluation. Therefore, I explored related fields to identify an appropriate model to help ground my work. Socio-technical system design from information science most closely aligned to the problem I am addressing and has a rich history of use.

Socio-technical system (STS) design was first developed in the 1970’s to better plan the design and re-design of management information systems. At its core, STS assumes that any work process consists of an independent, but highly interactive social system and technical system. The social system consists of the personnel and the incentives for these personnel, and the authority structure of the organization. The technical system consists of the technology being used, the processes of the technology, and the tasks carried out with the technology (Bostrom and Heinen 1977). Figure 1 below represents the first conceptualization of this framework in an article from 1977 in Management Information Systems (MIS) Quarterly (Bostrom and Heinen 1977). In this diagram, the arrows are important to signify that when designing a technical system, the relationships between all parts of the social and technical systems must be considered. In adapting this theoretical model, I am making an assumption that foundation evaluation capacity can be broken down into a social system of people and where they are placed as well as a technical system involving data collection tools and evaluation activities.
While articles on STS design discuss how structure, personnel, technology, and tasks interact with one another, I did not identify an article that formally labels these interactions on the original diagram. In addition, the role of policy is inferred but never explicitly represented in STS diagrams. Therefore, I adapted the original STS design for use in my dissertation. Figure 2 below represents a modified version of the original STS design that maintains the original principles and adds clarifications for my purposes.

The core labels in the STS design were modified to fit the technology that I will be focusing on and the primary task that will be carried out in an R&D evaluation approach once the data is collected. In addition, policy clearly undergirds all of the decisions that are made on personnel, structure, data collection tools, and evaluation activities. The diagram now indicates this relationship. Finally, I labeled all of the interactions between the central components of socio-technical systems.
These arrow labels clarify the specific components and characteristics of foundation R&D evaluation approaches that I will investigate. I am interested in understanding how leaders are able to implement new data collection tools; who they provide the authority to govern the way the tools are designed and operated; and how they use results of evaluation activities that are conducted with the outcomes and impact data that is collected. I am also interested in understanding how personnel are incentivized to use the system and how the analysis of the data feeds back to these personnel to inform future program planning. In the technical system, I will also investigate the processes that are used to move from raw impact data to robust, informative evaluations. Each of these components and characteristics of R&D evaluation approaches will be investigated in the firm interviews to identify a broad range of options.

**Approach and Methods**

To answer the proposed research questions, I developed a two-stage approach, which is described visually in Figure 3 below. In Stage 1, I identified a variety of foundation approaches to R&D evaluation through a literature review, initial interviews with a broad set of organizations, and in-depth interviews with select foundations. Stage 1 resulted in a variety of components and characteristics for evaluating R&D. In Stage 2, Federal science policy officials engaged in a collaborative design workshop to explore how to adopt and refine foundation options to benefit the Federal Government. Stage 2 resulted in a set of strategies that Congress,
the White House, and science agencies could use to strengthen the Federal Government’s capacity for conducting evaluations that inform future program planning. The following sections outline the goals, data sources, and methods for each data collection activity in Stages 1 and 2.

**Figure 3. Overall Approach to Dissertation**

![Diagram showing the overall approach to dissertation]

**Stage 1**

Stage 1 focused on identifying foundation approaches to R&D evaluation for use in strategy through three data collection activities. The following describes the literature review, exploratory interviews, and in-depth interviews that identified a myriad of approaches from foundations. In addition, the literature review and exploratory interviews explored whether insights from private firms could be captured to describe approaches for conducting R&D evaluations that inform future program planning and strategy.

**Literature Review**

Prior to designing and executing interviews with firms³ and foundations, I conducted an extensive literature review to understand what has been done in this area broadly over the last 60 years. The intent was to identify the types of issues in R&D evaluation that have been focused on, explore the evolution of R&D evaluation across firms and foundations, and identify key experts to potentially interview and include in the design workshop. The results of this review are captured in Chapter 1.

Academic journals, books, and reports from organizations that support firm and foundation R&D efforts were explored. Academic journals in economics, engineering, operations research, innovation, and management were reviewed using Google Scholar, Web of Science, and

---

³ Note: The research questions on firms were dropped after the exploratory interviews. Therefore, the literature review and the exploratory interviews focused on both firm and foundation evaluation approaches.
individual journal websites. Websites from organizations such as the Foundation Center and Industrial Research Institute also provided reports and publications that were reviewed to identify firm and foundation R&D planning and evaluation practices.

The search terms R&D, evaluation, program planning, impact assessment, etc. are too generic to allow for systematic exploration of the literature. Therefore, a snowball sampling approach⁴ was taken utilizing key literature identified from prior studies. Articles that were cited in these key studies were examined as well as articles that cited these studies using Google Scholar. Specific journals that focus on R&D evaluation in the private and philanthropic sector were also examined for relevant literature including Research-Technology Management. Finally, targeted literature searches were conducted to identify specific information on the characteristics and processes developed to implement evaluation systems within private and philanthropic R&D organizations. The search was not limited to a specific time period, allowing for review of assessments and new evaluation methods back to the early 1950’s.

Exploratory Interviews

Initially, exploratory interviews were designed and conducted with philanthropic foundations, private firms, and consultants who support these organizations. The interviews aimed to identify a range of foundations and firms that have robust evaluation approaches that are used to inform future program planning. In addition, these semi-structured interviews captured broad descriptions of how firms and foundations plan, monitor, and evaluate their R&D investments.

Interview Sample

Exploratory interviews were conducted with several foundations based on a multi-step selection process that is outlined in Figure 4 below. Given the limited number of philanthropic foundations that fund R&D, I initially aimed to develop a census of these organizations using the Foundation Directory Online (FDO). This database is maintained by the Foundation Center, a resource organization that tracks and supports philanthropies domestically and internationally.

---

⁴ Snowball sampling approaches identify relevant individuals or articles for research by first identifying a few key individuals or articles that are well known in the given domain and then using these resources to identify additional individuals or articles. See Lewis-Beck, Bryman, and Futing Liao (2004) for further information.
Using this database, I identified 41 foundations that appeared to fund U.S. science in some form. After manually reviewing these foundations’ websites and their information on FDO, 21 candidate foundations met the inclusion criteria of being based in the United States (U.S.), currently operating, and funding R&D specifically. Finally, contact information for all candidate foundations was sought using foundation websites, the Inside Philanthropy news website, and previous identified academic publications. Emails were sent to 17 foundations whose staff email addresses could be identified. Several foundations that did not respond to interview requests operate with only one or two staff members and likely do not engage in evaluation activities. Ultimately, staff from 10 foundations responded and were interviewed. These initial, exploratory interviews were conducted with evaluation staff where possible and conducted with science program officers when the foundation did not have any formal evaluation staff.

A sample of firms to interview were initially identified through the Global Innovation 1000 study by Strategy&. I selected the 15 U.S. firms that had the top R&D expenditures in 2016 to contact. This selection method was chosen with the rationale that the firms that spend the most on R&D are most likely to invest in systematic evaluation activities. These firms represented numerous industries, including retail, software, pharmaceuticals, and automobiles, which conduct R&D in different ways. Therefore, this sample could have potentially provided a variety of perspectives on R&D evaluation. Unfortunately, contact information for R&D leaders within these firms proved impossible to ascertain. While the proprietary database Company Dossier by Lexis Nexis listed email addresses for R&D executives within these firms, the overwhelming majority of emails sent to these addresses bounced back without delivery, indicating erroneous addresses. A deeper discussion of methodological issues with recruiting firms and potential mitigation strategies for future research is provided in Chapter 4.

5 Most of the 20 foundations that did not meet the inclusion criteria funded efforts related to science but did not explicitly fund R&D, making them less relevant to the Federal Government.

After failing to contact firms with top R&D expenditures, I reached out to the Industrial Research Institute (IRI) for assistance in developing a convenience sample of U.S.-based firms that conduct R&D. The IRI is a membership-based research collaborative of R&D firms that has conducted internal research on R&D best practices since the 1970’s. With the assistance of IRI, I conducted a small focus group\(^7\) at their annual spring meeting and seven individual interviews with R&D managers representing firms in varied industries including defense, pharmaceuticals, food chemicals, and professional services.

Finally, four exploratory interviews were conducted with consultants who advise firms and foundations on R&D best practices. Foundation consultants were identified through recommendations from foundation employees. Firm consultants were identified through the IRI.

**Methods for Semi-Structured Exploratory Interviews**

One-on-one semi-structured interviews were conducted with foundation, firm, and consultant staff over the telephone and recorded for notetaking purposes. I also took high-level notes during each of these interviews. Interviews lasted 30 to 60 minutes depending on the interviewees’ availabilities. The interview protocol was developed using the theoretical framework and the findings from the literature review. The foundation interview protocol is reproduced in Appendix 1. The firm interview protocol is reproduced in Appendix 2. These protocols focused on understanding the organization’s approach to planning, monitoring, and evaluating R&D investments. Standard human subjects protection practices were observed. To allow for open and honest discussions, participants were informed that their comments would not be attributed to them personally or to their organization.

Consultant interviews loosely followed the foundation and firm interview protocols to identify potential candidate organizations that I should further investigate. Standard human subjects protection practices were also observed. To allow for open and honest discussions, participants were informed that their comments would not be attributed to them personally or to their organization. The two foundation consultant interviews helped validate that I had identified all of the relevant foundations and provided contact information for several of these foundations. The two firm consultant interviews were conducted prior to the firm interviews and therefore aided in adapting and refining the firm interview protocols to use the correct terminology that firms would understand. Firm consultant interviews also provided further information on specific firms’ approaches, or lack thereof, to R&D evaluation.

\(^7\) This focus group utilized the firm interview protocol and effectively was a group interview.
Data Analysis

Interview notes were reviewed to identify organizations to further investigate. Prior to conducting the interviews, I had intended to use interview notes to identify the most diverse set of R&D approaches since this dissertation aimed to provide a myriad of methods for the Federal Government to increase its capacity for conducting R&D evaluations that inform strategy. However, the limited sample population and lack of evaluation activities within some foundations made this approach unnecessary. Of the 10 foundations that were initially interviewed, I selected seven as candidates for further interviews. The three that were not selected did not evaluate their R&D investments.

After reviewing the interview notes from the firm interviews, I made the decision to drop my research questions that focused on gathering insights from private firms. In addition to the difficulty in recruiting firms, none of the seven firms that I interviewed conducted formal evaluations of their R&D investments. Rather, they had built in evaluation-type assessments into continual reviews of whether to continue funding individual projects. Therefore, the firms I interviewed did not have R&D evaluation approaches that were designed to inform strategy and future program planning. Dropping these questions allowed me to focus on better capturing foundation approaches.

In-Depth Foundation Interviews

In-depth, multi-level interviews were conducted with five of the seven foundations selected from the exploratory interviews. The goal of the in-depth foundation interviews was to capture details on the components and characteristics of R&D evaluation approaches. Through interviewing multiple employees within each foundation, I was able to capture multiple perspectives to answer each of my secondary research questions. These interviews resulted in a menu of characteristics and components of foundations’ R&D evaluation approaches and a set of models for increasing R&D evaluation capacity.

Interview Sample

Within each foundation, I sought to interview science program officers, evaluation officers, and leadership. A few foundations in my sample did not have dedicated evaluation officers; for these foundations, I asked to interview additional science program officers to capture complete information on how they approach R&D evaluation. Interviewees were identified through foundation websites and through recommendations from initial interviewees. Ultimately, I interviewed 14 additional personnel across five foundations after the initial exploratory interviews. In each foundation, I interviewed at least two additional personnel beyond the initial

8 One foundation declined to further participate and one foundation did not respond to requests for further interviews.
staff member interviewed during the exploratory interviews. Additional interviews were conducted based on the number of evaluation personnel within the foundation and staff member willingness to participate. Willingness to participate in in-depth interviews varied widely across foundations with some initial interviewees noting that their lean personnel models left little opportunity to participate. The exploratory interviews found that foundations with lean personnel models did not have dedicated evaluation staff and had similar approaches to R&D evaluation. Leadership interviews were particularly difficult to conduct as they often referred me to evaluation personnel rather than agreeing to be interviewed. Rather than interviewing foundation presidents as originally planned, leadership interviews were conducted with senior science staff who oversaw several R&D programs.

Methods for In-Depth Interviews

Semi-structured interviews were conducted one-on-one via telephone and recorded for transcription purposes. Interviews lasted 30-60 minutes depending on the interviewees’ availability and their role within the foundation. The interview protocol was developed after conducting the exploratory interviews to ensure that the right terminology was used. In addition, the protocol explored each aspect of the theoretical model to ensure that interviewees described their foundation’s entire approach to R&D evaluation. Appendix 3 provides the full general protocol for all interviewees. Questions were often tailored slightly based on the exploratory interviews to ensure that the questions fit within the foundation’s context. While the protocol separated out questions for science program officers, evaluators, and leadership, interviewees occasionally fit into two categories and therefore were asked questions across multiple sets of questions. For each foundation, I aimed to ask all interview questions to at least two individuals. Given limitations in interview participants, most science program officer questions and evaluation questions were asked twice but leadership questions often were not. Standard human subjects protection practices were observed. To allow for open and honest discussions, participants were informed that their comments would not be attributed to them personally or to their organization.

Interview transcripts were generated for each interview and then I qualitatively coded using an inductive approach. Initial exploratory interviews were included in the qualitative coding process since they provided valuable general information on foundations’ approaches to planning, monitoring, and evaluating R&D investments. The software package NVivo was used to organize and code interviews. High-level coding was initially performed to identify the broad themes from the interviews. Given resource limitations, I coded these responses on my own. These codes identified the primary components and characteristics of R&D evaluation

---

9 For example, foundations often use unique names to define sets of similar R&D projects that I refer to as “programs”. I adapted the protocol to use foundation-specific terminology.
approaches within each foundation. This resulted in a menu of characteristics and components for R&D evaluation approach and three distinct models for increasing evaluation capacity. These results were used in the Stage 2 workshop and the senior interviews described below. After consultation with my committee, I conducted another round of coding to identify further variations across foundations’ R&D evaluation approaches that is described in Chapter 2.

**Stage 2**

Stage 2 focused on identifying which foundation approaches could be beneficial to the Federal Government and how they would need to be tailored to fit within the Federal context. The following describes the collaborative design workshop and the senior Federal interviews that answer the primary research question.

**Workshop with Federal Stakeholders**

While research indicates that there is considerable overlap between the functions and operations of R&D in the private, public, and foundation sectors, the Federal Government does face unique requirements and oversight (Rainey and Bozeman 2000). Additionally, unlike firms and foundations, the Federal Government’s science agencies do not report up to a single entity (Fealing et al. 2011). Therefore, some translation was required to adapt the components and characteristics from foundations to the Federal context. Through a day-long workshop, Federal Government employees participated in a design process that capitalized on their expertise to identify which foundation approaches could be adopted to strengthen the capacity for the Federal Government to conduct evaluations that inform strategy and future program planning. Unlike a traditional workshop, this collaborative design workshop focused on eliciting perspectives and opinions from participants through a series of small and large group activities. The following section describes the workshop participants that were targeted for this workshop and the process and methods for each of the sessions.

**Workshop Participants**

Workshop participants were sought from every major Federal science agency. Targeted email invitations were sent to evaluation leaders within each agency where they could be identified. These individuals were also asked to nominate potential participants in other agencies. In agencies where there are no explicit evaluation offices or personnel tasked with overseeing evaluation, broad invitations were sent to senior R&D leaders. Recommendations for participants were also gathered from researchers who have conducted Federal R&D evaluations. I aimed to include two participants from each science agency who either conduct or oversee R&D evaluations. I targeted employees who were being paid under the General Service (GS) pay scale levels 14 and 15 or their equivalent. GS-14 and GS-15 level employees are supervisors and senior government leaders who typically lead offices within agencies or act as senior advisors.
This group was chosen with the assumption that they were the highest-level agency leadership that would understand the nuances and complexities of R&D evaluation to be able to understand which foundation approaches could be beneficial.

Ultimately, nineteen individuals attended the workshop from most of the major science agencies (specific agencies noted in Chapter 3). The only science agencies that were not represented at the agency or department level were the National Oceanic and Atmospheric Administration (NOAA) and the U.S. Geological Survey (USGS). Invitations were sent to employees at both agencies but relevant and available staff could not be identified for the workshop. In addition to science agency participants, two individuals from the White House Office of Management and Budget were included at the recommendation of a senior science agency evaluation officer. These individuals were advised that they would be asked to provide their opinions throughout the day but that their comments would not be attributed to them personally or to their agencies. This allowed for participants to provide honest critiques of current efforts and potential issues with trying to change the way agencies conduct R&D evaluations.

Session 1 Process and Methods

In the first session, participants were asked to design cluster evaluations\(^{10}\) that aim to inform strategy by looking across multiple R&D programs that share similar characteristics. Cluster evaluations were a fairly new concept for the foundations that I interviewed. Yet, a few agencies had conducted similar evaluations, although the terminology was new. These evaluations focus on analyzing programs within the same area or programs that share the same structure.\(^{11}\)

After framing the concept of cluster evaluations, the workshop group discussed their experiences with cluster evaluations in the Federal Government, how such evaluations could benefit science agencies, and the challenges with implementation. This discussion helped to further define the concept of cluster evaluations before splitting into three small groups. Each small group was tasked with designing a research plan to conduct a cluster evaluation for a specific topic. The three guiding topics I chose were Science, Technology, Engineering, and Math (STEM) education, the Small Business Innovation Research (SBIR) program, and the Brain Research through Advancing Innovative Neurotechnologies (BRAIN) Initiative due to

\(^{10}\) Cluster evaluations systematically identify the outputs, outcomes, and impacts of a set of grant programs that have shared characteristics. These new evaluations aim to inform strategic decisions at the portfolio level such as whether to increase or decrease a program, how to leverage one program to support another, or how to change individual program requirements to improve the likelihood of program success.

\(^{11}\) Further descriptions of cluster evaluations are provided in Chapter 2.
their broad applicability across agencies. After guided small group discussions, each group reported out to the larger group.

**Survey Design**

Prior to Session 2, participants were asked to fill out a short survey selecting the components and characteristics of foundation R&D evaluation practices that they thought would best benefit their agencies. In addition, the survey asked participants to provide information on current evaluation practices and resources within their agencies. See Appendix 4 for the complete survey instrument. In contrast to the other sessions, the survey offered an opportunity to gather individual opinions on how foundation practices could best benefit individual agencies.

I gave a short presentation prior to administering the survey that outlined the preliminary findings from Stage 1 on foundation evaluation practices. This presentation outlined each of the components and characteristics discussed on the survey and provided examples to the group. After the presentation, participants were then asked to fill out individual paper surveys. Participants provided their personal opinions based on their expertise and experiences, not the views of their agency.

**Session 2 Process and Methods**

In Session 2, participants were asked to identify the relative strengths and challenges of different models for placing evaluation-focused employees within the Federal Government. Among the foundations interviewed, I identified three distinct models that foundations have used to strengthen their capacity for conducting R&D evaluations. Over all other efforts and changes, I found that hiring evaluation-focused personnel was the most common and concrete way that foundations sought to expand and improve their use of R&D evaluations. However, foundations are experimenting with different models for using evaluation personnel.

This session aimed to identify which foundation personnel models could be advantageous for science agencies if they sought to incorporate additional evaluation-trained personnel into their agencies. This session also aimed to identify how personnel models would need to be tailored to fit a Federal context. Participants were split into small groups and asked to develop a sales pitch for a specific personnel model. I asked participants to design the pitch to Congress and assume that the model they were advocating for would be applied uniformly for all science agencies. In the small group sessions, participants discussed the strengths, challenges, and potential changes needed for their model. Small group discussions were then summarized to the entire group.

---

12 Each of these topics is funded by multiple agencies using a variety of approaches. In addition, each of these topics have been discussed within the National Science and Technology Council, the interagency group of science agencies that coordinate related efforts.
In Session 3, participants were asked to develop a “learning agenda” for science agencies. The U.S. Agency for International Development (USAID) and the U.S. Department of Housing and Urban Development have developed learning agendas that guide senior leadership in developing evidence to improve their performance. USAID describes their learning agenda as “short- and long-term priority research and policy questions relevant to the department’s mission” (Commission on Evidence-Based Policymaking 2017). Within foundations, dedicated evaluation units described similar efforts that help guide the activities of learning officers and help to determine central evaluation activities that look beyond individual program needs. In the final session, I asked participants to engage in an ideation and design thinking session to brainstorm research questions that they would want to include in their agencies’ learning agenda.

The guiding focus of this session was, “What are the evaluation research questions that need to be answered in order to strengthen the outcomes and impacts of future research programs?” Participants were allowed to suggest questions that were specific to their agency or relevant across multiple agencies. Participants wrote each research question down on a post-it note. Ideas were collected as participants wrote and posted for everyone to view. In addition, a few repetitive questions were collapsed into a single idea. Participants were then given six stickers to vote on the questions that, if answered, could lead to the biggest improvements in R&D programs. Participants were allowed to vote for their own ideas and could use several “votes” for the same question.

After the workshop, I sorted these questions into three overall categories. Agency operations questions focused on how to improve the efficiency or effectiveness of agency operations, both within the agency and with external partners. The assessment questions focused on how to improve R&D evaluations and specific evaluations that participants are interested in answering. Finally, program design questions are focused on improving the structure, operations, or focus on R&D programs. These questions are similar to assessment questions, yet have some explicit or implicit focus on using assessment results to change future R&D programs.

Interviews with Senior Federal Leaders

Finally, three senior Federal science leaders were interviewed to gather their opinions on how to strengthen the Federal Government’s capacity to conduct complex evaluations that inform strategy. These 30-minute, semi-structured interviews were conducted with leaders who are one-level above the personnel that were included in the workshop. These specific individuals were selected due to their potential interest in the topic area. The aim of these interviews was to capture higher-level leadership views on ways that foundation R&D evaluation approaches can be adopted to strengthen Federal capacity to conduct evaluations that inform future R&D program planning. In addition, the interviews offered a way to validate that the workshop targeted the right individuals and captured the diversity of opinions on how, and if, foundation
approaches could benefit Federal science agencies. The interview protocol was adapted from the workshop survey. It is reproduced in Appendix 5. Standard human subjects protection practices were observed. Unlike prior interviews, I asked if participants were willing for their comments to be attributed to a senior leader within their agency or office. Their agency affiliation helps to provide context for their opinions. Interviews were transcribed and compared with existing findings from the workshop. Quotes were pulled to strengthen the policy implications described in Chapter 4.

As a supplement to these interviews, I also interviewed a workshop participant who works at the National Science Foundation to better document the agency’s approach to R&D evaluation. This interview helped to inform the policy implications and recommendations in Chapter 4 since the workshop and senior Federal leader interviews recommended an evaluation model that is akin to NSF’s approach.

**Overview of Dissertation**

The above introduction outlined the problem statement, motivations, research questions, methodology and theoretical framework that guide this dissertation. Chapter 1 expands upon the motivations to provide a literature review on R&D and evaluation efforts within the Federal Government, philanthropic foundations, and private sector firms. Chapter 2 describes the myriad of components and characteristics of foundations’ approaches to R&D evaluation. To a much lesser extent, it also describes firms’ approaches, or lack thereof, to R&D evaluation. Chapter 3 provides the key findings from the collaborative design workshop with Federal stakeholders on which aspects of foundations’ evaluation approaches could benefit the Federal Government. Finally, Chapter 4 defines a series of policy implications and recommendations for Congress, the White House, and science agencies on how to improve Federal capacity for conducting evaluations that inform future program planning. This final chapter also highlights potential areas for further research.
Chapter 1: Literature Review

Key Findings from Chapter 1:

- Federal efforts since the 1990’s have increased all agencies’ focus on evaluating the results of all Federal programs but have not consistently led to improved decision-making on the structure and focus of R&D programs.
- Science agency-focused efforts to develop the capacity to evaluate the results of R&D investments have failed to achieve their stated goals.
- Foundations have invested heavily in evaluation efforts since the 1990’s, including recent increases in funding for evaluation activities, and therefore offer the potential for learning.
- Since the 1950’s, academic literature has continuously explored how to optimize R&D project selection but have focused little efforts on evaluating the outcomes of funded R&D to inform future program planning.

The recognition of R&D’s role in fueling innovation has spurred researchers to study how to effectively manage R&D since the early 1950’s (Fealing et al. 2011). In academia, countless journal publications have explored how organizations can better select R&D projects (e.g., Remer and Nieto 1995), strategically develop scientific talent, and implement accountability systems to manage scarce resources. In business, the Industrial Research Institute\textsuperscript{13}, an industry group of R&D managers, has supported a research group since 1968 to study best practices in managing R&D (Industrial Research Institute 2013). In government, Federal-wide efforts to improve effectiveness and efficiency of programs began in the early 1990’s and individual R&D programs have funded ad hoc studies to evaluate the results of their programs. Given this extensive background, this chapter aims to succinctly highlight the relevant studies and reports on evaluation efforts while articulating the unique scope and contribution that this dissertation makes to the literature.

\textsuperscript{13} The Industrial Research Institute was renamed the Innovation Research Interchange in 2017.
This chapter begins with a brief overview of Federal R&D to highlight the large scope of this enterprise, the primary actors that shape R&D programs, and the opportunities to use evaluation to better manage Federal R&D. Next, I expand on the introduction’s discussion of Federal efforts to evaluate R&D outcomes and impacts to provide a historical context for this dissertation and emphasize the opportunities for improvement. Finally, I outline the history of philanthropic foundations and private firms’ efforts to evaluate the impact of their R&D investments to motivate the value of examining these organizations and the gaps in current information on these organization’s R&D evaluation approaches.

**Overview of Federal R&D Enterprise**

Since World War II, the United States Federal Government has invested significant resources into R&D. In fiscal year 2016 alone, the U.S. Government spent approximately $148 billion on R&D, representing more than 10% of total discretionary spending (Hourihan and Parkes 2016a). While yearly budgets vary, the percentage of funding for R&D relative to overall discretionary spending has remained steady over the past 30 years (Hourihan and Parkes 2016a). The vast majority of R&D investments (~96% of all R&D) are managed through various programs within seven departments and agencies: the Department of Defense, Department of Health and Human Services (including the National Institutes of Health), Department of Energy, National Aeronautics and Space Administration, National Science Foundation, Department of Agriculture, and Department of Commerce (Hourihan and Parkes 2016b). Each of these organizations is responsible for designing, managing, and evaluating their own R&D programs. However, budgetary guidance and coordination is provided by the White House Office of Management and Budget (OMB) and the Office of Science and Technology Policy (OSTP) (Stine 2009).

OMB is responsible for reviewing individual agency budget submissions, making necessary revisions to match administration priorities, and approving individual budgets for inclusion in the President’s annual budget submission (Sargent et al. 2016, Hourihan and Parkes 2016b). The role of the OSTP has shifted across administrations (Stine 2009) but it is typically responsible for coordinating R&D programs across agencies (Hourihan and Parkes 2016b, Lane, Evans, and Matthews 2016, Fealing et al. 2011). Jointly, the OMB and OSTP release a yearly memo to science agencies to guide interagency R&D efforts. This memo helps to define administration priorities and guide the development of individual agency budget requests (see for example Zients and Holdren 2012). Additionally, OSTP manages the National Science and Technology Council (NSTC), an interagency group of science leaders who are responsible for coordinating science policy across the major science agencies and departments (Stine 2009, Lane, Evans, and Matthews 2016).
Despite White House guidance and coordination, each science agency and department (henceforth referred to as “agencies“) operates fairly autonomously in designing and managing its approach to R&D (Stine 2009). Agencies fund scientific research and development in a variety of disciplines through a series of individually designed R&D programs. Federal R&D programs focus on different levels of scientific inquiry including basic research, applied research, and experimental development (see OMB Circular A-11, Section 84 for definitions). Additionally, these programs fund projects in different ways (e.g., one-time lump sum versus annual payments) and place varying requirements on scientists who receive funding (e.g., varying reporting responsibilities or collaboration requirements). Figure 5 below highlights these key variations in program design. R&D programs vary in how much funding they receive, the particular field or area of research they invest in, the structure of their investments, and the requirements they put on grantees (Heinze 2008).

Figure 5. Variations in R&D Program Design

Despite their differences, R&D programs are often developed and supported with the hope that they ultimately impact similar areas across government, academia, the economy, and society (Gottron 2017, Lane and Bertuzzi 2011). For example, the NIH aims to reduce morbidity and mortality across society through funding biomedical research. The National Institute for Occupational Safety and Health, part of the Centers for Disease Control and Prevention, also aims to reduce morbidity and mortality but through worker-safety research such as improving harnesses for construction works. Despite funding vastly different areas of research, and using different funding structures, these two agencies share similar impact goals. A recent study I led at RAND found common agreement on areas of intended impact across science agencies (Basco 2016). In 2016, with funding from the National Science Foundation, I convened a diverse group of program planning and evaluation officials from all major science agencies to explore whether they could reach a consensus on a core set of R&D impact metrics that are relevant across fields of science and important to policymakers. The study found that participants agreed on 33 core metrics that could be used to assess the impact of Federal R&D on government, academia, the economy, and society. These metrics ranged from “lives saved from new information, products or technologies developed from R&D” to “new products developed from R&D” (Basco 2016).
This decentralized structure of R&D with varying requirements across agencies but similar goals could enable comparisons across programs and agencies. Unfortunately, the Federal Government does not systematically track these investments to understand how differences in program structures lead to differentiated results (Stephan 2015). Even within individual agencies, there is little systematic data collected across programs on impacts of research (Research Technology and Development Evaluation Topical Interest Group of the American Evaluation Association 2015; henceforth American Evaluation Association). The following section will provide a historical overview of Federal efforts to improve evaluation use to provide context for the current state of evaluation capacity and approaches within science agencies.

**Federal R&D Evaluation Capacity and Approaches**

Starting in the early 1990’s, the Federal Government began enacting reforms to encourage all Federal agencies to focus on the results of their programs, rather than simply focusing on the activities they are undertaking (Stevens 1997). Prior to these efforts, science agencies had not invested in evaluation efforts or capacity. Rather, they had successfully argued to Congress that retrospective evaluations would not be possible or valuable for R&D (Cozzens 1995). However, Federal-wide evaluation efforts pushed science agencies to invest in evaluation activities and better connect these to program planning (American Evaluation Association 2015). This section highlights broad, Federal-wide evaluation efforts to provide a historical perspective on policies that have impacted science agencies’ evaluation approaches and to highlight ways that Congress has sought to improve Federal programs. In addition, this section expands upon the introduction to discuss R&D-specific evaluation efforts undertaken by science agencies to highlight efforts since Jack Marburger’s 2005 call to improve science policy decision-making.

**Federal-Wide Efforts to Improve Government Programs**

In 1993, Congress passed the Government Performance and Results Act (GPRA) (Rojas 2000)(Rojas 2000)(Rojas 2000)(Rojas 2000) to improve the efficiency and effectiveness of Federal programs by requiring all agencies to better articulate the goals of their programs and gather information on their performance (National Academy of Sciences, National Academy of Engineering, and Institute of Medicine 2001). This landmark legislation was the first congressional effort to formally require agencies to document the goals and performance of their programs (Brass 2012). The legislation noted that Congress believed that a lack of defined quantitative program goals hindered managers’ efforts to improve Federal programs. In addition, the lack of data on program performance hindered Congress’ ability to conduct oversight and make budget allocation decisions. Combined, these issues led to inefficiency and waste that undermined public confidence in government programs according to the legislation (1993). To address these concerns, GPRA required agencies to develop and share strategic plans and performance assessments with Congress and the OMB. Through connecting annual assessments
of program outcomes to yearly budget decisions, GPRA aimed to improve the management of Federal programs (National Academy of Sciences et al. 1999).

GPRA did not specifically target science agencies or R&D programs. Rather, it required all agencies and departments to submit strategic plans and assessments to Congress and the OMB (National Academy of Sciences et al. 2001). However, science agencies faced unique challenges in implementing GPRA’s requirements (National Academy of Sciences et al. 2001, 1999). The Government Accountability Office14 (GAO) noted in 1997, as GRPA was beginning to be implemented, that developing performance measures for scientific research programs is difficult given attribution concerns since the Federal Government is not the only funder of R&D. In addition, the GAO noted that the outcomes of R&D may occur several years later, making it difficult to identify annual performance indicators (Stevens 1997).

Rather than develop a unified approach to solving these issues, science agencies pursued multiple pathways for developing performance measures. This resulted in a set of quantitative and qualitative measures that could not be combined to provide an overall view of the performance of Federal R&D investments (National Academy of Sciences et al. 1999). In addition, agencies varied in the granularity of their performance data, making it difficult to determine which R&D programs were successful (National Academy of Sciences et al. 2001). A National Academies panel in 2000 could not clearly assess the degree to which these performance reports were used by Congress for decision-making. Some agencies noted that it was difficult to use their own performance results to inform future planning, due to the short timelines of annual budget cycles that necessitated budget planning before performance results were available (National Academy of Sciences et al. 2001). GPRA was viewed as a significant step in ensuring agencies assess the results of their programs (Mihm 2013). However, it fell short in addressing the unique challenges of assessing the performance of R&D programs (American Evaluation Association 2015).

President George W. Bush’s administration developed a new performance management tool called PART, Program Assessment Rating Tool, as part of the response to their public acknowledgements of the failures of GPRA (Moynihan and Lavertu 2012). Used from 2003 to 2008, this tool assessed the “overall effectiveness” of Federal programs. However, the definition of a “program” in PART was determined by the OMB, which was criticized for not following agencies’ definitions of programs and aggregating certain programs together for reporting purposes (Brass 2012). The GAO found that these differences in program definitions created conflicts with GPRA reporting and sometimes led to prioritizing PART over GPRA (Irving 2004). A survey of Federal managers indicated that participation in GPRA and PART reviews had little impact on the actual use of performance management data for decision-making

14 The Government Accountability Office, or GAO, was previously named the General Accounting Office.
(Moynihan and Lavertu 2012). PART was ultimately deemed a failure by the Obama Administration, who stopped using the questionnaire and supported legislation to reform GPRA (Moynihan and Lavertu 2012).

To address the shortcomings of GRPA, Congress passed the GPRA Modernization Act of 2010 (referred to as GPRAMA or GPRA Mod) in 2011. GPRA Mod placed new requirements on the OMB and agencies to develop government-wide and broad agency goals to address cross-cutting issues. In addition, the legislation formally designated agency officials and a council to focus on performance improvements, codifying previous executive orders (Mihm 2013). However, GPRA Mod did not address the unique issues that science agencies face when trying to assess the results of their R&D programs (American Evaluation Association 2015). The legislation did not address the attribution or time-lag issue with observing outcomes. The legislation also failed to standardize reporting across agencies. As a group of R&D evaluation experts noted, “[f]rom the program evaluation perspective however, GPRAMA does not clarify or improve on approaches to designing or using results from program evaluations that Congress described in GPRA” (American Evaluation Association 2015).

Efforts to Improve R&D Evaluation Capacity and Practices

While GRPA is acknowledged as a “watershed” moment in 1993 that resulted in new efforts across the Federal Government to better define and assess Federal programs’ performance (Brass 2012), the resulting changes were not enough to satisfy President George W. Bush’s Science Advisor Jack Marburger. In 2005, Marburger noted:

“Much of the available literature on science policy is being produced piecemeal by scientists who are experts in their fields, but not necessarily in the methods and literature of the relevant social science disciplines… that can be useful to policy experts.” (Marburger 2005)

He believed that difficult decisions on how to best design and structure Federal science efforts “will not be remedied by simple measures” (Fealing et al. 2011). Rather, Marburger believed that a new “science of science policy” academic field needed to emerge to develop new methods, data, and dedicated scholars to answer fundamental questions of how to best design R&D programs. In response to this call, the National Science Foundation (NSF) established two related efforts—a grant program and a project to collect outcomes and impact data across the Federal Government. This section will briefly describe these efforts and opportunities for further work.

The NSF developed the Science of Science and Innovation Policy (SciSIP) program\(^\text{15}\) in 2005 after being directed to include a budget request for the new program by the OMB (Cozzens

\(^{15}\) Full disclosure: I received a grant under this program.)
The SciSIP program follows the standard NSF funding model of openly soliciting individual research proposals that respond to a broad funding opportunity announcement. Rather than prescribe a research agenda that aligned with Marburger’s interests (or science policymakers’ interests), NSF followed its standard approach of gathering ideas for the program solicitation from the research community (National Research Council 2014). After three workshops with NSF-funded social scientists in related disciplines, the initial goals of the SciSIP program were defined as “(1) develop usable knowledge and theories of creative processes and their transformation into social and economic outcomes; (2) improve and expand science metrics, datasets, and analytical tools, yielding changes in the biannual science and engineering indicators and other data collections; and (3) develop a community of experts across the federal government, industry, and universities focused on SciSIP” (National Research Council 2014, Cozzens 2010). The initial cohort of research proposals was funded in 2007 and subsequent cohorts continue to be funded annually or semi-annually as of February 2018.

In 2015, then-SciSIP program officer Maryann Feldman funded a series of workshops to identify a path forward for the program. Implicit in the call for these workshops, and explicitly stated by workshop organizers, was the notion that that SciSIP program has failed to influence science policymaking. One workshop report noted that researchers funded under the SciSIP program are fundamentally disconnected from the policymakers in Congress and the executive branch. In many cases, science policymakers were unaware of SciSIP-funded research and researchers (Alexander, Hart, and Hill 2016). As of February 2018, NSF has not released a new roadmap for the SciSIP program or developed a revised solicitation to encourage better research proposals that can inform policy. This suggests that Jack Marburger’s call to develop a robust field of science policy has failed to be adequately answered. While the program has funded numerous proposals that have explored aspects of science policy (National Research Council 2014), it has failed to help policymakers make better decisions on how to design R&D programs to maximize the economic, social, academic, and governmental benefits of Federal investments.

In addition to the SciSIP program, the NSF experimented with a Federal-wide data collection system to gather information on the outputs and outcomes of Federal R&D beginning in 2010. This system, known as the STAR METRICS project, began as a pilot between the National Institutes of Health (NIH) and the NSF under the auspices of OSTP that aimed to collect disparate information from universities on the initial outputs and outcomes of federally funded projects (Lane and Bertuzzi 2011). After it launched, the Department of Energy and the Environmental Protection Agency also signed on as partners (National Research Council 2011). Its leaders had ultimately hoped to be able to utilize existing, disparate databases to track outcomes to impacts in order to assess the return on investments from Federal R&D (National

---

16 Full disclosure: I worked this project from 2010 to 2011 as an intern and then contractor.
Research Council 2011). In its first stage, it aimed to capture granular information on the scientific workforce that is funded under Federal R&D programs by gathering payroll and accounting data from universities. In the second stage, it aimed to gather input data from Federal science agencies and connect it to a variety of databases that track the outcomes and impacts of R&D. However, few Federal agencies were willing to provide their own data, not enough universities contributed their data, and there were difficulties in linking research funding data sets to output and outcome data sets. Therefore, the pilot project was ended in 2015 (Rockey 2015).17

**Current Approach to Federal R&D Evaluation**

In the absence of a cohesive program or system to measure impacts and evaluate R&D programs, individual R&D programs have continued to fund one-off studies. Unfortunately, there is no way currently to track how many of these studies are conducted each year or how much money is spent on evaluation activities. Additionally, there is no data on the number of evaluation personnel within science agencies that are helping to design and conduct these evaluations. Due to this lack of data, I conducted a survey as part of this dissertation to develop estimates of the number of personnel and funding devoted to evaluation. While systematic assessments of Federal R&D evaluation approaches are not available, an American Evaluation Association sub-group and a few academics have described certain aspects of current approaches to Federal R&D evaluation.

The de facto standard in Federal R&D evaluation has been the mixed-methods program evaluation (American Evaluation Association 2015). Mixed-methods program evaluations are process and summative evaluations that typically focus on identifying the specific goals of the program; developing a logic model18 to describe the program’s resources, activities, outputs, and outcomes; and collecting ad hoc data to assess the extent to which individual projects lead to programmatic outcomes (see Zuckerman et al. 2014, for an example). While formative and process evaluations address some aspects of the overall R&D process, they do not capture impacts. This makes it difficult to assess the overall success of a program and to compare to other programs. Also, these evaluations are conducted independently without coordination across programs, limiting their value outside of the individual program that is being evaluated.

Evaluations are typically conducted by third-party evaluators from think tanks or research institutions (American Evaluation Association 2015). Third-party evaluators enable program

---

17 Julia Lane, the original leader of the STAR METRICS program, left the Federal Government and re-launched this initiative as UMERTICS within a multi-university consortium. See Lane (2017) for further details on the new UMERTICS initiative.

18 The W.K. Kellogg Foundation (2004) defines a logic model as “a picture of how your organization does its work – the theory and assumptions underlying the program. A program logic model links outcomes (both short- and long-term) with program activities/processes and the theoretical assumptions/principles of the program.”
managers to justify investments to outside groups such as Congress and gather input from grantees on how the programs should be modified to promote further success. R&D program managers typically use short-term contracts for evaluations that specify research questions, data collection, and methods tailored to their unique program. The current system of ad hoc evaluation limits the ability for programs to compare outcomes and duplicates data collection efforts. It also limits policymakers from being able to use evaluation results to inform which program should be funded in the future or how new programs should be structured to achieve its goals (Fealing et al. 2011). Further work is needed across all science agencies to improve their ability to conduct complex, large-scale evaluations that can inform policymaking at agency, White House, and congressional levels.

**Philanthropic Foundation R&D Landscape and Evaluation Practices**

The level of experimentation and focus on evaluation activities within foundations that fund R&D indicate the potential for the Federal Government to learn from foundations. This section will begin by highlighting the growth of evaluation within philanthropic foundations broadly. I will then narrow the focus to philanthropic foundations who fund R&D to document their relative importance within the national R&D enterprise and highlight the connections between government and foundation R&D approaches. Finally, this section will discuss the limited research on evaluation practices specifically within foundations that fund R&D to provide the motivation for further exploring evaluation practices within these organizations.

As the Federal Government was implementing GPRA provisions, foundations also began to focus on evaluating the outcomes of their investments in the 1990’s. The professionalization of nonprofit management, the proliferation of new foundations, and high-profile evaluation champions helped lead this change (Coffman et al. 2013, Behrens and Kelly 2008). In addition, GPRA’s emphasis on the use of logic models influenced foundations’ use of the evaluation tool to connect goals, activities, outputs, and intended outcomes of programs (Behrens and Kelly 2008). Throughout the 1990’s and 2000’s, foundations experimented with a multitude of approaches for evaluation, with limited success in connecting evaluation results to strategic decisions (Behrens and Kelly 2008, Patrizi and Sedway 2006). However, foundations have maintained interest in improving practices through learning from one another. Many foundations have also recently been devoting more resources to evaluation according to researchers who have been tracking evaluation spending since 2009 (Coffman et al. 2013, Center for Effective Philanthropy and Center for Evaluation Innovation 2016). This suggests that foundations have a rich history of experimenting with ways to improve evaluation capacity to develop better evaluation results that can inform future program planning and strategy.

Foundations have also increased their focus on evaluation over the last decade. A 2016 survey indicated that 50% of the foundations surveyed had invested more funds into evaluation
over the last two years than they had previously invested in evaluation activities. Funding levels stayed the same at 45% of the foundations surveyed (Center for Effective Philanthropy and Center for Evaluation Innovation 2016). This survey was based on similar research conducted in 2009 and 2012. The 2009 survey indicated that 62% of the foundations surveyed had invested more money for evaluation activities over the past two-year period. Similarly, a survey in 2012 found that 50% of foundations surveyed had invested more money for evaluation activities over the past two-year period. In 2012, 30% of the foundations surveyed indicated that their evaluation spending had stayed relatively the same (Coffman et al. 2013). This indicates that a large group of foundations have recently been investing greater funding for evaluation activities.

The 2016 survey found that about half of the respondents had 1.5 or more employees dedicated to evaluation work. Additionally, more than one-third of the foundations surveyed had dedicated evaluation units (Center for Effective Philanthropy and Center for Evaluation Innovation 2016). However, the use of evaluation for future program planning purposes may be limited to the top funders of R&D. The 2016 survey asked respondents to describe what they hope will change with foundation evaluation approaches in the next five years. The top response was for better linkages between evaluation results and strategy (Center for Effective Philanthropy and Center for Evaluation Innovation 2016). In addition, the 2012 survey indicated that only large foundations spent more than 10% of their time learning from evaluation results. While these surveys focused broadly on all foundations, several large foundations that responded fund R&D (Coffman et al. 2013). This suggests that numerous large foundations may be able to provide advice and perspectives on how to effectively develop evaluation capacities to better utilize evaluation results for future program planning.

Foundations play a significant role in funding R&D in the United States (Murray 2013). In 2016, nonprofit organizations, including foundations, provided $4.6 billion for R&D conducted within academic institutions, accounting for more than 6% of all funding for academic R&D (National Science Board 2018). In certain scientific areas, foundation R&D programs are exemplars that the Federal Government tries to emulate. For example, the National Institutes of Health (NIH) funded an evaluation of their NIH Director’s Pioneer Award (NDPA) program in 2012 that compared the results of NDPA awardees with similar academics funded by the Howard Hughes Medical Institute, a large health R&D foundation (Lal et al. 2012). In other areas, foundations aim to play a unique role that supplements or complements Federal investments. In health research, a consortium of foundations noted that they often fund emerging areas of research, neglected diseases, underserved researchers, and high-risk projects that are not fully supported by industry or government (Myers et al. 2012).

---

19 The 2009 and 2012 surveys were conducted on much smaller samples of foundations than the 2016 survey but still support the conclusion that many foundations have steadily increased funding for evaluation activities.

20 Funding data on only foundation-funded R&D is not available.
Limited studies have focused specifically on the subset of foundations that fund R&D, although the foundation landscape studies that were outlined in the previous paragraphs included R&D funders within their study populations. In addition, case studies and academic literature on individual evaluation practices within R&D funding-foundations indicate similar levels of effort and focus on evaluating investments to inform future strategy compared to all foundations (Parker 2016, Dillman and Christie 2017). In fact, the Robert Wood Johnson Foundation (RWJF), which funds health research, is seen as a leader within the broader philanthropy community. Approximately 20% of RWJF’s funding is dedicated to evaluation activities (Dillman and Christie 2017). An in-depth case study of the Hewlett Foundation, who also funds R&D, highlighted a deep interest in evaluation and a continuous process to adjust its approach to make evaluation relevant and useful across the Foundation. In 2015, the Foundation spent approximately 1.4% of its programmatic dollars on evaluation, with leaders hoping to meet a 2% expenditure goal (Parker 2016).

The Federal Government may also be able to learn from foundations’ efforts to develop processes for “organizational learning” or “strategic learning.” The idea of a “learning organization” was initially promoted by Senge in 1990 but did not begin to be explored by foundations until the early 2000’s (Behrens and Kelly 2008). Case studies of foundation evaluation practices indicate that organizations view these as distinct activities from purely evaluating the results of their efforts and have developed specialized learning activities (Behrens and Kelly 2008, Parker 2016, Dillman and Christie 2017). This indicates that using evaluation results to inform future program planning is recognized as valuable amongst at least a subset of grantmaking foundations.

Unfortunately, the literature to date does not provide system-level perspectives on the policies, personnel, data collection methods or tools for foundations’ evaluations of R&D. Case studies provide some insight into the thought processes of foundations as they have designed and altered their evaluation systems. However, they do not provide sufficient detail to identify the components and characteristics of their systems that enable robust collection of impact data and ensure its use in program planning. In addition, the high-level surveys of foundation practices do not focus specifically on R&D organizations. Rather, many grantmaking foundations surveyed fund non-profit service delivery work such as providing healthcare to the poor. Therefore, more research is needed to understand the specific practices of R&D-focused grantmaking foundations.

Private Sector R&D Landscape and Evaluation Practices

Common assumptions about the differences between private, for-profit organizations and public organizations may lead researchers to question the value of examining private sector practices to inform public sector planning. However, decades of research highlight that there are
many commonalities between private firms and government organizations (Rainey and Bozeman 2000). In addition, the breadth of private firms that have experimented with different ways to design and implement evaluation systems since the 1950’s provides a rich set of experiences to learn from. This section will identify the connections between private and public R&D operations, the state of research on private sector evaluation practices, and organizations that may be valuable partners in conducting this dissertation.

There is a pervasive belief that private firms are motivated solely by profit while public institutions have a more complex set of goals. However, research on firms indicates that they have multiple goals, particularly when considering R&D, which mirror goals stated by government including training scientists (Gutjahr et al. 2010). Firms are not solely focused on profit as a singular goal. Rather, research that has compared general private firm goals and government goals, including non-R&D related efforts, has shown that the goals of private organizations are just as complex and ambiguous as government goals. Specifically, managers within private firms report that their firms’ goals are often as numerous and conflicting as goals reported by government managers. In addition, surveys indicate that managers’ perceptions of the vagueness and difficulty in measuring progress of goals is similar between private firms and government (Rainey and Bozeman 2000). When focusing on R&D-oriented organizations, similarities in goals are even stronger. Since the early 1950’s, operations researchers have continuously proposed a series of new methods that aim to optimize R&D project and portfolio selection in the private sector. In many of these methods, several objectives are used in addition to profit, including maximizing opportunities to train scientists on new methods, signaling a niche expertise within a market, and balancing risk between short-term and long-term projects (Stummer and Heidenberger 2003, Gutjahr et al. 2010, Meskendahl 2010, Baker 1974). These objectives map well to government roles of training the next generation of scientists, ensuring U.S. competitiveness globally, and supporting various projects along the R&D spectrum. The complex, sometimes incongruent objectives observed in the Federal Government are therefore not unique given the similar objectives found in the private sector.

Unlike the Federal Government, however, industry has invested significant resources into determining how to best manage its R&D activities and has supported activities by outside researchers (Chiesa and Frattini 2009). Since 1938, the Industrial Research Institute (IRI) has brought industry leaders together to discuss how to “advance the field of innovation management” (IRI webpage). The IRI also recognized the need for formal research on how to effectively manage R&D. In 1968 the IRI attempted, unsuccessfully, to convince NSF to develop a dedicated research program to identify strategies for better managing R&D. So the IRI started its own committee to guide action-oriented research funded and conducted by its member organizations (Industrial Research Institute 2013). Industry has also funded outside groups to conduct research on how it operates and has participated in numerous studies by academics and management consultants, including the Matheson and Matheson study described below.
The seminal 1998 book, *The Smart Organization*, is one example of industry participation in a study\(^\text{21}\) that highlights how to effectively manage R&D. The authors leverage years of experience as management consultants to businesses to identify the key components of effective R&D organizations. In their analyses, they identified 45 best practices of successful R&D organizations that fall within three pillars. This dissertation focuses on one of these pillars, “improving decision quality” (Matheson and Matheson 1998) that is outlined in Figure 6 below. The authors note that while the best organizations “learn from post-project audits”, “measure R&D effectiveness”, and “create frameworks for learning”, not all organizations have found a way to implement these practices. Yet, the simple recognition of these functions indicates that the best firms are significantly ahead of government when it comes to evaluating R&D and using evaluation results to alter future investments.

**Figure 6. Best Practices for Industry R&D**

Despite the recognition that firms need to develop ways to improve decision quality, there is little literature focusing on evaluating how the design and structure of their R&D investments shape the ultimate impacts that result. Rather, evaluation studies tend to focus on how to monitor immediate outputs and outcomes of research for feedback into how to better assess and manage

---

\(^{21}\) This study was launched by the Strategic Decisions Group, a business consultancy, with sponsorship from the R&D Decision Quality Association, an industry group (Matheson and Matheson 1994).
the personnel conducting research. These studies explore the uses of Performance Management Systems (PMS) and Balanced Scorecards that attempt to combine financial and non-financial indicators to compare personnel and projects to benchmarks (Chiesa and Frattini 2009). Rather than assessing if their system for funding R&D is effective, they assess the relative value of personnel and projects working within that system. These studies explore the appropriate metrics for use in PMS, how to effectively collect data, how to assess data, and how to adapt systems to different types of R&D and different settings (Chiesa and Frattini 2009).

Overall, the literature indicates that firms have experimented with various forms of evaluation and feedback. The continued proliferation of methods for selecting R&D also indicates that industry is still experimenting with the best ways to inform how R&D should be conducted. In addition, the continued discussions of how to measure R&D effectiveness indicate an opportunity to learn from firms about how they implement new evaluation systems. Unfortunately, the literature to date does not provide a comprehensive view of industry practices for evaluating its R&D. There is a lack of system-level perspectives on the policies, personnel, data collection tools, and methods for evaluating R&D to inform future program planning. In addition, there are limited case studies examining how the components and characteristics of an evaluation system work together to enable managers to learn from the impacts of their projects. This motivates the need for further research to understand private firms’ approaches to R&D evaluation.

**Gaps in Existing Literature**

This dissertation addresses two major gaps in the literature: data on Federal R&D evaluation capacity and foundation approaches to R&D evaluation. First, I developed baseline estimates of the number of evaluation-focused employees within science agencies and the amount of money spent on R&D evaluation activities each year. There are no comprehensive assessments of R&D evaluation capacity within the Federal Government. Evaluations are often funded through science program funds and therefore cannot be tracked using existing appropriation or spending data. Additionally, third-party contracted evaluations are not routinely collected across science agencies or disclosed to the public, making it difficult to track evaluation activities using public data. Similarly, there is no database that catalogs evaluation-focused personnel within science agencies. Agencies may refer to evaluation personnel using different terms and may not have unique job classifications for these individuals but rather refer to them as general health science administrators or other generic terms. I conducted a brief survey with senior Federal science policy, planning, and evaluation officials to collect estimates of the money spent on R&D evaluation activities each year and the number of full-time equivalents (FTEs) that are focused on evaluation activities within science agencies.
This dissertation also identified a variety of approaches that philanthropic foundations have taken to R&D evaluation. There is very limited literature on the policies, personnel, data collection methods, and tools that foundations use to assess the outcomes and impacts of their R&D investments. Current literature does not focus on foundations that fund R&D and does not provide system-level perspectives on the components and characteristics that enable foundations to have the capacity to evaluate R&D investments to inform future program planning. This dissertation provides new knowledge on the range of ways that foundations have strengthened their capacity to design and conduct R&D evaluations to inform strategy.
Chapter 2: Foundation and Firm Findings

Key Findings from Chapter 2:

- Foundations have heavily invested in personnel to strengthen their capacity to conduct R&D evaluations.
- Evaluations are used for multiple purposes ranging from accountability to strategy and can examine several different aspects of R&D programs from the broader context of the program to ultimate impacts of the program.
- Foundations have not pursued the development of organization-wide data collection tools to capture outcomes and impacts of R&D beyond collecting narrative project reports.
- Limited evaluation practices were identified within the small sample of interviews with large, U.S.-based private firms.

This chapter provides an overview of the different approaches that philanthropic foundations use to evaluate the outcomes and impacts of their R&D programs. A range of components and characteristics that enable R&D evaluation efforts were identified through interviewing a diverse set of foundations. The exploratory interviews with 10 foundations provided general information on how foundations plan, monitor\(^{22}\), and evaluate the outcomes and impacts of their R&D investments. In-depth interviews with five selected foundations expanded upon the exploratory interviews to develop a holistic view of foundation approaches and to explore how foundations use evaluation results to inform future program planning and strategy. These in-depth interviews were conducted with a mix of evaluation staff, science program officers, and senior leaders to triangulate information and ensure complete understanding of foundations’ approaches. Together, these two sets of interviews addressed the three secondary research questions in this dissertation. This chapter provides insights on each of these questions:

\(^{22}\) There was no agreed upon distinction between monitoring and evaluation by interviewees; therefore, I asked about both sets of activities and then reported on evaluation activities as defined as retrospective analyses of the results of R&D programs.
1. How do R&D grantmaking foundations plan and develop their R&D programs?
2. How do R&D grantmaking foundations evaluate the outcomes and impacts of their R&D portfolios?
3. How do R&D grantmaking foundations use evaluations of their past R&D projects to inform future planning of R&D investments?

Rather than try to identify best practices amongst foundations, this dissertation aims to identify a range of ways in which foundations approach R&D evaluation. Therefore, this chapter describes the myriad of components and characteristics that constitute R&D evaluation approaches within the foundations interviewed. Given the semi-structured nature of the interviews (see Appendixes 1 and 3 for the full protocols), each interviewee did not discuss all aspects of their foundation’s approach to R&D. In addition, varying numbers of interviews were conducted within each foundation based on availability of staff and the complexity of foundation’s approach to R&D evaluation. Therefore, the following chapter does not quantify how many interviewees use each component and characteristic that is described. Rather, this chapter provides a landscape of the different ways that foundations have addressed R&D evaluation and planning. As a way of structuring these findings, this chapter will outline options based off the adapted socio-technical framework that is reproduced below in Figure 7. Within each component and characteristic, I will also describe how foundations use the resource in planning and evaluating their R&D to answer the secondary research questions.

Figure 7. Adaptation of Socio-Technical Theory Framework
The following chapter describes combined findings from the exploratory and in-depth foundation interviews; the limitations of the theoretical framework to accurately outline foundation R&D evaluation approaches; and finally, the limited findings from firms that were interviewed before their associated research questions were dropped.

**Insights on R&D Evaluation from Foundations**

This section captures the range of components and characteristics that constitute foundations’ approaches to R&D evaluation as of fall 2017. Several foundations that were interviewed explained that their approaches to R&D evaluation continuously evolve as they identify best practices from other foundations and experiment with new ways to understand the impacts of their programs. As is noted in some of the sub-sections below, several foundations have devoted new resources and developed new evaluation activities in the past two years. The following section outlines how foundations have utilized evaluation personnel; leadership influences that shape planning and evaluation approaches; and the evaluation activities that foundations have invested in and how they have been funded.

**Use of Evaluation Personnel**

Above all other components and characteristics, the foundations that were interviewed have utilized personnel to increase their capacity to conduct R&D evaluations that inform strategic decisions. Several of the foundations that were interviewed have developed new evaluation units within the last five years to expand their ability and focus to conduct R&D evaluations throughout the lifecycle of their programs. These new efforts highlight the value that foundations realize in understanding the outcomes and impacts of their funded R&D as they seek to make strategic investments to shape science and society.

From the 10 foundations that were interviewed, three distinct models for evaluation personnel were identified: centralized evaluation services, evaluation consultants, and program officers as evaluators. Centralized evaluation services use evaluation and learning experts to design and oversee evaluations from an independent office. Evaluation consultants use internal evaluation experts embedded within program offices to advise program officers on how to design evaluations. Program officers as evaluators use program officers to develop and carry out monitoring and evaluation activities that focus on project-level continuous improvement. The following section describes the roles, skills, and stated benefits of personnel hired under each model. This section also outlines how these personnel fit within the overall organizational structure of foundations. Finally, this section briefly highlights the components and characteristics of foundations’ R&D evaluation approaches that are most commonly associated with each model.
Centralized Evaluation Service

Foundations with centralized evaluation services hire evaluation and learning experts within independent evaluation offices, separate from science programs, that directly report to senior foundation leadership. Personnel within this model design, lead, and fund evaluations that examine individual programs and groups of similar programs. Rather than science program officers supervising evaluations of their programs, foundations with centralized evaluation services use evaluation-trained personnel to oversee third-party evaluation contracts. In some foundations, personnel within this model also conduct ad hoc internal evaluations of foundation operations to identify potential changes to improve internal processes and procedures. Several evaluation personnel within centralized evaluation services who were interviewed noted that they explicitly do not play a monitoring role with grantees to track accountability. These evaluation personnel are able to engender trust with grantees and obtain better information on outcomes and challenges with their grants by explicitly not focusing on assessing individual grantee performance. Science program officers are responsible for monitoring and accountability in order to ensure that evaluation officers are not viewed as additional oversight. Rather, these evaluation officers are interested in conducting evaluations for broader learning.

Within these centralized evaluation services, personnel fill 4 distinct roles: evaluation officer, organizational effectiveness officer, learning officer, and evaluation leader. Evaluation officers are responsible for designing and overseeing R&D evaluations. Organizational effectiveness officers are responsible for developing reports and assessments of the overall health and performance of the foundation. Rather than focusing on individual evaluations, organizational effectiveness officers examine the broader organization using administrative data and feedback from grantees to track performance. Within the last two years, foundations with this personnel model have also developed “learning officer” positions that focus on developing activities and training sessions to improve foundation practices using the results of external evaluations or internal working group discussions. Finally, evaluation leaders provide central support and guidance on how to design evaluation and learning efforts to improve foundation practices.

Centralized evaluation service personnel typically have general interdisciplinary evaluation skills developed either through graduate training or on-the-job experience. Additionally, this model often enables foundations to hire PhD-trained methods experts who can advise evaluation staff and program officers on the best ways to design evaluations to collect actionable data. However, evaluation personnel are not content experts in the R&D that the foundation is funding.

---

23 Science program officers (also referred to simply as ‘program officers’) are the staff that are responsible for the design and execution of individual R&D programs within their field of expertise. These personnel often sit within broader science divisions and often report to a chief science officer (or similar role). Program officers are directly responsible for the solicitation, review, and awarding of individual R&D proposals.
Evaluation staff and program officers believed there were several benefits to using a centralized evaluation service model. Through operating in an independent office separate from science programs, evaluation staff are able to focus on cross-cutting evaluations that look beyond individual programs to understand how sets of related programs are supporting the mission of the foundation. In addition, these evaluation staff are viewed as more independent and objective than program officers when designing and overseeing program evaluations. Since they are not responsible for program decisions, they have less of a vested interest in the evaluation portraying the program in a positive light. Therefore, some interviewees viewed these staff as helping to ensure that third-party contracted evaluations accurately identify the strengths and challenges of the program. Finally, evaluation staff noted that the dedicated learning officer role is a fairly new position that would not exist outside of a dedicated evaluation unit. Learning officers inherently take a broad, cross-cutting view of the foundation and therefore need to be within a unit that is empowered and interested in foundation-wide improvements. This learning officer is able to explicitly focus on and plan activities to spread evaluation activities broadly and ensure that they are used to improve foundation operations and programs.

Interviewees who work within this model also discussed other characteristics and components of their R&D evaluation approach that this model encourages. For example, the development of central evaluation service offices also led to the establishment of centralized, dedicated funds for evaluation activities within some foundations that were interviewed. These independent budgets help to ensure that evaluation officers can carry out complex evaluations to benefit individual program offices in addition to benefitting the overall foundation. Interviewees within centralized evaluation offices also discussed designing more complex evaluations than program officers who also act as evaluators. Two foundations with central models described developing new types of evaluations that look beyond individual programs to inform strategy development. These new types of evaluations aim to provide more actionable information to program officers on how to improve their programs to achieve their intended impacts.

Internal Evaluation Consultants

Internal evaluation consultants are internal foundation employees with evaluation training who are embedded within individual science programs and initiatives to advise program officers on evaluation issues. Rather than hire external consultants to advise programs on an ad hoc basis, these foundations have recruited evaluation experts as internal foundation employees to continuously support specific science programs through advising them on the most effective ways to design, conduct, and use evaluations. These personnel report to the same senior evaluation leader who typically reports to the foundation president. This allows internal consultants to provide advice on ways to better incorporate evaluations without being supervised or censored by science program leadership. However, internal evaluation consultants do not operate as a single entity. Rather, each consultant supports a set of related programs as they
develop, implement, evaluate, and terminate R&D programs with loose connections to other consultants. These consultants develop common evaluation resources for program officers such as Request for Proposal (RFP) language for external evaluation contracts. In addition, they develop logic models and evaluation templates to assist program officers in carrying out evaluations. Typically, these consultants act as guides and resources for program officers but are not directly responsible for overseeing evaluations. Rather, program officers are responsible for managing external contracts and making the final decisions on evaluation design and execution. Within some foundations, consultants also participate in internal strategy reviews to rapidly assess the challenges and opportunities for improvement within individual programs. Similar to centralized evaluation service staff, evaluation consultants explicitly stated that they do not play a monitoring role.

Interviewees described evaluation consultants as having general interdisciplinary evaluation skills through academic or on-the-job training. However, these personnel were not content experts in the R&D that the foundation is funding. Additionally, these personnel are not PhD-trained evaluation method experts. Rather, personnel often possess masters-level credentials in evaluation or related fields.

Interviewees within foundations that use this model provided several examples of the value of these evaluation staff. By embedding themselves with program staff, evaluation consultants are able to develop familiarity with the scientific content of the programs they are supporting. Therefore, their evaluation recommendations better address the unique aspects of the science and grantees that the foundation is funding. In addition, this close working relationship enables evaluation consultants to develop trust with program staff who are more likely to take their advice when developing and executing evaluations.

Interviewees in foundations that use this model also discussed using cost-sharing mechanisms as a means for funding evaluations. Rather than having a large, centralized fund that is allocated sufficient funding to support all evaluation activities, internal evaluation consultants must convince science programs to provide partial funding for evaluations. The evaluation consultants have access to a centralized set of funds that can be used to support external, third-party evaluations. However, these funds cannot cover the full cost of these evaluations. Therefore, interviewees noted that program officers fund part of the evaluation through programmatic funds with evaluation consultants providing the rest. Cost-sharing models help to ensure that evaluations will be valuable to evaluation staff and program officers since each has a financial stake in the evaluation.

Program Officers as Evaluators

A few of the foundations that were interviewed require science program officers to also act as evaluators. Rather than devote limited foundation resources to evaluation officers, these foundations insert evaluation roles within program officer job descriptions. Under this model,
program officers often focus on conducting continuous improvement activities at the grant level. One program officer under this model described his role as acting as an advisor and resource for his grantees to help ensure their success. Through yearly site visits, he talks with his grantees about their research projects and helps to brainstorm solutions to any barriers or issues. He viewed his role as helping to ensure overall program success through supporting individual projects. In addition to this continuous improvement role, program officers who also act as evaluators monitor grantees for accountability purposes and develop RFPs for midterm and final evaluations. These program officers are typically content PhDs in the R&D field they are funding. However, they rarely have academic or job training in evaluation methods.

Interviewees within foundations that use this model identified three primary benefits to using program officers as evaluators. Program officers have direct relationships with their grantees and are therefore able to conduct oversight and problem solve with more efficiency and effectiveness than evaluation staff. In addition, this model helps to keep foundation operating costs as low as possible by not funding additional evaluation staff. Finally, interviewees noted that program officers who act as evaluators can directly provide incentives for grantees to achieve outcomes by linking success metrics to future payments or support.

Program officers who use this personnel model for evaluation also mentioned that evaluations are typically funded through ad hoc program funds. Rather than establish dedicated evaluation funds, these foundations simply pull from individual program funds when paying for external evaluations. These program officers also discussed conducting limited types of evaluations, primarily process and summative evaluations.

Comparing Personnel Models

Foundations are experimenting with different personnel models to determine the optimal staff needed to conduct evaluations that can inform strategy and learning. Centralized evaluation services and internal evaluation consultants are relatively new models, with foundations often previously relying on program officers to carry out evaluation activities. Table 1 below provides an overview of the similarities and differences between the personnel models that foundations described in the interviews.
Table 1. Comparing Personnel Models

<table>
<thead>
<tr>
<th>Primary Evaluation Roles</th>
<th>Centralized Evaluation Service</th>
<th>Internal Evaluation Consultants</th>
<th>Program Officers as Evaluators</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Design, lead, and fund evaluations</td>
<td>Advise program officers on how to design, lead, and fund evaluations</td>
<td>Conduct continuous improvement activities at the grant level</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Evaluation Personnel Training</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>General Interdisciplinary Evaluation Skills</td>
<td>✔</td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>Dedicated Evaluation Method Experts (PhD-trained)</td>
<td>✔</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Content Experts in R&amp;D Field (PhD-trained)</td>
<td></td>
<td></td>
<td>✔</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Stated Benefits</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Focus on Cross-cutting Evaluations</td>
<td>✔</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deep Knowledge of Individual R&amp;D Programs</td>
<td></td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Direct Relationships with Grantees</td>
<td></td>
<td></td>
<td>✔</td>
</tr>
</tbody>
</table>

| Funding for Evaluation                    | Central, independent fund         | Cost sharing between centralized fund and program funds | Ad hoc program funds |

**Leadership Influence on Evaluation Capacity**

Foundation leadership determines the resources that are devoted to evaluation and the role of evaluation in informing the design of new R&D programs. This section outlines the range of roles for foundation boards of advisors and presidents in planning and evaluating R&D programs. In addition, it highlights the lack of personnel oversight as a means of encouraging foundation staff to focus on R&D evaluations.
Role of Boards in Planning and Evaluation

Each of the foundations that were interviewed has a Board of Advisors or Board of Trustees (henceforth referred to as the board) that helps provide overall leadership and approval for science programs. In the foundations interviewed, boards play a large role in approving new programs but a small role in evaluating the outcomes of existing programs. When planning new programs, boards can identify new priorities based on the original foundation mission or founders’ expressed interests. In addition, boards often play a key role in formally approving new programs and associated budgets after being provided white papers by program officers that outline the rationale, focus, and value of new programs. Interviewees described board members as being strongly engaged in determining the focus and size of new programs. However, interviewees also acknowledged that once a program has been approved, boards provide little oversight on the evaluations that are conducted or the small changes that are made based on evaluation results. Rather, interviewees noted that the board trusts the president and program staff to monitor progress and make the necessary changes to ensure a program is successful. Some foundations require programmatic changes to be approved by the board but program officers indicated that the board often trusts staff to make the appropriate changes and therefore does not closely scrutinize these changes. Several evaluation staff that were interviewed stated that they desired stronger board interest in evaluation but had not identified a best practice for engaging board members.

Activating Energy of Presidents

Foundation presidents play key roles in designing, approving, and monitoring R&D programs according to interviewees. The foundations that were interviewed operate as fairly flat organizations. As such, presidents are often engaging directly with program officers to shape R&D programs. For example, when designing new R&D programs, program officers described iteratively developing white papers with presidents before presenting new ideas to the board. In addition, new foundation presidents often develop new or special initiatives. Special initiatives often do not fit within the broader categories of funding programs and therefore are directly supervised by the president. Several foundations explained that presidents are empowered to develop new or special programs without board approval if they are exploratory or under a specified funding amount. Therefore, presidents play a large role in approving new programs. Finally, presidents often require program officers and senior science leaders to provide yearly updates on the progress of science programs. Through formal memos and evaluations, presidents review the statuses of all of their programs to provide information to the board and to identify potential issues that need to be addressed. Depending on the foundation, presidents may be empowered to approve changes to the sub-goals of R&D programs if needed or approve programmatic changes to increase the likelihood that programs will have the intended impact.
In addition, foundation presidents have been the activating energy behind increased efforts to strengthen the capacity for R&D evaluation within foundations. Several foundations that recently created evaluation units credited new foundation presidents with initiating and supporting these efforts. For example, one foundation president hired a consultant to write a report on the best ways to structure evaluation approaches to ensure that foundations conduct timely evaluations to inform strategic decisions. This led to a new centralized evaluation service being launched and the recruitment of a well-known R&D evaluator from a smaller foundation to lead the effort. Interviewees indicated that presidents shape foundation agendas and priorities. In addition, presidents often approve yearly administrative budgets for activities such as evaluations. Therefore, presidents act as the activating energy when new evaluation efforts are launched since they can simultaneously provide leadership support and funding to ensure success.

Accountability and Personnel Performance Evaluations

Program officers were also asked to describe how they are held accountable for the successes of their R&D programs and how personnel performance evaluations are used for this purpose. All program officers that were interviewed noted that their personnel performance evaluations did not incorporate questions or assessments of the success of their R&D programs. Therefore, the performance of a program officer is not directly tied to the performance of the program or programs they administer. Rather, personnel performance evaluations focus on assessing how well program officers achieve the process goals that they define in previous years. For example, one interviewee noted that each year she has to document a series of personal process goals such as increasing her presence at conferences to engage more scientists or learning how to use social media to better advertise the foundation’s work. Personnel assessments then assess the extent to which employees met these process goals through a combination of a self-assessment and a supervisor-assessment. Program officers described varying levels of satisfaction with these personnel evaluations. However, they noted that personnel performance is not driven by personnel evaluations. Rather, foundations’ cultures of excellence and the innate drive within program officers help to ensure that R&D programs are successful. One program officer remarked that their organization is fairly small; therefore it is easy to observe colleagues’ performance without the need for formalized personnel evaluation assessments. In addition, two foundations that were interviewed hire program officers into time-limited positions. The defined length of time helps to fuel program officers to work quickly to make impacts and establish successful programs before their tenure ends.

Variety of Evaluation Activities Serve Several Purposes

Foundations undertake a myriad of evaluation activities to serve a variety of purposes. This section briefly describes the stated purposes of evaluation activities and then further describes the specific types of evaluations that foundations currently conduct.
Purposes of Evaluation Activities

The foundations that were interviewed engage in evaluation activities for accountability, advocacy, continuous improvement, and strategy development purposes. While some purposes lead to specific types of evaluations, foundations also employ a mix of types of evaluation activities for multiple purposes. The section below defines the variety of purposes of evaluation to highlight the different aims and intended outcomes that can be sought when engaging in evaluation activities. This section ends with a table demonstrating the intersections between purposes of evaluation and types of evaluation. The section afterward then describes the variety of types of evaluation activities to highlight the value of different types of evaluation and the different questions they can answer.

**Accountability-focused evaluations** aim to demonstrate that the R&D funds were used to carry out the mission of the foundation or to justify that investments were spent wisely. These evaluations may focus more on monitoring activities such as tracking financial expenditures and documenting grantee activities. Foundations who use evaluations for accountability purposes described these efforts as necessary for documenting the foundation’s financial responsibility to the board or outside funders. These evaluations are typically conducted at regular intervals as defined by foundation policy or practice.

**Advocacy-focused evaluations** aim to demonstrate the initial successes of a program or identify changes in the funding landscape as a means of arguing for additional funding. Program officers described using evaluations as a means of advocating for funding to either continue their program or to increase the size of their programs to award more grants. In some foundations, conducting evaluations for advocacy purposes is a standard tactic. In these foundations, program officers routinely engage outside, independent evaluators a year before they go before the foundation president or board to argue for continued or additional funding.

Interviewees described two types of evaluations that could be used for advocacy purposes. Process evaluations can be used to identify initial outputs, outcomes, and impacts of the program to justify its success. These evaluations also often focus on programmatic changes that could strengthen the success of the program. Alternatively, program officers also described using “landscape evaluations” to justify the need for further or greater funding. Landscape evaluations look at the foundation’s efforts within a broader ecosystem of funders and within a broader context of the problem they are trying to solve to identify where the foundation may best add value. For example, a program officer funded a landscape evaluation to support his argument that the overall funding landscape was not sufficiently addressing a problem that was growing in importance. This evaluation helped support the program officer’s request to the board for increasing the size of the program and increasing the duration of the program to try to impact this important but underfunded space.
**Continuous improvement-focused evaluations** aim to identify programmatic changes to improve an individual R&D program or set of related programs. The majority of interviewees described using evaluations for continuous improvement efforts. Process or developmental evaluations\(^{24}\) are explicitly designed for continuous improvement so interviewees’ discussions of continuous improvement were often combined with discussing process or developmental evaluations. These types of evaluations were described as being conducted on an ad hoc basis at the program officers’ discretion or built into standard program lifecycles to be conducted every few years or at major milestones. These types of evaluations were most often described as focusing on individual R&D programs. However, a few interviewees described multi-program evaluations for continuous improvement. These evaluations combine several programs with similar goals to explore how to better balance funding across the set of programs or to identify ways to better leverage one program to strengthen others.

**Strategy development-focused evaluations** aim to evaluate past successes and challenges within an R&D program or multiple programs to inform program planning and strategy to improve the success of future programs. These evaluations that are focused on single programs can be conducted as process or summative evaluations depending on the lifecycle stage of the program. Interviewees described multi-program strategy development-focused evaluations as “cluster evaluations” which are further described in a section below. These evaluations are relatively new to foundations who aim to better connect evaluation, strategy, and learning efforts. These evaluations were described as ad hoc efforts that were initiated by evaluation-focused staff within foundations.

Foundations use evaluation activities for a variety of purposes. Table 2 below provides an overview of the intersection between the purposes of evaluation and the types of evaluation activities that are used for these purposes. As the table illustrates, a single type of evaluation can be used for multiple purposes. The following sections describe each *type* of evaluation to provide further information on value of these evaluations and types of questions they answer.

\(^{24}\) These terms are often used interchangeably.
<table>
<thead>
<tr>
<th>Purposes of Evaluation</th>
<th>Landscape</th>
<th>Formative, Process and Developmental</th>
<th>Summative</th>
<th>Cluster</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accountability</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Advocacy</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Continuous Improvement</td>
<td></td>
<td>✓</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Strategy Development</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

Landscape Evaluations for Advocacy and Continuous Improvement

Interviewees described landscape evaluations as broad assessments of fields of science and related funding sources for R&D. These types of evaluations are used for planning purposes before an R&D program is initiated and to identify potential changes that are needed to existing R&D programs to address shifts in the science or funding landscape.

When planning new R&D programs, landscape evaluations are used to identify where the foundation can best support science and how it can uniquely add value to the R&D landscape. In looking for fields of science to fund, program officers described trying to identify nascent fields of science that needed support to develop and looking for underfunded fields that could address societal concerns. These explorations are driven in part by the mission of the foundation and the specific interests of the program officer. Some foundations have a mission to seed new fields of science by funding grants that aim to explore the intersection of science fields in a new way or apply existing technology to new problems. For these foundations, landscape evaluations are pivotal in determining where to start new R&D programs. Other foundations focus on improving specific aspects of society, such as improving child health, and use landscape evaluations to identify underfunded areas of science that could impact that societal goal.

In addition to identifying areas of science to invest in, landscape evaluations can also focus on identifying how to design the R&D program to leverage the foundation’s unique assets and approaches. Program officers at several foundations noted that their funding for R&D is relatively small compared to Federal funding or overall funding by foundations. Therefore, they must be strategic in what they fund and how they fund it to have an impact. Landscape evaluations can inform foundations on who is funding R&D within a specific field and the funding mechanisms they are using to help identify where the foundation can best support the field. For example, a program officer noted that other foundations and the Federal Government were funding academic centers of excellence. So, the foundation decided to fund researchers
who were unaffiliated with these centers at different universities and try to connect them with the leading centers. Through funding these individuals, the program officer hoped to widen the network of researchers working in this area and develop a hub and spoke model for existing university centers. Landscape evaluations can provide information to make these types of strategic planning decisions.

Landscape evaluations can also be used to identify necessary changes to existing programs when the funding or issue landscape changes. For areas of science that are rapidly evolving such as cybersecurity, landscape evaluations can help document shifts in the field that require the foundation to change its strategies. For example, a foundation conducted an ad hoc landscape evaluation for their cybersecurity program after the recent hacks to Equifax and other sensitive systems. This evaluation identified the need for greater funding to the program to address growing issues in cybersecurity and the need to narrow their program focus to target the key issues in the field today.

Landscape evaluations often require quantitative and qualitative methods. Quantitative methods may be used to track funding over time by multiple funders or quantify the scope of the challenges that the program will address. Qualitative methods may be used to capture expert opinions on the future of the field and where the foundation may best invest in R&D. These evaluations can be conducted internally by program officers and evaluation staff or can be contracted out to third-party evaluators. Internally conducted landscape evaluations are often built into the program planning process and result in a white paper or program proposal. Externally conducted landscape evaluations can be used to provide objective, third-party evidence to foundation leadership or the board. Foundations use landscape evaluations as a standard part of the program lifecycle and use them in an ad hoc fashion as they need to capture the shifting landscape.

Formative, Process, and Developmental Evaluations for Multiple Purposes

Interviewees described formative, process, and developmental evaluations as all similarly focused on improving existing programs through examining a combination of grantees activities and initial outputs along with foundation processes and procedures. While formal definitions within parts of the evaluation community distinguish these types of evaluations, interviewees used each of these terms to describe their efforts to improve existing programs at different points in programs’ lifecycles. Therefore, this section discusses these terms as one type of evaluation. These evaluations aim to identify small programmatic changes that could improve foundation operations or grantee activities to better achieve the ultimate goals of the program.

25 See Dehar, Casswell, and Duignan (1993) for a summary of the evaluation community’s disagreements on the differences between process and formative evaluations.
Interviewees described conducting these types of evaluations after the first year of a new program, at periodic intervals throughout the program’s lifecycle, and on an ad hoc basis as the program officer or leadership deem them necessary. These evaluations can involve qualitative and quantitative data collection and be conducted internally or externally by third-party contractors. Interviewees most often stated that these evaluations were designed to be directly used by the program officer to make programmatic changes. In a few foundations, evaluation results are also used to provide updates to foundation presidents on the progress of programs.

Evaluations that focus on improving internal foundation operations were described as qualitative examinations of how to improve the application and selection processes for R&D programs to better achieve the original aims of the program. These evaluations may explore whether changing to open solicitations from closed solicitations would result in more innovative proposals being submitted or whether changing application requirements would improve proposal submissions. Some foundations use third-party surveys to improve application processes. One foundation described an independent evaluation effort that included interviewing applicants who did not submit a completed application in order to identify challenges in the application process.

Evaluations that focus on grantee activities and initial outputs were described as mixed-method approaches that aim to adjust grantee requirements and foci to better achieve the original goals of the program. These efforts may compare the implementation of the program to its original goals to identify areas where changes to existing or future grant awards could improve the likelihood of program success. For example, a program officer that was interviewed used the results of a process evaluation to narrow the set of projects he was funding to boost funding for a select group of grantees that were seeing successes. The process evaluation helped him to recognize that he was funding too many awards in too many areas that would not help him achieve his ultimate goals. He used the evaluation results to justify programmatic changes to the board and was approved. These evaluations are typically undertaken by external, third-party contractors who are hired for individual evaluations.

A foundation interviewed also conducts internal process evaluations that they refer to as internal strategy reviews. These reviews are designed to assess the progress of the initiative, identify challenges, and contemplate adjustments to the program goals and timeline based on experiences so far. These ad hoc reviews are conducted by program officers and their staff in consultation with evaluation staff. Interviewees acknowledged that these reviews are less rigorous than external evaluations but help to reflect on successes and challenges of implementing the program as originally planned. These reviews may result in changes to process goals that were initially developed at the start of the program. For example, the program may have established at the beginning of the program 10 process or interim goals for a program that they believed were indicators of program success towards an ultimate goal. These process goals were made using logical assumptions about how an R&D program would progress. However, an
internal strategy review allows program officers to reflect on these goals to identify whether their assumptions held up and whether these process goals are still relevant for measuring the progress towards the ultimate goal. This may result in dropping a few of the process goals, extending the timeline for expecting these goals to be achieved, or shifting goals based on program experiences so far. Internal strategy reviews may also result in changes to the program structure such as shifting the emphasis on the types of grants that are given. While the program staff conducts these reviews, any programmatic changes must be approved by foundation leadership and the board. However, an interviewee noted that the board does not generally question these changes as they have faith in the science program staff.

Summative Evaluations for Advocacy and Strategic Planning

Interviewees also described conducting summative evaluations to assess the outcomes and impacts that result from funded R&D. These evaluations are often conducted at the end of a program’s initial lifecycle to either advocate for extending the program or to document lessons learned that can apply to future related programs. All foundations that discussed summative evaluations indicated that they are conducted by external, third-party contractors. These evaluations often use a mix of qualitative and quantitative methods to gather data from multiple sources to holistically capture the outcomes and impacts of the R&D program. A program officer that was interviewed noted that many summative evaluations are conducted too early given the long timeframe between initial investments and expected outcomes. However, he acknowledged that there may be little interest by foundations to conduct evaluations 10 to 15 years after the end of the program as the original program officers will likely have moved on and the foundation may not be funding similar research anymore. Therefore, summative evaluations often capture outcomes and impacts at a point in time but do not capture the full results of an R&D program.

Interviewees explained that summative evaluations can be used to advocate to the board that a program has been successful enough that it should be extended past its original end date. By demonstrating positive impacts of a program, program officers can argue that limited foundation funds have been well spent and extending the program would only lead to more impactful work. In some cases, foundation staff must argue for extending a program by comparing its realized impacts versus the potential impacts of a newly proposed program. Interviewees described this as a difficult process since proposed programs provide aspirational impacts while existing programs must demonstrate real impacts. However, objective, independent summative evaluations help to strengthen program officers’ advocacy arguments.

Foundations also use summative evaluations to document the achievements of the programs, the facilitators of these achievements, and challenges that limited successes. Some foundations have set durations for programs (e.g., 10 years) and do not allow for renewals or extensions. Therefore, summative evaluations are not needed for advocacy purposes. Rather, summative evaluations are used to wrap up a program and provide insights for future programs. If used by
other program staff, these evaluations can help to identify potential pitfalls to avoid and strategic mechanisms that helped to ensure program success.

Cluster Evaluations for Continuous Improvement and Strategy Development

Finally, two foundations that were interviewed described a new type of evaluation that is designed to inform strategy. Cluster evaluations, or bundled evaluations, identify the outputs, outcomes, and impacts of a set of grant programs that have shared characteristics. These new evaluations aim to inform strategic decisions at the portfolio level such as whether to increase or decrease a program, how to leverage one program to support another, or how to change individual program requirements to improve the likelihood of program success. Rather than relying on expert opinion from program officers or advisory groups, cluster evaluations are systematic examinations of the outcomes and impacts of a set of programs with related goals or approaches. These evaluations may also examine the interactions between programs or grantees funded within programs to identify ways to better leverage the investments from one program to strengthen another program.

The concept and unique nature of cluster evaluations are potentially best explained with an example. Suppose a foundation aims to decrease the morbidities that result from poor diabetes care management. To affect this change, the foundation invests in three programs: A) studies that identify promising practices within health insurance companies to incentivize patients to manage their disease, B) new educational campaigns for physicians to enable them to provide better counseling to diabetes patients, and C) app development that explores the impact of automated tips and reminders to help diabetes patients manage their disease. A cluster evaluation may examine the extent that each program is leading to the intended impact and determine changes that need to be made to each program and to the overall strategy. For example, the cluster evaluation could identify that apps only work when physicians help patients set up the technology. Therefore, grantees in program B could be directed to consult with grantees in program C to identify opportunities for collaboration. Alternatively, the cluster evaluation could identify that program A has quickly discovered best practices and therefore funding can be diverted to programs B and C that need further support. The cluster evaluation is designed to help inform the overall strategy for the foundation as it aims to impact a specific area of science or society.

The two foundations who discussed cluster evaluations were in the planning stages of their first cluster evaluation at the time I was interviewing them. Both foundations started designing and planning these evaluations in early 2017 through work by dedicated evaluation units. These evaluations were being designed internally but would be carried out through independent, third-party contractors. Like process and summative evaluations, cluster evaluations will use a mix of qualitative and quantitative methods. These evaluations should result in a series of changes at the individual program and portfolio level to better achieve the overall aims of the foundation.
Growing Emphasis on Learning

In addition to evaluation activities, many of the foundations that were interviewed have begun to invest in “learning” activities and personnel. These foundations explicitly separate learning from evaluation to emphasize that it is a new set of efforts to connect the dots between evaluation and strategy. Rather than simply hope that evaluation results are used to improve programs, learning activities are designed to share evaluation results internally and explore how other programs across the foundation can leverage results from another program to improve their own programs. Learning activities may result in individual program changes or in foundation-level policy and practice changes such as changes in types of grantees that are funded, how programs interact with grantees, and the mix of types of research that are funded (fundamental research, applied research, capacity building, etc.). In addition to programmatic changes, foundations are using learning activities to help improve support services such as grants management and finance.

In addition to using evaluation results for learning, some foundations are convening internal working groups to identify potential learning activities. One foundation described an internal process of developing activities for “learning days” where all foundation staff have to be in the office and participate. To prepare for this yearly event, the learning officer works with an internal group of foundation staff from different programs to identify potential lessons that could benefit the entire foundation. These lessons could come from evaluations or from working group suggestions. The learning officer then develops training or discussion sessions for all foundation staff that aim to change the way the foundation functions to improve business practices across the foundation. Since all staff must attend these activities, including program and support services, the learning officer aims to develop sessions that are relevant to everyone. Therefore, learning is not just focused on improving science programs but has a broader purpose.

Several foundations indicated that they have only devoted personnel and resources explicitly to learning activities in the last two years and are therefore still defining what learning entails and how to best encourage learning across the foundation. However, several interviewees noted that developing a culture of learning was important in addition to the learning activities that are provided. Evaluation staff within foundations were very interested in learning from other foundations and experimenting with the best ways to conduct evaluations. By emphasizing “learning,” these staff appear to now be aiming to spread that innate interest in continuous improvement to the rest of the foundation staff. No foundations that were interviewed claimed to have best practices for developing this culture or conducting learning activities but the foundations that have expanded their evaluation units to include learning believe it is important as strategic grantmaking foundations.

26 See Senge (1990) for original discussion of learning organizations and activities.
Funding for Evaluations

The foundations that were interviewed fund evaluation through three different sources—centralized funds, cost sharing between evaluation and program funds, or ad hoc program funding. Additionally, foundations use three different means for determining how much money to devote to evaluation activities. The following sections outline the three different sources of funding for evaluation and the differing means for determining funding amounts.

Types of Evaluation Funding: Centralized, Cost Sharing, and Ad Hoc

Centralized, dedicated funds for evaluations have been set up by some foundations in the last few years. These funds are either centrally managed as a single pot of money for evaluating all programs or are distributed to individual programs to carry out their own evaluation activities. Centrally managed funds are typically overseen by evaluation staff within an independent evaluation unit that is separate from science programs. Distributed funds are provided to science program officers as part of their yearly budget process and must be used on evaluation activities. In one foundation with this distributed model, program officers also have an administrative budget in addition to a science program budget that can be used for additional evaluation activities such as hiring consultants to help set up evaluation efforts. Since these funds are centrally allocated, one program officer indicated that he did not view the administrative budget or the evaluation budget as a “tax” on its program budget but rather an additional set of resources to carry out his program.

Cost-sharing mechanisms between evaluation units and science program units have also been set up by some foundations to fund evaluations. These foundations have dedicated, central funds for evaluation that are managed by evaluation staff. However, the foundation hopes to fund more evaluations than can be conducted using these funds alone. By partnering with science programs, they can leverage program funds to conduct a great number evaluations and larger-sized evaluations. In addition, evaluation staff that were interviewed that use this mechanism noted that cost sharing encourages science program officers to “buy into” the evaluation and use the results.

Ad hoc funding is also used by some foundations to fund evaluations. Rather than develop a dedicated evaluation fund, these foundations pull from individual science program funds to conduct evaluations. A program officer whose foundation uses this source of funding for evaluations noted that evaluations are not needed every year and the amount of money needed for evaluation changes depending on the type of evaluation and the stage of the R&D program.
Funding Amounts: Set-Asides, Aspirational Goals and Tailored Funding

Foundations also use several different means for determining how much money to spend on evaluation activities including establishing set asides, setting aspirational goals, and tailoring funding based on the type of evaluations. Some foundations have established explicit set asides amounts for evaluations. For example, one foundation interviewed devotes 1% of all program funds for evaluations. Another foundation interviewed devotes 2% of program funds for evaluation while a third foundation that was interviewed sets aside specific amounts by individual program. These foundations indicated that they highly value the role of evaluation and typically manage their evaluation budget through centralized funds. Other foundations set aspirational goals for how much money the foundation overall should spend on evaluation activities. For example, a foundation that was interviewed described having a goal of spending 10% of program funds on evaluation each year. Rather than set aside 10% of funding, this foundation encourages science programs to fund their own evaluations or to cost-share with their evaluation unit. The overall 10% goal does not require each program to conduct evaluations every year but aims to encourage science programs to devote significant resources when appropriate to understanding how to improve and to document the outcomes and impacts of their R&D programs. Finally, other foundations that were interviewed vary the amount of funding for evaluation based on the types of evaluations they expect to conduct in that year. So, a centralized fund or ad hoc program funds for evaluation could shift dramatically from year to year depending on program needs.

Limitations of the Socio-Technical Framework for Describing Foundations

While the socio-technical theoretical framework helped to guide interviews with foundation staff on the myriad of components and characteristics that define their approaches to R&D evaluation, it does not ultimately represent the central features of foundations’ R&D evaluation approaches. The following section describes the limitations of this framework, reproduced below in Figure 8, and the failed assumptions that were made in adapting this framework.
Lack of Data Collection Tools

In selecting the socio-technical framework, I believed that technology would play a large role in shaping how foundations are able to conduct R&D evaluations. After reviewing foundation websites, survey results from nonprofits who study foundations, and academic papers on foundation evaluation efforts, it was clear that foundations were investing significantly more resources than the Federal Government, by percentage of funding, to evaluation activities. I assumed the personnel and financial resources devoted to evaluation within foundations would have led to investments in common data collection tools that collect standardized information on the outcomes and impacts of R&D programs. With dedicated evaluation staff, foundations could consider how to better align individual evaluations across multiple programs. In addition, the financial resources for evaluation could have made it possible to fund common data collection to reduce overall spending on independent evaluations that require extensive data collection. If foundations are continuously funding evaluation efforts, it was logical to assume that these were not treated as independent, unrelated activities. However, interviews with foundations indicated that this assumption was incorrect.

None of the foundations that were interviewed have invested in organization-wide data collection tools to track outcomes and impacts of investments. Foundations have invested in developing or buying grant management systems that include narrative reports from grantees on the activities and outputs of the grant but these systems do not track outcomes and impacts.
Rather, outcome and impact data are collected de novo for each evaluation that is conducted. Through interviews with evaluation staff, three reasons for the lack of common data collection tool development were identified: program officer autonomy, emphasis on third-party evaluators, and limited coordination across evaluations.

Across evaluation staff, program officers, and senior leaders, interviewees described the relative power of individual program officers to independently decide how to shape their programs and subsequently their individual program evaluations. Program officers are viewed as experts in the field who are trusted to design, implement, oversee, and evaluate R&D programs with little oversight after initial program approval. In several foundations that use evaluation consultant or program officer as evaluator models, program officers are ultimately in charge of the design and execution of R&D evaluations. Evaluation staff indicated that this leads to program officers creating highly tailored evaluations that collect very nuanced data that relate only to their R&D program. In addition, these data are often collected in different ways across programs. One program may collect outcome data through structured surveys while another may use structured interviews to collect deeper information. Program officer autonomy to determine which data elements to collect and how to collect them leads to large variations across programs. Therefore, there is little value in standardizing data collection as program officers would likely continue to ask for nuanced information for their specific programs.

In addition, all interviewees expressed a strong preference for independent, objective, third-party evaluations. One interviewee stated that evaluation by default meant third-party, independent assessments. Therefore, each evaluation is carried out through individualized requests for proposals (RFPs). As an evaluator, there is no incentive to develop common data collection tools across programs when you are funded just to assess a single program. Even when the same evaluator conducts multiple evaluations for a single foundation, there is no incentive to develop a common tool since there is no funding for these efforts and additional evaluation contracts cannot be guaranteed. In fact, contractors get paid to conduct the same manual data collection for different programs so by developing a common tool, they may actually be limiting the fees they are able to charge the foundation. Without incentives or funding, there is no reason to expect third-party contractors to invest the resources or time to develop common data collection tools, even when they continuously collect the same data for different programs.

Finally, interviewees indicated that little coordination occurs between evaluations as they are being developed. Evaluation staff do not look at common data being collected across proposed evaluations and do not coordinate when using the same evaluator. Rather, each RFP is treated as a separate contract. This lack of coordination inhibits the ability for evaluation staff and program officers to recognize overlapping data collection efforts that could be combined to reduce time and funding requirements.
Fundamentally, a socio-technical system assumes that technology helps drive tasks that are conducted by personnel within a particular organizational system. Therefore, personnel, organizational structure, and evaluation activities would all influence and be influenced by the data collection tools (the technology) that a foundation has developed. Unfortunately, foundations have not invested in common data collection tools. So, there were no policy implementation issues identified to describe how the organizational structure of the foundation influences the tools that are developed. Conversely, there were no incentives identified for personnel to use the data collection tools they developed as shared solutions since none exist. Therefore, the socio-technical framework as originally proposed does not accurately capture the overview of foundations’ approaches to R&D evaluation. The revised diagram in Figure 9 highlights how eliminating the data collection tools ultimately breaks down the value of the diagram. As is discussed earlier in the chapter, the social system of personnel and where they are placed within the foundation ultimately drives how foundations approach R&D evaluation. This diagram helped in exploring foundation practices but the lack of data collection tools limits its utility to display how the components and characteristics of R&D evaluation approaches interact.

**Figure 9. Revised Theoretical Model**

![Revised Theoretical Model Diagram]

**Limited Policy Underpinnings**

In adapting the socio-technical framework, I added in policy as an underlying feature that I assumed drives actions and changes within all components and characteristics of R&D evaluation approaches. As a public policy researcher, I aim to identify how rules and regulations
can best guide intended actions and behavior. Therefore, I was interested in understanding the policies that govern the organizational structure, personnel, data collection tools, and evaluation activities within foundations. However, policy is not a primary driver of change or growth in foundations’ approaches to R&D evaluation.

Foundations are inherently small organizations with limited-sized staffs. In most foundations that were interviewed, there was only one level of personnel between science program officers and the president of the foundation. In addition, interviewees described several of these foundations as fairly flat organizations with large amounts of program officer autonomy. In interviews, evaluation staff and program officers described the approval process for new activities (short of starting new R&D programs) as simple conversations with the foundation president. Therefore, changes to foundation practices can occur rapidly without multiple levels of approval or the need for formal policy documents. In addition, there is limited bureaucracy to formally create policy.

Rather than drive change or development of R&D evaluation capacity through policy, foundations have used evaluation-focused personnel to shape foundation approaches. As is discussed earlier in this chapter, evaluation personnel work closely with program officers to design, implement, and use evaluations. Through personnel interactions, funding support, and leadership support, evaluation personnel have been able to increase the complexity and use of evaluations to drive strategic change.

While the lack of policies underpinning R&D evaluation approaches was common among most foundations, one foundation that was interviewed uses a formal document to guide evaluation. This foundation was among the larger foundations that were interviewed and has a large, independent evaluation unit. In interviews with both evaluation staff and program officers, interviewees highlighted the importance of this evaluation guide for dictating how programs carry out their evaluation activities. According to the program officer who was interviewed, this guide is still followed closely within the set of R&D programs he was a part of despite the guide being written several years ago. It is not clear why this policy document was viewed as pivotal to the foundation’s approach to R&D evaluation and whether this document ensures greater focus on evaluation than personnel efforts by other foundations.

**Limited Insights on R&D Evaluation from Firms**

Exploratory interviews were conducted with a small sample of U.S.-based firms that have dedicated R&D units to identify organizations that conduct R&D program evaluations that can inform strategic decisions and future program planning. However, these interviews did not identify any candidate organizations for further in-depth interviews. Rather than conduct retrospective analyses of the outcomes and impacts of their R&D programs, these organizations have embedded evaluation into standard reviews to assess whether to continue funding a project.
Six of the seven firms that were interviewed used some form of stage-gate process to periodically assess whether R&D projects should continue or be terminated. Stage-gates refer to specific project milestones that all projects are expected to reach as they develop. Some firms define stage-gates as specific time periods between reviews while others have goal-oriented milestones such as number of experiments fielded. This stage-gate review process requires senior managers to assess the initial successes of the project, its potential value to the company, and its likelihood of continual success. These assessments could change dramatically from review to review as the R&D matures, the industry landscape changes, and firm resources change. These reviews result in a binary decision to continue funding or terminate the R&D project and re-allocate the resources to other efforts. Since these firms are conducting R&D with internal personnel, projects can easily be eliminated and resources redistributed. Stage-gate reviews can also identify changes to personnel or methods or focus of the project to improve its likelihood of success for projects that receive continued funding. These reviews are a type of individual project evaluation but are focused on continuous improvement at the project level and do not look at overall R&D programs.

Among the seven diverse firms that were interviewed that come from different industry sectors, only one firm conducted retrospective analyses to learn from prior R&D efforts. This firm, who provides specialty R&D services to other firms who do not have the capacity or expertise to carry out specific projects, takes a simple approach to evaluation and learning. After each project is terminated due to success or failure, the project leaders write up a brief synopsis of the project and potential lessons learned for future projects. These write-ups are aggregated together into a single document to allow the R&D program manager to easily search through all reports to conduct a meta-analysis to identify broader lessons that can be applied to all projects. For example, this firm discovered that projects that are given large amounts of initial funding do not work well. Rather, funding small amounts of work with continual review results in better outcomes. This led to changes in the stage-gate review process to require all projects to be reviewed every two weeks. The R&D manager who utilizes this method strongly believed that organizations should only collect as much information as they are able to use. He believes that any further data collection beyond these simple project write-ups would not be valuable since he does not have the resources to analyze additional data.

Given the limited amount of R&D evaluation conducted within firms that were recruited for this dissertation, I decided to drop the research questions associated with studying firms. The literature review revealed that there was potential for the Federal Government to learn from firm practices. However, the insurmountable recruitment issues, further discussed in Chapter 4, and the lack of valuable information from the firms recruited limited the value of conducting in-depth interviews with these firms.
Chapter 3: Workshop Findings

Key Findings from Chapter 3:

- Adapting a “centralized evaluation service model” from foundations with evaluation-trained personnel and dedicated budgets could help strengthen the capacity for the Federal Government to conduct R&D evaluations that inform future program planning.
- Significant financial and human capital resources would be needed to conduct cluster evaluations that require systematic assessments of multiple programs with similar characteristics across several agencies.
- There are multiple opportunities for cross-agency collaboration to share data and conduct evaluations.

To understand which foundation approaches could be adapted to the Federal Government, I convened a group of Federal science policy, planning, and evaluation officials from across the major science agencies. This dissertation aims to identify collective Federal solutions to Federal-wide problems. Therefore, I pulled together a diverse mix of senior leaders to collaboratively explore which approaches from foundations could broadly apply to science agencies. This chapter outlines the findings from this convening to identify a path forward for potentially strengthening the capacity for the Federal Government to conduct R&D evaluations that can inform future program planning.

On November 16, 2017, I facilitated a one-day workshop in Washington, D.C. The workshop aimed to 1) Design research plans for conducting cluster evaluations of multiple programs with similar characteristics that could inform agencies’ strategies; 2) Identify the strengths and challenges of different models for placing evaluation-focused employees within the Federal Government; and 3) Develop a preliminary research agenda that would enable science agencies to better plan future investments using the results of prior R&D investments. To accomplish these goals, I developed three inter-related group sessions that enabled workshop participants to share their experiences with evaluation, collaboratively explore new ideas, and brainstorm collective solutions.
Each session addressed one of the aims. In Session 1, participants discussed ways to design evaluations that cut across programs and potentially across agencies to inform strategy development. In Session 2, participants debated the relative merits of different models for placing new evaluation-focused staff within the Federal Government. In Session 3, participants developed a prioritized list of research questions that could help to inform the future design of R&D programs. The method and process for each session is described in further detail in the Introduction.

Nineteen individuals participated in the workshop. The majority of participants were policy, planning, and evaluation leaders from across the major Federal science agencies. The majority of these participants were being paid at General Schedule (GS) 14 or GS 15 levels. These individuals were asked to give their personal opinions from their years of experience overseeing and conducting R&D evaluations but they were not asked to act as representatives for their agencies. To ensure open and honest discussions among participants, I informed them that their names would not be publicly shared and their comments would not be attributed to their home agency. The following chapter does not identify individual participant concerns but discusses overall conclusions from group discussions and survey findings. Participants came from the following agencies:

- Advanced Research Projects Agency-Energy (ARPA-E)
- Department of Defense Office of the Under Secretary of Defense for Personnel and Readiness (DOD)
- Department of Energy (DOE)
- Department of Health and Human Services (HHS)
- Department of Homeland Security- Science and Technology (DHS S&T)
- Millennium Challenge Corporation (MCC)
- National Aeronautics and Space Administration (NASA)
- National Institutes of Health (NIH)
- National Institute of Standards and Technology (NIST)
- National Science Foundation (NSF)
- Naval Research Laboratory (NRL)
- Office of Management and Budget (OMB)
- Office of Naval Research (ONR)

---

27 One workshop participant missed the first session and another workshop participant declined to participate after the survey.

28 One individual was from the Millennium Challenge Corporation and two individuals were from the Office of Management and Budget. These individuals provided outside perspectives that aided in science-focused discussions.
In each workshop session, participants explored multiple aspects of foundation approaches. Therefore, the chapter below highlights participants’ insights by the themes identified in Chapter 2, rather than discussing findings session-by-session. Initially, I present the evaluation research questions that participants were interested in pursuing. These questions help provide context for understanding the issues that Federal science agencies face and help to frame the rationale behind some of the participants’ opinions that are discussed in later sections. Next, I describe participants’ views on the various resources and approaches that could be adapted from foundations to the Federal Government to execute evaluations to answer these questions. Participants’ views on third-party cluster evaluators are then discussed, followed by views on the financial resources necessary to conduct these evaluations. Finally, the majority of the chapter presents participants’ views on the personnel and supports around personnel that are needed to increase the Federal Government’s capacity for conducting R&D evaluations that inform strategy.

**Potential Evaluation Research Questions**

In Session 3 of the workshop, participants identified a series of potential evaluation questions that could be investigated by individual agencies or groups of agencies to strengthen the outcomes and impacts of future programs. These questions were initially framed as the beginnings of a “learning agenda” for science agencies. The U.S. Agency for International Development (USAID) and the U.S. Department of Housing and Urban Development have developed learning agendas that guide senior leadership in developing evidence to improve their performance. USAID describes its learning agenda as “short- and long-term priority research and policy questions relevant to the department’s mission” (Commission on Evidence-Based Policymaking 2017). The guiding focus of this session was, “What are the evaluation research questions that need to be answered in order to strengthen the outcomes and impacts of future research programs?” Participants were allowed to suggest questions that were specific to their agency or relevant across multiple agencies. After the workshop, I sorted these questions into three overall categories:

- **Agency operations questions** focused on how to improve the efficiency or effectiveness of agency operations, both within the agency and with external partners.
- **Assessment questions** focused on how to improve R&D evaluations and specific evaluations that participants are interested in conducting.
- **Program design questions** focused on improving the structure, operations, or focus of R&D programs. These questions are similar to assessment questions, yet have some explicit or implicit focus on using assessment results to change future R&D programs.

---

29 A limited-release workshop summary was also developed for workshop participants that highlights findings session by session.
Participants voted for the questions that could lead to answers that would have the greatest impact on R&D programs. The finalized, rank-ordered list of questions by category is provided below. Only questions that received at least one vote are included.

**Agency Operations Questions**

1. How do we accomplish better data sharing across agencies (DoD, CDC, HHS, etc.)?
2. How many levels of management are required to optimally perform the agency’s mission?
3. Are we “ready“ to engage in evaluation and learning at the agency level?
4. How can the Federal Government better prioritize investment in Science and Technology?
5. Have the outcomes of projects/programs been captured and shared with other projects/programs?
6. What actions do we need to start, stop, and keep in order to _________________?
7. What is the appropriate mix of support staff (%) to accomplish the [overall] mission of the agency?
8. What evaluation capabilities are needed at various levels of the agency to begin [new evaluation and learning activities]?

**Assessment Questions**

1. What are effective models for evaluating science, technology and innovation progress and impact?
2. How do we justify funding R&D (Innovation) when we may not see “results/advancements“ for a while?
3. Is R&D funding leading to innovations and technologies to advance the (specific) fields?
4. Identify best strategies for collecting outputs for research awards
5. How does the evidence produced by NIH and other researchers of HHS impact policies, programs and practice, and ultimately public health?
6. What outcome/data should we collect from grantees/recipients to demonstrate our “impact/results“?
7. Does the current intellectual property (IP) policy framework weaken or slow technology development from Federal R&D?
8. Best strategies for longitudinal evaluations? [increase responsiveness and avoid survey overload]
9. What percent of projects in a program’s portfolio transitioned to new products, programs, or new knowledge?

---

As stated in the methods section, participants were then given six stickers to vote on the questions that, if answered, could lead to the biggest improvements in R&D programs. Participants were allowed to vote for their own ideas and could use several “votes” for the same question.
**Program Design Questions**

1. Are we making investments that are “risky” enough?
2. How to solicit, identify, and fund innovation?
3. What are “best practice” R&D strategies (selection of projects, research integration, funding, etc.) that maximize R&D return on investment?
4. Is the recent focus on public-private partnerships strengthening or diluting the Federal R&D enterprise?
5. What are the barriers/facilitators for broadening participation and diversity in STEM enterprise?
6. How do we leverage “lessons learned” when the technology itself or innovation fails?
7. What key factors contribute to collaborations (academia, industry, Federal Government)?
8. What are the barriers/facilitators of research (STEM) competitiveness across academic institutions/states?
9. When does investing in single PI awards yield diminishing returns?

These questions highlight the value of looking for cross-agency solutions to increase the capacity for the Federal Government to conduct R&D evaluations that inform future program planning. The answers to most of these questions would provide actionable information to a wide range of science agencies. In addition, most questions would require data from multiple agencies. Throughout the workshop, participants indicated a strong interest in cross-agency collaboration. The question that received the most votes across all categories was, “How do we accomplish better data sharing across agencies (DoD, CDC, HHS, etc.)?” In Session 1, a participant commented that it was easier for her to assess whether a firm received private sector follow-on funding than to assess whether a firm received follow-on government funding after an initial Federal grant.

As a set, these questions represent a wide-ranging, bold agenda for helping government agencies develop and integrate R&D program evaluations. Each of these questions would require a mixed-methods approach to answer and most would require new data collection. This helps to motivate the need for government agencies to increase their capacity for conducting evaluations. Without additional resources, or new approaches for supporting evaluations, it is not clear that these evaluation questions will be answered. Throughout the workshop, participants explored the impacts of applying foundation approaches uniformly across science agencies in an effort to identify potential Federal-wide solutions to clearly pervasive issues. The rest of the chapter outlines participants’ views on adapting foundation approaches to R&D evaluation to provide resources and approaches to answering difficult, but solvable, questions to inform future strategies.
Cluster Evaluation Findings

In the first workshop session, participants were asked to develop research plans to conduct cluster evaluations\(^{31}\) for research programs that cut across multiple agencies. The term “cluster evaluation” was new to almost all participants but the general approach and focus of cluster evaluations were familiar to many participants. After a general discussion to develop a common understanding of a cluster evaluation, participants were separated into small groups to discuss cluster evaluations for specific topics that cut across agencies. I chose to use Science, Technology, Engineering, and Math (STEM) education, the Small Business Innovation Research (SBIR) program, and the Brain Research through Advancing Innovative Neurotechnologies (BRAIN) Initiative to ground the conversations. The following highlights the overall findings from the large group discussion and the small breakout group discussions.

Several participants provided examples of Federal evaluation activities that are akin to cluster evaluations. These participants did not believe the concept of cluster evaluation was new but had not previously had a specific term for these types of evaluations. One agency routinely asks advisory panels to review sets of related R&D programs to determine if they are using the right grant mechanisms and if the programs are providing what was intended when they were created. However, the participant who described this process noted that these are expert panels and they do not use quantitative measures to assess sets of programs. Another participant described launching a recent cluster evaluation for six related programs, some of which have been running for six years. The intended goal of the evaluation is to inform strategic planning in the future. While some participants were familiar with the concept of cluster evaluations, most had not conducted one.

Differences in agencies’ missions and operations may hinder the value of cross-agency evaluations. Cluster evaluations are inherently designed to inform strategy. However, participants noted that there is no overarching science strategy within the Federal Government. Rather, agencies take different approaches to achieving tailored missions. Therefore, the concept of a cluster evaluation may not be as applicable across agencies. Even in the SBIR group, where all agencies are theoretically administering a similar program, participants struggled to come up with a cluster evaluation research question that would help inform future program planning in multiple agencies. Participants noted that while each agency’s SBIR program similarly targets small businesses, the way agencies have integrated their own SBIR programs into their overall R&D programs varies greatly. For example, the R&D portfolio within one agency focuses heavily on funding small businesses. For that agency, the SBIR program is viewed as completely integrated with its other R&D programs. A participant from another agency expressed a

---

\(^{31}\) Cluster evaluations are defined as systematic assessments of the outcomes and impacts of multiple programs to identify potential strategic changes to one or more of these programs that share similar goals or characteristics.
completely opposite view. To him, the SBIR program does not align with his agency’s mission and therefore is seen as a waste of funds for a mismatched program. In his view, he could not imagine a SBIR cluster evaluation that would inform his agency’s strategy.

The methods and data used for cluster evaluations are similar to other evaluations already commissioned by Federal agencies. While cluster evaluations take a wider view and are focused on informing an agency’s strategic direction, the methods and data necessary to conduct these evaluations are not unique. Participants discussed the need for new data collection and better linkage of existing data but noted that these needs are common for all R&D evaluations. Similarly, participants discussed a myriad of common evaluation methods including network analysis, road mapping, and return-on-investment analyses.

Cross-agency cluster evaluations would require a central hub for coordination and data collection. Each of the groups discussed a need for greater collaboration and coordination across agencies to develop and undertake a cluster evaluation. This hub could be centrally run within the Federal Government or contracted out to a third party. However, each group believed a central group was needed to engage in pre-evaluation activities such as defining the scope of the cluster evaluation, sharing relevant data, and developing the Request for Proposals (RFP). The BRAIN Initiative group suggested that grantees from the National Science Foundation’s Science of Science and Innovation Policy (SciSIP) program could be a valuable resource in developing a hub. However, other participants disagreed with the utility of SciSIP researchers to support evaluations that directly inform government program planning.

Respondents indicated that third-party evaluators should be evaluation methods experts to provide the most actionable information for improving R&D programs. In the individual survey, I asked participants to select the type of third-party evaluators that their agency should employ to provide valuable information that can be used to improve R&D programs. Participants could select A) subject-matter experts who have deep knowledge of the science being conducted; B) evaluation experts that are methods experts without specific domain expertise; or C) advisory committee members that have an understanding of the science but are typically subject-matter experts in related scientific fields. No respondents selected “C”, advisory committee members as the preferred evaluators for agencies. The overwhelming majority of respondents selected “B”, agreeing that evaluation experts should be part of the third-party evaluation teams that support agencies. However, over a quarter of those respondents also selected “A”, that subject-matter experts should be included in evaluation teams. Another quarter of respondents believed that “A”, subject-matter experts alone, would provide the most actionable information to agencies on how to improve R&D programs.
Financial Resource Findings

Few science agencies have dedicated funds for conducting evaluations. In addition, science agencies devote relatively small amounts of funding to evaluation activities each year compared to foundations. Dedicated, centralized evaluation funds could help to ensure that more R&D evaluations are in fact conducted. Additional cost-sharing mechanisms with program offices may help ensure buy-in for these evaluations. Unfortunately, cluster evaluations that are designed to impact strategy require millions of dollars to execute and would require several years to design and implement. Significant additional resources would be needed to undertake this type of evaluation. The following section describes participants’ views on the financial resources needed to strengthen the capacity for Federal R&D evaluation.

Significant resources would be needed to conduct cluster evaluations. Participants spent a significant amount of time at the beginning of the small group discussion in Session 1 trying to descope their cluster evaluation to a manageable size. Despite this work, each small group concluded that cluster evaluations would require millions of dollars and years of effort to complete. In addition, participants noted that significant time would be required by agency staff, even if the evaluation was conducted by an independent, third-party contractor. As a very rough estimate, the BRAIN Initiative group thought the pre-evaluation activities would cost $2,000,000 and the full evaluation would cost $30,000,000. This estimate was based on funding needed for an intermediary group to manage the various evaluation activities, funding for convening agencies, grantees and evaluators, and money for grant making to conduct evaluation activities. This group noted that the BRAIN Initiative did not uniquely require more funding to evaluate; rather, all cluster evaluations of cross-agency initiatives would require significant funding. The SBIR and STEM Education groups agreed that cluster evaluations would require multimillion-dollar investments but struggled to come up with an estimate, in part because they had not agreed on a specific scope for the evaluation. One participant suggested that cluster evaluations should be designed into modules since multi-year funding cannot be ensured.

Most respondents believed that funding for evaluations should be centralized in a dedicated fund. In the individual survey, I asked participants how evaluation funding should be structured to ensure efficient and effective use. Participants chose between A) a centralized, dedicated fund; B) cost sharing between an evaluation fund and program funds; and C) ad hoc funding using program funds. The majority of respondents who chose a centralized fund indicated that this structure would ensure independence and that funds would not be re-allocated for other efforts. The quarter of respondents who chose cost sharing indicated that this structure would help to ensure that program officers buy into the evaluation and collaborate on its design. The two individuals who chose ad hoc funding indicated that flexibility was important.
Personnel Findings

Science agencies vary widely in the number of evaluation-focused personnel they employ and where they are placed within agencies. Some agencies have no dedicated staff but split evaluation responsibilities across numerous individuals. Other agencies have centralized staff that focus exclusively on R&D evaluations. In Session 2, participants were asked to identify the relative strengths and challenges of different models for placing new evaluation-focused employees within the Federal Government. Over all other efforts and changes, I found that hiring evaluation-focused personnel was the most common and concrete way that foundations sought to expand and improve their use of R&D evaluations. However, foundations are experimenting with different models for using evaluation personnel. In addition, agencies are immensely larger than any of the foundations I interviewed. Therefore, Session 2 was designed to explore how, and if, the foundation approaches to evaluation personnel could be adapted to the Federal Government.

Overall, participants believed that if they were able to hire new evaluation-focused employees, they should be located within a centralized, independent evaluation office to be most successful. Foundations have developed centralized evaluation services that place evaluation-trained staff within a separate office from science programs that reports directly up to the foundation president. Participants noted that in order for this office to impact program officers’ designs of R&D programs, they will have to find a way to reach out to program officers and engender trust. While a centralized approach was most highly endorsed, there are clear challenges with this model and it may not benefit all agencies. In addition, participants were skeptical that Congress would provide funding for new personnel. The following section outlines participants’ views on the relative strengths and weaknesses of each personnel model.

Centralized Evaluation Service

Before the small-group discussions of specific personnel models, participants filled out a survey on individual preferences for their agency. In the survey, most respondents indicated that a centralized evaluation service within agencies would be the most effective in increasing their agency’s capacity for conducting R&D evaluations. More than half the group selected a centralized evaluation service model. However, several respondents noted that having evaluation consultants in addition to centralized staff would be desirable. Only three respondents preferred additional program officers solely.

In the small group discussions, participants noted several strengths of this model including:

- Ability to coordinate activities across the agency

32 Foundations that were interviewed described centralized evaluation services, internal evaluation consultants, and program officers as evaluators models. Full descriptions of these models are provided in Chapter 2.
Opportunity to act as a knowledge broker to connect strategy and implementation efforts
Opportunity to share data, insights, and evaluation results across the agency
Independence in determining which programs to evaluate and how to evaluate

Participants also noted several challenges including:

- Success depends on the culture of the agency
- Personnel may not understand the programs they are evaluating
- Evaluators may not be seen as credible if they are not also subject matter experts
- Agencies whose missions do not solely focus on funding research (e.g., include other activities such as service delivery) may not benefit from a centralized model
- Determining the appropriate placement for this group to be effective (e.g., department, agency, or agency office level)

**Internal Evaluation “Consultants”**

Other foundations have hired internal evaluation “consultants” that act as resources embedded within science programs but report up to a single senior evaluation leader. These consultants advise program officers on the design of monitoring and evaluation activities (but do not oversee evaluations); develop common evaluation resources such as logic model templates; and participate in internal strategy reviews conducted within the science programs. Similar to the centralized model, evaluation consultants intentionally do not play a role in monitoring grantees. These personnel often have general interdisciplinary evaluation skills but are not content or evaluation method experts.

Participants noted several strengths to this model including:

- Consultants will generate buy-in from Program Officers to use evaluation results by working with them daily
- Consultants will better understand program needs than a central office
- Consistent support throughout the lifetime of the program from program conception through closeout
- Consultants will develop content expertise through working with program officers

Participants also noted several challenges including:

- Consultants will not be able to generate trust with program officers if they report to higher-level management. Without this connection, consultants may wield little influence.
- Higher-level support will be required to incentivize program officers to listen to evaluation consultants
- Consultants would not be considered independent evaluators since they work so closely with program staff
In several agencies, it would be difficult to justify hiring internal consultants without making them responsible for overseeing the contracts for evaluations.

**Program Officers as Evaluators**

Finally, a few foundations have used program officers to carry out monitoring and evaluation activities. Rather than hire dedicated evaluation staff, these foundations have incorporated evaluation responsibilities into program officer roles. Most of the evaluation work they conduct though is continuous improvement at the grant level. Rather than act as a neutral third party, program officers actively engage with grantees to act as a resource to connect them to additional supports to ensure success. In addition, program officers monitor grantees for accountability purposes. These personnel are also responsible for developing RFPs for midterm and final evaluations that are conducted by third-party entities. Program officers are PhD-trained content experts but not evaluation method experts.

Participants noted several strengths to this model including:

- Program officers (POs) understand how to design research questions
- POs would have a stake in the evaluation and therefore are more likely to use the information that is provided
- It is possible to build evaluation capacity within existing program officers

Participants also noted several challenges including:

- Lack of evaluation training may lead to lower-quality evaluations
- The dual role of POs does not allow for objectivity. POs would be required to document the flaws of the program while simultaneously advocating for continued funding.
- Individual program evaluations would be highly tailored, limiting their potential value for broader learning across the agency
- Program officers may tend to focus on evaluations for accountability purposes rather than focusing on continuous improvement or learning
- POs would be less likely to assess the transitions of R&D efforts across agencies to understand the broader strategy implications at the Federal level

**Hybrid Approach Combining Centralized Evaluation Service and Internal Evaluation Consultants**

In discussing strengths and challenges of each of the personnel models, several participants commented that a hybrid approach would best serve agencies. They argued that some central personnel were needed to develop evaluation resources and coordinate across programs. These staff could also assist in developing common metrics and data collection across programs. However, participants noted that this unit cannot operate in isolation. To be truly effective, they
believed that evaluation consultants were needed at the program level to generate buy-in from program officers and ensure that evaluations address program needs.

The survey results offer a potential alternative to consultants. Several respondents noted that cost-sharing mechanisms for evaluations would push program officers to value evaluations since they would have a financial stake in them. Since survey results were analyzed after the workshop, I was not able to ask participants whether the cost-sharing mechanism could replace the need for evaluation consultants at the program level. However, it offers a potential path forward if Congress does not allow agencies to hire evaluation consultants.

Finally, a few participants noted that with any of the models, additional training for program officers would be needed to develop evaluation consumers. These participants expressed that program officers need to better understand how to use evaluation results and the insights they can offer. While participants believed that program officers can be taught to use evaluations, few participants believed they should be responsible for designing R&D evaluations. While a centralized approach has challenges, participants expressed that this model could best strengthen science agencies’ capacity to conduct R&D evaluations. However, there may be cultural and governance issues that need to be addressed before a centralized model could be most effective, discussed next.

**Culture for Evaluation**

Throughout the workshop, participants noted potential cultural issues that may hinder efforts to strengthen evaluation capacity to inform future program planning. Agency beliefs in the unique nature of each R&D program, limited learning efforts to apply evaluation findings broadly, and disconnects across science agencies may hinder efforts to better understand the results of R&D investments.

**Science agency personnel believe that their program is unique and cannot be compared with other R&D programs.** In Session 1, multiple participants agreed with an assertion that program officers believe that their programs contain a “secret sauce” that cannot be captured or compared. This view would create pushback to the concept of cluster evaluations that inherently look across multiple programs with similar characteristics. Even amongst workshop participants who expressed exasperation with this concept, an underlying belief in every program being unique was pervasive. In each of the small-group discussions on designing potential cluster evaluations, participants continually noted the differences between their agency’s programs and another agency. This underlying culture may hinder any efforts to look at programs from a strategic level.

**Agencies conduct a variety of evaluations but seldom use individual evaluations for broader learning.** Most survey respondents indicated that their agencies currently conduct process/developmental and summative evaluations. In addition, about 40% of respondents
indicated that their agencies also conduct a combination of formative, landscape, and cluster evaluations. However, few agencies engage in learning activities that are designed as intentional points of reflection that utilize evaluation results to improve agency operations. The agencies that engage in learning activities utilize leadership retreats and advisory committee meetings to reflect on lessons from individual programs that can be applied broadly.

**Cluster evaluations may have greater impacts within agencies than across agencies due to variations in agency approaches and missions.** Finally, several participants suggested that cluster evaluations should first be explored within agencies before venturing into cross-agency evaluations. The small-group activity inherently led participants to discuss cross-agency evaluations. However, participants noted that approach, mission, and leadership differences across agencies create difficulties when trying to design a cluster evaluation that will be used to inform changes in agencies’ strategies. Therefore, cluster evaluations may be easier to implement and ultimately have more impact if they are focused on programs within a single agency.

**Governance and Leadership**

Participants noted several concerns with the governance for R&D evaluations and the leadership dynamics that would enable agencies to design and use evaluations for future program planning. Throughout the workshop, participants noted significant challenges with accessing data from other science agencies when trying to track the follow-on funding from their investments. There are multiple connections between agencies’ R&D programs but there has been little effort to intentionally map these connections to an overall ecosystem or cross-agency strategy (e.g., connecting health information technology investments across DOD, VA, NIH, and NSF). Yet, participants were keenly interested in data sharing and collaboration to provide better data to enable individual agencies to track how their initial investments lead to follow-on funding from other agencies. There are a multitude of opportunities for science agencies to collaborate to share data and engage in shared evaluation projects given similar evaluation interests. However, top leadership must directly support these activities and agencies need a formal mechanism for collaboration. The following section outlines the governance and leadership issues that participants organically brought up throughout the workshop.

**Most respondents indicated that agency leadership should have the discretion to select which programs were “major initiatives” that required evaluations.** The survey asked participants to define which programs Congress should require agencies to evaluate to ensure efficient and effective use of R&D evaluation funds. Participants could set the definition of required programs by a specific funding amount, a specific number of years the program existed, or allow the agency to set its own definition. Most respondents preferred a definition that allowed each agency to define “major initiatives” that need to be evaluated. In justifying this choice, respondents believed that flexibility was important to ensure agencies focus on the
highest priority programs that are not always the most expensive or longest lasting. One-third of respondents chose a definition based on specified funding amounts. Only one respondent preferred a definition set by a specific number of years.

**Agencies need to develop and agree on intra- and inter-agency strategies before designing cluster evaluations.** In each small group in Session 1, participants struggled to come up with a cohesive evaluation question that would be valuable to several agencies. Each agency approaches the topics in a slightly different way and little work has been done to connect the dots between R&D programs within and across agencies. There is no cohesive strategy that aligns science programs together with particular topics. The BRAIN Initiative group noted that most of the programs that fall under this initiative were already being pursued by individual agencies. Therefore, the initiative was less of a cohesive strategy and more of a re-packaging of existing programs. Even in agencies with multiple programs in the same topic area, participants had a hard time identifying the common thread that could be assessed. Throughout each group, it was clear that cohesive strategies are needed that link together R&D programs to agencies’ intended impacts. Without these initial strategies, it is difficult to identify how a cluster evaluation could be valuable.

**Connections to the Commission on Evidence-Based Policymaking**

While designing the workshop, the Commission on Evidence-Based Policymaking (CEP) released its final report that provided a series of recommendations to Congress and the President on ways to increase the collection and use of evidence in policymaking. One set of these recommendations that focused on “Federal Evidence-Building Capacity” aligns well with the foundation strategies identified in Chapter 2. Specifically, the CEP recommended that Congress and the President enable agencies to:

1. Identify a Chief Evaluation Officer and develop the authorities necessary to build an evaluation workforce
2. Develop multi-year learning agendas
3. Coordinate evidence-building activities across departments
4. Streamline approval processes for data collection and procurement
5. Ensure sufficient resources for evidence building activities

Given the overlap, the timeliness of the report, and the potential impact it could have on science agencies if the recommendations become legislation, I decided to incorporate the CEP findings into the overall workshop design and presentation. For example, Session 3 focused on developing a set of evaluation research questions that were framed as a “learning agenda.” This helped provide context as to the questions that agencies would answer if they had greater capacity to conduct Federal R&D evaluations. Incorporating the CEP recommendations into the
workshop also added further incentive for Federal stakeholders to participate in the workshop since the findings could have direct implications on legislation.

Finally, the workshop sessions provide further evidence to support the CEP’s recommendations and potentially offer a rationale for special focus on science agencies. In the following policy implications and conclusions chapter, I will outline the recommendations for Congress that resulted from the workshop and indicate how these recommendations align with the CEP recommendations.
Chapter 4: Policy Implications and Conclusions

Key Findings from Chapter 4:

- Congress should explore the best approach to centralizing oversight of R&D evaluation activities and the White House should centralize support of evaluation activities.
- Science agencies should hire evaluation-trained experts with funding support from Congress.
- Science agencies and the OMB should track expenditures on R&D evaluation activities and Congress should develop dedicated funding streams for R&D evaluations.
- Science agencies and the OSTP should improve evaluation designs to better inform future program planning.

How can we develop better information to guide the design of R&D programs to encourage faster, cheaper, and greater innovation? How can we facilitate Federal science agencies to improve their capacity to conduct evaluations that inform their future program planning? This dissertation aimed to identify approaches from philanthropic foundations\(^{33}\) that could be adopted by the Federal Government to strengthen its capacity to conduct complex evaluations that could inform future R&D program planning. Initially, the literature review framed current challenges with Federal R&D evaluation capacity and the potential for learning from foundations. The foundation interviews then identified a range of approaches for developing evaluation capacity to enable the use of evaluations to inform strategy. The design workshop then leveraged interview findings to identify which foundation strategies can benefit the Federal Government. The workshop also explored how these strategies need to be refined and tailored to fit a Federal context. In this chapter, I interpret and integrate findings from the literature review, foundation interviews, and design workshop to motivate a series of policy recommendations to improve the Federal Government’s use of R&D evaluations for informing future program planning.

\(^{33}\) Evaluation practices within private firms were also investigated but ultimately dropped due to lack of focus on evaluations and limited sample size.
The policy recommendations described in this chapter were developed through a six-step process, summarized above in Figure 10 and discussed in detail below:

1. **Identify Foundation Approaches**—Initially, a range of foundation approaches to R&D evaluation were identified through a series of interviews with foundation employees in Stage 1 of this dissertation. Rather than identify best practices, these interviews aimed to identify a variety of options for strengthening R&D evaluation capacity. For example, the interviews identified three potential personnel models to strengthen R&D evaluation capacity. The full interview findings are described in Chapter 2. To determine which of these findings could benefit the Federal Government, and ultimately be turned into policy recommendations, these findings were shared with a group of Federal stakeholders.

2. **Solicit Opinions from Federal Stakeholders**—A design workshop was developed and conducted in Stage 2 of this dissertation to engage Federal stakeholders in a design process to determine which foundation approaches could strengthen Federal R&D evaluation capacity. The workshop was designed as three interactive, overlapping sessions that each focused on exploring a subset of foundation approaches. In each session, foundation approaches were presented to a diverse group of science policy, planning, and evaluation experts from across various Federal science agencies. These individuals then participated in a series of small and large group exercises to explore the benefits, challenges, and opportunities of foundation approaches. In addition, these experts provided their opinions on how foundation approaches would need to be adapted and tailored to fit a Federal context. The full workshop findings are described in Chapter 3. Through analyzing the workshop notes, a clear set of recommendations emerged.

3. **Develop Preliminary Recommendations**—A preliminary set of recommendations was developed directly from the workshop notes. Foundation approaches that received broad endorsement from the workshop participants were re-phrased and tailored based on participants’ comments into actionable policy recommendations. For example, workshop participants had strong views on the personnel model that could most benefit Federal
science agencies. Foundation approaches that received broad disapproval from the workshop participants were not developed into policy recommendations. In addition, participants’ comments about unique Federal R&D evaluation issues that were discussed at length in the workshop, such as data sharing difficulties, were developed into policy recommendations. Finally, the final report from the Commission on Evidence-Based Policy Making and results described in Chapter 2 of this dissertation were re-reviewed to identify additional policy recommendations that could strengthen the initial recommendations identified from the workshop. For example, the personnel model that was endorsed in the workshop is highly-correlated with the centralized leadership approaches that were identified by foundations but not explored in-depth in the workshop. These related approaches were adapted into recommendations. The preliminary set of recommendations was then further explored through select interviews with senior Federal science leaders.

4. **Interview Senior Leaders for Input**—Three senior Federal science leaders were interviewed to gather their input on the foundation findings and preliminary recommendations. Two leaders were chosen based on their relative positions that were more senior than the workshop participants and one leader was chosen due to her role in advising evaluation efforts at several agencies within a department. The results of these interviews are shared throughout this chapter to provide additional context and support for the final recommendations. These interviews led to clarifications of the preliminary recommendations and helped to identify the need for further review of academic and policy literature to ensure that the preliminary recommendations were targeted to the right level of actor (Congress, the White House, or science agencies) and fit with current legislation.

5. **Review Literature on Federal Context**—Finally, literature that is discussed in Chapter 1 was re-reviewed to ensure that the preliminary recommendations target the right Federal actors and fit within the current Federal context based on existing legislation. Small additional literature reviews were conducted for specific recommendations to strengthen the explanation of the need for change and its potential impact. Preliminary recommendations were further clarified and adjusted based on the literature reviewed.

6. **Finalize Recommendations**—Recommendations were finalized based on input from each of the preceding steps. I debated whether these recommendations should be further tailored due to the current political climate. Ultimately, I decided that this public policy dissertation is intended to be an academic pursuit; therefore, the recommendations should reflect all of the recommendations that developed out of the research on foundations and discussions with Federal stakeholders. Some of these recommendations may not be viable
options given the current Administration and Congress but the upcoming election in 2018 could create new opportunities. Also, science has remained a relatively bipartisan issue (Fealing et al. 2011) and therefore these recommendations could be acted upon by either party, or ideally as a joint-effort similar to the Murray-Ryan Commission.

The final recommendations that resulted from this multi-stage process are organized below by the fundamental foundation strategies that were identified in the foundation interviews. To date, philanthropic foundations that fund R&D have identified four key strategies to strengthen the use of evaluation for strategy development:

1. Hire evaluation-trained experts
2. Centrally oversee and support evaluation activities
3. Invest in evaluation activities
4. Improve evaluation designs

The following chapter describes how these strategies need to be tailored to fit the Federal context. In addition, this chapter provides a series of actions that Congress, the White House, and science agencies could take to increase evaluation capacity to conduct meaningful evaluations that can inform strategy. No single action, in isolation, will be sufficient. Rather, it is the synergy between hiring evaluation personnel, centralizing support, investing in evaluations, and improving evaluation designs that will lead to greater evaluation capacity within science agencies. I note where specific policy recommendations will influence the success of others. In addition, this chapter will integrate perspectives from high-level Federal officials to provide nuance on how recommendations may influence individual departments and agencies.

After providing recommendations for Federal policymakers, this chapter will provide potential paths forward for academic researchers, including caution on researching firm practices to inform government approaches. Finally, this chapter ends with a brief conclusion summarizing the focus, findings, and implications of this dissertation.

**Hire Evaluation-Trained Personnel**

Science agencies should hire additional personnel who are trained in evaluation methods and understand the complexities of evaluating R&D investments. These individuals should be tasked with improving R&D evaluation activities across the agency and introducing new evaluation activities to better inform future program planning. Through hiring additional evaluation-trained personnel, Federal science agencies will be able to increase their ability to design and conduct evaluations that can provide actionable information to policymakers on how to better structure future R&D programs. This section will summarize previous discussions on the value of evaluation personnel from foundations’ perspectives and current capacity limitations in the Federal Government. Additionally, this section will discuss how increased evaluation-trained
personnel could potentially influence Federal evaluation capacity. This section will also provide specific recommendations to Congress and science agencies on hiring additional evaluation-trained personnel.

**Value to Foundations**—Above all other means, foundations have used new personnel hires as a way to increase evaluation capacity. By hiring evaluation-trained experts, foundations noted that they have been able to increase the rigor and value of R&D evaluations. In addition, these staff have the expertise to develop new evaluation activities that aim to provide greater insight on how to structure future investments to maximize desired impact. These staff also have the capacity to focus on cross-program learning, ensuring that evaluation results provide insights beyond the individual program that they focused on.

When compared to foundations that used science program officers as evaluators, foundations that invested in evaluation-trained staff described conducting more complex evaluations, emphasized learning as a primary intended outcome of evaluations, and explicitly stated the need to connect evaluation results to future program planning. The combination of training and explicit job responsibilities focused on evaluation helps these personnel to elevate the quality and quantity of R&D evaluations. Science program officers are able to undertake some evaluation activities but interviewees most often noted that these activities focus on accountability efforts or continuous improvement at the individual project level. Program officers do not have the time, training, or potentially the interest in looking beyond their individual program to assess how to improve overall foundation strategies through assessing sets of similar programs. Several foundations that have hired evaluation-trained staff previously relied on program officers to act as evaluators. In hiring evaluation-trained staff, they have acknowledged the deficiencies in using program officers as evaluators and the unique value that evaluation staff can bring to foundations.

Foundations that have hired evaluation-trained personnel placed them within centralized evaluation services or have embedded them as internal evaluation consultants. Under both models, foundations hire personnel with formal training and education in evaluation methods. In both models, evaluation-trained experts help science program officers understand the value of evaluation and how it can be incorporated into broader learning. Additionally, evaluation personnel in both models focus on specific sets of R&D programs to become familiar with the unique aspects of those programs. Internal evaluation consultants are particularly focused on working alongside program officers to ensure that their evaluation recommendations are tailored to the unique needs of individual R&D programs. Evaluation staff in both models support program officers and help to develop better evidence to inform future program planning.

**Current Federal Capacity Limitations**—While the number and qualifications of Federal evaluation-focused personnel are difficult to ascertain, it is clear from the survey conducted at
the design workshop\textsuperscript{34} that many agencies have few full-time, dedicated staff that focus on R&D evaluation activities. In a few agencies, respondents reported that no full-time staff are devoted to evaluation; rather, evaluation responsibilities are divided among several employees whose primary job responsibilities are not related to monitoring or evaluation. In interviewing two evaluation employees employed by two of the largest R&D-funding agencies, it is also clear that many of the limited evaluation staff within science agencies are not trained in evaluation or measurement methods. Rather, these employees are PhD-trained scientists with deep content knowledge in specific areas of R&D who made transitions from conducting science to evaluation and policy roles. In an interview with a senior evaluation advisor who was originally trained as a health researcher, she acknowledged that she is missing the training necessary to understand how to effectively translate evaluation results into policy prescriptions. The interviewee stated that PhD scientists are inherently trained to be cautious about over-interpreting findings and strongly believe in providing extensive information on the limitations of evaluation studies. These inclinations do not fit with policymakers’ desires to quickly understand the policy implications of evaluation studies. The lack of evaluation-focused personnel and the lack of evaluation-trained personnel limit the Federal Government’s ability to design and conduct evaluations that can inform future program planning.

The Commission on Evidence-Based Policymaking also observed that there is a lack of evaluation staff necessary to conduct evaluations within some parts of the executive branch. The Commission noted that “Federal departments require a range of expertise including in-house expertise to design and conduct evaluations or surveys, advanced technical knowledge required for combining or analyzing data, and expertise in writing and managing highly technical evidence-building contracts” to develop and use evidence to better inform policymaking. As part of its investigation, the Commission conducted a survey of Federal offices that “collect or use data for statistics, evaluation, research, or policy analysis or spend a portion of their budget for such purposes.” Over half of respondents noted difficulties in hiring staff with the skills necessary to conduct evaluations and evidence-building activities. The Commission recommended that agencies “build a high performing evidence-building workforce,” which aligns directly with the following recommendations on hiring evaluation-trained personnel within science agencies (Commission on Evidence-Based Policymaking 2017).

**Potential Influence of Additional Evaluation-Trained Personnel**—Federal science agencies would greatly benefit from hiring evaluation-trained experts to assist science program officers in designing, conducting, and using evaluations to inform future program planning. A senior official at the National Science Foundation, who has recently hired similarly trained evaluation personnel, noted that evaluation-trained personnel are better at assessing the link between what an R&D program is doing and what the Federal Governments is getting as a result of those R&D

\textsuperscript{34} See Introduction for discussion on survey findings related to personnel.
programs. Evaluation staff are able to look at the program from an outsider’s perspective to assess how the decisions that were made on the program’s design and implementation affect the outcomes and impacts of the R&D that was funded. He further noted that evaluation-trained personnel bring a slightly different perspective and mindset to R&D evaluation than PhD-trained scientists and engineers. Evaluation personnel frame issues differently and are able to take a more systems view of understanding the intent of a program and how it connects to the program activities that then influence the ultimate program outcomes and impacts. The senior official also noted that evaluation-trained personnel are better equipped to handle the complexity and ambiguities involved in conducting R&D evaluations. Within philanthropic foundations, these personnel are increasing the rigor of evaluations, developing new evaluation designs, and encouraging leadership to look beyond individual programs to use evaluations to inform broader strategy development. Evaluation-trained personnel could play similar roles within Federal science agencies.

Program officers as evaluators is an alternative model to hiring evaluation-trained personnel. However, workshop participants were adamant that science program officers as evaluators was not a valuable model to adopt from foundations to the Federal Government. While a few foundations have used this model given the desire for lean staffing, this model limits the types of evaluation activities that can be undertaken, particularly hampering evaluations that could inform future program planning. Program officers are able to conduct project-based evaluation activities that ultimately improve individual R&D projects but cannot be used to inform broader questions about how to tailor future programs to strengthen their likelihood of success. In the survey at the design workshop, several respondents advocated that science agencies should pursue any model except science program officers as evaluators. Evaluation-trained personnel could bring greater expertise to Federal science agencies to raise the rigor and value of current R&D evaluation activities. In addition, these personnel can help introduce new activities to ultimately strengthen the systematic evidence needed to inform the design and structure of future R&D programs to maximize the likelihood that they impact society and the economy.

Recommendations—This section describes four key actions that science agencies and Congress should undertake to increase the number of evaluation-trained experts within science agencies:

1. Develop Centralized Evaluation Services with Evaluation-Trained Staff—Science agencies should adopt foundations’ centralized evaluation service model and hire evaluation-trained personnel to staff these offices.
   a. Seek Advice from NSF’s Evaluation and Assessment Capability Office—Science agencies who pursue this approach should seek advice from the National Science Foundation given NSF’s recent experience developing a similar model.
2. **Develop Evaluation Capacity at Lower Levels**—In addition, agencies should develop evaluation capacity at lower levels to ensure that program officers understand the value of evaluation and can effectively design and implement evaluation activities.

3. **Increase the Number of FTEs for Evaluation Experts**—Congress should support these efforts by increasing the number of full-time equivalents (FTEs) for evaluation experts to provide additional agencies the ability to hire additional personnel.

4. **Designate Learning Officers**—Finally, science agencies should designate some evaluation personnel to act as learning officers to ensure that evaluation results are used to inform future program planning.

These actions will significantly increase science agencies’ capacity to conduct R&D evaluations that can inform future strategy.

**1. Develop Centralized Evaluation Services with Evaluation-Trained Staff**

Science agencies should develop centralized evaluation services modeled based on the foundation approaches discussed in Chapter 2. These centralized services should be staffed by evaluation-trained personnel who are given the responsibility and authority to coordinate evaluation activities across the agency, connect evaluation efforts to program planning, and develop standardized evaluation plans. Agencies should dedicate sufficient staff to fill the four distinct evaluation roles that were discussed in Chapter 2 of evaluation officer, organizational effectiveness officer, learning officer, and evaluation leader. These personnel should have evaluation expertise gained either through academic degrees or intensive post-graduate training. In addition, it would be helpful, for knowledge and credibility purposes, for these personnel to have experience in conducting prior evaluations of R&D programs.

These centralized offices should be placed close to senior leadership who oversee R&D for the agency. In some agencies, this will require developing an entirely new office and hiring new personnel. In other agencies, this will require consolidating existing personnel and offices into a single entity that is responsible for R&D evaluation. For example, the NIH does not have a centralized evaluation service within its Office of the NIH Director. Rather, personnel within multiple divisions, and multiple offices within those divisions, provide evaluation capacity to the agency. The NIH should consolidate these personnel into a newly created office of R&D evaluation and ensure that personnel have substantial training in evaluation methods and measurement.

35 Some agencies that fund R&D also perform other functions. The centralized evaluation service should be placed at the highest level in the agency that oversees R&D programs, which may not be the director-level for these agencies.

36 Descriptions of offices on NIH.gov indicate dispersed evaluation roles across the Office of the Director.
In discussing the benefits and challenges of adopting this model to the Federal Government, workshop participants were keenly focused on where to place these centralized evaluation offices to be the most effective. They believed that this office would an important coordinator of disparate evaluation activities across an agency. Similarly, the Commission on Evidence-Based Policymaking advocated for centralized evaluation offices to act as coordinators for disparate existing evidence-building activities (Commission on Evidence-Based Policymaking 2017). To be an effective coordinator, workshop participants argued that this office needs to be placed between program offices and strategy offices. While OSTP and OMB provide general guidance on R&D strategy through their yearly budget memo and budget discussions, R&D strategy is primarily developed at the agency level. Therefore, centralized evaluation services should be placed at the agency level close to senior leadership who oversee R&D policy and planning, rather than having a single centralized evaluation service for all science agencies.

Along with hiring evaluation-trained personnel, these offices would best be supported through increased investments in evaluation activities and improvements in evaluation designs. These recommendations are further described in later sections.

1a. Seek Advice from NSF’s Evaluation and Assessment Capability Office

Agencies that seek to develop centralized evaluation services should seek advice and input from the National Science Foundation given NSF’s recent efforts to develop a similar office. While this office is relatively new and still developing its capabilities, its staff can offer potential insights for agencies interested in developing similar offices. NSF’s Evaluation and Assessment Capability (EAC) office aims to “enhance NSF capabilities for evaluation, knowledge management, and decision-making” by “1) harness[ing] high quality data to promote learning and improve agency results, 2) introduce[ing] and use of timely cost effective evaluations, and 3) strengthen[ing] agency-wide use of evidence for program and policy, decision-making, and action” according to its website. I interviewed an NSF employee familiar with the EAC to gather further insights on its origins and operations. This office is still in its infancy; therefore, it would be premature to advocate for agencies to completely adopt the EAC model or develop agreements to enable the EAC to support smaller science agencies that do not want to develop their own capabilities. Rather, this recommendation focuses on using NSF as a knowledge resource. The following section describes the EAC as it was reported to highlight the similarities with centralized evaluation services and to motivate the value in other agencies seeking advice from the NSF.

The EAC office was formally created in June 2015 within the broader Office of the NSF Director. However, the creation of a dedicated evaluation office had been discussed since 1995.

37 This interviewee was asked to provide her perspectives but was not asked to represent the National Science Foundation.
Ultimately, the EAC was developed as an experiment with agreement from senior leaders from across the NSF to test the value and effectiveness of a centralized evaluation service. Individual directorates (major R&D units) within NSF had invested in developing local evaluation capacities to expand their use of evaluations and therefore were keenly interested in defining the roles of this new central office and how it would interface with directorate-level evaluation efforts. Unlike foundation approaches to developing centralized evaluation capacities, NSF used a democratic process to develop the structure and focus of the EAC.

As of January 2018, the office was staffed by five individuals—an evaluation leader, two evaluators, a data scientist, and support professional. The evaluation leader and evaluators all have formal training in evaluation methods and extensive experience in designing and conducting evaluations (OIA's EAC Section at the National Science Foundation website). NSF decided to use temporary hiring authorities to bring in new personnel to populate the EAC as a way to allow for the opportunity to shut down the office (and not be forced to retain the personnel) if this experiment fails. Agencies could use similar hiring authorities as they explore developing centralized evaluation services.

Currently, the EAC is not funded through a dedicated line item in the budget. Rather, each directorate contributes funds from its appropriation to support the personnel and activities that the EAC undertakes. The EAC uses these funds to support cross-directorate evaluations and covers 50% of the costs for these types of evaluations. NSF has developed several cross-directorate programs to fund R&D in areas that intersect multiple parts of the organization. The EAC helps evaluate these efforts and acts as the contracting representative when using third-party contractors to conduct these evaluations. In addition, EAC evaluators provide guidance to program officers who are designing and managing individual R&D program evaluations. This is similar to the way in which some foundations have set up their centralized evaluation services.

The office is still relatively new and still defining its role within NSF and proving its value. Currently, it is generating awareness and interest in its abilities and expertise through providing quick, low-cost evaluation support. As the office develops, it aims to develop common evaluation resources and provide educational services on the value and use of evaluation activities. Agencies should engage EAC staff to learn from their experiments and efforts to date. If OSTP develops an NSTC subcommittee on R&D evaluation, NSF could periodically present to this group to develop a community of practice.

---

38 It is not clear how success or failure of the office will be assessed. However, the office is working to provide valuable resources to the Foundation. So, a future assessment may focus on opinions from directorate leadership on the value of the EAC.
2. Develop Evaluation Capacity at Lower Levels

Science agencies should also develop evaluation capacity at lower levels (i.e., within science divisions) within science agencies through hiring internal evaluation consultants or training program officers on evaluation designs and uses. Internal evaluation consultants within foundations provide direct support to science program officers as they plan, monitor, and evaluate their R&D programs. These personnel act as advisors throughout the lifecycle of R&D programs to ensure that programs are implemented well, to identify adjustments to improve programs, and to help evaluate programs for learning. This support helps to ensure that program officers appropriately use evaluation methods to inform their decision-making. Alternatively, science program officers could be trained on appropriate uses of and designs for evaluating their programs. Rather than providing direct support on a daily basis, centralized evaluation staff could educate program officers on evaluation basics and when to ask for support to implement evaluation efforts.

Multiple workshop participants advocated for increasing evaluation capacity at lower levels through hiring internal evaluation consultants in addition to setting up centralized evaluation services. They believed that in order for program officers to buy into the value of evaluation for informing program planning, they would need repeated exposure to personnel that could help them tailor evaluation approaches to their programs. In a way, NSF’s EAC is developing buy-in through similar interactions with program officers. However, it is unclear whether science agencies will be able to invest in additional personnel to act as internal evaluation consultants. Funding allocations from Congress will ultimately determine the viability of hiring additional personnel unless agencies decide to re-allocate current FTEs for this purpose. Additionally, workshop participants were wary of consultants reporting up to a senior evaluation leader. Agencies would need to develop local solutions on who should oversee these personnel to ensure they are trusted by program officers but guided by a leader who understands evaluation methods and purposes.

Alternatively, agencies could develop training materials to educate program officers on evaluation methods and techniques. Centralized evaluation staff could reach out to program officers to help educate them on the resources that are available to them and why they might be useful. While this approach would not allow for constant, continual support, it could provide a less expensive approach to developing evaluation capacity at lower levels within science agencies. Through education, program officers can become better consumers of evaluation knowledge and better understand how to structure future evaluations to provide actionable information on how to improve their programs. These options could also be pursued sequentially, with agencies first focusing on developing educational resources and then investing in internal evaluation consultants if science program officers express an interest in further support. Agencies should consider developing evaluation capacity at lower levels by either
initially hiring personnel to work directly with program officers on a daily basis or developing educational materials to create better evaluation consumers.

3. Increase the Number of FTEs for Evaluation Experts

Congress should allocate new funding to support more full-time equivalents (FTEs) to enable science agencies to hire additional evaluation experts. If agencies are going to undertake new evaluation activities to better inform strategy, they will either need to hire new staff or re-direct existing staff. In interviews with two senior Federal leaders, they acknowledged that little new evaluation work would likely be undertaken without investments in additional personnel. Simply, current Federal employees already have full-time job responsibilities that are unlikely to be easily dropped. Therefore, Congress should allocate new funding to enable agencies to hire new evaluation-trained experts.

This recommendation deviates slightly from the recommendations from the Commission on Evidence-Based Policymaking. The Commission recognized that some agencies will “opt to strengthen their human capital by hiring staff or establishing units with particular expertise” but stopped short of recommending that Congress allocate new funding for additional personnel. Rather, the Commission recommended that the Office of Personnel Management ensure that agencies have the appropriate hiring authorities and strategies to hire the personnel to strengthen evidence-building capacity (Commission on Evidence-Based Policymaking 2017). While this is an admirable recommendation that may address hiring issues within some aspects of the Federal Government, it is not clear that new authorities and strategies would result in new evaluation personnel within science agencies without accompanying funding to pay for these personnel. The Commission does not offer any suggestions for where agencies could find funding for new personnel or types of personnel that can be re-assigned to these roles. This is a deficiency with the Commission’s report and recommendations. The Commission advocates for agencies to undertake new evaluation activities and build evaluation capacity. To support this goal, Congress should allocate new funding for additional evaluation-focused FTEs.

The Foundations for Evidence-Based Policymaking Act that unanimously passed the House November 15, 2017, in response to the Commission’s recommendations, is intended to encourage agencies to further develop evaluation capacity (The Foundations for Evidence-Based Policymaking Act of 2017). However, the bill has no specific provisions for science agencies and provides no authority for agencies to hire additional personnel. The Senate, which is currently reviewing the bill as of April 2018, should introduce new provisions to increase the number of FTEs that agencies are able to fill with evaluation personnel.

4. Designate Learning Officers

Finally, science agencies should designate a select number of evaluation personnel to be responsible for organizational learning across the agency. Learning officers are a fairly new
concept within foundations but show promise in helping to develop cross-organizational learning that leverages the results of evaluations of individual R&D programs. These personnel are responsible for identifying opportunities for learning and growth across the organization and develop seminars or trainings in response. Rather than focusing on individual program needs, learning officers use their centralized position to improve programs and processes that impact multiple R&D programs. By creating learning officers, science agencies can ensure that existing evaluation activities are utilized more effectively to inform broader learning. Additionally, learning officers can help identify new areas for evaluation that could benefit the entire agency. Several workshop participants expressed interest in these positions and were interested in playing similar roles within Federal agencies. Science agencies should identify a few evaluation personnel who could best serve in this role to facilitate cross-agency learning and growth.

**Explore Centralizing Oversight of and Centralize Support for R&D Evaluation**

Congress should take steps to explore how to best centralize oversight of R&D evaluation activities and the White House should develop centralized support for R&D evaluation activities. Congress should develop a commission or joint committee to explore how it can best centralize oversight to develop unified guidance for all science agencies on ways to strengthen R&D evaluation capacity. This entity should explore the value of Congress centralizing its own oversight authority through committee structure revisions compared to the value of empowering OSTP and OMB to have greater control over directing R&D evaluation activities. The entity should advise Congress on the best path to centralizing oversight given the complexities in evaluating R&D programs and the need for strong, uniform guidance on how to strengthen R&D evaluation capacity.

Additionally, OSTP should develop a dedicated National Science and Technology Council subcommittee to support R&D evaluation activities within and across science agencies. This subcommittee should be tasked with exploring ways to design evaluations that better inform strategy and developing better data to track the outcomes and impacts of Federal R&D investments. This section will outline the value of centralized oversight and support from foundations’ perspectives; current oversight and support limitations in the Federal Government; how exploring centralization of oversight and centralizing support could potentially influence Federal evaluation capacity; and specific recommendations to Congress to explore how to best centralize oversight and to the White House to centralize support for R&D evaluation activities.

**Value to Foundations**—The foundation interviews clearly indicate the need for, and value of centralized support and oversight for R&D evaluation. As is described in Chapter 2, foundations that have increased their evaluation capacity reported leveraging and benefiting from central oversight and support from senior leadership. Specifically, many foundation presidents have
funded new, dedicated evaluation units and directed funding specifically for R&D evaluation activities. These presidents act as change agents within their organizations to increase focus on R&D evaluation and provide top-level support for evaluation personnel. As leaders, they are able to unilaterally change the focus of foundation operations and champion the successful implementation of better R&D evaluation practices. Operating in relatively small, flat organizations, presidents are also able to provide one-on-one guidance to individual R&D programs and to monitor progress. Foundation presidents have also designated executive-level evaluation leaders to oversee centralized or distributed staff who focus on evaluation. These senior leaders oversee the development of common evaluation resources; education of program officers on the value and proper design of evaluation activities; and facilitation of cross-department learning. Together, presidents and evaluation leaders help to standardize and improve R&D evaluation activities through oversight and support. Federal agencies could benefit from similarly-concentrated high-level support and oversight.

**Current Federal Oversight and Support Limitations**—Unfortunately, there is no natural corollary in the Federal Government for a single individual to champion evaluation reform and a senior leader to oversee R&D evaluation. There is no equivalent head of all Federal R&D or an office that could claim authority to oversee all R&D evaluations. Even within a single agency, often oversight and R&D evaluation support is distributed. The decentralized structure of the Federal R&D enterprise is an amalgamation of multiple departments, agencies, and offices who report to a wide range of congressional committees. Congress ultimately has the authority to decide the size and focus of R&D programs. However, Congress does not examine the overall R&D enterprise as a singular effort; rather individual agencies and programs are reviewed by a variety of committees (Fealing et al. 2011). In fact, almost all congressional committees are involved in R&D policymaking through authorizing or appropriating roles (Stine 2009). While Congress’ “disaggregated, overlapping, and somewhat random arrangement of jurisdictions has turned out to serve science well” in terms of funding allocations and diversity of R&D approaches, it has also impeded the ability for Congress to systematically address deficiencies across the R&D enterprise (Fealing et al. 2011). The workshop highlighted the Federal-wide need to improve evaluation capacity to inform future program planning at agency, White House, and congressional levels. However, no committee has broad enough jurisdiction to oversee the entire R&D enterprise. Even when focusing on an individual agency, often multiple committees claim partial jurisdiction (National Academy of Sciences et al. 1999). When implementing GPRA provisions, the lack of a central oversight authority led to large variations in science agencies’ reporting on performance measures that made it difficult to compare similar programs (National Academy of Sciences et al. 2001). Additionally, the lack of a dedicated committee on R&D has led to few members of Congress developing the expertise necessary to provide strong oversight of R&D programs (Fealing et al. 2011). While Congress has the constitutional authority to provide oversight of R&D programs,
its current structure and lack of specialized knowledge on R&D hinder its ability to appropriately
guide science agencies on how to increase R&D evaluation capacity.

The White House Office of Science and Technology Policy could take a more active role in
oversight of R&D evaluation activities but does not currently have the statutory authority to
force science agencies to follow their oversight requests. The 1976 legislation that established
OSTP mandates that the office:

- “Serve as a source of S&T analysis and judgment for the President with respect to major
  policies, plans, and programs of the federal government.
- Advise the President and others within the Executive Office of the President on the
  impacts of S&T on domestic and international affairs.
- Lead an interagency effort to develop and implement sound S&T policies and budgets.
- Work with the private sector to ensure federal investments in S&T contribute to
  economic prosperity, environmental quality, and national security.
- Build strong partnerships among federal, state, and local governments, other countries,
  and the scientific community.
- Evaluate the scale, quality, and effectiveness of the federal S&T effort.”

Yet, the legislation did not grant OSTP the authority to oversee and direct actions within
R&D agencies. Rather, OSTP relies on its position in the White House and the clout of its
director to persuade agencies to follow their recommendations, with varying success depending
on the administration and OSTP Director (Stine 2009). OMB has also sought to change agency
evaluation practices by introducing new reporting requirements, such as President Bush’s PART
questionnaire. However, it is not clear that these efforts have resulted in meaningful changes
within agencies (Moynihan and Lavertu 2012).

Through executive order, OSTP also manages the National Science and Technology Council,
which is tasked with coordinating science policy across science agencies (Stine 2009, Lane,
Evans, and Matthews 2016). The NSTC helps to support cross-agency R&D efforts and acts as
the interagency convening body for science agencies. The NSTC can change with each
administration since its operations are set by executive order (Stine 2009).

**Potential Influence of Centralizing Support and Oversight**—Through establishing a
commission or joint committee, Congress can explore options for centralizing oversight of R&D
evaluation activities to standardize and maximize the advice it gives to science agencies on how
to best develop evaluation capacity to answer congressional questions on the effectiveness and
efficiency of R&D programs. This commission or joint committee can explore options for
concentrating congressional oversight such as developing a new science committee with

---

39 These bullet points summarize the legislative language and were directly copied from Lane, Evans, and Matthews
(2016).
jurisdiction over all R&D programs. Additionally, this entity could explore the value of granting oversight authority to OSTP and OMB. This could lead to new legislation to centralize oversight and provide a more unified guiding voice for science agencies on how to develop better evaluation capacities. If agencies each develop their own centralized evaluation services, a central oversight authority could help guide their development to ensure that these services standardize their capabilities and provide consistent value to agencies.

The development of a centralized support for R&D evaluation activities through an NSTC subcommittee could help support cross-agency R&D evaluation activities and improve the overall evaluation resources for science agencies. The NSTC subcommittee could help to develop a healthy community of practice\textsuperscript{40} to support individual agency efforts to improve the design of evaluations that aim to inform future program planning.

**Recommendations**—This section describes the two key actions that the Federal Government could take to move towards centralized support and oversight of R&D evaluation activities. These actions will help to standardize and maximize advice to science agencies on approaches to strengthen R&D evaluation capacity and help support the development of better evaluation activities and resources.

1. **Explore How to Concentrate Oversight to Standardize and Maximize Advice to Science Agencies**—Congress should explore how to concentrate oversight of R&D evaluation activities to standardize and maximize advice to science agencies.

2. **Develop a Dedicated NSTC Subcommittee to Support R&D Evaluation Activities**—Finally, OSTP can act without congressional action to develop a dedicated National Science and Technology Council (NSTC) subcommittee to encourage and support cross-agency R&D evaluation efforts.

These policy actions are inspired by foundation approaches but are uniquely Federal recommendations based on the existing structure of the Federal R&D enterprise.

**1. Explore How to Concentrate Oversight to Standardize and Maximize Advice to Science Agencies**

Congress should develop a commission, joint committee, or another mechanism to explore the value of concentrating its oversight authority of science agencies. Commissions and joint committees have been used by Congress in the past to explore potential revisions to congressional committee structures (Welsh 2008). These entities have allowed for Congress to explore the potential value and ramifications of congressional structure changes. While there have been few changes to committee structures in recent years, recommendations from prior

\textsuperscript{40} A community of practice is a group of individuals who share similar interests in an activity and collectively learn how to improve their ability to do an activity through periodic discussions and resource sharing (Wenger 2000).
joint committees have resulted in drastic changes to committee structures and rules including the consolidation of committees and subcommittees (Welsh 2008). A new joint committee could identify the relative value of keeping the current distributed oversight authorities against concentrating oversight within Congress or granting oversight authority to OSTP and OMB.

In order to have the requisite knowledge to provide additional oversight and guidance to science agencies, Congress would need to develop specialized expertise in R&D and evaluation beyond what it currently possesses. Congress has traditionally relied on the executive branch to manage the scientific enterprise and few members of Congress focus their efforts primarily on science (Fealing et al. 2011). Congress would need to invest in developing staff and legislators that are science policy experts, which is feasible if a new R&D committee were developed. Concentrating the oversight of R&D programs will encourage specialization among staff and members of Congress. In addition, congressional rules dictate that all committees receive funding for dedicated staff (Welsh 2008).

Congress could alternatively grant OSTP and OMB with greater authority to direct agency R&D evaluation activities. OMB and OSTP have directed agencies to develop better evidence on the effectiveness of their R&D programs through their annual budget guidance (Zients and Holdren 2012). However, they have limited authority to direct how agencies design, conduct, and use evaluations. This limits their ability to ensure that agencies conduct evaluations that can inform strategy and limits their ability to standardize evaluation practices across agencies. Congress could give OSTP and OMB greater authority to shape these activities.

This joint committee, or similar body, could also determine that the challenges of concentrating oversight authority to standardize and maximize advice to science agencies on how to best increase evaluation capacity may outweigh the benefits. It is possible that the committee could determine that the NSTC subcommittee described below is sufficient to support R&D evaluation activities. Further study is needed to determine the best way to use oversight authority to strength R&D evaluation capacity.

2. Develop a Dedicated NSTC Subcommittee to Support R&D Evaluation Activities

Finally, the White House Office of Science and Technology Policy should develop a dedicated NSTC subcommittee to support R&D evaluation activities within and across science agencies. At the end of the design workshop, participants expressed interest in staying in touch with each other. Additionally, several noted that they had not met each other prior to the workshop despite performing similar duties across different agencies. NSTC is well positioned to bring together evaluation leaders across science agencies to collaborate on ways to help each other strengthen evaluation capacity. Currently, there is no designated working group or
subcommittee on R&D evaluation that bring these individuals together.\footnote{The Research Business Models Working Group under the Subcommittee on Social, Behavioral, and Economic Sciences of the Committee on Science has previously focused on developing common reporting requirements across science agencies on the activities and outputs of Federal R&D grants (Orszag and Holdren 2010). However, its official scope and function does not include supporting R&D evaluation practices or capacity (Tornow and Kaplan 2014).} However, the administration is currently in the process of revising the structure of the NSTC (American Institute of Physics 2018) and should develop a dedicated subcommittee as part of any revisions.

In many ways, this subcommittee could pick up where the design workshop ended to explore ways to design cluster evaluations, share evaluation best practices, and implement data sharing agreements to provide better data on the follow-on results of Federal R&D funding. Cluster evaluations are a fairly new concept for many agencies and foundations. The two foundations that were interviewed were still developing designs for these evaluations when they were interviewed. The NSTC subcommittee could continue to explore how cluster evaluations could be beneficial to science agencies and how they could best be designed. In addition, the subcommittee could act as a community of practice to share best practices in designing, conducting, and using evaluations. Finally, several participants noted that agencies often struggle to track the initial follow-on successes of their grantees due to data sharing issues across agencies. This subcommittee could develop memorandums of understanding (MOUs) for data sharing across agencies to help individual agencies better track follow-on Federal funding that resulted from their initial investments in R&D.

**Invest in Evaluation Activities**

Science agencies and the OMB should track spending on evaluation activities and Congress should develop dedicated funding streams for additional evaluation activities. This section will outline the value of investing in evaluation activities from foundations’ perspectives, current capacity limitations in the Federal Government, how increased investments in evaluation activities could potentially influence Federal evaluation capacity, and specific recommendations to Congress and the OMB to invest in greater evaluation activities.

*Value to Foundations*—Foundations have developed dedicated funding streams for evaluation activities to ensure that sufficient funding is spent on evaluation to inform future program planning. By separating out evaluation funding from general program funds, program officers no longer view evaluation spending as a burden that detracts from funds available for R&D projects. Rather, program officers described these funds as additional support to ensure the success of their program. Evaluation officers who control dedicated evaluation funds indicated that they were able to better direct the design of evaluations than when funding was done on an ad hoc basis using program funds.
**Current Federal Capacity Limitations**—As the literature review and workshop survey indicate, it is difficult to determine the amount of funding that science agencies spend on evaluation activities. However, respondents that were able to provide estimates indicate low levels of spending when compared to foundations. There is no dedicated funding for R&D evaluations and current reporting on funding expenditures does not allow for tracking of evaluation activities.

**Potential Influence of Investing in Evaluation Activities**—By first tracking evaluation expenditures, OMB can provide insights into the varying amount of funding for evaluation across science agencies. This may provide insights for agencies, OMB, and Congress on particular agencies that need assistance in developing dedicated evaluation funding streams. In addition, this information will help to set a baseline to compare against if Congress were to pass legislation instituting the Commission on Evidence-Based Policymaking’s recommendation on increasing evaluation capacity. By tracking evaluation expenditures, Congress would be able to objectively assess whether their efforts result in additional efforts to develop evidence-building capacity. Additionally, dedicated evaluation funding streams will ensure that agencies have the capacity to expand their evaluation activities. New funding may help to answer some of the questions proposed by workshop participants that were outlined in Chapter 3. Answers to these questions could help develop more effective and efficient R&D programs.

**Recommendations**—This section describes two key actions that OMB, science agencies, and Congress should undertake to invest in evaluation activities:

1. **Track Evaluation Expenditures**—OMB and science agencies should better track expenditures on evaluation activities. Without a solid baseline, it will be difficult to track improvements.
2. **Develop Dedicated Evaluation Funding Streams**—Additionally, Congress should develop dedicated funding streams for evaluation activities to ensure that science agencies are able to carry out evaluations that can inform future program planning.

1. **Track Evaluation Expenditures**

   The OMB should direct science agencies to track and report on R&D evaluation expenditures. While Congress seeks to increase agencies’ evaluation capacity and use through its Foundations for Evidence-Based Policymaking Act, there will be minimal ways to measure whether this legislation in fact impacts expenditures. The Act would require agencies to report on a series of activities but senior officials at two major science agencies indicated that agencies could easily re-package existing activities into a new report without making meaningful change. By tracking evaluation expenditures, Congress will have a more objective metric on current evaluation efforts and be able to track changes after the legislation is enacted. Additionally, this data can help to justify the need for additional dedicated funding for R&D evaluation activities.
2. Develop Dedicated Evaluation Funding Streams

Congress should fund increased evaluation activities through dedicated line items in agency appropriations. Several of the foundations that were interviewed are beginning to experiment with separate, dedicated funding streams for evaluation. These funds can either be centrally managed by evaluation offices or distributed directly to individual science programs. Treating evaluation funding separately from program funding encourages program officers to conduct program evaluations and to view these studies as value-added rather than as obligations. Workshop participants noted that dedicated funding streams would help to protect funding for R&D evaluations that could end up being re-allocated towards R&D projects or other administrative functions if it were part of a general fund. Congress should add budgetary line items for evaluation activities that it wants science agencies to carry out and should focus on multi-year funding.

These funds could be considered as a tax on existing funding allocations or additional funding specifically for R&D evaluation activities. If viewed as a tax, science agencies and science advocacy organizations may argue that evaluations take away from directly funding science projects, which is their primary mission. In the design workshop, several participants noted that the Small Business Innovation Research Program funding set-aside is similarly perceived as an unnecessary expense that detracts from the primary missions of some agencies. Therefore, it will be important to highlight the potential value of evaluation in providing information to improve the effectiveness and efficiency of R&D funding. While evaluation funding may decrease the number of scientific projects that can be funded, the insights from evaluation should provide actionable information that strengthens the value of the projects that are funded. Additionally, Congress should consider framing evaluation funding as additional funding that is solely dedicated to evaluation activities. In practice, this framing may not alter where the funding comes from but may help to convince science agencies and science advocates that evaluation benefits R&D programs.

Improve Evaluation Designs and Resources

Finally, science agencies and OSTP should improve evaluation designs and the resources needed to conduct evaluations that can inform future program planning. This section will outline the value of improved evaluation designs and resources from foundations’ perspectives; current capacity limitations in the Federal Government; how improved evaluation designs and resource could potentially influence Federal evaluation capacity; and specific recommendations to the White House science agencies to improve evaluation designs and resources.

Value to Foundations— Foundations are experimenting with new ways to design evaluations to look beyond individual programs to inform strategy. In particular, cluster evaluations offer the potential to examine sets of related programs to identify which components
of programs are leading to intended outcomes and impacts. Rather than focusing on individual programs, cluster evaluations aim to examine multiple programs to determine how to leverage successful programs to support programs that are not as successful. These evaluations may identify changes to grantee requirements, funding mechanisms, or types of projects that are funded to increase the likelihood that a program is successful. While these evaluations are relatively new efforts by a few foundations, their approach to informing strategy is promising.

**Current Federal Capacity Limitations**—Workshop participants noted several examples of evaluation activities that aim to look at sets of related programs. However, these are often expert-driven discussions that do not systematically look at the relative outcomes and impacts of similar programs. Rather, these expert panels review current funded projects to assess portfolio balance. While expert opinions are valuable, systematic examinations of R&D programs could provide better evidence on changes that could positively influence intended outcomes and impacts.

**Potential Influence of Improving Evaluation Designs and Resources**—Implementing cluster evaluations within agencies may offer new insights on how to structure R&D programs to increase the likelihood that they influence academia, government, society, and the economy. These evaluations are explicitly focused on informing strategy that cannot be achieved by individual program evaluations. In discussing cluster evaluation designs, workshop participants also identified several areas where science agencies and OSTP can act to improve the data needed for evaluations.

**Recommendations**—This section outlines a series of actions that science agencies and OSTP should undertake to improve evaluation designs and resources:

1. **Develop Intra-Agency Cluster Evaluations**—Science agencies should develop intra-agency cluster evaluations to better understand which programs are leading to intended impacts and how to shape other programs to achieve better results.
2. **Develop Interagency Data Sharing Agreements**—OSTP should also facilitate the development of interagency data sharing agreements to enable more complete assessments of the outcomes of R&D.
3. **Explore Viability of a Single Data System on R&D Outcomes and Impacts**—Finally, OSTP should also explore the viability of developing a single data system to track R&D outcomes and impacts.

**1. Develop Intra-Agency Cluster Evaluations**

Science agencies should develop intra-agency cluster evaluations to identify ways to improve R&D programs. Science agencies aim to positively influence academia, government, society, and the economy. By conducting cluster evaluations, science agencies will develop better evidence to inform the design of R&D programs to achieve their intended impacts. The National Academies concluded in 1999 that the most effective way to evaluate Federal R&D programs...
was through expert review (National Academy of Sciences et al. 1999). Yet, foundations and science agencies have developed evaluation methods and strategies since then to provide more objective quantitative and qualitative analyses on the impact of R&D programs. Science agencies should continue to advance evaluation methods through developing intra-agency cluster evaluations that examine similarly focused or structured R&D programs. Workshop participants noted that these evaluations will likely be complex, costly, and time intensive. Yet, their potential for informing strategy warrants agency effort. Additionally, by focusing these evaluations on programs within a single agency, evaluators may overcome some of the unresolved issues that were raised in the design workshops such as issues with linking data from multiple agencies.

2. Develop Interagency Data Sharing Agreements

OSTP should facilitate science agencies to develop interagency data sharing agreements to enable better tracking of outcomes from federally funded research. The notion that it is easier for a Federal evaluator to track follow-on funding from the private sector than from other Federal agencies is tragic and fixable. Agencies should be able to track their grantees across agencies to understand how initial funding in one R&D program can spur follow-on work funded by another agency. Tracking these types of successes will help agencies better understand early successes of their R&D and provide clues on where to look for later stage impacts. OSTP should bring together agencies to develop agreements, protocols, and technology for sharing R&D data for evaluation purposes. While OMB and OSTP previously tried to standardize data collection from grantees across science agencies through the Research Performance Progress Report (RPPR) effort (Orszag and Holdren 2010), they did not focus on developing data sharing agreements or similar technology to enable data sharing. New efforts must focus on the practical and technical aspects of data sharing. This could best be achieved through an NSTC subcommittee as discussed earlier in this chapter but also could be pursued without formal NSTC involvement.

3. Explore Viability of a Single Data System on R&D Outcomes and Impacts

OSTP should explore the viability of creating a single data system to track the outcomes and impacts of Federal R&D. Foundations have yet to develop common data systems to track outcomes and impacts across their various R&D programs. Yet, logic would suggest that organizations could greatly benefit from reduced costs and time burden of third-party contractors collecting data each time they conduct an evaluation. Rather than paying for data collection each

42 During the design workshop, one participant commented that she often was able to track follow-on private funding but was unable to track follow-on Federal funding if it came from another agency. Multiple workshop participants agreed with this comment and lamented the difficulty in tracking Federal funding.
time an evaluation collects data on patents, for example, a single data system could be designed to aggregate together common public databases that are used to identify patent data. The previous STAR METRICS system demonstrated the potential value of such a system and several potential pitfalls to its successful implementation. OSTP was a signatory on the STAR METRICS effort but did not oversee its development. OSTP should use its authority and purview to help agencies develop a viable cross-agency data system, taking lessons from the prior effort. In particular, OSTP should explore what support it would need from Congress and OMB to successfully design and implement a system that all science agencies would actively use.

Further Areas for Research to Strengthen Recommendations

While this dissertation provided new knowledge on R&D evaluation approaches, there were limitations to this work and there are multiple ways to explore how to improve Federal R&D evaluation capacity. This dissertation aimed to gain insights from U.S.-based philanthropic foundations who fund R&D and apply those insights to the Federal Government to strengthen its capacity to conduct R&D evaluations that inform future program planning. In pursuing that aim, I ran into a series of methodological issues that could be addressed in further research. Most significantly, my attempts at gaining insights from private firms were unsuccessful and offer caution to future researchers. Additionally, my research focused on R&D evaluation approaches within U.S.-based organizations and did not capture potential lessons to be learned from international organizations. This section describes potential further areas for research to build off this dissertation to identify ways to strengthen Federal R&D evaluation capacity to inform strategy development.

Address Methodological Issues

Researchers interested in reproducing or expanding this study could resolve numerous methodological issues with additional resources, time, or foresight. The Approach and Methods section in the Introduction provides a clear description of the research process and its results. The following list highlights potential modifications and additions to the research process that may, in hindsight, have improved the results. Future research in this area should consider these recommendations.

- **Develop Partnerships with Commercial Support Organizations to Gain Better Access**—This dissertation relied on the willingness of foundation and firm employees to discuss detailed information on how they plan, monitor, and evaluate R&D investments. In several cases, employees were unwilling or uninterested in being interviewed. Yet, there are several support organizations and consultancies that routinely interact with foundations and firms to conduct research on how to improve their R&D practices. These organizations provide evaluation support; encourage networking through conferences and
communication platforms; and conduct research that can benefit their members. Through providing support services, these organizations have developed rapport with firms and foundations over several years and have demonstrated their value to these organizations. Future researchers should explore how to partner with these organizations to leverage their credibility, access, and knowledge to increase the number of organizations that are willing to participate in a study. This could enable future researchers to identify a broader range of R&D evaluation practices that could be beneficial to the Federal Government.

- **Secure Federal Leadership Approval to Increase Participation in Workshops**—The design workshop included Federal employees from most of the agencies that heavily invest in R&D. However, a few agencies were not represented. If another design workshop were organized, researchers should secure endorsements from senior Federal leaders as an additional incentive to ensure agencies designate personnel to participate. Researchers could secure these endorsements through engaging senior Federal leaders early in the project to seek out their advice and input on the study design. If the leaders feel that they have a stake in the research, they might be willing to encourage their subordinates to participate in a later workshop. A broader range of voices may help generate new ideas on how to apply foundation approaches to the Federal Government and could strengthen the representativeness of the workshop conclusions.

- **Conduct an Implementation Workshop based on Findings of the Design Workshop**—The design workshop identified a series of foundation approaches to R&D evaluation that could benefit the Federal Government. However, the workshop did not identify ways to implement these approaches. Future researchers, potentially in collaboration with a new NSTC subcommittee on R&D evaluation, could convene a second workshop to develop a consensus on actions science agencies could take, in absence of congressional directives, to strengthen their evaluation capacity. This could strengthen agency buy-in and help to identify potential issues with the recommendations provided above before they are implemented.

These actions could help to identify a broader range of R&D evaluation approaches, ensure all science agencies are able to provide input, and increase the odds that this research results in increased evaluation capacity to conduct evaluations that can inform future program planning. However, these actions do not necessarily ensure that different conclusions would be reached. Rather, these recommendations could help to ensure that future research achieves its original aims.

**Cautiously Use Private Firms for Federal Insights**

Researchers should be cautious in attempting to identify insights from private firms for use in research to improve Federal processes. As the literature review indicates, academics and private firms have spent years exploring how to best conduct and manage R&D. These articles and
books clearly indicate the potential for learning from private industry. However, my experience trying to work with private firms now makes me cautious about the ease of, and value in, identifying insights based on processes and activities of private firms.

Recruiting large, U.S.-based firms to participate in this research proved insurmountable. Initially, I tried to contact R&D leaders for 30-minute exploratory interviews to describe their process for planning, monitoring, and evaluating R&D investments. The firms with the greatest investments in R&D often did not provide detailed public information on their websites of who their R&D leaders were or provide their contact information. Organizational structures were also difficult to ascertain using publicly available information, making it difficult to identify the right position to contact and then track down contact information.

In addition, proprietary databases that claim to track senior leadership and their associated contact information proved to be unreliable. Researchers should ensure they will have access to private firms and identify specific individuals to interview before initiating their research. Even in firms where I had contacts that were willing to connect me with individuals, I faced difficulties in recruiting specific personnel to participate in my research. There is little incentive for private firm employees to talk with researchers and potential risks to sharing private, non-public information on company operations.

I recommend that researchers partner with existing organizations that work regularly with the private firms they are interested in interviewing. In addition, I would identify a clear value proposition for these firms to ensure that they have an incentive to positively respond to interview requests. For example, researchers could provide written feedback to the firm that summarizes the interview findings from that firm and compares them to similar firms. Even among private sector employees who agreed to be interviewed, there appeared to be a lack of interest in helping provide insights to inform changes in the Federal Government. Researchers may want to develop dual purposes for their research to ensure that it will directly benefit those who are interviewed.

Additionally, it is important to quickly identify firms that will be able to provide valuable advice. As noted earlier, few of the firms that I interviewed engage in formal evaluation practices to assess the outcomes and impacts of their funded R&D. If possible, researchers should develop short screener questionnaires or interview protocols to ensure that interviewees’ and interviewers’ time is not wasted. These questionnaires and protocols should be piloted with consultants or others familiar with private firm R&D operations before widely distributed to ensure that the language will resonate with firms. In asking about R&D evaluation, firms often believed I was asking about evaluating initial proposals to determine which projects to fund.

43 Most notably, Lexis Nexus’ Company Dossier provides contact information for senior leaders in large, U.S. based companies. However, almost all of the email addresses I tested from that database resulted in email bounce-backs.
After consulting with consulting firms that support businesses, I was able to adjust my language to better clarify my interests and focus to firms. Support organizations and consulting firms may also be able to offer insights on firms that fit your research inclusion criteria. Additionally, they may be able to provide insights on whether firms will be valuable in providing alternative perspectives or approaches that could benefit the Federal Government. Researchers should engage these experts early in the planning process to ensure that studies do not face the similar issues that I encountered.

**Explore a Broader Diversity of Evaluation Approaches from International Organizations to Inform Recommendations**

Finally, future work could explore R&D evaluation approaches within science agencies and philanthropic foundations in other countries. While U.S. science agencies have resisted evaluation efforts, science agencies in other countries, notably from the European Union, have been more willing to explore how to best evaluate the outcomes and impacts of R&D investments (see for example Morgan Jones et al. 2016). In addition, several interviewees noted that international philanthropic foundations, such as the Welcome Trust, have invested significantly in developing R&D evaluation capacities. Future research could leverage the interview protocols that I developed (provided in the Appendixes) to explore whether there are additional approaches to R&D evaluation that were not identified in this dissertation. These organizations may also be able to provide longer-term evidence on the effectiveness of the R&D evaluation approaches that U.S. philanthropic foundations are currently employing.

**Implementation Challenges to Monitor and Address**

This dissertation offers a series of ambitious policy recommendations that could significantly increase evaluation capacities across Federal science agencies. However, as with all policy solutions, there are several potential challenges to successful implementation. As Congress, the White House, and science agencies implement recommendations, they should carefully monitor and address implementation challenges to minimize the negative externalities that could arise. This section will briefly outline actions that can be undertaken to mitigate potential implementation challenges.

Science agencies should focus on developing trust between newly-hired evaluation staff and existing science program officers. Program officers drive programmatic design decisions on how to structure their R&D programs. If science agencies want to use evaluations to better inform future program planning decisions, they will have to convince program officers to trust the evaluation data that are provided by evaluation staff. Evaluation-trained personnel will inherently have different perspectives and expertise than program officers. It is possible that program officers may not immediately understand or appreciate the value of evaluation-trained personnel. Additionally, centralizing evaluation staff within a new office that reports directly to senior
leadership may lead to initial distrust of evaluation staff by program officers. Science agencies should actively work to ensure that new evaluation staff engender trust with program officers. Science agencies could seek out advice from the National Science Foundation, who has actively worked to develop communication and trust between their new centralized evaluation service and science program officers.

Congress and the White House should work together to appropriately frame the value of increased spending on evaluation activities. Science advocacy organizations have continually argued that Federal spending on R&D needs to be increased. These organizations could argue that investments in new evaluation activities would divert away funding from science agencies’ core missions of funding R&D. However, if framed correctly, scientists and science advocacy organizations could be convinced that increased evaluation efforts can only help to strengthen the R&D enterprise. Evaluation activities can potentially better inform the design of programs, which could lead to reduced administrative burdens on scientists, limits on grantee requirements that do not encourage better science and increases in the likelihood that R&D projects will lead to innovations. The White House Office of Science and Technology Policy should work with Congress to present a unified front on the value of increasing the use of evaluation activities to better inform future program planning.

Implementation challenges are unavoidable given the complexity and size of the Federal Government. However, active monitoring can identify these challenges as they rise. Congress, the White House, and science agencies should immediately address these issues as they are identified to mitigate their negative impacts and help to ensure that these recommendations lead to better capacity to conduct evaluations that can inform policymaking.

**Conclusion**

Science agencies seek to fund R&D that will positively impact science, society, the economy, and academia. Yet, little evaluation work has been done to assess how to optimally structure R&D programs to maximize the likelihood of their impact. This dissertation explored how to strengthen the Federal Government’s capacity to conduct R&D evaluations that can develop the evidence needed to inform the design and structure of future R&D programs to achieve faster, cheaper and greater innovation. After identifying a range of R&D evaluation approaches from philanthropic foundations that fund R&D, I convened a set of Federal evaluation experts to explore which approaches could be adopted and tailored to strengthen Federal R&D evaluation capacity. This chapter concludes this research by describing a series of policy recommendations to implement the top four strategies that foundations have adopted to increase their R&D evaluation capacity. The Federal Government should:

1. Hire evaluation-trained experts within science agencies,
2. Explore centralizing its oversight of and centralize support for R&D evaluation activities,
3. Invest in additional evaluation activities, and
4. Improve the design of and resources for evaluations.

Foundation interviewees reported that these four strategies have strengthened foundations and these strategies offer the potential to improve the personnel, governance, technology and evaluation activities that constitute Federal evaluation capacity. Congress, the White House, and individual science agencies should pursue a series of activities to support these four primary strategies.

This dissertation has primarily focused on new activities and resources that could be leveraged to increase Federal evaluation capacity to conduct evaluations that can better inform future program planning to maximize the likelihood of positively impacting science, society, and the economy. This focus on new activities and resources is due, in part, to the approach of this dissertation. I aimed to understand to identify strategies from philanthropic foundations and adapt them to the Federal Government. Foundations have recently invested in new evaluation-trained personnel and evaluation activities to better inform strategy. These organizations have recognized the need for new personnel, resources, and activities to better answer the evaluation questions that have been repeatedly proposed since the early 1990’s.

I have also focused on new efforts and resources, in part, due to the varied purposes of existing evaluation efforts within science agencies. As is described in Chapter 2, evaluation activities can serve a variety of purposes from tools to advocate for greater funding to accountability checks to ensure finances are spent on appropriate activities. The survey conducted during the workshop confirmed that science agencies often conduct evaluations for similar purposes. Rather than try to advocate for reforming existing evaluation activities, that may serve these alternative purposes, I have recommended new activities.

Finally, I focused on new activities and resources due to the dearth of evaluations that are designed to inform strategy. While baselines are difficult to assess due to lack of tracking of evaluation activities and evaluation personnel, the workshop survey results indicate little spending on evaluation activities broadly and few evaluation-focused employees within science agencies. The Commission on Evidence-Based Policymaking also broadly acknowledges the limited funding, personnel, and activities devoted to developing the evidence necessary to effectively and efficiently design government programs. In today’s resource-constrained environment, it is difficult to advocate for new resources. However, increasing investments in evaluation to inform strategy could result in faster, greater, and cheaper innovations. What if the Federal Government could identify ways to better structure medical R&D to find faster cures to cancer? What if we could better inform the design R&D programs to develop new technologies? By strengthening R&D evaluation capacity, the Federal Government will be able to pursue a variety of evaluation questions that could increase the likelihood that Federal R&D investments positively impact academia, government, the economy, and society.
Congress can strengthen evaluation capacity by exploring how to concentrate oversight and support to provide clear advice on R&D evaluation efforts, increase the number of FTEs to allow for agencies to hire additional evaluation-trained experts, and provide greater funding for R&D evaluation activities. The White House, through the OMB and OSTP, can further strengthen evaluation capacity by developing an interagency group to coordinate evaluation activities, track evaluation expenditures across R&D agencies, develop interagency data sharing agreements on R&D outcomes, and explore the viability of a single R&D system to track impacts.

Finally, science agencies can strengthen their own evaluation capacity through a series of actions that do not require congressional or White House intervention. Agencies can develop centralized evaluation services with evaluation-trained personnel, develop evaluation capacity at lower levels within the agency, and designate learning officers to ensure that evaluation results are used for broader learning and strategy. Additionally, agencies can track their spending on R&D evaluation activities and develop intra-agency cluster evaluations to inform future program planning. These efforts should be pursued concurrently across all levels of government and, by doing so, will help strengthen one another as they are implemented.

Science agencies face unique challenges when trying to develop and analyze the evidence needed to inform the design of future R&D programs. Yet, foundations have demonstrated that these challenges can be overcome with investments in governance structures, personnel, funds for evaluation, and evaluation design advancements. The Foundations for Evidence-Based Policymaking Act indicates that Congress is interested in encouraging reform. This dissertation provides a series of recommendations specifically to improve science agencies. The Federal Government should capitalize on current interests in developing evidence-based policy to enact a series of science policy reforms to strengthen its capacity to conduct evaluations that can inform the structure and design of future R&D programs.
References


Zients, Jeffrey D., and John P. Holdren. 2012. Science and Technology Priorities for the FY 2014 Budget. edited by Executive Office of the President of the United States.

Appendix 1: Exploratory Interview Protocol

This interview is part of a study that is being conducted by the RAND Corporation, a research organization. The focus of the study is to understand the components and characteristics of R&D evaluation systems within private firms and foundations that could be adopted by the Federal Government to design a new R&D evaluation system. We are interviewing senior leaders in a variety of organizations that fund and conduct R&D to understand a broad range of ways that R&D can be planned, monitored, and evaluated. We will then present these findings to a group of Federal science policy officials to explore how to adapt these components into a single Federal system.

In these initial exploratory interviews, we are aiming to identify a broad range of organizations that manage R&D in different ways. Based on the results of these interviews, we may ask if you are willing to participate in further research efforts. However, by participating today, you are not committing to these later-stage activities.

Your participation in this interview is voluntary, and you can skip any questions that you don’t feel comfortable answering. This data will only be used by the research team at RAND for the study, and your responses will not be shared with others at your organization, or anyone else outside of the research team in a way that might individually identify you.

Do you have any questions before we get started?

Do you have any objections to me recording this session for my notes?

1. Let’s start with some general questions.
   a. What is your role at your organization and how long have you held this role?
   b. How does your role fit into the broader structure of your organization?

2. We’re interested in understanding how foundations plan, monitor, and evaluate their R&D investments. So, let’s focus first on how your organization plans its R&D investments.
   a. What is the process for developing a new program or project at your organization?
   b. What types of information do you use to inform the design of a new program/project?
   c. Who is involved in determining the structure of the program? [i.e., who is eligible, funding amounts, focus areas, grantee requirements, etc.]
   d. How do you make tradeoffs between a new program versus existing programming?
3. Let’s now focus on monitoring activities.
   a. How do you monitor the progress of your current investments?
   b. What types of data do you collect?
   c. Who is involved in monitoring and tracking?
   d. What types of interventions may occur based on the monitoring results?

4. Finally, we’re interested in evaluation activities. To start, can you give me an overview of how Your organization approaches evaluation?
   a. What types of evaluation activities does your organization typically undertake?
   b. What’s the purpose of your evaluation activities?
   c. How do you compare evaluation results across programs/projects?
   d. How are evaluation results used?

[Thank you, wrap-up, follow-on options]
Appendix 2: Firm Interview Protocol

[30-minute semi-structured protocol]

This interview is part of a study that is being conducted by the RAND Corporation, a research organization. The focus of the study is to understand the components and characteristics of R&D evaluation systems within private firms and foundations. We are interviewing different levels of employees within a variety of organizations that fund and conduct R&D to understand a broad range of ways that R&D can be planned, monitored, and evaluated. We will then present these findings to a group of Federal science policy officials to explore how to adapt these components into a single Federal system.

We have conducted an initial set of interviews with philanthropic foundations to explore how they evaluate their investments. We’re now interested in identifying a broad range of ways that private firms plan and evaluate their R&D investments.

Your participation in this interview is voluntary, and you can skip any questions that you don’t feel comfortable answering. This data will only be used by the research team at RAND for the study, and your responses will not be shared with others at your organization, or anyone else outside of the research team in a way that might individually identify you.

Do you have any questions before we get started?

Do you have any objections to me recording this session for my notes?

Intro: Let’s start with some general questions.

5. What is your day-to-day role at your organization and how long have you held this role?
6. How does your role fit into the broader structure of your organization?
7. What types of R&D does your organization undertake?
   a. Do you invest in basic and applied research?
   b. Who do you fund? [internal researchers, universities, external labs, etc.]

Intro: We’re interested in understanding how your firm evaluates the outcomes and impacts of its R&D investments.

8. What types of evaluation activities does your firm typically undertake?
   a. Who is responsible for overseeing these activities?
   b. Who funds evaluation activities?
   c. What types of evaluation activities are you involved in?
9. From your perspective, what is the purpose of evaluation at your organization?
10. How are evaluation results used?
11. What role does evaluation play in comparing the relative success of different programs?
Intro: Looking at the other end of the spectrum, we're also interested in understanding how you plan your R&D investments.

12. What is the process for developing a new program or project at your firm?
13. What types of information do you use to inform the design of a new program/project?
14. Who is involved in determining the structure of the program? [i.e., who is eligible, funding amounts, focus areas, grantee requirements, etc.]
15. How do you make tradeoffs between a new program versus existing programming?

Intro: Finally, we're interested in how you monitor the progress of your investments.

16. How do you monitor the progress of your current investments?
17. What types of data do you collect?
18. Who is involved in monitoring and tracking?
19. What types of interventions may occur based on the monitoring results?

[Thank you, wrap-up, follow-on options]
Appendix 3: In-Depth Foundation Interview Protocols

[30-60-minute telephone interview protocol]

Informed Consent

This interview is part of a study that is being conducted by the RAND Corporation, a research organization. The focus of the study is to understand the components and characteristics of R&D evaluation systems within private firms and foundations. We are interviewing different levels of employees within a variety of organizations that fund and conduct R&D to understand a broad range of ways that R&D can be planned, monitored, and evaluated. We will then present these findings to a group of Federal science policy officials to explore how to adapt these components into a single Federal system.

After an initial set of exploratory interviews, we have selected your organization to further study. In these in-depth interviews, we are speaking with several employees within your organization to better understand your approach to R&D evaluation.

Your participation in this interview is voluntary, and you can skip any questions that you don’t feel comfortable answering. This data will only be used by the research team at RAND for the study, and your responses will not be shared with others at your organization, or anyone else outside of the research team in a way that might individually identify you.

Do you have any questions before we get started?

Do you have any objections to me recording this session for my notes?

Questions to Ask Everyone

Intro: Let’s start with some general questions.

1. What is your day-to-day role at your organization and how long have you held this role?
2. Please describe the organizational structure at your organization.
   a. How does your role fit into the broader organizational structure?

Intro: We’re also interested in understanding your role in evaluation at your organization.

3. What types of evaluation activities are you involved in?
4. From your perspective, what is the purpose of evaluation at your organization?
5. How do you personally use evaluation results?
Science Program Officer Protocol

Communication, Policy Implementation, and Feedback Mechanism

Goals: to understand...

- How program officers’ effectiveness is assessed
- How much is spent on evaluation
- How decisions are made on the structure of R&D programs
- Interest in analyses on how to best structure investments
- Differences between project and program evaluations

Intro: Now I’d like to discuss how your organization decides on the structure of your R&D programs. By structure, we mean who is eligible, the types of awards given, grantee requirements, etc.

1. To start, what is the standard structure of your R&D programs?
2. How was this standard structure initially determined?
3. When creating new programs, how do you determine the program structure?
   a. What types of information do you use to determine the program structure?
   b. Who is involved in determining the program structure?
4. How valuable would it be if the Federal Government could provide quantitative evidence on the impact of different program structures on later stage outcomes?
   a. How would this data influence your decisions when creating new programs?

Intro: We’re also interested in understanding how you plan and budget for evaluation efforts.

1. What is the process for designing and executing a program evaluation?
   a. How are program evaluations distinct from grant or project focused evaluations at your organization?
2. How are evaluations factored into your overall program budget?
3. How much do you spend on evaluation activities each year?
   a. How does that compare to your overall operating budget?
   b. In your opinion, does your organization spend too much, too little, or just the right amount on evaluating its investments?
4. If you wanted to spend more money on evaluation, what is the process for requesting additional funds?
   a. Who would be involved in that process?
   b. What types of information would you need to provide to justify the increased spending?
5. When in the program lifecycle do you decide what types of to conduct?
Personnel, Governance, and Feedback Mechanism

Intro: Shifting focuses, I’d like to ask how your organization evaluates program officers’ effectiveness. This may be a sensitive subject but as a reminder, we’re simply trying to identify how R&D program evaluation is used at your organization.

1. To start, how is the performance of program officers evaluated?
2. What types of information are used in these evaluations?
3. What is the purpose of performance reviews at your organization?
4. What types of actions can result from personnel performance evaluations?
5. How is the success of a program officer’s portfolio assessed?
   a. How does this information feed into their personal performance evaluation?
6. What are the incentives for a program officer to run an effective program?

Evaluator Protocol

[Sub-Program Officer If No Evaluator]

Data Collection Tools

Goals: to understand...

- What data collection tools/systems they use
- How those tools/systems were developed and specifically the role of outside organizations
- The specific types of data collected

Intro: We’re also interested in understanding the tools and systems you use to collect evaluation data and the specific data you collect.

1. To start, what types of evaluation activities does your organization typically undertake?
2. What tools and systems do you use to assist in collecting data for these evaluation activities?
   a. [Follow-up with each activity to determine if there are additional tools/systems used] b. What tools and systems do your outside contractors use when collecting evaluation data for your programs?
   c. Are there any program-specific tools or systems that are used for evaluation?
3. What are the purposes of these tools/systems?
4. How were these tools developed? (e.g., purchased, developed-in house, commercial product w/enhancements, etc.)
5. Who was involved in deciding which tool to buy/build?
6. What types of data are collected using these tools?
   a. Which of these data are collected for all of your programs?
7. How frequently do you use these tools to collect data?
8. Who is responsible for data collection?
9. How are these data collected?
   a. What are the incentives for providing quality data?
   b. How are these data validated?
10. What types of evaluation data do you collect outside of these tools?
    a. Who is responsible for data collection?
    b. How are these data collected?
    c. What are the incentives for providing quality data?
    d. How are these data validated?

Analysis of Data

*Goals: to understand...*

- How they evaluate program success
- How they feed evaluation results into future planning decisions
- How they compare programs
- How they compare to outside efforts
- How they account for their role within a broader ecosystem

*Intro: We’re also interested in understanding how evaluation results are used.*

1. How does your organization determine if individual projects are successful?
2. How does your organization determine if a program is successful?
3. How do you determine the causal impact of your program given other efforts in similar areas?
4. How does your organization determine whether to continue a program?
5. How does your organization compare the success of different programs?
6. What types of comparisons do you make between your programs and programs run by other foundations? Government institutions?

Incentives

*Goals: to understand...*

- Internal vs. External evaluation services
Intro: Finally, we’re interested in understanding how you determine whether to conduct evaluations internally or to contract to external partners.

1. What are the advantages of external evaluators?
2. What are the advantages of internal evaluators?
3. When planning evaluations, how do you determine whether to contract out the work?

Leadership Protocol

Organizational Structure

Goals: to understand...

- How leadership views evaluation
- How leadership assess the organization’s success

[continuation of prior series of overall questions]

1. How do you factor in evaluation activities when designing new initiatives and programs?
2. If a donor provided funding solely for evaluation activities, how would you spend those funds?
   a. What kinds of questions would you want to answer?
   b. How would it change how you run your organization?
3. How do you evaluate the success of your organization?
   a. How do you identify changes that could improve the success of your organization?
4. How do you evaluate the success of individual programs?
   a. How do you identify changes that could improve the success of your programs?

Personnel

Goals: to understand...

- The role of the board in shaping programs, directing approach to evaluation, and the resources focused on evaluation
- The autonomy of science program officers in making decisions on program planning and evaluation
- The role of evaluators (if they exist)
- The skills needed of evaluators and science program officers

Intro: Next, we’re interested in better understanding how your organizational structure influences R&D evaluation.

1. Please describe the leadership structure of your organization.
2. What is the membership composition of your board?
3. When recruiting new board members, what characteristics are you searching for?
4. What role do advisory boards play in your organization?
   a. What role does your board play in making decisions about R&D programs?
   b. What role does your board plan in evaluating the successes of your R&D programs?
5. What role does your senior management play in evaluating the successes of your R&D programs?
6. How do you assess their individual performance?
   a. What information do you use?
   b. Who is involved in assessing their performance?
7. What skills do you look for when recruiting senior managers? Program officers? Evaluators?

Communication, Governance, Feedback Mechanism

Goals: to understand...

- How the leadership views the role of R&D evaluation
- Interest in evidence indicating the best ways to structure investments

1. If a donor provided funding solely for evaluation activities, how would advise the organization spend those funds?
   a. What kinds of questions would you want to answer?
2. How do you evaluate the success of your organization?
3. As a foundation, what role should this organization play relative to its peers? To Federal agencies?
4. How valuable would it be if the Federal Government could provide quantitative evidence on the impact of different program structures on later stage outcomes?
Appendix 4: Workshop Survey

Name: __________________________________________

Title: __________________________________________

Agency: ________________________________________

Imagine you are tasked with drafting legislation that aims to increase the resources for R&D evaluations to better inform future program planning. This legislation would apply only to your agency and provide additional resources for evaluation. Please answer the following questions under this scenario.

1. If your agency was able to hire new employees using the foundation personnel models discussed, which model would be the most effective in increasing your agency’s capacity to conduct R&D evaluations?
   - Centralized evaluation service
   - Evaluation consultants
   - (Additional) program officers as evaluators

2. When contracting out R&D evaluations, what type of evaluators should your agency employ to ensure that they receive the most actionable information on how to improve R&D programs?
   - Subject-matter experts
   - Evaluation experts
   - Advisory committee members

3. If Congress included a requirement for your agency to evaluate all “substantial” programs, which definition should be used to ensure efficient and effective use of R&D evaluation funds?
   - All programs over a specified funding amount
   - All “major initiatives” as defined by the agency
   - All programs that last more than a specified number of years

   a. Why did you choose this definition?

   __________________________________________________________________________
   __________________________________________________________________________
   __________________________________________________________________________
4. How should the funding be structured to ensure efficient and effective use of funds?
   - Centralized, dedicated fund
   - Cost sharing between an evaluation fund and program funds
   - Ad hoc funding using program funds
   a. Why did you choose this funding structure?
      __________________________________________________________
      __________________________________________________________
      __________________________________________________________

   We are also interested in understanding your agency’s current capacity for conducting R&D evaluations. The information provided below will be aggregated together and individual responses will not be shared in our reports.

5. What types of evaluations does your agency currently conduct?
   Check all that apply.
   - Formative
   - Landscape
   - Cluster
   - Process/Developmental
   - Summative
   - Do not know.

6. What types of “learning” activities does your agency engage in to use the results of R&D evaluation for strategy formation?
   __________________________________________________________
   __________________________________________________________
   __________________________________________________________
   __________________________________________________________

7. From your perspective, what are the purposes of evaluation within your agency?
   Check all that apply.
   - Accountability
   - Advocacy for additional funding
   - Continuous improvement
   - Informing strategy
8. How much does your agency spend on R&D evaluation activities each year?  
   Please provide your best estimate. Exclude monitoring activities.

   ________________________________

9. How many full-time equivalents (FTEs) does your agency employ to conduct and oversee evaluations of R&D programs? 
   Please provide your best estimate. Exclude personnel focused on monitoring activities.

   ________________________________

10. Comments:__________________________________________________________
    ________________________________
    ________________________________
    ________________________________
Appendix 5: Senior Federal Leader Interview Protocol

Intro: I’d like to start by talking about the Commission on Evidence-Based Policymaking. Coincidentally, many of their recommendations align with foundation approaches to R&D evaluation.

1. Are you familiar with the house legislation that was develop out of the final recommendations from the Commission?

[If yes…]

2. How do you envision the legislation will impact practices at your agency if it is enacted?
3. What you recommend that Congress add to the existing legislation to ensure that agencies have the capacity and structure to conduct R&D evaluations that inform program planning?

[If no…]

4. Are you planning any changes to the way your agency designs or conducts R&D evaluations?
5. What you recommend to Congress if they were interested in developing legislation to ensure that agencies have the capacity and structure to conduct and use R&D evaluations that inform program planning?

Intro: Through my conversations with foundations, it’s become clear that their primary approach to improving evaluation capacity has been to hire evaluation officers that are explicitly tasked with supporting program officers to design and implement better R&D evaluations.

6. If your agency was able to hire new employees who focus on R&D evaluations, where should they be placed within your agency to be the most effective?
7. What skillsets and training should these employees have?
8. Who should oversee these employees to ensure that their advice is used and evaluation results are acted upon? Why?
9. If you could advise Congress on where to provide additional personnel billets within science agencies, where would you advocate that evaluation-focused personnel be placed broadly within science agencies?

Intro: Foundations are also experimenting with different funding structures for R&D evaluations. Some foundations have dedicated, centralized funds while others rely on cost sharing between evaluation and program units. These foundations previously just used ad hoc program funding for evaluations but found the need for alternative funding structures.
10. How should the R&D evaluation funding be structured at your agency to ensure efficient and effective use of funds?

11. If you could advise Congress on a funding structure for R&D evaluations for all science agencies, which funding structure would you advocate for? Why?

Intro: Finally, I’m interested in your views on the value of external evaluators versus internal evaluators.

12. What are the advantages of external evaluators?
13. What are the advantages of internal evaluators?

14. When planning evaluation activities, how do you determine whether to contract out the work?

Intro: If you have a few more minutes, I’d like to ask one final question.

15. If a donor provided funding solely for evaluation activities, how would you spend those funds?
  a. What kinds of questions would you want to answer?