Informing Research Choices: Indicators and Judgment

The Expert Panel on Science Performance and Research Funding

Council of Canadian Academies
Conseil des académies canadiennes

Science Advice in the Public Interest

INFORMING RESEARCH CHOICES: INDICATORS AND JUDGMENT

The Expert Panel on Science Performance and Research Funding

THE COUNCIL OF CANADIAN ACADEMIES
180 Elgin Street, Suite 1401, Ottawa, ON Canada K2P 2K3

Notice: The project that is the subject of this report was undertaken with the approval of the Board of Governors of the Council of Canadian Academies. Board members are drawn from the Royal Society of Canada (RSC), the Canadian Academy of Engineering (CAE), and the Canadian Academy of Health Sciences (CAHS), as well as from the general public. The members of the expert panel responsible for the report were selected by the Council for their special competences and with regard for appropriate balance.

This report was prepared for the Government of Canada in response to a request from the Natural Sciences and Engineering Research Council via the Minister of Industry. Any opinions, findings, or conclusions expressed in this publication are those of the authors, the Expert Panel on Science Performance and Research Funding, and do not necessarily represent the views of their organizations of affiliation or employment.

Library and Archives Canada Cataloguing in Publication

Council of Canadian Academies. Expert Panel on Science Performance and Research Funding
Informing research choices [electronic resource] : indicators and judgment / The Expert Panel on Science Performance and Research Funding.

Issued also in French under title: Éclairer les choix en matière de recherche.
Includes bibliographical references and index.
Electronic monograph in PDF format.
Issued also in print format.

ISBN 978-1-926558-42-4

1. Research – Evaluation. 2. Engineering – Research – Evaluation.
I. Title.

Q180.55.E9C68 2012 507.2 C2012-902679-4

Disclaimer: The internet data and information referenced in this report were correct, to the best of the Council's knowledge, at the time of publication. Due to the dynamic nature of the internet, resources that are free and publicly available may subsequently require a fee or restrict access, and the location of items may change as menus and webpages are reorganized.

Council of Canadian Academies
Conseil des académies canadiennes

Canada This assessment was made possible with the support of the Government of Canada.

The Council of Canadian Academies

Science Advice in the Public Interest

The Council of Canadian Academies (the Council) is an independent, not-for-profit corporation that supports independent, science-based, expert assessments to inform public policy development in Canada. Led by a 12-member Board of Governors and advised by a 16-member Scientific Advisory Committee, the Council's work encompasses a broad definition of "science," incorporating the natural, social, and health sciences as well as engineering and the humanities.

Council assessments are conducted by independent, multidisciplinary panels of experts from across Canada and abroad. Assessments strive to identify emerging issues, gaps in knowledge, Canadian strengths, and international trends and practices. Upon completion, assessments provide government decision-makers, academia, and stakeholders with high-quality information required to develop informed and innovative public policy.

All Council assessments undergo a formal report review and are published and made available to the public free of charge in English and French. Assessments can be referred to the Council by foundations, non-governmental organizations, the private sector, or any level of government.

The Council is also supported by its three founding Member Academies:

The Royal Society of Canada (RSC) is the senior national body of distinguished Canadian scholars, artists, and scientists. The primary objective of the RSC is to promote learning and research in the arts and sciences. The RSC consists of nearly 2,000 Fellows — men and women who are selected by their peers for outstanding contributions to the natural and social sciences, the arts, and the humanities. The RSC exists to recognize academic excellence, to advise governments and organizations, and to promote Canadian culture.

The Canadian Academy of Engineering (CAE) is the national institution through which Canada's most distinguished and experienced engineers provide strategic advice on matters of critical importance to Canada. The Academy is an independent, self-governing and non-profit organization established in 1987. Members of the Academy are nominated and elected by their peers to honorary Fellowships, in recognition of their distinguished achievements and career-long service to the engineering profession. Fellows of the Academy are committed to ensuring that Canada's engineering expertise is applied to the benefit of all Canadians.

The Canadian Academy of Health Sciences (CAHS) recognizes individuals of great achievement in the academic health sciences in Canada. Founded in 2004, CAHS has approximately 400 Fellows and appoints new Fellows on an annual basis. The organization is managed by a voluntary Board of Directors and a Board Executive. The main function of CAHS is to provide timely, informed, and unbiased assessments of urgent issues affecting the health of Canadians. The Academy also monitors global health-related events to enhance Canada's state of readiness for the future, and provides a Canadian voice for health sciences internationally. CAHS provides a collective, authoritative, multidisciplinary voice on behalf of the health sciences community.

www.scienceadvice.ca
@scienceadvice.ca

The Expert Panel on Science Performance and Research Funding

Rita Colwell, (Chair), Distinguished University Professor, University of Maryland, College Park; Johns Hopkins University Bloomberg School of Public Health; Senior Advisor and Chairman Emeritus, Canon U.S. Life Sciences, Inc.; President and CEO, CosmosID, Inc. (College Park, MD)

Max Blouw, President and Vice-Chancellor, Wilfrid Laurier University (Waterloo, ON)

Linda Butler, Visiting Fellow, Australian National University (ANU); Conjoint Professor, University of Newcastle (Soldiers Point, Australia)

Susan E. Cozzens, Professor of Public Policy, Director of the Technology Policy and Assessment Center, Associate Dean for Research, School of Public Policy, Ivan Allen College, Georgia Tech (Atlanta, GA)

Irwin Feller, Senior Visiting Scientist, American Association for the Advancement of Science (AAAS) (State College, PA)

Yves Gingras, Professor and Canada Research Chair in History and Sociology of Science, Université du Québec à Montréal (UQAM) (Montréal, QC)

Jacques Hurtubise, FRSC, Chair, Department of Mathematics and Statistics, McGill University (Montréal, QC)

Gretchen Jordan, Principal Member of Technical Staff, Sandia National Laboratories (Albuquerque, NM)

John S. MacDonald, O.C., FCAE, Chairman, Day4 Energy, Inc. (Burnaby, BC)

Marja Makarow, Chief Executive, European Science Foundation (ESF) (Strasbourg, France); Professor of Applied Biochemistry and Molecular Biology, University of Helsinki (Helsinki, Finland)

James (Jim) McGroddy, Retired, International Business Machines (IBM) (New York, NY)

Tim McTiernan, President and Vice-Chancellor, University of Ontario Institute of Technology (UOIT) (Oshawa, ON)

Sir Keith O'Nions, Rector, Imperial College London (London, UK)

René Simard, O.C., FRSC, FCAE, Professor Emeritus, Université de Montréal (Montréal, QC)

Alan E. Winter, FCAE, President and Chief Executive Officer, Genome British Columbia (Vancouver, BC)

Ronald Woodward, President, Clockbuilder Consulting, Ltd.; President Emeritus, Red Deer College (Red Deer, AB)

Acknowledgements

This report provides the guidance of the 16-member Expert Panel on the use of indicators as decision-support tools. To ensure wise and effective national investment in research that will contribute to the welfare of citizens, difficult choices must be made. Discovery research is the manifestation of human creativity and ingenuity from which social stability, economic vitality, and national and international security derive. This report embodies the Panel's wisdom and counsel offered in the spirit of international collaboration and best judgment.

I am sincerely grateful to my fellow Panel members for their dedication and the time and energy they so generously volunteered. Their combined efforts and expertise resulted in a valuable compendium of practical lessons that provide wise guidance for the global community of science policy. I would particularly like to thank Dr. Max Blouw for his excellent service on behalf of the Panel.

Preparing this assessment has been a pleasing challenge because of its specific relevance to Canada where I was fortunate to spend two productive post-doctoral years of study earlier in my career. Finally, I wish to thank the Council staff for their extraordinary support and for ensuring that the Panel followed precisely Council practices and policies.

Rita Colwell, Chair
Expert Panel on Science Performance and Research Funding

Project Staff of the Council of Canadian Academies

Assessment Team:	Renata Osika, Program Director Christina Stachulak, Senior Program Director R. Dane Berry, Research Associate Fiona McArthur, Research Associate Andrea Hopkins, Program Coordinator Christina McMahon, Program Coordinator Claire Chesnais, Intern
With assistance from:	Clare Walker, Editor Jean-Pierre Toupin, Certified Translator, English-French Yves Gingras, Reviewer, French Language Christine Gonthier, Certified Translator, English-French Accurate Design, Report Interior Design

Report Review

This report was reviewed in draft form by the individuals listed below — a group of reviewers selected by the Council of Canadian Academies for their diverse perspectives; areas of expertise; and broad representation of academic, policy, and non-governmental organizations.

The reviewers assessed the objectivity and quality of the report. Their submissions, which will remain confidential, were considered fully by the Panel, and many of their suggestions were incorporated into the report. They were not asked to endorse the conclusions nor did they see the final draft of the report before its release. Responsibility for the final content of this report rests entirely with the authoring Panel and the Council.

The Council wishes to thank the following individuals for their review of this report:

Claire Donovan, Reader, HERG at Brunel University (London, UK)

Stevan Harnad, Canada Research Chair, Cognitive Science, Université du Québec à Montréal (Montréal, QC)

Diana Hicks, Professor and Chair, Georgia Institute of Technology (Atlanta, GA)

Kari Kveseth, Science Counsellor, Royal Norwegian Embassy (Beijing, China)

Cheryl Oros, Founder and CEO, Oros Consulting (Washington, DC)

Pekka Sinervo, FRSC, Professor, University of Toronto (Toronto, ON)

David Strangway, O.C., FRSC, President and Chancellor, Quest University (Squamish, BC)

Robert Tijssen, Professor, Science and Innovation Studies, Leiden University (Leiden, Netherlands)

Don Wright, President, British Columbia Institute of Technology (Burnaby, BC)

The report review procedure was monitored on behalf of the Council's Board of Governors and Scientific Advisory Committee (SAC) by **Judith Hall, O.C., FRSC, FCAHS,** Professor of Pediatrics and Medical Genetics, University of British Columbia (Vancouver, BC). The role of the report review monitor is to ensure that the Panel gives full and fair consideration to the submissions of the report reviewers. The Board of the Council authorizes public release of an expert panel report only after the report review monitor confirms that the Council's report review requirements have been satisfied. The Council thanks Dr. Hall for her diligent contribution as review monitor.

Elizabeth Dowdeswell, President & CEO
Council of Canadian Academies

Executive Summary

Discovery research in the natural sciences and engineering (NSE) is a key driver in the creation of many public goods. Scientific advances help catalyze innovation, create new knowledge, foster economic prosperity, improve public health, enable better protection of the environment, strengthen national security and defence, and contribute in myriad other ways to national and sub-national policy objectives. For all of these reasons, most governments around the world wisely invest substantial public resources in supporting discovery research in the NSE. Canada is no exception. The Natural Sciences and Engineering Research Council (NSERC) spends approximately one billion dollars a year on scientific research. Over one-third of that goes directly to support discovery research through its Discovery Grants Program (DGP). Many influential Canadian discoveries and research breakthroughs stand as testimony to the value of these investments, and past evaluations of the DGP have found it to be a vital and highly effective component of Canada's research funding landscape.

Public funding organizations like NSERC often struggle with how best to allocate funding across research fields and programs. Once these allocation decisions are made, funding organizations must then determine how to best communicate and justify them to the research community, policy-makers, and the public at large. Thus funding organizations are increasingly looking to science assessment tools and quantitative science indicators for guidance in informing these decisions. New indicators and an emerging "science of science policy" can potentially improve the overall effectiveness and transparency of how funding agencies allocate resources and monitor the performance of their research investments. The growing abundance of indicator and assessment choices, however, can also make it difficult for policy-makers and research funders to know which assessment methods and indicators are most appropriate in a given context.

THE CHARGE TO THE PANEL

To help guide future funding reallocations for the DGP, in 2010 the federal Minister of Industry, on behalf of NSERC, posed the following question to the Council of Canadian Academies (the Council):

> *What do the scientific evidence and the approaches used by other funding agencies globally have to offer, in terms of performance indicators and related best practices in the context of research in the natural sciences and engineering, carried out at universities, colleges, and polytechnics?*

In response to the charge, the Council convened an Expert Panel of 16 Canadian and international experts from diverse fields such as public policy, economics, research funding and administration, mathematics and statistics, science history and sociology, bibliometrics, and other NSE fields. The Panel, which met four times over the course of 2011, reviewed a wide range of evidence from published studies and examined science assessment practices in 10 countries in detail.

SCIENCE INDICATORS AND ASSESSMENT STRATEGIES FOR DISCOVERY RESEARCH

Existing science indicators and assessment strategies can be categorized in many different ways. They include those based on deliberative methods, such as peer or expert review, and those based on quantitative indicators, including publication and citation counts, numbers of researchers or students, research funding amounts, and grant applications. NSE research funding allocation decisions require sets of indicators that capture information on research quality, research trends, and research capacity.

For each of these assessment types, the Panel developed a taxonomy of potential methodologies and indicators, and assessed the validity of these indicators with respect to the assessment objective. The Panel focused exclusively on science performance at the national level of research fields in the NSE (rather than at the level of individual scientists or research teams), and on the indicators and methodologies most relevant to discovery research, such as that funded by NSERC's DGP.

MAIN FINDINGS

Many science indicators and assessment approaches are sufficiently robust to be used to assess science performance in the NSE at the level of nationally aggregated fields. For example, bibliometric indicators based on weighted publication counts can be useful in assessing research output at the level of a research field. Citation-based indicators — when appropriately normalized by the field of research and based on a sufficiently long citation window — can be useful metrics in assessing the overall scientific impact of research in a given field at the national level. Many other types of quantitative indicators, such as those based on student or researcher population, research funding amounts, and the state and quality of available scientific infrastructure and equipment, can be useful in characterizing research trends or national research capacity in certain assessment contexts.

Quantitative indicators should be used to inform rather than replace expert judgment in the context of science assessment for research funding allocation. Although many types of quantitative indicators can be reliable and informative in science assessments at the national field level, these indicators should not be used to support research funding allocation without expert judgment. The body of evidence now available recognizes that the most promising strategies rely on a balanced use of quantitative indicators and expert judgment. A review of recent experiences in selected countries and research funding organizations globally lends further support to this conclusion. In the United Kingdom, the long-standing Research Assessment Exercise (RAE) is scheduled to be replaced with the Research Excellence Framework (REF). The REF will retain core reliance on peer review, but will allow for use of quantitative indicators. In Australia, a recently adopted national research assessment system relies on a model of expert judgment informed by quantitative indicators. Many countries — including the United States, Finland, and the Netherlands — have employed science assessment strategies combining indicators and expert judgment in various contexts. For national research assessment in the NSE at the field level, the weight of the evidence suggests the best approach is a combination of quantitative data and expert judgment.

International "best practices" offer limited insight with respect to science indicator use and assessment strategies. Construction and application of indicators are context dependent. Whether an indicator is informative or reliable depends as much on the specific context as on the nature and construction of the indicator. No single indicator, set of indicators, or assessment strategy offers an ideal solution in research assessment contexts for NSE discovery research. The individual circumstances of the assessment and the research funding context must be considered. For NSERC, these decisions will necessarily take into account both the overarching federal S&T strategy as well as the mandate of NSERC and the specific objectives of its programs. The assessment must reflect proximal goals (in terms of desired outcomes or results) and the ultimate objectives of the funding program or organization.

Mapping research funding allocation directly to quantitative indicators is far too simplistic, and is not a realistic strategy. Indicators may reveal useful information about science performance, but funding allocation decisions are complex. In most respects, neither the existing body of evidence nor the experience of international funding processes justifies a simplistic funding allocation based solely on quantitative indicators. Funding agencies may choose to increase the allocation of resources to an area of research weakness to bolster performance, or, alternatively, direct resources away from areas of research weakness and towards strengths. These choices are driven by the strategy of a funding agency and program. In addition, for discovery research, past performance is not always a strong predictor of future performance. In most areas of scientific work, there is no compelling reason to believe that past successes will inevitably lead to future successes or past failures to future failures. As a result, science indicators — essentially measures of past performance — may not provide a reliable guide to future prospects. Overall, the Panel found no evidence that there is a single correct funding response to any assessment results.

GUIDELINES AND PRINCIPLES FOR SCIENCE ASSESSMENT

It was not the Panel's mandate to provide policy recommendations for national NSE assessment strategies. It did, however, formulate some general guidelines for developing an approach to assessments, which are presented here (see Summary of Methodological Guidelines). In addition to methodological guidelines, the Panel developed the following general principles for defining a process for NSE assessment in the context of informing research funding allocation:

- **Context matters:** Effective use of indicators or assessment strategies, as applied to research fields in the NSE, is context dependent. Thus any approach should take into account national science and technology objectives as well as the goals and priorities of the organization and funding program.
- **Do no harm:** Attempts to link funding allocation directly to specific indicators have the potential to lead to unintended consequences with negative impacts on the research community. Promising strategies identified by the Panel to mitigate this risk include relying on a balanced set of indicators and expert judgment in the assessment process.
- **Transparency is critical:** Assessment methods and indicators are most effective when fully transparent to the scientific community. Such transparency should include both the assessment methods or indicators (e.g., indicator construction and validation, data sources, criteria, procedures for selecting expert reviewers) and the method or process by which the indicators or assessments inform or influence funding decisions.
- **The judgment of scientific experts remains invaluable:** Many quantitative indicators are capable of providing useful information in the assessment of discovery research at the national and field level. In the context of informing research funding decisions, however, quantitative indicators are best interpreted by scientific experts with detailed knowledge and experience in the relevant fields of research, and a deep and nuanced understanding of the research funding contexts in question, and the scientific issues, problems, questions, and opportunities at stake.

Summary of Methodological Guidelines

Context is critical in determining whether any science indicator or assessment strategy is appropriate and informative. As a result, it is impossible to provide a list of universally applicable best practices. With respect to assessing scientific research in the NSE at the level of nationally aggregated research fields, however, the following general methodological guidelines may be of assistance.

Assessments of Research Quality

Indicators associated with monitoring research quality often relate to different aspects of quality or different timeframes. As a result, the best approach relies on a combination of assessment strategies and indicators.

- For an assessment of research quality of a field at the national level, a balanced combination of deliberative methods and quantitative indicators is the strongest approach.
- For an assessment of the scientific impact of research in a field at the national level, indicators based on relative, field-normalized citations (e.g., average relative citations) offer the best available metrics. At this level of aggregation, when appropriately normalized by field and based on a sufficiently long citation window, these measures provide a defensible and informative assessment of the impacts of past research in the NSE.
- Quantitative indicators of research quality should always be evaluated by informed expert review because accurate interpretation of data from available indicators can require detailed contextual knowledge of a field.

Assessments of Research Trends

As with research quality, the best approach associated with monitoring research trends relies on multiple assessment strategies and indicators to create a composite perspective on emerging research trends across fields. Such an approach should rely on a combination of assessment strategies and indicators that includes one or more metrics from each of the following types:

- Trends in grant applications by research topic: Capturing research trends that directly pertain to funding requests ensures that trends related to the direct demand for resources across fields are factored into the process.
- Bibliometric methods: Advanced bibliometric approaches based on keyword analysis and identifying emerging clusters of highly cited research provide useful insights at a more detailed level. These can be used to flag active areas of research, which may span multiple fields, as targets for possible added support.

continued on next page

- Student population: Trends in student population, captured by indicators such as PhD enrolment rates by field, can be useful in anticipating longer-term research trends and monitoring changes in the levels of training and expertise over time.

Quantitative indicators of research trends should always be evaluated by informed expert review because accurate interpretation of data from available indicators may require detailed contextual knowledge of a field.

Assessments of Research Capacity

The best approach associated with monitoring research capacity relies on multiple, diverse indicators to create a composite of underlying features that determine capacity in a field. As a general guideline, one or more indicators from each of the following categories is suggested:

- Funding: Measures of the level of research funding are informative in analyzing research capacity, particularly in comparison to past funding levels and other research sectors. The diversity of funding sources can also be important.
- Infrastructure: The extent and quality of research infrastructure and facilities (e.g., laboratory space, capital investment) are direct determinants of capacity. Measures related to information and communication technology infrastructure should also be considered where appropriate.
- Numbers of researchers and students: The student and researcher populations are a key determinant of research capacity, and metrics based on these populations are consequently an important aspect of this type of assessment.
- Networks and collaborations: Patterns of research collaboration and networks (e.g., co-authorship of papers) within a field can also be tracked to provide insights into research capacity.
- Field characteristics: Assessments of research capacity should also include measures such as the average research team size, average size and duration of research grants, material and equipment intensity, cost of research, and access to research facilities.

As with the assessment of research trends, research capacity in NSE fields should always be assessed through informed expert review because interpretation of data from quantitative indicators may require contextual knowledge.

Table of Contents

1

Introduction and Charge to the Panel

- Why this Assessment Matters
- The Charge to the Panel
- Assessment Methodology
- Report Structure

1 Introduction and Charge to the Panel

1.1 WHY THIS ASSESSMENT MATTERS

The problem of determining what areas of research to fund permeates science policy. Nations now invest substantial sums in supporting discovery research in natural sciences and engineering (NSE). They do so for many reasons. Discovery research helps to generate new technologies; to foster innovation and economic competitiveness; to improve quality of life; and to achieve other widely held social or policy objectives such as improved public health and health care, protection of the environment, and promotion of national security. The body of evidence on the benefits that accrue from these investments is clear: in the long run, public investments in discovery-oriented research yield real and tangible benefits to society across many domains.

These expenditures, however, are accompanied by an obligation to allocate public resources prudently. In times of increasing fiscal pressures and spending accountability, public funders of research often struggle to justify their funding decisions — both to the scientific community and the wider public. How should research funding agencies allocate their budgets across different areas of research? And, once allocations are made, how can the performance of those investments be monitored or assessed over time? These have always been the core questions of science policy, and they remain so today.

Such questions are notoriously difficult to answer; however, they are not intractable. An emerging "science of science policy" and the growing field of scientometrics (the study of how to measure, monitor, and assess scientific research) provide quantitative and qualitative tools to support research funding decisions. Although a great deal of controversy remains about what and how to measure, indicator-based assessments of scientific work are increasingly common. In many cases these assessments indirectly, if not directly, inform research funding decisions.

In some respects, the primary challenge in science assessment today is caused more by an overabundance of indicators than by a lack of them. The plethora of available indicators may make it difficult for policy-makers or research funders to determine which metrics are most appropriate and informative in specific contexts. Assessments of scientific work are conducted for many reasons. They may be broad or narrow in focus, and may target different levels of aggregation — ranging from individual researchers to research teams and institutions, and all the way to nationally aggregated research fields. Some assessments are carried out with an understanding that they will directly impact research funding allocation. Others are intended to provide an analysis of current performance, which may or may not impact any funding decisions. Finally, evaluations may be retrospective in nature (*ex post*) or prospective (*ex ante*), though nearly all science indicators are fundamentally retrospective since they are tied to past research accomplishments or outcomes. All of these parameters affect which types of indicators may be appropriate or informative in an assessment context.

Assessment systems tied to the allocation of public funds can be expected to be contentious. Since research funding decisions directly affect the income and careers of researchers, assessment systems linked to those decisions will invariably have an impact on researcher behaviour. Past experiences with science assessment initiatives have sometimes yielded unintended, and undesirable, impacts. In addition, poorly constructed or misused indicators have created scepticism among many scientists and researchers about the value and utility of these measures. As a result, the issues surrounding national science assessment initiatives have increasingly become contentious. In the United Kingdom and Australia, debates about national research assessment have been highly publicized in recent years. While such attention is testimony to the importance of these assessments, the occasionally strident character of the public debate about science metrics and evaluation can impede the development and adoption of good public policy.

In response to these trends, there is a growing demand among research funding organizations for clear guidelines on effective science assessment strategies and indicators. High-profile efforts have been undertaken in several countries to discern "best practices," and the renewed interest in metrics for assessing discovery research has catalyzed new scholarship. Technological and methodological advances have led to a "profusion of measures" (Van Noorden, 2010), and new metrics continue to emerge with the increasing popularity of publishing and accessing research findings on the internet. As new technologies and analytical techniques continue to reshape science measurement and evaluation, policy-makers and research funders struggle to ensure their assessment systems and funding allocation processes reflect the latest advances and best available knowledge.

1.2 THE CHARGE TO THE PANEL

The Discovery Grants Program (DGP) of the Natural Sciences and Engineering Research Council (NSERC) is Canada's flagship program for the funding of discovery research in the natural sciences and engineering. In 1994 NSERC undertook the first of three funding reallocation exercises to ensure that the DGP remained "dynamic and responsive to changes […] in the research environment" (NSERC, 2006a). These reallocations aimed to reset the historical budget envelopes used by field-level committees to award research grants. A 2006 evaluation of these reallocation exercises, however, concluded that the cost of the exercises, in both economic and human terms, outweighed any benefits from the modest changes in funding allocations that were enacted (NSERC, 2006a).

Since that time, NSERC has explored various options to ensure responsiveness of the DGP funding allocations to the evolving scientific landscape. As part of this ongoing effort, in 2010 NSERC referred the following question and sub-questions to the Council of Canadian Academies (the Council):

> *What do the scientific evidence and the approaches used by other funding agencies globally have to offer, in terms of performance indicators and related best practices in the context of research in the natural sciences and engineering, carried out at universities, colleges, and polytechnics?*
>
> 1. *What existing qualitative and quantitative indicators and metrics are relevant to budget allocation in the context of support for research in the natural sciences and engineering, and how can they be categorized (e.g., shelf life; cross-disciplinary and international comparability; relevance to interdisciplinary vs. focused disciplinary areas; and applicability to emerging vs. established research areas)?*
>
> 2. *What are international best practices in the construction, methodological review, and use of quantitative and qualitative indicators for research evaluation and budget allocation in support of basic research in the natural sciences and engineering?*
>
> 3. *Considering the foregoing, and in light of the Government of Canada Science and Technology Strategy and NSERC's objectives for the support of research, what key considerations (e.g., risks, advantages/disadvantages, behavioural and institutional consequences) and principles emerge in determining defensible use and balance/weighting of performance indicators/metrics for budget allocation?*

In response, the Council appointed the 16-member Expert Panel on Science Performance and Research Funding (the Panel). Panel members come from Canada and abroad, and have a diverse range of backgrounds and areas of expertise. The Panel possesses a depth of experience in research administration and management (university presidents and administrators), and in the management of research funding in the public and private sectors. The Panel met four times over the course of 2011 to consider and address the charge. This report is the result of these deliberations.

It is important to clarify several key points relating to the scope of the Panel's charge. First, the Panel was specifically asked to examine practices relating to science assessment and science indicators *in the NSE*. It is widely recognized that the challenges associated with research evaluation and the use of quantitative indicators are particularly acute with respect to fields in the humanities, arts, and social sciences (e.g., Hicks, 2004). These issues, however, are excluded from detailed consideration in this report.

Second, NSERC's charge was motivated in part by past experiences with the DGP Reallocation Exercises and the need for better assessment strategies pertaining specifically to discovery research. Adhering to its charge to consider the role of indicators in the context of NSERC's programs, and with emphasis on the DGP, the Panel focused on indicators related to *ex post* and *ex ante* measures of inquiry-driven, research performance (and capability). Although the Panel recognized the existence and importance of indicators related to socio-economic impacts of scientific and technological research, extensive treatment of them was beyond its charge.

Third, the Panel was asked to concentrate on issues relating to the use of indicators or assessment methods in informing research funding allocations at the level of nationally aggregated research fields. The report does not cover issues relevant to other levels of aggregation (e.g., individual scientists, research groups, or institutions) in any detail. (See Figure 1.1 for a schematic of the various levels of science evaluation.)

Fourth, the charge requests an examination of science indicators as relates to assessing scientific research occurring in both universities and colleges and polytechnics. In response, the Panel has focused on issues relating to the assessment of discovery research in the NSE , irrespective of where that research occurs (see Box 1.1).

Finally, the charge specifically requests an examination of international experience with science indicators in the context of funding allocation. While the Panel has carefully considered the Canadian context for this study, its treatment in the report emphasizes recent experiences of other countries.

Evaluation of research grants (single/multiple principle investigators) – This occurs when it is necessary to evaluate the work of a single researcher or research program, such as evaluating whether a researcher will get tenure.

Evaluation of funding schemes – This occurs in order to evaluate whether a particular program is meeting society's objectives.

FOCUS OF THIS REPORT

Evaluation of research fields/disciplines – This occurs in order to evaluate the quality, trends, or capacity of a certain field.

Evaluation of funding policies or particular strategic issues – This may occur as a review of funding policies or strategic issues such as in an evaluation of national S&T objectives.

Evaluation of funding agencies – This may occur when it is necessary to evaluate how a funding agency is meeting its national objectives.

Adapted from ESF, 2009

Figure 1.1

Levels of science assessment and evaluation

This figure shows the various levels of aggregation for which evaluations of research can be carried out. The focus of this report is on assessment and evaluation activity targeted at the field/discipline level.

Box 1.1
Assessing Discovery Research in Canada's Colleges and Polytechnics

The Panel was asked to examine science assessment practices and indicators related to research undertaken by Canadian universities, as well as colleges and polytechnics. Traditionally, discovery-oriented, non-directed scientific research in the NSE has been carried out primarily within universities in Canada. Research activity in colleges and polytechnics was historically focused more on development of specific technological applications in collaboration with the private sector. University and college research roles, however, have evolved in recent years. Increasingly, all types of higher education institutions in Canada are participating in research across the spectrum from discovery to technology development, deployment, and commercialization.

continued on next page

As shown in Figure 1.2, there is inevitable overlap between and among universities, colleges, government, and the private sector with respect to research roles, and all participate, to some degree, in discovery research in the NSE. In recognition of this evolution in undergraduate institutions, researchers are eligible to apply for Discovery Grants from NSERC.

All varieties of research conducted along the continuum, from discovery to applied, contribute to the creation of knowledge. This diverse continuum should be reflected in the choice of assessment practices and indicators. For field-level assessment of discovery research in the NSE, selected measures and evaluative approaches should reflect the nature of the research being assessed, regardless of the institution where it is performed. Therefore, to the extent that colleges participate in discovery research, field-level assessment of discovery research in the NSE should capture participation with the same approaches as are applied to universities.

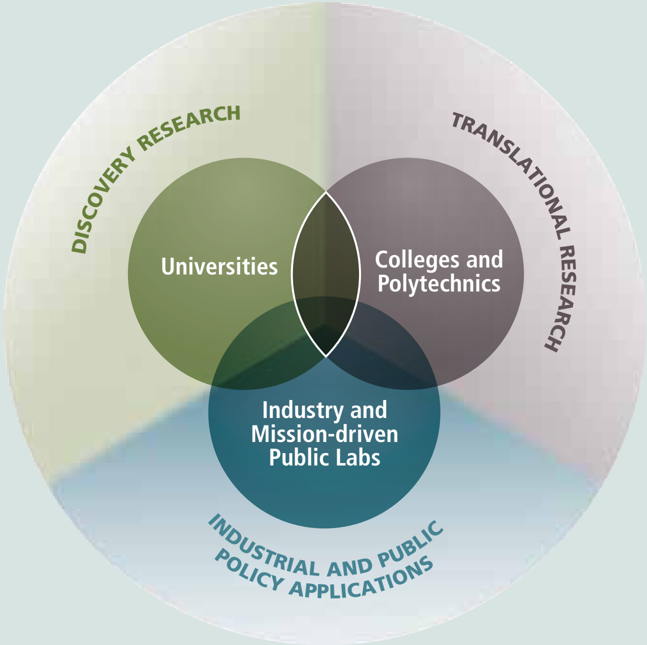


Figure 1.2

Overlapping research activity in Canada's research performing institutions

This figure shows the roles of various types of research performing institutions, all of which participate in discovery research to some degree.

1.3 ASSESSMENT METHODOLOGY

The Panel relied on two primary sources of evidence to address its charge:

- **Survey of the available literature:** The Panel's comprehensive review included literature on the use of science indicators and assessment methodologies for informing research funding allocation in the NSE. The Panel scrutinized studies on scientometrics and bibliometrics, the use of peer review and expert review in different models of science evaluation, the history and current practice of research evaluation at various levels, and past literature on research funding allocation. The Panel aimed to capture insights from rigorous scholarship that illuminated key aspects of the charge. Evidence that emerged from this literature review, which is referred to throughout the report, played a major role in informing the Panel's conclusions.
- **International case studies:** The Panel was charged with determining what the approaches used by funding agencies around the world had to offer about the use of science indicators and related best practices in the context of research in the NSE. As a result, the Panel developed detailed case studies on 10 selected countries. The purpose of these case studies was two-fold: (i) to ensure that the Panel had a fully developed, up-to-date understanding of indicators and practices currently used around the world; and (ii) to identify useful lessons for Canada from the experiences of research funding agencies in other countries. Findings and instructive examples drawn from these case studies are highlighted and discussed throughout this report. Summaries of the 10 case studies are presented in Appendix A.

The 10 countries selected for the case studies satisfied one or more of the following four criteria established by the Panel:

- **Knowledge-powerful countries:** countries that have demonstrated sustained leadership and commitment at the national level to fostering science and technology and/or supporting research and development in the NSE.
- **Leaders in science assessment and evaluation:** countries that have notable or distinctive experience at the national level with use of science indicators or administration of national science assessment initiatives related to research funding allocation.
- **Emerging science and technology leaders:** countries considered to be emerging "knowledge-powerful" countries and in the process of rapidly expanding support for science and technology, or playing an increasingly important role in the global context of research in the NSE.

- **Relevance to Canada:** countries known to have special relevance to Canada and NSERC because of the characteristics of their systems of government or the nature of their public research funding institutions and mechanisms.

The 10 countries selected by the Panel are as follows: Australia, China, Finland, Germany, the Netherlands, Norway, Singapore, South Korea, the United Kingdom, and the United States. The research undertaken for each country consisted of a number of components. A review of previous studies, along with relevant government documents, reports, and websites, provided an initial base of information for each country. The Panel also used a short online questionnaire to collect data from research funding agencies on the types of science indicators used and their relevance to particular research funding allocation processes. Finally, it conducted a series of interviews with representatives from research funding agencies (and other relevant organizations) to gain insights into their experiences with science indicators and assessment methodologies, and to validate the information and insights gained from previously published sources.

It should be noted that the Panel's survey of international experiences in the use of science indicators and assessment methods was not limited to the 10 countries featured in the case studies. The Panel also reviewed examples of relevant practices from many jurisdictions, all of which were considered in its deliberations but are not reported in detail here.

1.4 REPORT STRUCTURE

The report is structured as follows:

- Chapter 2 provides an overview of the Canadian research funding context, focusing on the federal S&T strategy and main features of NSERC's DGP. The chapter also describes recent international experiences with science funding and assessment, and summarizes key lessons that emerge from the Panel's international case studies.
- Chapter 3 discusses the funding allocation decision process at the field level and describes the roles played by indicators, expert judgment, and policy decisions.
- Chapter 4 presents a general review of available science indicators and assessment strategies. It highlights key aspects of deliberative and quantitative approaches to science assessment and discusses general issues related to indicator types.
- Chapter 5 introduces the Panel's set of evaluative criteria for assessing the validity of quantitative indicators used for research quality and research trends. The chapter reviews available indicators and methods for assessing research quality, and summarizes the Panel's evaluation of these indicators.

- Chapter 6 reviews available quantitative indicators and evaluative methods for assessing research trends.
- Chapter 7 presents the types of indicators available for assessing of research capacity, provides examples of these indicators, and summarizes their key features.
- Chapter 8 summarizes the Panel's overall findings and responses to the questions comprising its charge.

Three appendices provide additional technical information related to the report.[1] Appendix A provides summaries of the 10 international case studies undertaken for this assessment, while Appendix B contains the full case studies and reference list. Appendix C contains details of the review of quantitative indicators.

A Note on Terminology

The language used by policy-makers sometimes differs from that used by scientists. Even within the literature on science assessment, there can be inconsistency in the use of terms. For purposes of this report, the Panel employed the following definitions:*

Discovery research: inquiry-driven scientific research. Discovery research is experimental or theoretical work undertaken primarily to acquire new knowledge of the underlying foundations of phenomena and observable facts, without application or intended use (based on the OECD definition of "basic research" in OECD, 2002).

Assessment: a general term denoting the act of measuring performance of a field of research in the natural sciences and engineering relative to appropriate international or global standards. Assessments may or may not be connected to funding allocation, and may or may not be undertaken in the context of the evaluation of programs or policies.

Scientometrics: the science of analyzing and measuring science, including all quantitative aspects and models related to the production and dissemination of scientific and technological knowledge (De Bellis, 2009).

Bibliometrics: the quantitative indicators, data, and analytical techniques associated with the study of patterns in publications. In the context of this report, bibliometrics refers to those indicators and techniques based on data drawn from publications (De Bellis, 2009).

continued on next page

1 Appendix A is available at the end of this document. Appendix B and C are available online at www.scienceadvice.ca.

Deliberative methods: science assessment strategies based on the use of judgment and deliberation (NRC, 2006). Deliberative methods include both expert review and peer review.

Quantitative indicators: any indicators constructed from quantitative data (e.g., counts of publications, citations, students, grants, research funding).

Fields: broadly related areas of research. This use of "field" is not specific to the level of detail; fields can be broad (e.g., chemistry) or narrow (e.g., spectroscopy). In the context of this report, the term generally implies broad research areas unless otherwise specified.

Research quality: the quality or calibre of research as determined by the values and standards of the scientific community. Research quality is a complex, multidimensional attribute that takes into account various factors such as originality, rigour, and scientific impact (but does not include consideration of broader socio-economic impacts of research) (REPP, 2005; Butler, 2007).

Research trends: the various trends related to evolution of scientific research such as emerging or declining fields of study, changing research foci, new patterns of collaboration, etc.

Research capacity: the overall capacity for undertaking or performing scientific research in a field, as determined by factors such as available infrastructure and facilities, funding levels, human resources, cost of research, and nature of existing partnerships and collaborations, etc.

---

* Where other sources have been drawn upon in the development of these definitions, these are noted.

2

Science Funding and Assessment: The Canadian Context and International Experience

- The Canadian Context: Federal Support for NSE Research
- The International Context: A Review of Science Assessment Practices Abroad
- Conclusions

2 Science Funding and Assessment: The Canadian Context and International Experience

Key Points

- Public support for discovery research in the NSE in Canada occurs within the context of the federal S&T strategy.
- The Natural Sciences and Engineering Research Council's Discovery Grants Program is the main federal mechanism for supporting discovery research in the NSE in Canada. An international review of the DGP found it to be highly effective in meeting its goals; however, concerns have been raised in the past that allocation of DGP funding across fields is overly dependent on historical funding patterns.
- Recent international experiences with science assessment initiatives reveal that (i) the national research funding context invariably affects the nature of any assessment activity undertaken, and (ii) many countries increasingly rely on a combination of deliberative methods and quantitative indicators in national research assessment.

Nearly all national governments around the world provide support for discovery research to some degree, within the capacity of their fiscal and scientific circumstances. Accounting for the full, public value of this support in concrete terms, however, is invariably problematic, and justifying continued support for scientific research is a perennial challenge. Several central features of scientific research drive these challenges. First, the benefits from scientific advances often accrue only in the long run and are difficult to link back to specific funding programs or projects. The full extent of the social or economic benefits that result from any scientific advance may be known only decades (perhaps centuries) after the research is undertaken (see Box 2.1).

Second, driven by the need for public accountability, public research funding agencies (along with other government departments and agencies) are often required to set measurable performance goals and quantify the extent to which those goals are achieved. It is difficult, however, to define appropriate performance objectives for research funding agencies that are both meaningfully connected to the fundamental objectives and mandates of those organizations and empirically measurable.

Third, public research funding agencies are under continual pressure to justify their expenditures in light of future expectations about the benefits of particular lines of research. But predictions of benefits likely to result from any specific investment in discovery research are unreliable. As stated in a 2006 U.S. National Research Council report on assessing science: "No theory exists that can reliably predict which research activities are most likely to lead to scientific advances or to societal benefit" (NRC, 2006). As a result, funding agencies cannot reliably differentiate between the values that will eventually result from investments in different research types, fields, or areas.

Box 2.1
The Unpredictable Benefits of Scientific Discovery

It might seem that the path of scientific discovery is direct: identify a problem, and then find a solution. A 1916 speech by J.J. Thomson, winner of the 1906 Nobel Prize in physics, extolled the benefits of encouraging discovery research — research unencumbered by a driving need for industrial application. He highlighted the use of x-rays in surgery: "Now, how was this method discovered? It was not the result of a research in applied science starting to find an improved method of locating bullet wounds. This might have led to improved probes, but we cannot imagine it leading to the discovery of x-rays in surgery" (Rayleigh, 1942).

Rather, x-rays were discovered by chance in November 1895 while Wilhelm Roentgen was experimenting with a Crookes tube. And a young man shot on Christmas Day 1895 in Montréal, Quebec became the first patient to have surgery connected to an x-ray. Traditional methods had been unable to locate the bullet, and doctors chose to let the wound heal naturally. A Crookes tube assembled at McGill University, however, was used to generate x-rays according to Roentgen's discovery. In February 1896 it was used to locate the position of the bullet, which was then successfully removed (Sullivan, 2011; Gingras, 1987).

The liquid crystal display (LCD) flat-screen television provides another relevant example. The basic science behind LCD began in 1888 when Austrian botanist and chemist Friedrich Reinitzer discovered liquid crystals. At the time, all Reinitzer knew was that cholesteryl benzoate, a compound extracted from carrots, melted at 145˚C, yet stayed cloudy until 179˚C (Gray, 2009). A year later, German physicist Otto Lehmann discovered that even though cholesteryl benzoate moved like a liquid at 145˚C,

continued on next page

it retained a crystal structure at temperatures up to 179°C (Nobelprize.org, 2012). It took another 80 years for this discovery to move into application, when Richard Williams discovered that passing a voltage through a thin layer of liquid crystals generated a pattern, and then another 20 years before this invention began to show up in living rooms (Magoun, 2007).

Finally, consider the invention of magnetic resonance imaging (MRI), a technique that has saved countless lives because of its ability to detect tumours, heart problems, and other ailments. Its origins trace to the 1930s when Isidor Rabi, a New York physicist, decided to investigate atomic properties of sodium — nuclear spin and magnetic moment. He discovered that when exposed to magnets, nuclei will line up parallel or anti-parallel to a magnetic field and will flip if exposed to an electromagnetic wave of a frequency specific to the material. When organisms such as the human body are exposed to a strong magnetic field, many of the hydrogen nuclei in the body align with the direction of the magnetic field. Depending on the intensity of the field, this information can be used to develop 3-D spatial imaging. It wasn't until 1977 that the first images of a human were produced using this technique. And when Rabi himself was imaged in 1988, he said: "I never thought my work would come to this" (Rigden, 2000).

None of these challenges, however, have prevented governments from continuing to support discovery research in the NSE. The latest evidence on the extent of world scientific activity indicates that public spending on scientific research continues to grow worldwide (OECD, 2011). But there is increased pressure on research funding agencies to improve science assessment tools and methods, thereby strengthening the evidence base on which research funding allocation decisions are made.

As science assessment practices around the world continue to evolve, research funding agencies are presented with more options for supporting, informing, and assessing resource allocation decisions. This chapter sets the stage for an examination of these developments. The first section provides an overview of the key features of the Canadian NSE funding context, as related to the current interest of NSERC in assessment practices and indicators. The second section describes recent international experiences related to science assessment and research funding allocation, drawing heavily on the Panel's 10 international case studies.

2.1 THE CANADIAN CONTEXT: FEDERAL SUPPORT FOR NSE RESEARCH

In Canada the majority of grant-based research funding from the federal government is channelled through three funding agencies known collectively as the Tri-Council. Each granting agency has a mandate to fund a broad base of research in its sphere of activity. The Canadian Institutes of Health Research (CIHR) provide opportunities for biomedical, clinical, and health systems services, and social, cultural, environmental, and population health. The Social Sciences and Humanities Research Council of Canada (SSHRC) supports research in the social sciences and humanities, and NSERC supports research activity in the NSE.

Two aspects of the Canadian NSE funding landscape are particularly relevant to the Panel's charge: (i) the overall policy context for NSE funding in Canada as determined by the federal government's S&T strategy, and (ii) the objectives and characteristics of NSERC's Discovery Grants Program.

2.1.1 The Federal S&T Strategy

Federal support for Canadian discovery research in the NSE occurs within the context of Canada's S&T strategy, as presented in *Mobilizing Science and Technology to Canada's Advantage* (Industry Canada, 2007). This strategy, created in 2007, is intended to provide general direction and an overarching policy framework for all federal programs that support science and technology in Canada. It defines three broad policy goals for federal departments and agencies involved in the support of S&T in Canada, including NSERC. The goals are described in terms of fostering three Canadian S&T "advantages:"

- **Entrepreneurial advantage:** translating "knowledge into commercial applications that generate wealth for Canadians."
- **Knowledge advantage:** ensuring that Canada is "positioned at the leading edge of important developments that generate health, environmental, societal, and economic benefits."
- **People advantage:** ensuring that Canada remains "a magnet for the highly skilled people we need to thrive in the modern global economy with the best-educated, most-skilled, and most flexible workforce in the world."

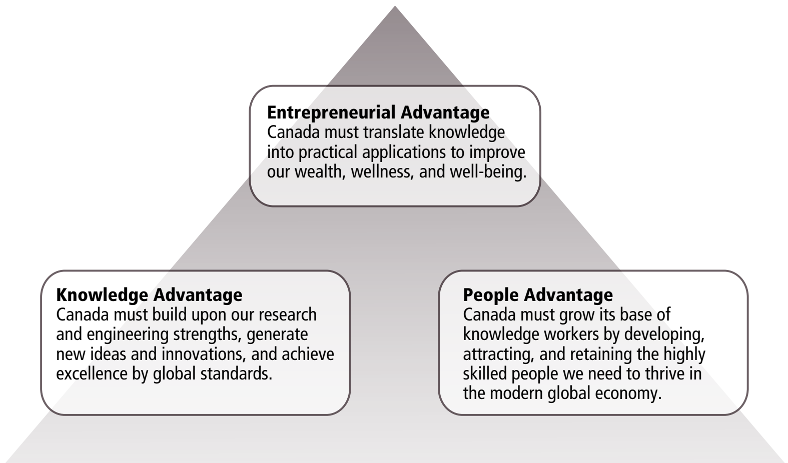
(Industry Canada, 2007)

The federal S&T strategy also articulates four core principles intended to guide government policy with respect to science and technology: i) promoting world class excellence; ii) focusing on priorities; iii) encouraging partnerships; and iv) enhancing accountability. The first of these principles reiterates the government's commitment to ensuring that "Canadians perform at world-class levels of scientific and technological excellence" (Industry Canada, 2007). The second states the government's intention to be strategic about focusing support on areas of strength and opportunity. The third expresses its intention to support cross-sectoral (e.g., business-university) S&T collaboration; and the fourth to strengthen governance and reporting practices related to federal support for S&T. Together, these principles and "advantages" comprise the current policy framework that governs federal support for discovery research (and S&T in general) in Canada (see Figure 2.1).

CANADA'S FEDERAL SCIENCE AND TECHNOLOGY FRAMEWORK

Vision: We will build a sustainable national competitive advantage based on science and technology and the skilled workers whose aspirations, ambitions, and talents bring innovations to life.

To achieve this vision, we will create three S&T Advantages for Canada:



Government actions will be guided by four core principles:
- Promoting world-class excellence
- Focusing on priorities
- Encouraging partnerships
- Enhancing accountability

Reproduced from Industry Canada, 2007

Figure 2.1
Canada's Federal S&T Framework

Understanding the federal S&T strategy is important for this report for two reasons. First, NSERC's funding programs (along with most federal programs that support S&T in Canada) are broadly organized to reflect the federal strategy's three main goals. NSERC's suite of programs is divided into three categories, corresponding to three strategic outcomes: those that focus on innovation, those that focus on discovery, and those that focus on people (i.e., building human capital) (NSERC, 2011a). While not identical to the three overarching goals of the federal strategy, the choice of these strategic outcomes was clearly driven in part by the federal policy context. NSERC's programs, which fall under the "discovery" heading, are aligned with the "knowledge advantage." The DGP is the largest of these discovery-oriented programs.

Second, the principles articulated in the federal strategy have implications for science assessment activities undertaken by NSERC. These principles express many of the values governments typically espouse for science assessments of publicly funded research: for example, the desire to improve accountability for public expenditures on scientific research, to prioritize research support in areas of national strength or opportunity, and to ensure the results of this support are world leading (see OECD, 2010) for a general discussion of these motivations). As a result, these principles provide a direct motivation for NSERC (and other federal departments and agencies that support research activity) to assess and evaluate research investments on an ongoing basis.

2.1.2 NSERC and the Discovery Grants Program

The Discovery Grants Program is NSERC's primary funding mechanism for supporting discovery research in Canada and its oldest and largest funding program.[2] The DGP is also one of the most important general sources of support for discovery research in the NSE in Canada. Receiving its support is often critical for researchers in attracting additional funding from other sources, which, in many cases, substantially exceeds DGP funding (NSERC, 2008).

As mentioned in Section 2.1.1, NSERC's funding programs are currently organized around three strategic outcomes: building human capital, strengthening discovery research, and supporting industry-academic partnerships in transforming knowledge into innovation (NSERC, 2011a). Figure 2.2 shows the current breakdown of NSERC's grants and scholarships programs according to these categories. Roughly one-third of NSERC's current spending is linked to supporting discovery research in Canada, and the DGP comprises the bulk of that support.

2 The DGP, which has gone by several names since NSERC's creation in 1978, was previously known as Operating Grants and Research Grants (NSERC, 2008).

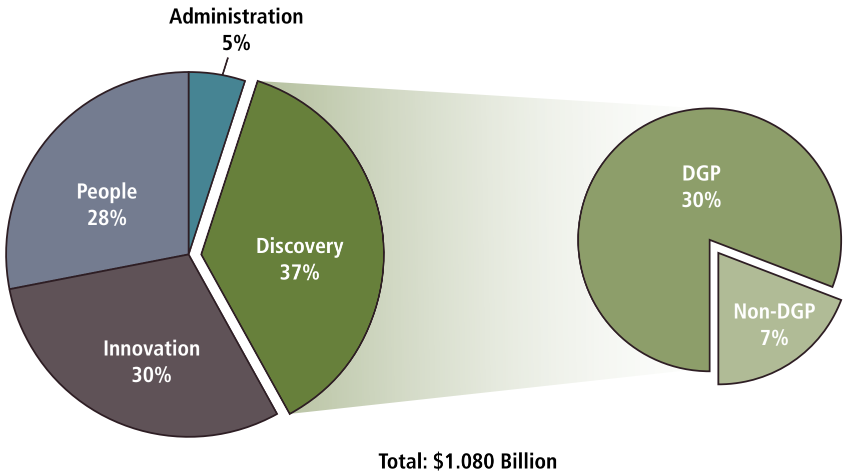


Figure 2.2

NSERC Budget Allocation 2010–2011 by Strategic Outcome (millions of dollars)

This figure shows the division of NSERC's budget across the organization's three strategic outcomes for 2010-11. Over one-third of NSERC's annual budget is invested in discovery research and funding for the Discovery Grant Program accounts for the majority of those expenditures.

The stated goals of the DGP are to:

- promote and maintain a diversified base of high-quality research capability in the natural sciences and engineering in Canadian universities;
- foster research excellence; and
- provide a stimulating environment for research training.

(NSERC, 2011b)

Individuals and teams of researchers can apply for Discovery Grants, which are normally five years in duration (longer than typical international project-based funding) (NSERC, 2008). One of the most notable features of the DGP is that it targets support towards research programs, rather than projects. Once a researcher is funded through a Discovery Grant, the funding is not limited by the activities described in the grant application. Rather, the researcher can shift resources to new research interests if they arise — provided they fall within NSERC's mandate (NSERC, 2011b). Discovery Grants have a discipline-specific minimum amount to ensure that each recipient receives a level of support sufficient to support a research program that "can have a meaningful impact on the field of study" (NSERC, 2011b). Discovery Grants are distinctive among research funding programs in that they can only be applied to direct research costs and not to faculty salaries or institution overhead or indirect costs (NSERC, 2008).

Past evaluations undertaken by NSERC have generally found that the DGP is a highly effective funding program. For example, an international review conducted in 2008 concluded that the DGP had "supported Canada's best researchers at an internationally competitive level" and struck an appropriate balance between its two objectives of "promoting and maintaining a diversified base of research" and "fostering research excellence." The review committee concluded that the DGP "is an exceptionally effective model for supporting Canadian research in the NSE fields" (NSERC, 2008).

The DGP has periodically been criticized since allocation of funding across fields of research relies inordinately on historical funding patterns as a baseline, overall funding levels by field are relatively inflexible to change. Individual Discovery Grants are awarded to applicants through a typical peer review process (i.e., applicants are reviewed and judged by a volunteer committee of their peers). Prior to the award of individual grants, however, funding is first divided among various evaluation groups corresponding to particular fields of research. (NSERC recently moved to a system of 12 discipline-based evaluation groups; previously, funding was allocated across 28 grant selection committees.) This initial step in the allocation process is driven primarily by a reliance on historical funding patterns, with the committees (now evaluation groups) receiving similar allotments to what they had received in previous years.

NSERC first responded to these concerns in 1991 with the creation of the DGP Reallocation Exercise to ensure that the DGP remained "dynamic and responsive to changes in the various disciplines and in the research environment" (NSERC, 2006a). It intended to accomplish this by periodically (once every four years) reallocating a certain percentage (up to 10 per cent) of DGP funding across fields. Three rounds of the DGP Reallocation Exercise were carried out (in 1994, 1998, and 2002) with the methodology evolving significantly between rounds. A 2006 evaluation of the Reallocation Exercise concluded that this aspect of the DGP was generally not meeting its objectives. Despite the completion of the three rounds of reallocation, funding for most grant selection committees changed very little, typically staying within four per cent of pre-reallocation budgets (NSERC, 2006a). Given this minimal level of change, the program's evaluators concluded that costs of the reallocation program, as currently designed, outweighed its benefits (NSERC, 2006a), and it was subsequently suspended.

Nevertheless, the review of the DGP Reallocation Exercise reiterated that the original rationale of the reallocation exercise was both still relevant and supported by members of the NSERC community. As a result, NSERC has maintained an interest in periodically reassessing how DGP funding is allocated across fields, and what assessment practices, methods, or mechanisms might be used to inform this type of allocation.

2.2 THE INTERNATIONAL CONTEXT: A REVIEW OF SCIENCE ASSESSMENT PRACTICES ABROAD

The Panel considered a wide range of international practices in science assessment and funding allocation. In particular, it focused on 10 countries: Australia, China, Finland, Germany, the Netherlands, Norway, Singapore, South Korea, the United States, and the United Kingdom (see Section 1.3 for the Panel's selection criteria and Appendix A for more detailed information on the case studies).

Two general findings emerged from the Panel's review of the international evidence:

The context in which science assessments are carried out — including the research funding landscape, the institutional and organizational context, and the national policy environment — has a significant impact on the design of assessment practices for discovery research. Diverse assessment practices are used around the world. This diversity is reflected in the way these practices are incorporated in funding allocation processes. Given differences in the characteristics of research funding systems and research performing institutions, as well as differences in national policy frameworks and organizational mandates of research funders, any expectation of a single set of assessment practices or indicators ideal for all circumstances would be misplaced. Field-level research assessment for discovery research in the NSE will necessarily vary in accordance with the national research funding context. As a result, the concept of international "best practices" offers only limited insights with respect to science indicator and assessment practices in the context of informing research funding decisions.

There is a general trend towards national science assessment models that rely on a combination of expert judgment and quantitative indicators. A number of national research assessments have separately evolved towards this type of model, often in response to extensive reviews of available assessment methodologies sponsored by the government departments and agencies managing these initiatives. There is a growing consensus — reflected in both methodological recommendations made by research evaluation experts and assessment practices adopted by governments — that the strongest approach to research assessments at the field level relies on a combination of indicators and expert judgment.

The following sections explore these and other related findings in more detail.

2.2.1 Institutional versus Project-based Research Funding

Public research funding mechanisms can be broadly divided into two general types: (i) project-based research funding, which consists of grant-based funding awarded to individual researchers, teams, or research projects; and (ii) institutional research funding, which typically consists of block transfers from governments to research performing institutions such as universities and colleges. Project-based funding is usually awarded through publicly funded research councils (such as NSERC), with individual grant applications evaluated primarily in a peer review process. Institutional funding is allocated directly to universities from central or regional governments, with allocations often driven by funding formulas.

Project-based Research Funding

Information on the allocation of project-based research funding at public research funding agencies tends to be difficult to obtain. Since high-level resource allocation decisions (e.g., the allocation of funding across DGP evaluation groups in Canada) are often made at the discretion of agency executives or managers, descriptions of the allocation and decision-making processes involved may not be available in public documents. In addition, there is very little peer-reviewed literature on these decisions and allocation processes. The Panel was able to gather information on this type of decision for 4 of the 10 countries selected for the case studies: Australia, Germany, Finland, and Norway.[3] As a result, no firm conclusions can be drawn from this limited base of evidence.

The evidence does suggest, however, that other project-based funders typically rely on indicators to establish baseline levels of research funding (similar to NSERC), which can then be supplemented through discretionary allocations made at the executive or management level within those agencies. This is the case in both Germany and Australia, where a large portion of initial funding allocations across fields is decided on the basis of historical funding patterns, and a smaller portion

3 The major non-directed discovery research funding programs in these countries examined by the Panel were the National Competitive Grants Program (NCGP) at the Australia Research Council (ARC), the Individual Grants Program at the German Research Foundation (DFG), the Independent Grants Program (FRIPRO) at the Research Council of Norway (RCN), and the Academy projects at the Academy of Finland.

is reserved for discretionary allocation by agency management. The reliance on historical funding data is used to maintain a certain level of consistency with past funding levels, and to take into account differences in the cost of research across fields, which are sometimes presumed to be reflected in past allocation levels (Margaret Sheil, personal communication, June 15, 2011; Asbjørn Mo, personal communication, June 30, 2011).

Institutional Research Funding

Information on the use of indicators and assessment practices to inform or determine the allocation of institutional research funding is more readily available. The allocation of a substantial share of the overall funding for higher education institutions in England, based on the results of the United Kingdom's Research Assessment Exercise (RAE), is perhaps the best-known example of this type of funding scheme, though a number of other nations have adopted "performance-based" models for the allocation of institutional research funds (reviewed in OECD, 2010). Block transfers to higher education institutions in these models are partially determined by some form of research assessment, generally directing additional funding to institutions that show a higher level of performance based on selected indicators. Of the 10 case study countries, 6 have institutional funding models where a portion of funding to the higher education system is distributed based on research performance: Australia, Finland, Germany, the Netherlands, Norway, and the United Kingdom. Although these funding contexts are not directly comparable to those faced by project-based research funding organizations (such as NSERC), assessment practices related to the allocation of institutional research funding have been more widely studied and thus may hold important lessons for NSERC and Canada.

Performance-based models for allocating institutional research funding differ in both the processes by which assessments of research performance are undertaken and the means by which assessments are incorporated into resource allocation processes. In general, even where these mechanisms exist, the "performance-based component" of institutional research funding is usually a relatively small portion of total funding allocations. A large portion of this type of funding stream is usually based on historical funding levels; for example, historical funding levels account for 60 per cent of Norway's funding to institutions (OECD, 2010). In addition, changes in the overall levels of funding provided to institutions under these arrangements are often limited by historical funding levels. For example, in Australia, one institutional block grant (the Joint Research Engagement, JRE) does not allow funding levels to drop below 95 per cent of the previous year's funding level (DIISR, 2011a).[4] Most German states (German state governments control institutional funding) also implement a "tolerance band," linked to the previous year's budget, which limits the maximum potential gain or loss of funds from the performance-based portion of funding (OECD, 2010).

Several types of quantitative indicators are commonly used in these funding systems — most prominently measures of external (i.e., from other sources) research funding, measures of research output, and measures of student or faculty populations. Indicators based on external research funding are used in formulas for the allocation of block grants to higher education institutions in Australia, Norway, Finland, and Germany (see OECD, 2010; Geuna & Martin, 2003; DIISR, 2011b). In Norway, 30 per cent of research funding for universities is allocated based on indicators of this type, including external funding amounts from the Research Council of Norway (RCN) and the EU Research Framework Programme (OECD, 2010).

Many of these funding systems also allocate funding in proportion to one or more indicators of research output. For example, in Norway, 30 per cent of the research-based portion of institutional funding is based on publication output adjusted on three levels: quality of the journal, share of authorships, and publication form (e.g., book, article) (OECD, 2010). Publication output is 10 per cent of Australia's JRE funding budget, including books, book chapters, journal articles, and conference papers, averaged over the two most recent years (DIISR, 2011a). Finland also allocates a small amount of its research performance-based funding based on publication counts (five per cent)[5] (OECD, 2010). Many of these systems use similar indicators. Finally, some of these systems, most notably the U.K. RAE, rely on expert judgment to inform assessments of research quality, which then determine funding allocations.

The use of performance-based funding schemes for the allocation of institutional research funding appears to be increasing (OECD, 2010), and there is a growing body of evidence on how indicators and assessment practices are affecting national

4 On December 15, 2011 the Department of Industry, Innovation, Science, Research and Tertiary Education was established, replacing DIISR.

5 This indicator may be given greater weight in the future and adjusted based on journal quality. It is a focus of current discussions within Finland's Ministry of Education's working group on institutional reform (Anita Lehikoinen, personal communication, June 15, 2011).

research systems. The following sections discuss two salient issues: the potential behavioural impacts of national research assessment systems on the research community, and the role of expert judgment in assessment and funding allocations.

2.2.2 The Behavioural Impacts of Research Assessment and Evaluation Practices

Linking research funding decisions directly to indicators or assessment results in a formulaic manner invariably produces incentive effects, both positive and negative, which can have unintended consequences for the behaviour of individual researchers and research institutions. This topic has mostly been explored in the context of institutional research funding, since these allocations are often based directly on quantitative indicators.[6]

One of the most frequently examined types of behavioural impact is the effect of assessment systems or indicators on research output. Correlations have been observed between the use of indicators of research volume in a funding formula and subsequent increases in research output. Correlation does not imply causation, but presumably some of these increases arise due to behavioural incentive effects. Most of these observations stem from the use of publication count indicators. For example, the Ministry of Education and Research in Norway found the use of a publication indicator in the research performance-based portion of institutional funding had contributed to an increase in Norwegian publications, even though the publication indicator determined less than two per cent of total funding for tertiary education institutions (OECD, 2010). An Australian study concluded that the boost in the total output of Australian research articles in the mid-1990s was likely a result of institutional funding formulas incorporating publication output indicators (Butler, 2003). In the United Kingdom institutional funding is allotted in part based on the RAE. Although the RAE is based on expert review, panel members take into account publication output. Moed (2008) reported evidence suggesting that the RAE has had substantial impacts on U.K. research output, and that those impacts have varied through time depending on the different emphases of successive assessment rounds.[7]

6 See Chapter 4, OECD (2010) for a review of such impacts.

7 In the early years following the introduction of the RAE, total output of research publications in the United Kingdom increased. In response to a greater emphasis on quality over quantity in the 1996 RAE, the United Kingdom's share of world publications declined, but its share of articles in highly cited journals increased. In recent years the overall share of research publications appears to be increasing again, possibly in response to attempts to increase the number of staff included as research active (Moed, 2008; OECD, 2010).

The potentially negative aspect of this impact, however, is that when funding is explicitly linked to research output, researchers will be tempted to produce a higher *quantity* of publications at the expense of their *quality*. Some evidence bears out these concerns. For example, one study found that in the mid-1990s, a period in which Australia's total output of research articles increased, average relative citations declined (Butler, 2003). Governments have adopted various strategies to address this problem. Norway's Ministry of Education and Research believes this risk is sufficiently reduced by adjusting publication counts based on journal quality (OECD, 2010).[8] The U.K. RAE uses an expert review process informed by data on publication outputs, but not directly determined by those outputs. This strategy appears to have been effective, as there is a general consensus within the United Kingdom that the RAE has resulted in increased research quality (Government of the United Kingdom, 2006; OECD, 2010). Bibliometric evidence appears to confirm this in so far as the overall impact of U.K. research, as measured by citation patterns, has increased since the introduction of the RAE (Adams & Gurney, 2010; Moed, 2008).

Although weighting publication indicators to reflect journal quality may help to mitigate the problem of reduced output quality, this practice can result in new types of potentially undesirable incentive effects. For example, it has been noted that the RAE has likely incentivized U.K. academics to publish in well-known, highly cited scientific journals over other outlets (McNay, 1998; OECD, 2010; RIN, 2009). But labelling well-known, highly cited scientific journals as better quality than specialized or regional peer-reviewed journals is debatable. Although a stronger international orientation may be desirable in some cases, prioritizing international journals in research assessment processes could lead to the neglect of nationally (or regionally) important topics (REPP, 2005). In addition, although journals in highly specialized fields are more likely to have lower impact factors due to their smaller readership, that does not mean that the research published in them is of lower quality or less importance.

One example of this is seen in China where there are strong incentives for academics to publish in international journals. Although China does not have performance-based institutional funding formulas, many Chinese universities and research institutes frequently provide financial bonuses to academics based on journal publications, with larger rewards provided for publications in more prominent (or more highly cited) journals or journals included in Thomson Reuters' *Science Citation Index*. For example, at Zhejiang University, first authors

8 Finland's institutional funding formula will likely soon be changed to follow suit (Anita Lehikoinen, personal communication, June 15, 2011).

who publish in *Nature* or *Science* are rewarded 200,000 RMB (about C$30,000),[9] while authors publishing in any other journal are rewarded with lower sums of between 600 RMB (about C$90) and 14,000 RMB (about C$2,000) depending on the impact factor of the journal (Shao & Shen, 2011). Incentive effects from this type of practice may have contributed to the tremendous growth in China's record of scientific publications in the past decade, but policies that incentivize publication in international journals likely bias researchers against working on regionally specific issues and/or publishing in Chinese language journals (Shao & Shen, 2011).

Another type of incentive effect arising from the use of indicators in funding formulas is goal displacement where achieving high scores in an indicator become the goal rather than achieving a performance objective or level that is then measured by the indicator (REPP, 2005). Researchers and institutions can end up trying to find loopholes in order to inflate their performance on assessment indicators. As one example, indicator definitions for Australian institutional funding used to be defined such that conferences had to be "international" and journal editorial boards had to extend past a single institution. One university overcame the first restriction by including at least one international participant in conferences; one researcher overcame the second restriction by forming a journal editorial board composed of former graduate students (OECD, 2010).

A recent study has suggested that the institutionalized practice of providing scientists and researchers with financial bonuses based on publications has had serious ramifications for Chinese scientific integrity. Surveys have found that as many as one in three researchers working in Chinese universities admit to having plagiarized, falsified, or fabricated data in order to publish more quickly and in more prominent journals (Qiu, 2010). Research has also revealed the emergence of a substantial market for ghost-writing papers on nonexistent research, with illicit websites providing services such as fictional research papers, "bypassing peer review for payment, and forging copies of legitimate Chinese or international journals" (Qiu, 2010). The Chinese experience with publication-based bonuses may therefore argue for caution when considering a direct linkage between financial incentive for researchers and indicators based on research outputs.

Other types of impacts on researcher behaviour are also possible, and, in some cases, equally important. Besides the incentive effects noted above, the U.K. RAE has had substantial, unintended impacts on institutional hiring practices and

9 Canadian dollar figures calculated based on the exchange rate for May 2, 2011.

departmental structure, as well as effects on researcher morale, collaboration, autonomy, and research focus (OECD, 2010). For instance, the RAE has led to a "transfer market" in research faculty between and within institutions. Universities compete for leading researchers in advance of the assessments, which has led to an increasing variety of positions and contracting options for researchers. Evidence suggests that the RAE has led some U.K. universities to focus on hiring younger staff with research potential, while others have taken a more conservative approach and focused on hiring well-established researchers (HEFCE, 1997; OECD, 2010). Several universities have closed specific departments in response to poor ratings (OECD, 2010). Researcher morale and collegiality were found to be adversely affected in some cases due to sensitivities around which personnel within a department were selected as "research-active" staff in RAE submissions (HEFCE, 1997). Institutional preoccupation with RAE scores and their funding implications has led university administrators and managers to increasingly control the overall research directions of their staff and departments (McNay, 1998; OECD, 2010). The RAE may discourage high-risk research, especially in cases where researchers worry that they would not be able to generate research outputs in advance of the next round of assessment (Evaluation Associates Ltd., 1999; McNay, 1998; OECD, 2010). Researchers have also expressed concerns that the RAE does not do enough to encourage research collaboration, particularly with researchers outside higher education institutions (OECD, 2010; Evaluation Associates Ltd., 1999).

2.2.3 The Role of Expert Judgment in Research Assessment

The international case studies developed for this report reveal important lessons about the role of expert judgment in research assessment. A general finding from this review points to an evolution towards increased reliance on both deliberative methods and quantitative indicators in national research assessment exercises.

Perhaps the most illuminating examples of this trend come from Australia and the United Kingdom. The long-standing U.K. RAE is now scheduled to be replaced by the Research Excellence Framework (REF) (HEFCE, 2011b). Past reviews of the RAE vindicated its core reliance on peer review (e.g., Roberts, 2003), and newer research undertaken on bibliometric indicators for the REF concluded: "Bibliometrics are not sufficiently robust at this stage to be used formulaically or to replace expert review in the REF. However there is considerable scope for citation information to be used to inform expert review" (HEFCE, 2009). As a result, while the REF process as currently planned will include bibliometric indicators and benchmarks, these will not replace core reliance on the judgment of experts. Australia's recently developed national assessment process, Excellence in Research for Australia (ERA), relies on a similar combination of metrics and deliberation: bibliometric indicators and benchmarks are used, but ultimately are subject to review by panels of experts with the final responsibility for assessment outcomes (ARC, 2011).

Many other examples of international assessment practices combine elements of both quantitative analysis and deliberation. For example, while the United States does not have a national research evaluation exercise similar in scope to that of the United Kingdom, the U.S. National Research Council (NRC) has in the past undertaken international benchmarking exercises of research fields where expert panels assessed U.S. research performance and capacity by field using a variety of underlying quantitative data (e.g., NRC, 2000). A similar model of expert panel evaluation informed by metrics is found in Finland (e.g., Academy of Finland, 2011). The U.K. Engineering and Physical Sciences Research Council (EPSRC) uses a similar, informed panel-based format for its international reviews (e.g., EPSRC, 2011).

Research related to development of the U.K. RAE and the German Council of Science and Humanities (WR) *Research Ratings* is also informative in this regard. In both cases, expert review informed by quantitative indicators was found to be the best method for assessing research quality. In 2003 the U.K. government sponsored an independent review of the RAE, which validated its overall approach and emphasized future evaluations should remain founded on an expert review process. The review stated: "We [the panel] are now convinced that the only system which will enjoy both the confidence and the consent of the academic community is one based ultimately upon expert review" (Roberts, 2003).

In its 2004 report, *Recommendations for Rankings in the System of Higher Education and Research in Germany*, the WR concluded that in order to be effective and accurate, comparisons of research quality require informed expert review:

> In the light of international experience, the Science Council rules out research assessment systems that are solely based on quantitative indicators, as well as mere reputation-based ratings. A comparison of the quality of research performance requires a research area-specific assessment in the form of a peer review carried out on the basis of harmonised data and quantitative indicators ('informed peer review') on a predefined assessment scale (WR, 2004).

Four years later, this conclusion was supported by an analysis of its pilot *Research Ratings*, which found research quality would have frequently been rated differently if the grades had been computed by weighting quantitative indicators without further review (WR, 2008a).

In general, incorporation of expert judgment adds value to research assessments in several ways. Peers or experts in the field come equipped with background knowledge about characteristics of the research field, and, in the WR *Research Ratings* for example, take into account contextual information, such as innovative achievements or periods of fundamental change, that may undermine conclusions drawn from quantitative indicators (WR, 2008a). The WR also noted that the assessment of qualitative information is valuable, such as reading published work or self-characterizations provided by research groups (WR, 2008b). In addition, quality assessed by an expert review process often results in fewer incentive effects. Linking funding decisions to more complex assessment methods prevents focusing on one or several indicators to improve performance. Furthermore, the Roberts (2003) review of the U.K. RAE stated: "We are […] convinced that only a system based ultimately upon expert judgment is sufficiently resistant to unintended behavioural consequences to prevent distorting the very nature of research activity."

The implication is not that quantitative indicators should be eliminated from research assessments, but rather they should be used to inform expert deliberations. The WR pointed out several benefits of quantitative metrics. For example, their use prevents a review exercise from becoming an assessment based on reputation. The WR's analysis of the pilot *Research Ratings* found that, upon appraisal of the data, reviewers were likely to assess well-known units more critically than would have been expected on the basis of reputation (WR, 2008a). In addition, the WR noted that an analysis employing quantitative data saves reviewers time and helps ensure reliability (WR, 2004). Aksnes (2009) further notes that bibliometric analyses can save time, and that basing deliberations on data increases the credibility of the panel by increasing objectivity. Finally, quantitative data can also be used as a trigger to the recognition of anomalies, serving to flag areas where further expert scrutiny is warranted (Butler, 2007).

International experience does suggest, however, that an assessment process including expert review can be expensive and time consuming, and that such models are perhaps not suitable in every context. Although the recommendation was not implemented, the Roberts (2003) review of the RAE suggested that peer review should be used only for top research universities receiving the bulk of research funding, due to the administrative burdens created by the evaluation system.

An international expert review of the state of the Finnish innovation system, inspired by the U.K. RAE, recommended incorporating a "light" expert review into Finland's performance-based institutional funding scheme in order to better determine the research quality at Finnish institutions (Ministry of Employment and the Economy, 2009). The Finnish Ministry of Education and Culture's working group on institutional funding reform discussed adding an expert review component to the quality dimension, but was concerned that this approach would be too complicated for Finland's relatively small university system.

2.3 CONCLUSIONS

The focus of NSERC on science assessment practices is directed partly by a long-standing concern that the allocation of DGP funding across fields is overly dependent on historical funding patterns, and that future allocations should incorporate other factors such as research quality, changes in the scientific landscape, and the emergence of research fields.

This review of international science assessment reveals a diverse landscape of assessment methods and practices. Two of the lessons emerging from the review are especially relevant to the Panel's charge. First, the national research context is significant in defining a given science assessment, and no single set of indicators for assessment will be ideal in all circumstances, though evidence gathered from examining experiences of other countries may help inform the development of a science assessment strategy for Canada. Second, there is a global trend towards national science assessment models that incorporate both quantitative indicators and expert judgment.

3

The Research Funding Allocation Process

- The Research Funding Ecosystem
- The Funding Allocation Decision Process
- The Role of Indicators and Evaluation Methods
- The Role of Expert Judgment
- The Role of Policy Decisions
- Conclusions

3 The Research Funding Allocation Process

Key Points

- The mandate of a funding program must be clearly defined as it will drive the selection of indicators and evaluation approach, as well as application of indicator-based information in the allocation decisions.
- Transparency, timeliness, credibility, and efficiency are important characteristics of the funding decision process, requiring clear communication of the strategy employed to address competing demands.
- Funding allocation decisions require expert judgment combined with a relevant and valid suite of indicators and a thorough understanding of the policy and funding context.
- Indicators and expert judgment guide the collection of information to be interpreted for allocation decisions, which flow back to the policy domain where the final funding allocation is decided. Overall effectiveness of a funding program is predicated on this closed feedback decision loop.
- Research funding allocation decisions in the natural sciences and engineering require sets of indicators that capture information on research quality, research trends, and research capacity.

While this report focuses on indicators and deliberative methods for discovery research funding decisions, indicators and methods are applied within a larger context that includes performance management and public policy, as well as the funding allocation decision itself. Structured and systematic evaluation and measurement are essential components of a modern performance management system, and can provide a practical and efficient mechanism for both data collection and assessment of the effectiveness of investment in science. Indicators, when developed and used properly, play a pivotal role in illuminating the complexity of the research funding ecosystem. They can provide answers to various information needs, though mostly with retrospective accounts. Scientists may look to available indicators for information on research in their field: how their work, or that of their colleagues or institution, compares to others in the field; their impact; and what new opportunities are on the horizon. Similarly, policy-makers routinely use indicators to seek insights about and evidence of the various socio-economic consequences of publicly funded scientific research, or to evaluate the relative strengths and weaknesses of their national research systems (e.g., STIC 2009, 2011; CCA, 2006).

Understanding how indicators can be used effectively in research funding allocation decisions at the field level requires a closer look at the allocation decision process. This chapter first examines the broader national and specific funding program contexts in which these decisions occur — the research funding ecosystem. It then outlines the basic steps of the decision process, using a logic model approach to organize the information relevant to the decision and delineate the role of indicators and expert judgment in the process. The final section offers a brief discussion of the responsibilities of policy decision-makers in deciding how best to allocate funding across fields within the broad research funding ecosystem.

3.1 THE RESEARCH FUNDING ECOSYSTEM

There is a temptation in public policy to equate priority setting with defining preferred areas of research for funding (Stewart, 1995). Setting priorities for science, however, enables countries to develop specific policy direction for science funding in general. For discovery research, these priorities translate into various funding programs with specific funding objectives and goals. The practices observed in other countries suggest that both top-down (i.e., mission-driven funding) and bottom-up (i.e., competitive grants) approaches are necessary to reflect the multifaceted nature of the scientific research process. The interplay among science funding programs, and their correlation to other national policies (especially education policies), limits the ability to determine how specific science policy instruments work. Thus a broadly accepted view is that in funding discovery research in the NSE, a pluralistic approach (i.e., a mix of policies and programs) is more desirable than a centralized process when choosing among competing scientific priorities (Popper, *et al.*, 2000; Stewart, 1995; Bernanke, 2011).[10]

This pluralistic approach creates a complex multi-stakeholder research funding ecosystem with overlapping priorities and mandates pursued by many funding programs in parallel. Making choices about what to fund requires a better understanding of this ecosystem with its many, and often competing, pressures on funding allocation decisions. These decisions occur within the complicated context of a country's science policy, which is shaped by its societal, technological, economic, and historic experiences (see Figure 3.1). Social and political values, as well as economic events, all influence support for research. A country's choice of

10 In almost every country there are various programs for funding basic research. Each program and agency is free to make its own decisions within the purview of its mandates. In its 1995 report, the National Research Council (NRC) Committee on Criteria for Federal Support of Research and Development called for a specific methodology for priority setting and coordination of government-funded basic research in the United States (NRC, 1995). This recommendation has not been universally accepted within the United States or globally.

priorities communicates urgency in specific areas, and therefore likely influences the direction of science funding in general. For the purpose of funding decisions, the merits of a scientific field have been seen from both within the field and from an external perspective (scientific merits versus technological merits were first described by Weinberg, 1963).

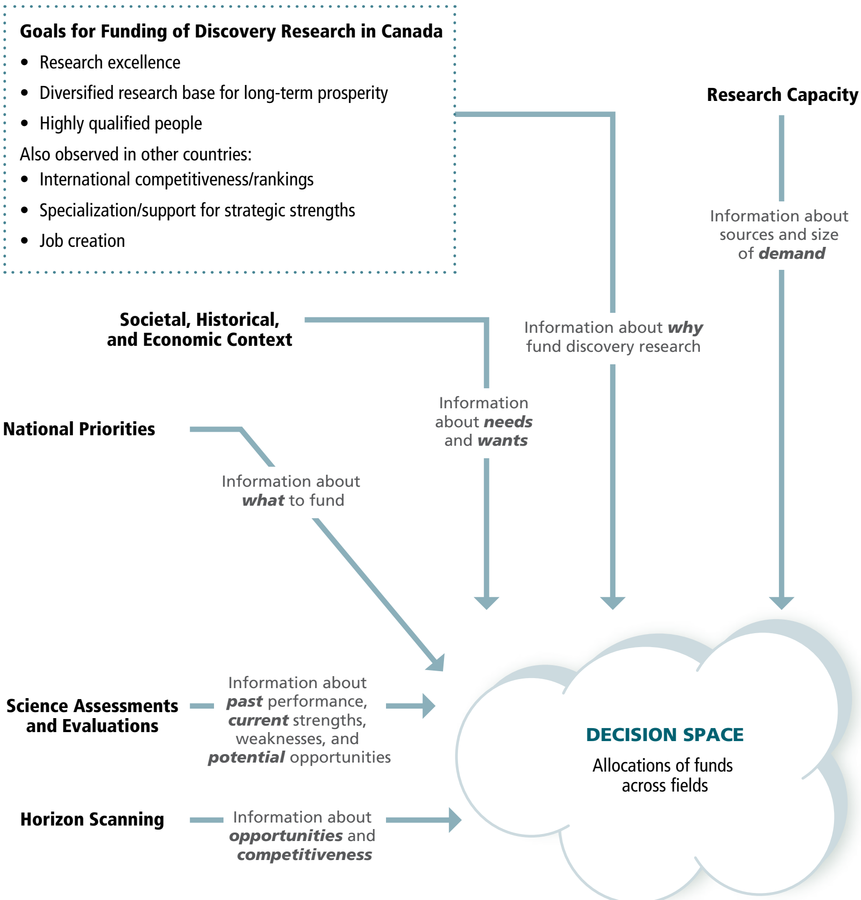


Figure 3.1

Decision space and the research funding allocation context

Funding decisions are influenced by many factors that reflect the individual context of a country, some of which are listed in this figure. The goals of funding discovery research, which are codified by national science policies, are highlighted in the dotted box (e.g., the Canadian S&T strategy). This figure also shows other types of information that may have an impact on funding decisions (e.g., research capacity, societal context, national priorities).

The mandate of a funding program must be clearly defined around its policy objectives within the larger research funding ecosystem. Doing so will dictate information needs, the selection of appropriate metrics and evaluation approaches, and how indicator-based information will be used to assist in allocation decisions. As investment into scientific research continues to grow, the choice of what to fund remains the central issue for both policy and scientific agendas (Popper, *et al.*, 2000; Stewart, 1995; Press, 1988; Weinberg, 1963; Bernanke, 2011).

3.2 THE FUNDING ALLOCATION DECISION PROCESS

Given the competing pressures of the various factors that influence funding decisions, policy-makers are increasingly turning to objective criteria and metrics to help support rational choices (Dutton & Crowe, 1988). Since not all aspects of the many and complex relationships in the funding ecosystem have been studied to date, policy-makers often have a fragmented view, limited by retrospective accounts that do not fully represent the complexities of the ecosystem. This limitation can lead to inappropriate or incorrect assumptions for funding decisions (Jordan, 2010; Sarewitz, 2011; Koizumi, 2011).[11]

In the Panel's view, translating program goals into funding decisions at the field level is a major determinant of which science will be conducted, and thus which societal needs will be addressed. Public investment (both nationally and globally) in discovery research influences a nation's overall research capacity and affects the trajectories of discovery research activities (Freeman & Van Reenen, 2009). Even though discovery research is constantly reinvented, and therefore self-correcting, public funders can use funding reallocations as a means of intervening: to address both which research is conducted (e.g., in expanding disciplinary diversity), and how it is done (e.g., early career support, collaborations and networks, shared infrastructure, and so on).

The challenge for the policy-maker is to translate strategic goals into specific criteria of merit that can be measured by objective and reliable evidence, and guide funding allocation decisions. This is not intended to be a competition among NSE fields; instead it is a consideration of the relevance of scientific activity to the potential contributions to knowledge and society (Dutton & Crowe, 1988).

11 As called for by Marburger (2005), a relatively new field of scholarship, the science of science policy, promotes this perspective in the United States (Fealing, *et al.*, 2011). This emerging field aims to define and standardize terminology and related knowledge base, and to improve the data, tools, and methods for rigorous science and technology policy development. The *Science of Science Policy Roadmap* presents an organized approach to studies in science policy (see http://scienceofsciencepolicy.net/page/sosp-roadmap).

The funding allocation decision process is made up of several broad phases (see Figure 3.2). First, allocation decisions should reflect the most effective investment strategies for achieving a program's mandate and priorities:

> If we conceive of science as a kinetic system and of policy as an effort to redirect, amplify, or block its energies, then the central question is always this: where can the judicious application of resources and effort lead systems to move in favourable directions? And the corollary question is: what are the larger ramifications of nudging poised systems toward a particular course? (Powell, *et al.*, 2011).

The goal in allocating, or reallocating, funds may be to maintain a *status quo* or to alter the dynamics within a research system — e.g., how the research community functions, its environment, and the research infrastructure. Funding allocations should also address incentives to modify or influence behaviours — e.g., scientists respond to incentives and funding opportunities. The clarity of the program mandate underpins the success of the funding decision process.

Next, funders need to define the information needs for the funding allocation decision and determine how these data can be obtained to provide the required information. This is the stage in which indicators and evaluation methods are chosen (see Section 3.3). Clarity is of paramount importance in both policy and evaluative objectives; indicators can be ambiguous and lack meaning unless the context is clearly defined. Since allocation decisions are evaluated by the fulfillment of the program's objectives, the choice of relevant indicators and evaluation methods needs to be aligned with the program mandate (Thomas & Mohrman, 2011). The criteria must also capture how a program interacts with the larger research funding ecosystem. The judgment of experts is key in the determination of information needs and selection of indicators.

Once indicators are selected, data are collected and analyzed. Evaluative assessments of scientific research are important to the cyclical process of policy decisions because the outcome feeds into future decision-making. Data can be obtained from many sources including bibliometric databases, expert panels and surveys, field-level evaluations, OECD and other government statistics databases, and funding program records. Some summary and integrative analysis of the multiple sources of data by experts is necessary (see Section 3.4).

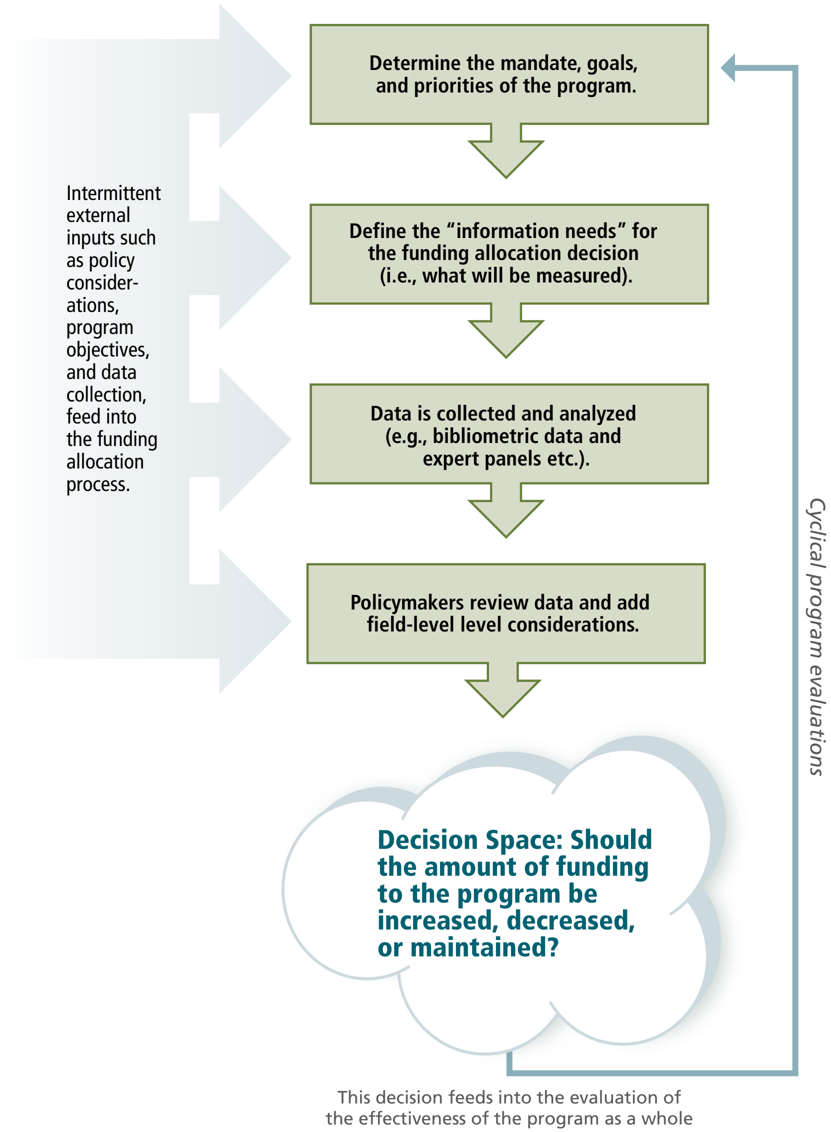


Figure 3.2

The broad phases of funding allocation

This figure presents a flow diagram of the broad phases of the funding allocation process.

The final stage of the decision process involves policy-makers (see Section 3.5). Adding their considerations to the data analysis and expert judgment is a crucial component of the iterative process. Data gathering and analysis inform broader judgment and decisions about policy and program design and funding allocations. Throughout the decision process, there are intermittent external inputs, resulting in a cyclical closed-loop from data to judgment, which makes management decisions responsive to changing societal priorities and to a continuously evolving state of scientific knowledge.

An effective funding decision process features a number of key characteristics:

- **Transparency:** All stakeholders must see the process as transparent and fair. The program mandate and its goals must be clearly defined and drive allocation decisions (i.e., must support the intended impacts). Decision-makers must clearly communicate the funding strategy for addressing competing demands — for example, maintaining broad diversity, supporting high-quality research, supporting emerging research, and building world-class research capacity. An explicit statement of the criteria used for decision-making (the evaluative criteria) can help facilitate a transparent decision-making process (Cozzens, 1999; Jordan & Malone, 2001), and ensure consistency in the evaluative inquiry and meaningful comparisons across fields (Dutton & Crowe, 1988). Both the public and the increasingly diverse research communities expect the process to be open (Cozzens, 1999).
- **Timeliness:** The process must be practical, effective, and capable of resulting in a decision in a timely manner (Research Council U.K., 2006).
- **Credibility:** Any evaluation and measurement must be grounded in credible and robust indicators and respected expert judgment. Availability and quality of data for identified measures may limit the set of appropriate indicators (Research Council U.K., 2006).
- **Efficiency:** The efficiency of the process depends on how effectively quantitative and qualitative sources can be operationalized (Bonaccorsi *et al.*, 2007; Werner & Souder, 1997). To be useful, the evaluative process must generate information and data that "enhance management practices and achieve goals, provide accountability for the stated goals, improve performance, allocate resources, and lead to informed policy decisions" (Jordan & Malone, 2001).

3.3 THE ROLE OF INDICATORS AND EVALUATION METHODS

A viable approach to science assessment and measurement begins with asking the right questions: "Far better an approximate answer to the right question, which is often vague, than an exact answer to the wrong questions, which can always be made precise" (Tukey, 1962). There is a trade-off between the need for high-quality and relevant information and the high cost and effort associated with measuring performance (Jordan & Malone, 2001). This renders the selection of relevant and robust indicators a crucial step in the overall decision process. The choice of indicators and metrics must provide a balance between which indicators are methodologically the most defensible and what information is captured. In assessments and evaluations of science, both quantitative and qualitative, specific criteria can be formulated as a series of questions (Dutton & Crowe, 1988) to aid the selection of appropriate indicators and guide their interpretation through expert judgment.

3.3.1 Using a Logic Model to Select Indicators and Evaluation Methods

An explicit model or conceptual framework that describes the purpose of the evaluative analysis can help research funders narrow down the appropriate indicators and available tools and methods (Jordan & Malone, 2001; Morgan, 2011). A logic model (see Figure 3.3), a common policy tool, can provide an instructive organizing structure for theoretical linkages between funding and the expected impacts and societal benefits from investing in discovery science:

- Discovery research activities require various *inputs*, the "investment" side of the research system. They include current and retrospective measures, which are generally reliable and easily available at different levels of aggregation from data collected and refined over many years (Jordan & Malone, 2001).
- The *outputs* of research activities refer to the knowledge produced and the training of highly qualified people. Outputs provide intermediate evidence of incremental developments and contributions to impacts on science and towards the expected long-term impacts (Jordan & Malone, 2001).
- *Impacts* assess whether the results achieved meet the goals of the funding program. Unlike input or outputs, impacts occur over multiple timeframes, and reflect the dynamic, complex, and unpredictable research process. The potential socio-economic benefits involve looking at the broad relevance of the research (Jordan & Malone, 2001).

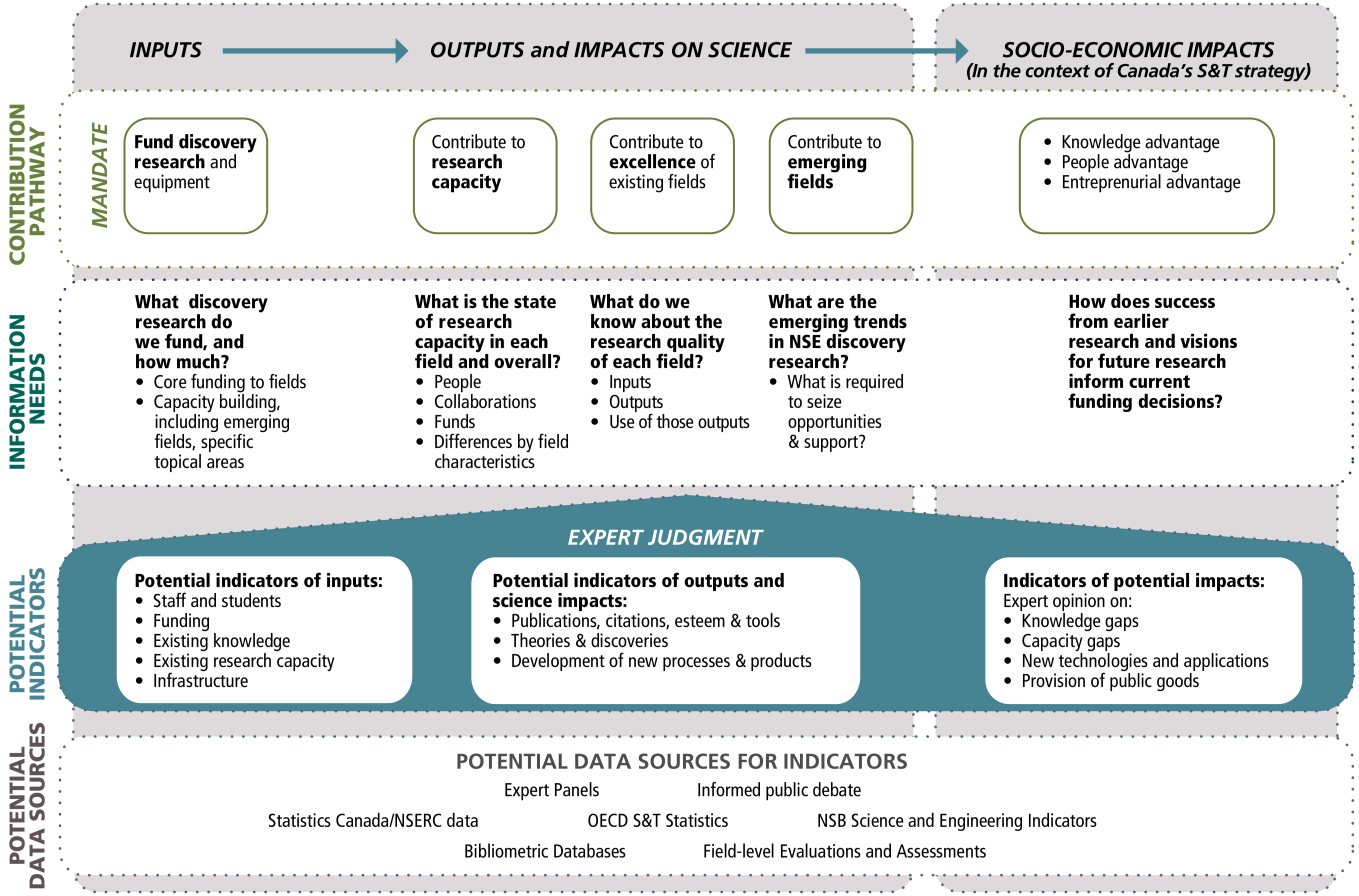


Figure 3.3

Logic model for the selection of appropriate indicators

This logic model illustrates how the mandate of the Discovery Grants Program can be applied to determine the appropriate indicators for collection in funding allocation decisions.

A logic model is not intended to be used as a reporting or accountability structure for a funding program, nor as a traditional linear view of discovery research at the forefront of the innovation lifecycle.[12] Rather, the top row of Figure 3.3 outlines the mandate of the funding program; and the second row defines a generic library of information needs in the form of relevant questions, including the following:

- What discovery research is funded, and what is the funding for each field?
- What is the state of research capacity in each field and overall?
- What do we know about the research quality of each field?
- What are the emerging trends in NSE discovery research?
- How do success from earlier research and visions for future research inform current funding decisions?

Indicators are selected to help answer these questions under the auspices of expert judgment, as demonstrated in the third row of the figure. Potential data sources can be used to construct indicators that follow from there (as shown in the bottom row).

Which indicators should be used will vary with the program mandate, the focus of the evaluation, the unit or level of assessment, the funding allocation mechanisms involved, the nature of the funding decisions at stake, and other factors. What to measure depends upon the information needs of the decision-maker and the characteristics of the system being measured (Jordan & Malone, 2001; Jordan, 2010). Although many funding agencies use a variety of data for evaluations, indicators, while robust and informative for one purpose, may in fact be inadequate when aggregated or repurposed. For example, aggregate data pertinent to national research fields are clearly not appropriate for the assessment of the merits of a research institution or individual researcher. Similarly, data on individuals may not always be aggregated to provide meaningful national statistics.

A single indicator cannot fully capture and represent the diversity and complexity of the relationships within a research funding ecosystem. Quantitative measures are a conduit of information that represents only very specific aspects of that ecosystem. This dissonance between what information can be conveyed by a one-dimensional metric and the multifaceted nature of discovery research continues to be the core challenge for any complex measurement of scientific activities (Martin, 1996). Indicators are usually constructed to help analyze some theoretically

12 The linear model of innovation, institutionalized as a result of V. Bush's influential essay, "Science, the endless frontier," has defined the underlying datasets (metrics of inputs and outputs) still in use in the science policy paradigm. One of the many implications is that metrics used for basic research have been focused on economic benefits only (Bush, 1945).

defined abstract property of reality that can be empirically observed (Van Raan, 2004). It is therefore beneficial and desirable not only to use multiple indicators, but also to select a basket of indicators that convey a range of perspectives, and help address different questions. The use of multiple, relevant indicators leads to a greater likelihood of converging on an accurate understanding of research performance (Martin, 1996).

The choice of indicators, to a large extent, represents a compromise between the desire to find the right answers to complex evaluative questions and the practical underlying policy and data considerations (Bonaccorsi *et al.*, 2007). The availability of indicators does not render them relevant or important to field-level allocation decisions for a program. Good indicators answer specific questions (i.e., are relevant to evaluative inquiry); are well founded in a theoretical understanding of how a program contributes to both proximate and future impacts related to the program's strategy, goals, and mandate; are feasible in terms of available high-quality data; and are transparent in terms of the underlying limitations of the data or data collection approaches (ESF, 2009).

3.3.2 Identifying Indicator Categories for Field-level Allocation Decisions

Indicators of relevance to field-level funding allocation decisions for NSE discovery research can be grouped into three categories, which are based on three main information needs:

- **Research quality:** The notion of research quality — a term that is often used yet rarely or uniformly defined — has long been a prominent feature of policy discussions and research funding decisions (Butler, 2007). In the context of research evaluations, research quality typically denotes the overall calibre of research based on the values, criteria, or standards inherent in the scientific community. In this sense, it is loosely understood as pertaining to the scientific merit of a research field, but not necessarily to the wider or immediate importance or impact of that field on society. Funders of discovery research seek information on research quality because they aim to optimize the development of incremental and breakthrough knowledge, and thus the potential for impact across all fields in the research they fund.
- **Research trends:** The dynamic natures of knowledge and science are embodied in a constant, though admittedly often slow, redefinition of scientific fields. Many factors drive the emergence of new areas of research — for example, an increasing number of global and cross-disciplinary collaborations, new tools and technologies, and investment from governments and industries. While new areas of knowledge diffuse across traditional disciplinary boundaries, as well as those between discovery and applied activities, the scientific merit of

a field is often expressed in terms of its relevance to related fields and of the new opportunities created (Press, 1988; Dutton & Crowe, 1988). Funders of discovery research analyze the evolution of scientific fields to ensure that funding allocations remain responsive to changes occurring in the research enterprise.

- **Research capacity:** This category encompasses the resources available in the internal and external environment in which research occurs, including human capacity as well as infrastructure and cost of research (Meek & van der Lee, 2005). The concept captures a range of diverse aspects that often cannot be limited to a specific field. Capacity is a shared good; for example, the classic NSE fields (e.g., mathematics, biology, chemistry) provide the knowledge foundation leveraged by other disciplines, or the digital infrastructure that is now an essential component of knowledge capacity in almost all scientific research. Field-level measurement of research capacity can be difficult; however, various indicators of research capacity have been standardized and are widely used by international and national organizations (e.g., the OECD, the European Commission, the U.S. National Science Foundation (NSF)). The challenge lies in meaningful field-level attribution and aggregation of various data sources. Funders of discovery research should seek information on research capacity in general, and field-specific capacity in particular, because world-class research cannot be conducted without access to adequate and sustainable capacity.

These three categories are intertwined: for example, research capacity is a crucial determinant of both the current and future quality of existing and research trends.

3.4 THE ROLE OF EXPERT JUDGMENT

Various evaluations and assessments of scientific research can provide valuable insights about past and current performance, while consideration of the future direction for research can help identify opportunities for growth ensuring long-term competitivness. But even with multiple approaches and a multitude of indicators, absolute quantification of discovery research is not feasible (Martin, 1996) and cannot provide accurate *ex ante* predictions or *ex post* assessment. Because most indicators are retrospective, it cannot be assumed that what has happened in the past is necessarily a reliable predictor for what might happen in the future. Therefore research evaluations need to combine quantitative indicators with expert judgment. For example, a National Research Council report (2006) recommended that funding allocation decisions should primarily rely on processes that combine deliberation and metrics:

> We recommend a strategy that combines analysis and deliberation, in which processes of open, explicit dialogue are organized to raise all the major decision-relevant issues, allow for input from all relevant perspectives, and provide for iterative discussion between researchers and science managers and for orderly reconsideration of past decisions.

When funding allocation decisions draw from formal evaluations, the central and essential evidence comes in the form of expert judgment, an assortment of various quantitative measures, and a combination of both. Though not free of bias (see Chapter 4 for further discussion), metrics and judgment provide a mechanism for incorporating both quantitative and qualitative evidence into funding decisions at all levels, including field-level allocations. Importantly, when used together they are seen as reliable and robust decision-support instruments (ESF, 2009). Expert judgment is what empowers prospective and multidisciplinary deliberations of otherwise retrospective measures.

Retrospective, or *ex post*, evaluations help identify potential improvements and operational cost savings; they can also help clarify responsibilities within the broadly considered research funding ecosystem, as well as deepen the understanding of how science benefits are realized (OECD, 1997; ESF, 2009; Rip, 2000). Increasingly, however, assessments and evaluations are expected to address prospective, or *ex ante*, considerations (Georghiou & Roessner, 2000; Jordan & Malone, 2001). These are much more difficult to assess as they demand answers to potential results expected in the future (in terms of both scientific and broad socio-economic impacts). These answers build on *ex post* evaluations — what investments were made, what outputs were produced, and why they occurred (i.e., the interplay between inputs and outputs, and their contribution to proximate impacts).

3.5 THE ROLE OF POLICY DECISIONS

The essence of scientific expert judgment is the interpretation of incomplete, but measureable, sets of information. The Panel believes that, even with the use of expert judgment, no existing approach obviates the need for policy decisions. Funding allocation choices balance the measurable evidence with the pressures of scientific and national priorities. Therefore, it is important to distinguish between policy issues and assessment of scientific merit and quality. The Panel's view is consistent with the findings of other expert panels, in that blurring the boundaries between the responsibilities of policy-makers (e.g., funding agencies) and those of experts burdens the capacity of the research system (e.g., SSHRC, 2008). There is no one preferred governance model when it comes to field-level funding decisions. The specific societal and political context, and its values, seems to be one of the major factors. Although expert review committees are seen as the best means for assessing the merits of research, the funding decision itself is made by a designated group or institution different than the one responsible for judging the scientific merit of research proposals.

For example, one question that policy-makers might address is to look at current infrastructure and expert capacity to determine if they will meet future needs. Decision-support tools, including visual tools such as bubble charts, can help policy-makers to assimilate large amounts of multidimensional information and plot multiple strategic considerations into one diagram (Cooper *et al.*, 2012). They can provide more comprehensive understanding of the *status quo* and help guide future allocation decisions by exploring various policy scenarios.

Although the overall decision process (shown in Figure 3.2) can be transparent with respect to information inputs, anecdotal evidence suggests that the responsibility for making the funding allocation decision rests with a small number of appointed individuals. Not surprisingly, the details of how, or the degree to which, formal procedures drive these decisions are poorly documented, with the implication that decisions are highly sensitive, rather than lacking in accountability. The appointed individuals with responsibility and accountability for these decisions embody the values and hold the trust of both the policy and the research communities. Thus, the underlying values of the decision will change with evolving societal, economic, technical, and institutional contexts. There is evidence, however, that, to varying degrees, these decisions are primarily driven by expert evaluation's that use both quantitative metrics (mostly retrospective) and qualitative analysis.

3.6 CONCLUSIONS

This chapter has examined the question of "choice" in research funding allocation decisions. From measurement and assessments of national strengths in science to the evaluations of individual scientists, decision-makers rely on quantitative and qualitative data to gain insights into aspects of the complex research system. As such, indicators and expert judgment are important to facilitate a decision-making process.

The information needs of a decision-maker determine the evidence of recent and past performance in fields of NSE research to be captured. Indicators of relevance to field-level funding allocation decisions for discovery research are based on information about research quality, research trends, and research capacity (e.g., people, tools, costs, and funds). Whether an indicator is reliable or informative for a given decision depends on the intended use, methodological construction of the indicator, and quality of the underlying data. Though no single indicator can capture the complexity of a research ecosystem, suites of indicators can present robust evidence of recent and past performance over a range of characteristics. Nevertheless, the judgment of scientific experts and policy decision-makers empowers the prospective and multidisciplinary deliberations of measures that are retrospective. It also addresses both scientific merit and societal relevance of the research, and illuminates options and trade-offs among alternative choices. In short, clarity of the program mandate will underpin success of the entire process.

4

Science Indicators: Understanding the Options

- Deliberative Approaches and Other Qualitative Assessment Methods
- Bibliometrics
- Other Quantitative Science Indicators
- Conclusions

4 Science Indicators: Understanding the Options

Key Points

- Science assessment strategies at the field level can be divided into two general groups: those based on deliberative methods (e.g., peer or expert review), and quantitative indicators.
- Assessments at the national field level often rely on combinations of deliberative methods and quantitative indicators.
- Deliberative assessment methods, such as expert and peer review, are commonly used in evaluations of individual researchers. While not without limitation, these assessment methods are also relevant to assessing performance at the level of a field of science.
- Bibliometric indicators are paradigmatic quantitative science indicators. While there are inevitably limitations associated with bibliometric indicators, many are now sufficiently robust to provide meaningful input to science assessment at the level of nationally aggregated research fields.

Technological and methodological advances related to the management of scientific research have led to a proliferation of science indicators (Van Noorden, 2010). This abundance of choices has made determining which science metrics are appropriate and informative more challenging. Faced with diverse possibilities, policy-makers increasingly rely on the guidance of research evaluation experts in the selection of indicators and assessment methods to inform decisions about the direction and extent of public investment in science.

This chapter provides an overview of existing science indicators and assessment methodologies at the level of nationally aggregated research fields.[13] Science assessment strategies are divided into two major types: those based on deliberation and expert judgment, and those based on quantitative data and analysis. While deliberative approaches remain the dominant method of science evaluation in most contexts (i.e., peer review of scientific papers and grant applications), reliance on quantitative data and indicators is increasingly prevalent in many types of research assessment, especially those focused on higher levels of aggregation.

13 This discussion does not deal with the distinct challenges and complexities that characterize the use of science indicators in the evaluation of individual scientists or researchers. On this topic, see a recent report from l'Academie des sciences (2011) in France.

The relative merits of each type are explored in this chapter, followed by a brief discussion of the most common quantitative indicators, most notably those based on bibliometric data. Chapters 5 through 7 provide more detailed examination and assessment of indicators related to three specific objectives: assessing research quality, research trends, and research capacity (information categories identified in Section 3.3.2).

This Panel is not the first to review indicators and approaches used to assess science. Van Noorden (2010) provides a useful, short review of science metrics. De Bellis (2009) and Moed (2005) offer extensive surveys of bibliometric and citation-based indicators. The Research and Evaluation Policy Project (REPP) at the Australian National University conducted a literature review related to quantitative research evaluation indicators, which included a detailed catalogue of indicators (REPP, 2005). A Canadian Academies of Health Sciences report on evaluating the return on investment in health research summarized many of the existing indicators used in research evaluation, particularly as related to health research (CAHS, 2009). And the National Research Council in the United States commissioned a study that reviewed various strategies for assessing science (NRC, 2006).

4.1 DELIBERATIVE APPROACHES AND OTHER QUALITATIVE ASSESSMENT METHODS

Following a study on science assessment carried out by the National Research Council (NRC) in the United States, this Panel used the concept of deliberative methods to distinguish between science assessment strategies based on peer review or expert judgment, and those based on quantitative analysis (NRC, 2006). Deliberative approaches use "discussion, reflection and persuasion to communicate, raise, and collectively consider issues, increase understanding, and arrive at substantive decisions" (NRC 2006, as cited in NRC, 1996a). In the context of research assessment, deliberative approaches can be generally divided into two related groups of practices: peer review and expert review (OECD, 2008).

4.1.1 Peer and Expert Review

Peer review, the most common form of deliberative approach, has historically been the dominant model of evaluation in science (Chubin & Hacket, 1990). Evaluation practices based on peer review are used in the context of reviewing papers for publication in scientific journals, assessing grant applications for research funding, and evaluating individual scientists or researchers for selection or promotion within research institutions (OECD, 2008). Given this diversity of applications, the varieties of peer review differ significantly in key aspects such as number and type of output provided, selection of reviewers, and criteria and guidance provided in the process, as well as the organization and structure of the deliberation (OECD, 2010). All peer review models, however, share a core reliance on the judgment of scientists working either in the same field or in a closely related field (i.e., peers) to that of the work or scientist being evaluated. Other characteristics common to most forms of peer review include in-person meetings and deliberation; procedures for reviewer recruitment to ensure inclusion of the required depth and breadth of expertise; procedures to ensure reviews are independent and free of outside influence; and procedures for structuring the deliberation or evaluation process, such as assessment criteria or ranking systems (NRC, 2006).

Expert review is an analogous but more general process. Following the definition provided by the OECD (2008), the Panel uses the term "expert review" to refer to deliberative evaluation processes based on expert judgment used in the context of evaluations of broader research fields or units. Where peer review is understood to pertain to the evaluation of individual research outputs, such as a single paper or grant application (or a specific individual in the case of hiring or promotion decisions), expert review relates to the assessment of larger aggregations such as a field or institution (see Box 4.1 for an illustration of this distinction).

Peer and expert review share common strengths. Most obviously, deliberative evaluation processes are capable of assessing research characteristics that are challenging to quantify (e.g., methodological rigour and appropriateness, the importance of specific content or ideas, the aesthetic dimension of research quality) (Moed *et al.*, 1985, as quoted in Butler, 2007). Other strengths include the ability to take into account nuanced and detailed understandings of (i) the context in which research is being performed (e.g., institutional and environmental characteristics that may relate to the viability of certain research avenues); and (ii) recent research trends that may suggest more or less fruitful avenues of inquiry.

Deliberative methods, however, are not without limitation. An OECD (2008) review of peer review processes identified two main weaknesses. First, the quality of peer review is dependent on both the quality and objectivity of the reviewers. Most peer review processes have, therefore, developed mechanisms and procedures to screen for potential sources of bias or conflicts of interest. Such mechanisms, however, are not infallible, particularly for highly specialized research fields or small countries, or in cases where the demand for reviewers strains the capacity of the research community. The second weakness identified, the risk that reviewers with well-established and recognized areas of expertise will tend to perpetuate orthodox or conservative research paradigms, is contentious. Some studies have suggested that peer review may be biased against interdisciplinary research

(NRC, 2006; Porter & Rossini, 1985; Travis & Collins, 1991; Rafols *et al.*, 2012), while others have proposed that it favours low-risk proposals (Langfeldt, 2001). This challenge can be mitigated, however, with mechanisms that ensure interdisciplinary representation in the selection of review panel members (Laudel, 2006) or explicit incentives to reward higher-risk proposals.[14]

Box 4.1
An Example: Peer Review in the United Kingdom and Expert Review in Australia

The national research assessment exercises undertaken by Australia and the United Kingdom illustrate the distinction between peer review and expert review. The U.K. RAE is essentially based on a peer review process. Panels of scientists and researchers are established for fields and sub-fields, and asked to assess the overall quality of a set of research outputs (most often papers) submitted by the institutions and units being evaluated. These panels operate very much as traditional peer review panels. Once the results from reviews of individual outputs are complete, they are aggregated to assess an institution's overall level of performance relative to its peers.

The recently created research assessment initiative, Excellence in Research for Australia (ERA), is best understood as a model of expert review. Expert panels are established to assess research performance in specific fields across institutions. With the exception of some panels in the humanities and social sciences, however, no reviews of individual research outputs are undertaken. ERA panels rely primarily on already aggregated data or indicators to inform their judgments; they do not read or review any publications from the institutions and research groups being evaluated.

For further information, see the case studies on Australia and the United Kingdom in Appendix A.

14 For example, both the National Science Foundation (NSF) and National Institutes of Health (NIH) in the United States have specific programs in place to support "high-risk" research projects. For more information, see http://commonfund.nih.gov/highrisk/ and http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=5319

Other potential criticisms of peer and expert review focus on lack of transparency in the process (Reinhart, 2010), especially where there is little or no public record of the nature or course of the deliberations; and on a tendency to fail to consider relevance of the research with respect to social objectives outside the remit of the specific research agenda (e.g., Nightingale & Scott, 2007).

4.1.2 Other Qualitative Methods of Assessment

The case study is perhaps the most common example of other types of qualitative methods used in research assessment. Case studies are often used to explore the wider socio-economic impacts of research. For example, the U.K. Research Excellence Framework (REF) (the proposed replacement for the RAE) will rely on case studies to assess research impact. A REF pilot project on use of case studies demonstrated the effectiveness of this method in analyzing and communicating these types of impacts (HEFCE, 2010). Public value mapping is another example of a case study-based technique used to explore questions about the larger social value of public research investments (Bozeman & Sarewitz, 2011; Meyer, 2011; Slade, 2011).

Project Retrosight is a Canadian example of the case study approach used in research assessment. Undertaken as part of a multinational study to evaluate the impact of basic biomedical and clinical cardiovascular and stroke research projects, Project Retrosight measured payback of projects using a sampling framework. Despite several limitations to the analysis (e.g., the number of case studies limiting the sample pool from which to draw observations, potential inconsistencies in reporting and comparability), the case study approach provided an effective platform for evaluating both the *how* and the *why* of evidence to demonstrate impact. The key findings of the study revealed a broad and diverse range of impacts, with the majority of broader impacts, socio-economic and other, coming from a minority of projects (Wooding *et al.*, 2011).[15]

4.2 BIBLIOMETRICS

Bibliometric indicators are the paradigmatic quantitative indicators with respect to measurement of scientific research. They are based on data drawn from trends in publication of scientific research, in particular peer-reviewed journal articles published in academic journals. Bibliometric indicators now come in many forms and varieties including counts of scientific publications;

15 For more information, see the Canadian Project Retrosight website via the Canadian Institutes of Health Research at http://www.cihr-irsc.gc.ca/e/43251.html

counts of citations to publications; and analysis based on other bibliometric variables such as authorship, keyword use, and patterns in cross-citation between research units or fields. Bibliometric techniques are also used in conjunction with various visualization techniques in the study of relationships among different domains of research. Finally, complex variables and analytical techniques may combine various types of bibliometric measures to yield novel forms of analysis (e.g., Klavans & Boyack, 2010).

The use of bibliometrics has evolved substantially in recent decades. As noted in the NRC (2006) report, the original use of bibliometric techniques, as applied to scientific research, tended to be descriptive in nature rather than evaluative (see also Godin, 2002; Van Raan, 2004). Science historians, scholars, and sociologists looked to data from scientific publications to understand better the development and course of scientific research. Bibliometric techniques began to be used for evaluative purposes in the mid-1970s (e.g., Narin, 1976). The evaluative use of bibliometric analysis was further developed when applied to research groups (Martin & Irvine, 1983). Bibliometric techniques have since continued to be developed and expanded, often in parallel with technological developments and expanding coverage of the scientific literature in bibliometric databases. They now represent one of the standard tools in science assessment. Many countries undertake national bibliometric studies that compare research outputs based on publication counts and other bibliometric indicators (see Box 4.2 for selected examples).

Box 4.2
Using Bibliometric Indicators at Higher Levels of Aggregation

Bibliometric data and indicators are routinely used at various aggregated levels to assess research performance across fields. Examples from national S&T reports from around the world are highlighted below.

- *The State of Science and Technology in Canada* analyzed a number of bibliometric indicators at the national level, including measures of publication output by field and international benchmark citation data (CCA, 2006). This report influenced the development of the federal S&T strategy (Industry Canada, 2007), and thus definition of Canada’s research priorities.
- The Institut de la statistique du Québec has published an *Annual Compendium of Scientific and Technological Activities* since 2005 that contains bibliometric indicators compiled by Canada’s Observatoire des sciences et des technologies (e.g., ISQ, 2011).

continued on next page

- U.S. *Science and Engineering Indicators, 2012*, published every two years by the National Science Board, provide a comprehensive dataset on the state of science and engineering in the United States, including comparison of publication outputs across fields and countries, based on bibliometric indicators (e.g., National Science Board, 2012).
- *Science and Technology Indicators Reports* in the Netherlands provide a periodic, comprehensive assessment of aggregate research output in the Netherlands, which presents a variety of bibliometric data and indicators (e.g., NOWT, 2010).
- *The State and Quality of Scientific Research in Finland* reports, published by the government, are part of a comprehensive evaluation of the Finnish science system undertaken by the Academy of Finland every three years (e.g., Academy of Finland, 2010). This analysis includes total scientific output in Finland, as well as basic comparisons of publication and citation measures by field of research.

See the international case studies in Appendix A for further information on these assessment initiatives.

The many variations of bibliometric indicators are too numerous to survey here in detail. Those looking for comprehensive overviews of current bibliometric practices can find useful surveys in De Bellis (2009), REPP (2005), Van Raan (2004), and Moed (2005). These metrics, however, can be broadly organized into three main classes: (i) those based on publication counts, (ii) those based on citations, and (iii) those based on other variables.

4.2.1 Indicators Based on Publication Counts

The simplest bibliometric indicators are those based on publication counts. In principle, such counts can be generated for many different types of publications (e.g., books, book chapters). In practice, due to the limitations of coverage in indexed bibliographic databases, existing indicators are most often based on counts of peer-reviewed articles in scientific journals. Basic publication indicators typically take the form of absolute counts of the number of journal articles for a particular unit (e.g., individual, research group, institution, or field) by year or for a period of years. Such indicators are typically framed as a measure of research output.

Additional indicators based on publication counts can be derived from shares of publication counts (e.g., a research group's share of total publications in an institution, a field's share of total publications in a country). These share-based indicators generally are used to capture information about the relative importance of research output originating from a particular unit or field. More advanced indicators based on weighted publication counts can also be created when publication output is typically weighted by some measure of the quality of the research outlet. For example, journal impact factors (a measure of the relative citedness of a journal) may be used to give a higher weight to publications in more prestigious or competitive journals. Unlike straight publication counts, these metrics also depend on some other measure of quality, either based on citation or on some other assessment of the relative quality of different journals. One example of this is the ranked outlet measure used in the Excellence in Research for Australia (ERA) initiative (see Box 4.3).

Box 4.3
Ranked Journal Outlets in Australia's ERA

One distinctive element of Australia's recent experiences with research assessment relates to the use of a comprehensive system of ranked journal outlets. In preparation for the first round of the Excellence in Research for Australia (ERA) initiative, the Australian Research Council undertook a large-scale assessment of more than 20,000 academic journals. Journals received one of four quality ratings. Journals deemed to be "one of the best in their field" received an A* rating; journals of lesser importance received lower rankings (A, B, or C). The journal rankings were developed based on an expert review process carried out with the assistance of Australia's four learned academies, and in consultation with the research community. The distribution of publications across the four ranks was then used in the first round of the ERA in 2010 as one of the measures for evaluating the quality of university research in each discipline (ARC, 2008).

Despite the considerable effort that went into devising them, ranked journal outlets were dropped from the ERA in 2011 and will no longer be a part of future rounds of assessment (Australian Government, 2011). Journal rankings were always a controversial aspect of the ERA. Interestingly, the government and the Australian Research Council eventually made the decision to drop the journal rankings not only because of concerns about the accuracy of the ratings, but also because some universities and research institutions were using the rankings inappropriately. For example, research managers at institutions had begun to set publication targets for the number of publications in A* and A journals (Australian Government, 2011), pressuring their faculties to focus publication efforts on certain outlets. Ultimately, it was an undesirable behavioural response from the research community that led to the elimination of this aspect of Australia's national research assessment process.

4.2.2 Citation-based Indicators

Citation-based indicators are used to capture information about the influence or impact of research on the scientific community. This practice is predicated on the notion that citations capture valuable information about the extent to which a journal article has influenced later scientific research (Moed, 2005). A wide variety of citation-based indicators are used at different levels of research assessment including simple citation counts (e.g., number of citations for an individual article or author or country); averaged citations (e.g., average citations by field or journal); or more derivative measures such as the h-index and its variants (Hirsch, 2005).

Indicators based on the distributions of citations across a body of work are also commonly used. For example, one well-known study calculated and compared the share of the top cited one per cent of world scientific papers by field for selected countries (King, 2004). Citation-based indicators are also sometimes used to assess relationships between different fields of research, or as interdisciplinary measures. These may take the form of measures based on the frequency of cross-citations between different research fields or units (Larivière & Gingras, 2010).

One of the most important citation-based indicators for comparing research performance across fields at the national level is the relative citation impact (sometimes referred to as average relative citations or ARC). This indicator is based on comparing the level of citedness of the unit being evaluated with the general level of citedness of research in that field internationally. For example, the average number of citations received by papers in clinical medicine in Canada is compared with the average number of citations received by all papers in clinical medicine worldwide.

4.2.3 Other Bibliometric Indicators

The third class of bibliometric indicators includes all those based on other variables. One common type of indicator is based on paper co-authorship, which is now analyzed to study patterns in scientific collaboration (e.g., Royal Society, 2011). Such methods can be used to analyze collaboration between individual researchers, research groups, institutions, fields, and countries (for Canada, see Larivière *et al.*, 2006). Other bibliometric indicators are based on the use of keyword searches in scientific papers. Most scientific papers are published with a small number of key words selected by the author to capture information about the subject. Key words can also be drawn from article titles and abstracts. Keyword searches in bibliometric databases can be used to define research topics, assess interdisciplinary research areas, and track research trends over time. They can also be used as parameters in the construction of novel discipline groupings or field- or topic-specific bibliometric indicators. Finally, other types of information contained in bibliographic records can serve as fodder for additional study. For example, address information for paper authors may be used to analyze the geographic distribution of research activity in a region or country. Acknowledgement data have been recently incorporated into some bibliometric databases, and can also be used in bibliometric analysis.

4.2.4 Limitations of Bibliometric Indicators

The conceptual and technical limitations of bibliometric indicators have been well documented (for example, see Moed, 2005; REPP, 2005). These limitations are not discussed at length here. Box 4.4 reviews key methodological issues related to use and construction of these indicators. In the Panel's view, none of these issues are insurmountable. While there are inevitably certain limitations associated with the data and construction of these indicators, they are in general sufficiently robust to allow for meaningful use in many assessment contexts at the field level.

Box 4.4
Some Facts about Bibliometric Analysis

The following parameters must be considered in undertaking any bibliometric analysis:

Selecting the unit of analysis: Bibliometrics can be used to study research output on many levels: individuals, research groups, institutions, fields, and regions or countries. Many indicators, however, are appropriate or reliable only at specific levels of aggregation. In general, bibliometric analysis is more reliable at higher levels of aggregation, and indicators that are inappropriate for evaluating individual scientists may be useful when applied to fields of research at the national level.

Field differences: Both publication and citation patterns vary by research field. Some fields traditionally publish and/or cite other research more often than others. Direct comparisons across fields are therefore rarely valid, and it is critical to use field-normalized metrics when using bibliometric indicators for evaluative purposes. A research group's output, for example, can be meaningfully assessed only by reference to the average research output *for that field of research*.

Data coverage: Ultimately, the usefulness of bibliometric indicators depends on the extent to which the relevant research outputs are covered in bibliometric databases, and this coverage varies by research field. Past studies have found that coverage tends to be high in the natural sciences, which place a high priority on journal publications.

continued on next page

In other fields, where publication of books, book chapters, monographs, etc. is more traditional, the extent of coverage is reduced (Archambault *et al.*, 2006). Although this problem is most severe with respect to the humanities, arts, and social sciences, it can also be a concern for some areas of the natural sciences and engineering where publication in conference proceedings may play an important role e.g., engineering and computer sciences (HEFCE, 2009, Annex H).

Multiple authorship: Most scientific papers have multiple authors. One extreme example is a recent paper in *Physics Letters B*, related to research at the European Organization for Nuclear Research (CERN) Large Hadron Collider, with over 3,000 contributing authors (Royal Society, 2011). Multiple authorship must be taken into account in the construction of bibliometric indicators. There are two main approaches to this: (i) an author may be given full credit for a paper (i.e., full counting); or (ii) credit for each article may be divided by the number of authors (i.e., fractional counting). Alternative strategies involving counting only first authors may also be employed; however, this is an imperfect solution as the conventions around the order of authors on scientific papers differ across fields (Pontille, 2004; RIN, 2009).

Negative citations and self-citations: Some scientists argue that indicators based on citations are misleading because a portion of citations constitute refutation of past work, rather than affirmation. Others argue that the frequency of self-citations (citations to an author's own past work) often weakens the validity of the indicators. Although these are reasonable concerns, empirical evidence shows the issues to be relatively minor at high levels of aggregation. Past studies have found that the large majority of citations (more than 90 per cent) are confirmative, and that faulty research is more often ignored than cited (see REPP, 2005; Herbertz & Muller-Hill, 1995). Self-citations can be identified, and one can construct indicators with or without self-citations (REPP, 2005). Many self-citations are valid, however, and distinguishing between legitimate and illegitimate self-citations is impractical in large-scale evaluations. Journal editors also act to some degree as "gatekeepers" to guard against flagrant abuses of self-citation. As a result, both the U.K. RAE (when it was contemplating a central role for bibliometrics) and the Excellence in Research for Australia (ERA) initiative made a decision to retain self-citations.

Timeliness: Bibliometric indicators are of varying degrees of timeliness. Research results may be dated by months, or even years, by the time they are published. More importantly, for citation-based indicators, it takes time for knowledge of an article to permeate the research community. These facts are taken into account through the concepts of citation windows and the citation "half-life" of a journal, the latter

continued on next page

being defined as the number of years going back from the current year required to account for 50 per cent of the citations in that journal. Appropriate citation windows vary by field; however, past research has generally found that in the natural sciences, a period of three to five years is sufficient since citations to most articles in these fields peak within three years (REPP, 2005; Moed *et al.*,1985).

For comprehensive reviews of these issues, see Moed (2005) and REPP (2005).

4.2.5 Data Sources for Bibliometrics

Three sources of bibliometric data are commonly used in research assessment exercises: (i) the Web of Science database and family of citation indices maintained by Thomson Reuters (formerly the Institute for Scientific Information or ISI); (ii) Elsevier's Scopus; and (iii) Google Scholar. Scopus and Web of Science have both been extensively used and tested in bibliometric analyses, and are sufficiently transparent in terms of their content and coverage to be generally useful in assessments of research performance at the field level. Although the overall coverage of the two databases does differ significantly, evidence suggests that, with respect to comparisons at the national level in the natural sciences, the use of either source yields similar results (Archambault *et al.*, 2009).

Due to ease of access and freely available data, Google Scholar is now frequently used in amateur bibliometric analyses. It is, however, a dynamic, proprietary search engine, and Google does not release static counts of the results of its searches at a given time, nor details of what is covered. Therefore, bibliometricians are unable to assess search results and coverage methodically. Furthermore, Google Scholar is constantly updated, and the same query run on consecutive days may return different citation counts. Cleaning the data available is also time intensive. One study found that it took over 30 times the number of hours to collect, clean, and standardize data from Google Scholar as it did to clean data from Web of Science (Meho & Yang, 2007). As a result, in the view of this Panel, Google Scholar should not be used as a data source for rigorous bibliometric assessment.

Increased pressure on scientists and institutions to prove the societal value of their work has led to rapid proliferation of amateur — or *savage* — bibliometrics (Gingras, 2008). In such cases, bibliometric indicators are applied without expertise or understanding of either the underlying data or appropriate use of indicators. The practice has become widespread with instant access to the Web of Science through university subscriptions and exacerbated with Google Scholar, which is a free source of data, undermining the legitimate use of sophisticated, validated techniques.

4.3 OTHER QUANTITATIVE SCIENCE INDICATORS

There are many types of quantitative indicators aside from bibliometrics. Common varieties include those based on external (i.e., from other sources) research funding, numbers of grant applications, researcher and student populations, and measures of esteem (e.g., counts of academic honours or awards) (for surveys of these other types of indicators, see REPP, 2005; De Bellis, 2009). Indicators based on online activities such as paper views or downloads have recently been developed. Finally, there are also quantitative indicators used to capture information about research infrastructure, such as annual investment in equipment and facilities or available laboratory space. These other quantitative indicator types are highly heterogeneous; they are based on different data sources depending on the national and research funding context, and used for many kinds of evaluation. Due to this diversity, a general discussion of their strengths and weaknesses is impractical; however, these indicator types are reviewed in the following chapters, which focus on indicator use in relation to three common science assessment objectives.

4.4 CONCLUSIONS

This chapter has provided a brief overview of available methods and indicators used to assess discovery research, emphasizing two prominent strategies: deliberative methods and bibliometric indicators. Both of these types of science assessment strategies have strengths and weaknesses. Deliberative approaches are traditionally associated with the assessment of individual researchers and research outputs, but can be applied at the level of research fields as well. While there are fundamental limitations associated with many quantitative indicators based on the nature and extent of the data, these data are, nevertheless, sufficiently robust to allow meaningful use of the indicators in many assessment contexts. In considering the design of any assessment, how indicators are used is as important as how they are constructed in determining whether they are likely to be valid and informative in a specific context. The Panel therefore concluded that ultimately it is not productive to discuss strengths and weaknesses of indicators without taking into account their intended use.

The next three chapters therefore provide a more detailed analysis of science indicators and evaluation methods in relation to three assessment objectives pertinent to informing research funding allocation: (i) research quality; (ii) research trends; (iii) and research capacity. (see Section 3.3.2 for discussion of information needed for research funding allocation).

5

Indicators for Assessing Research Quality

- Review of Quantitative Science Indicators
- A Taxonomy of Indicators for Assessing Research Quality
- Reviewing the Indicators
- Conclusions

5 Indicators for Assessing Research Quality

Key Points

- Quantitative indicators for assessing research quality can be divided into six major types: external research support (funding), student population, weighted publication counts, citations, esteem measures, and webometrics.
- Weighted publication counts and citation-based indicators are the only two types of indicators that are valid quantitative measures of research quality appropriate for use at the field level.
- Deliberative methods are also a valid strategy for assessing research quality at the field level, and the strongest approach to assessing research quality combines quantitative metrics and deliberative methods.

"Research quality," and related concepts such as research excellence and research strength, refers to the scientific merit of research judged employing values of the research community. The concept of research quality is widely recognized as being both complex and multidimensional, and many scientometric studies have emphasized that it is best understood being comprising several underlying characteristics, such as the scientific importance of the work, the rigor of the methods employed, and the elegance or aesthetic qualities of the research design and findings (e.g., Martin & Irvine, 1983; Moed *et al.*, 1985; CCA, 2006).[16]

Research quality is not necessarily synonymous with the scientific impact of work. The impact of research is inevitably a function of its quality, but impact is also dependent on other factors, including location of the author(s), reputation and personal network of the author(s), language of publication, and availability and prestige of the journal in which it appears (Martin & Irvine, 1983). These factors are important because bibliometricians are generally confident about the ability of bibliometric indicators to assess the *impact* of a scientific publication on other research, but question the ability of bibliometrics to judge overall *quality* of research (which some argue should be left to peer review) (see Butler, 2007; Nederhof & Van Raan, 1987; Van Raan, 1996). Indicators based on citations can provide a

16 It should be noted that the concept of "research quality" as used in many research assessment exercises typically focuses exclusively on the scientific value or impact the research, and excludes consideration of wider socioeconomic impacts. These factors are often considered under the somewhat misleading heading of "research impact" in many science assessment initiatives (e.g., HEFCE, 2011b).

reasonably accurate depiction of the patterns of impact as reflected by trends in scientific publication, however may not capture the full range of characteristics relevant to an assessment of quality. Nevertheless, such indicators can be useful in helping to inform judgements about quality.

This chapter reviews quantitative indicators useful for assessing research quality in the NSE, and presents the Panel's conclusions on validity of indicator options.

5.1 REVIEW OF QUANTITATIVE SCIENCE INDICATORS

The Panel carried out an extensive review of the different types of quantitative indicators in two assessment contexts: research quality (the focus of this chapter) and research trends (see Chapter 6). The aim of the review was to identify indicators valid for field-level research assessments in the NSE. To be deemed valid, the Panel concluded an indicator must be well researched and internationally recognized, with an application validated by existing research and past experience; and be able to support cross-field comparisons of research quality (some indicator types may be valid measures at the field level, but not capable of supporting comparisons across fields). The Panel also considered other secondary criteria that contribute to indicator validity:

- **Timeliness:** The indicator must relate to recent activities (data that relates to research undertaken many years previously does not reflect the current dynamics of the research environment and may lead to inappropriate funding decisions).
- **Behavioural Impact:** The indicator should not present a high risk of resulting in unintended and negative behavioral responses in the research community.
- **Level of aggregation:** The indicator should be relevant and valid in assessments at the field level.
- **Transparency:** The indicator should be transparent and based on publicly available methodologies and data.

5.2 A TAXONOMY OF INDICATORS FOR ASSESSING RESEARCH QUALITY

Figure 5.1 shows a taxonomy of indicators of research quality developed by the Panel. At the highest level, these indicators comprise six types: those based on (i) external research support; (ii) student population; (iii) weighted publication counts; (iv) citations; (v) esteem measures; and (vi) webometrics, such as online paper views and downloads. This group of indicators can also be categorized by measures of research input (e.g., funding); research output (e.g., scientific papers); or research impacts (e.g., citations).

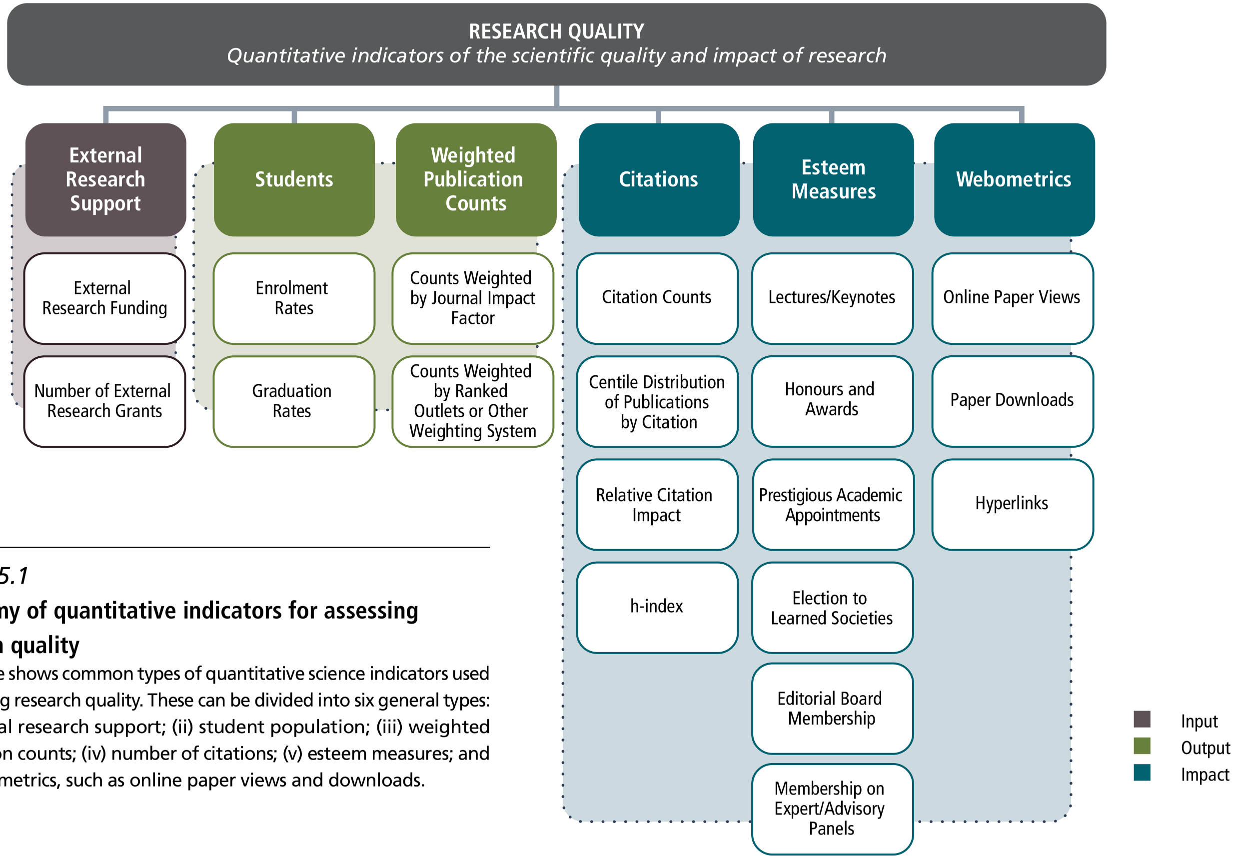


Figure 5.1

Taxonomy of quantitative indicators for assessing research quality

This figure shows common types of quantitative science indicators used in assessing research quality. These can be divided into six general types: (i) external research support; (ii) student population; (iii) weighted publication counts; (iv) number of citations; (v) esteem measures; and (vi) webometrics, such as online paper views and downloads.

The taxonomy identifies sub-types for each of the six indicator types. Indicators based on student populations typically use graduation and enrolment rates by program level and field of study. Metrics based on external research funding are usually constructed using dollar amounts or number of grants awarded from other sources. Webometrics typically consist of simple counts of online activities such as paper views, downloads, and hyperlinks between pages. The most common types of esteem measures include counts of keynote addresses, scientific awards and prizes, memberships in honorary societies, and appointments to editorial boards or advisory committees.

5.3 REVIEWING THE INDICATORS

Table 5.1 summarizes the main conclusions of the Panel concerning indicators for assessing research quality. In addition to deliberative methods, only two of the six types of quantitative indicators were found to be both valid in assessing research quality at the field level and able to support cross-field comparisons.[17]

5.3.1 Valid Indicators of Research Quality

Weighted Publication Counts

Weighted publication counts are counts of publications (see Section 4.2.1) that incorporate some other measure of quality and then weight publications accordingly. The most common form is based on journal impact factors: an article is weighted in accordance with the impact factor of the journal in which it is published (i.e., articles in highly cited journals are weighted more heavily than those in less cited journals). There are theoretically many different ways, however, of constructing weighted publication counts, not all of which will result in reliable or robust indicators. One interesting example of the use of weighted publication counts is the publication indicator used in the allocation of institutional research funding in Norway. This Norwegian bibliometric indicator relies on a two-tier scheme whereby publications in more prestigious outlets are weighted more highly than those in standard outlets. In this case, decisions about which outlets were included in the more prestigious tier were made by the publishing committee at the Norwegian Association of Higher Education Institutions, with input from the national councils in each field of research (Schneider, 2009; Sivertsen, 2009).

17 Some indicators are useful measures of quality within one field, but unable to support comparisons of quality *across* fields given fundamental differences in the relevant data sources or patterns of research activity.

Table 5.1

Validity of Types of Research Quality Indicators

Indicator Type	Is the indicator generally valid for assessing research quality at the field level?	
Weighted publication counts	Yes	Indicators based on unweighted publication counts are typically based on the quantity of research output across fields rather than on the quality of that research and therefore are not valid measures of research quality. Weighted publication counts, however, can be valid measures of research quality. These indicators generally weight publication counts based on some measure of quality such as journal impact factors or other citation-based variables.
Citations	Yes	Citation-based indicators can be a valid measure of research quality at the field level to the degree that they capture information about the scientific impact of research in that field. To be valid, any indicator of this type must be field normalized and based on an adequate citation window, and should only be used where a large proportion of the output is captured by the database.
External research support	No	Metrics based on external research support can be a valid measure of research quality, but not generally at the level of research fields. Such measures are often derivative measures from external peer review processes. Comparisons across research fields, however, are problematic because different fields have different underlying cost structures and base capacities. As a result, these indicators are more applicable in assessments at the level of research institutions or groups.
Student population	No	Student data are valid in many assessment contexts, and student training is an explicit objective in many NSE research funding programs. Since the output of students does not capture information directly about research quality, it should not be used as a primary indicator of research quality at the field level. Comparisons of student enrolment or output, however, may be useful as a secondary source of information, particularly as related to monitoring research trends and research capacity.
Esteem measures	No	Esteem-based indicators may be valid for use at lower levels of aggregation. Since they are not comparable across research fields, they are not applicable at the field level.
Webometrics	No	Webometrics have the potential to be valid in assessing research quality at the field level in the future. These types of indicators, however, are largely experimental and have not yet been validated through use in prominent, national research assessment exercises.
Deliberative approaches	Yes	Deliberative approaches, such as peer and expert review, are the only available approach to assessment of research quality capable of taking into account dimensions of research quality not amenable to quantification. Such approaches can be valid and applicable to field-level assessments; when applied at the field level, however, they are typically supplemented with quantitative data.

The Norwegian Ministry of Education viewed the adoption of this two-tier indicator as a way of protecting against the risk of an increase in publication quantity at the expense of publication quality (European Commission, 2010).

Citation-based Indicators

There are many types of citation-based indicators, including those based on straightforward citation counts, those based on an analysis of the distribution of citations in a given publication set, those based on relative citation impact, those based on journal impact factors, and finally the h-index, which is based on both publication and citation history. Not all of these indicators, however, are valid or informative in the context of research quality at the national field level. In reviewing indicators of this type, the Panel identified three key conditions related to construction and use of indicators that are instrumental in determining whether any given citation-based indicator is valid.

First, any citation-based indicators used to assess and compare research fields should be field normalized. Different fields of research have different citation cultures and practices; therefore, the baseline levels of citations differ across fields (Garfield, 1979; Moed *et al.*,1985; Butler, 2007). For example, papers published in biomedical research fields typically cite many more sources than those published in mathematics (Leydesdorff & Opthof, 2010). As a result, comparing absolute counts of citations across fields is misleading. Instead, citation-based indicators must be constructed to take into account the baseline level of citations in a field. The most prominent example is the average relative citation indicator (sometimes referred to as relative citation impact), which compares the average level of citations in a particular field in a particular country to the world average level of citations in that field.

Second, any citation-based indicator of quality should be based on a sufficiently long citation window (the time span in which citations to a publication or set of publications are counted). Citations take time to accrue. It may be a matter of months or even years before the impact of published articles becomes evident in new citations. Past research suggested that, for the natural sciences and engineering, an appropriate citation window is typically between three and five years (REPP, 2005; Van Raan, 1993; Moed *et al.*, 1985). More recent evidence, however, has proposed that a citation window as short as two years may be appropriate in some cases (e.g., Van Rann *et al.*, 2007). This evidence implies that citation-based indicators should be limited to assessing research published at least two years previously. Any attempt to use citation-based indicators for more recent research may result in spurious or misleading findings.

Third, a sufficient amount of the relevant research in a field must be included in the data source used to support the analysis. Typically, citation indicators are constructed based on citations in peer-reviewed, academic journals, which are captured in existing bibliometric databases such as Thomson Reuters' Web of Science and Elsevier's Scopus (see Section 4.2.5). The validity of indicators using this data then depends on the degree to which the total output of research in that field is captured and reflected in these databases. If a large percentage of research output in a field is not captured in this way, it implies that a large number of the total citations to other work in the field are not included in the analysis. A useful guideline can be drawn from U.K. research relating to the use of bibliometrics in support of the future Research Excellence Framework (REF). A scoping study on the use of bibliometrics in this context suggested that, in general, for citation-based indicators to be valid in a field, at least 40 per cent of the relevant research outputs in that field should be journal articles captured in the available commercial databases (HEFCE, 2009, Annex H). Since a large percentage of NSE research is accounted for by scientific journal articles, existing databases tend to have fairly good coverage of research output in these fields (HEFCE, 2009; REPP, 2005). Some fields, such as computer science and some engineering sub-fields, may have more limited coverage (see Box 5.1).

Box 5.1
Variation in the Use of Indicator by NSE Field

In general, there are few significant differences in the application of indicators across NSE fields. Peer-reviewed journal articles remain the primary mode for disseminating research output in most NSE fields. Although the required citation windows tend to be broadly similar, evidence suggests that citations tend to accumulate more slowly in mathematics and engineering than in the natural sciences (HEFCE, 2009).

Robust bibliometric indicators can be constructed for broad application in a majority of NSE fields, with the possible exception of some areas of engineering and computer science where journal articles may be a less significant outlet for research dissemination (Butler & Visser, 2006; Glänzel *et al.*, 2006; Lisée *et al.*, 2008). In these areas, if coverage of research output in standard bibliometric databases is insufficient, research evaluators may consider alternative measures that capture conference proceedings, patents, or citations in patents, but with the understanding that such measures are not perfect substitutes for citation-based indicators, and therefore have different implications (Lisée *et al.*, 2008).

Several citation-based indicators satisfy these three conditions, including the average relative citation impact indicator (see Section 4.2.2); and various indicators based on citation distributions (e.g., a country's share of the top one per cent or five per cent of cited publications in a particular field). Given that different types of citation-based indicators are capable of meeting these conditions, the subsequent choice of which indicators are used in an assessment context is dependent on the specific objectives of that assessment and what information is most relevant.

5.3.2 Invalid Indicators of Research Quality at the Field Level

In the Panel's view, the other four types of quantitative indicators identified in Table 5.1 are not valid for assessing research quality at the level of nationally aggregated fields.

The h-index

There are many types of invalid citation and publication based indicators as well. For example, the well-known h-index (Hirsch, 2005) is based on a combination of publication and citation information. The h-index is noted here only to caution against its increasing popularity. Bibliometric experts generally do not consider it to be a valid or useful indicator (Gingras, 2008; Van Leeuwen, 2008). H-indices are highly correlated with overall publication counts (Van Leeuwen, 2008), indicating that they are heavily influenced by the simple quantity of papers produced. This metric was also designed for use at the level of individual researchers. While there has been some experimental calculation of h-indices at higher levels of aggregation, this use has not been validated through rigorous testing or application in any large-scale national research assessment exercise.

External Research Support

External research support is commonly used as a basis for quantitative metrics in assessments of research quality. These indicators are relatively easy to obtain from data about funding or shares of funding received, and can be grouped by various characteristics such as source of funding or time period. They are often used as a proxy for research quality under the assumption that the competitive ability to attract research funding is reflective of the calibre of past research and the overall reputation of the recipient of the funding (e.g., institution, research group, researcher) (OECD, 2010).[18]

18 One example of the use of such measures at the field level is in the Australian Excellence for Research in Australia (ERA) initiative (see Appendix A).

As a measure of quality, however, these indicators have significant limitations. Their reliability depends on past funding decisions and the underlying expert judgment used in these decisions. All funding decisions, as described in Chapters 2 and 3, are influenced by a variety of factors, such as national priorities, that may not correlate with recent or relevant dimensions of research quality. In addition, comparisons across research fields or institutions are problematic because the amount of required funding invariably differs across fields. The use of such indicators in support of field-level funding allocation decisions raises the possibility that new funding decisions largely reproduce and further entrench past decisions. Although these indicators are not valid for assessing research quality, they can provide useful information about research capacity (see Chapter 7).

Student Population

Another group of readily available quantitative indicators sometimes used for assessments of research quality combines the various aspects of student population dynamics within and across fields. Although these measures can be easily compiled and analyzed at various levels of aggregation (e.g., research groups and labs, institutions, fields of study, regions, countries) or study (e.g., undergraduate and graduate, Masters and PhD), comparisons across fields are problematic. These measures are not valid for assessing research quality at the field level because student choices are influenced by many factors other than perceptions of quality, such as geographic location, personal interest, and perceptions of labour market outcomes. Student population data are most informative in general assessments of research capacity (see section 7.2.2).

Esteem Measures

Professional honours, such as awards, prizes, invitations to keynote addresses, prestigious academic appointments, and memberships in honorary societies, are examples of the recognition that scientists may receive for outstanding and pivotal work in their fields. Esteem-based indicators quantify these means of recognition, and have been used to describe research quality at the individual level (e.g., when assessing suitability for positions or professional advancement). A pilot study undertaken in Australia (Donovan & Butler, 2007) rejected esteem-based measures for use in rigorous evaluations of research quality; however the study did find that these types of indicators may have some usefulness as secondary markers of reputation. The Panel found no evidence of rigorous studies of the use of such indicators at higher levels of aggregation.[19] Timeliness presents by far the most critical constraint for these indicators. Since honours and awards are typically granted with a considerable delay after the achievements, they do not provide current information (Donovan & Butler, 2007). The Nobel Prize is the preeminent example, often awarded several decades after the research in question is performed and recognized. Similarly, data aggregation, which requires custom databases, is resource intensive and subject to data quality concerns. Finally, esteem measures may not have equivalents for all fields, thus invalidating cross-field comparisons (OECD, 2010).

Webometrics

Webometrics represent a new frontier in scientometrics as they measure internet-based production or usage of scientific output (Thelwall, 2009). The most common of these include tracking of online journal access, or views and downloads of research papers. Other more experimental possibilities include analysis of hyperlinks, frequency of use of certain identifiers in search engines, and metrics based on usage or access patterns in particular web-based platforms with a research focus.[20] Webometrics, however, are an inherently less precise indicator of scientific impact of research because they capture all views or downloads, introducing ambiguity and increasing the need for careful interpretation.

One advantage of webometrics in comparison with traditional bibliometrics, is that they have the potential to be more timely (i.e., impact can be measured instantaneously from the number of views or downloads) and comprehensive (in theory, they capture information about a wider set of research outputs than journal articles e.g., presentations, patents, software, and other websites) (Thelwall, 2009). The Panel, however, found no evidence of the systematic use of webometrics in any large-scale research assessment exercise. Since they promise to account for a broader impact of published research, they may in the future offer new information for evaluations of scientific research (Van Noorden, 2010). Thelwall (2008) also suggested that webometrics are unlikely to replace traditional bibliometrics in the near future, but may be used more commonly in preliminary studies or dedicated assessments of online scientific activity.

---

19 There are some exceptions with respect to national research assessments targeted at research institutions. For example, the Australian ERA used these metrics in 2010 and will retain them for 2012; and the United Kingdom has explored the usage of these types of indicators in the context of the RAE (see http://www.inf.ed.ac.uk/admin/rae/esteem.html).

20 As an example, hyperlinks have been used to create and analyze web impact factors, an indicator analogous to journal impact factors (Ingwersen, 1998).

5.3.3 The Role of Deliberative Methods in Assessing Research Quality

While both weighted publications and citation-based indicators are useful measures of the relative impact of scientific research across fields, they cannot be expected to capture all the characteristics of a multidimensional assessment of research quality at the field level. Some research characteristics, such as the relevance of a specific research output to broader socio-economic objectives or the originality and uniqueness of a research program, will not necessarily be reflected by use of either of these indicator types.

Several in-depth studies on methods of research evaluation have concluded that quantitative indicators cannot replace deliberative processes for research assessment and evaluation, such as peer or expert review. For example an NRC (2006) report came to the following conclusion:

> None of the available analytical methods of science assessment is sufficiently valid to justify its use for assessing scientific fields or setting priorities among them. Judgment must be applied to interpret the results from these methods and discern their implications for policy choices. This situation seems unlikely to change any time soon.

This report states that the best way to assess scientific fields is to use an approach where deliberative processes (such as expert review) are informed by analytical methods, including the use of quantitative indicators. A literature review of the role of quantitative indicators for research assessment, conducted as part of the Research Evaluation and Policy Project (REPP) in Australia, also found a widespread consensus in the literature that "...quantitative analyses of research performance should enhance rather than replace peer review"(REPP, 2005). Butler (2007) and Moed (2007) also argue that quantitative indicators have a valuable role in assessing research activity at higher levels of aggregation, but that those indicators should be considered within the context of expert review. As expressed by Butler (2007):

> The character of research quality is complex and multidimensional… No single quantitative measure, or even a 'basket' of indicators, can address all its facets. Nor can a small panel of peers be expected to combine sufficient knowledge of the performance of all a nation's institutions and all a nation's researchers active in their discipline to enable them to arrive at error-free judgments. The most sensible approach is to combine the

two methods, by assembling a group of highly qualified experts in the discipline and arm them with reliable, discipline-specific data to assist their deliberations.

These studies suggest that the most promising assessment approach for assessing research quality at the field level is a model of *informed* expert review, which balances reliance on quantitative indicators with expert judgment.

5.4 CONCLUSIONS

The Panel concluded that most types of quantitative indicators are not valid options for assessing and comparing research quality across fields. Only certain kinds of citation-based indicators and weighted publication counts are capable of providing valid, useful insights into the overall level of research quality at this level of aggregation. Other quantitative indicators may be valid for other purposes, but not for assessment of quality across fields of science at the national level. Combining deliberative and quantitative approaches has particular strength. In assessments of research fields, the use of quantitative indicators based on citations or weighted publication counts can serve as a valuable check on expert opinion, while expert opinion can help ensure that any quantitative information is not misinterpreted. As a result, when it comes to assessing research quality at the field level, the most promising approach is one that relies on both deliberative methods and appropriately constructed quantitative indicators. In Chapter 6 the Panel applies its assessment process to available measures of research trends.

6

Indicators for Assessing Research Trends

- A Taxonomy of Indicators for Assessing Research Trends
- Reviewing the Indicators
- Deliberative Approaches to Assessing Research Trends
- Conclusions

6 Indicators for Assessing Research Trends

Key Points

- Quantitative indicators for assessing research trends can be divided into five major types: grant applications, research funding, researcher population, student population, and bibliometric methods.
- All five indicator types represent potentially valid, informative choices for assessing research trends at the field level in the natural sciences and engineering. Since they capture information about different aspects of research activity on different timescales, a set of indicators offers the most effective approach.
- Deliberative methods are valid for assessing research trends at the field level, and are essential to ensure indicator-based information for decision-making is appropriately interpreted in the context of relevant factors.

In addition to assessment of research quality, research funding agencies require information on research trends and/or the evolution of scientific research (i.e., emerging or declining fields of study, changing research directions, and new patterns of collaboration) to make effective funding allocation decisions. This chapter reviews and evaluates quantitative indicators and deliberative approaches used in the context of assessing research trends. As in Chapter 5, the Panel's discussion focuses primarily on the overall validity of these indicator types (see Section 5.1 for a brief explanation of the Panel's review of quantitative indicators and Appendix B for additional details).

6.1 A TAXONOMY OF INDICATORS FOR ASSESSING RESEARCH TRENDS

A diverse set of indicators (see Figure 6.1 for a taxonomy) is required to capture the various aspects of the changing interests and activities of discovery research over time. Many different types of quantitative indicators can be used in this type of assessment, including: (i) grant applications; (ii) research funding; (iii) researcher population; (iv) student population; and (v) trends identified using bibliometric measures. These indicators, as well as measures of research quality, can be classified as input, output, and impact measures, with each serving a different role in informing research evaluation.

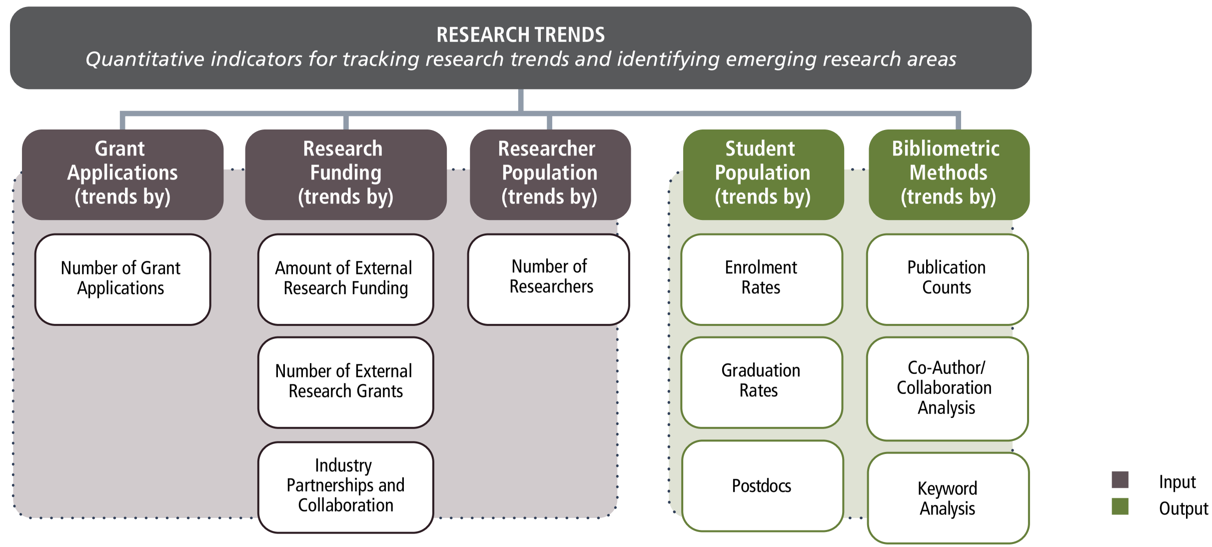


Figure 6.1

Taxonomy of indicators assessing research trends

This figure shows major types of quantitative indicators used in assessing research trends. As can be seen, these can be sub-divided into five major types: i) grant applications; ii) research funding; iii) researcher population; iv) student population; and v) bibliometric methods.

6.2 REVIEWING THE INDICATORS

The Panel concluded the five categories of quantitative indicators described in the taxonomy are valid for assessing research trends (see Table 6.1), provided that they are interpreted correctly. The indicators provide meaningful and valuable insight with respect to the evolving and dynamic nature of fields in the NSE. These metrics capture different and complementary aspects of research trends by targeting different underlying factors and relating to different timescales.

Table 6.1
Validity of Research Trend Indicators

Indicator Type	Is the indicator valid for assessing research trends at the NSE field level?	
Bibliometric methods	Yes	Many advanced bibliometric techniques can be used to monitor research trends and identify emerging research areas. Simple monitoring of publications over time by field or topic of work provides useful information on the level of activity by field. More advanced methods based on co-citation clustering, keyword analysis, and visualization are also valid approaches in many cases.
Grant applications	Yes	A change in a grant application trend provides a benchmark of researcher activity directly relevant to research funding decisions. In general, these indicators provide useful information on overall trends and direction of research in specific fields, but interpretation requires expert judgment and consideration of various field characteristics and variables.
Research funding	Yes	Trends in research funding over time can provide a valid and informative indication of changes in resources flowing to support research in specific NSE areas. Since changes in funding levels, however, may or may not be a response to scientific developments in a field, they must be interpreted with care.
Researcher population	Yes	Indicators based on changes in researcher population can provide valid and informative measures of research trends. Emerging or expanding research areas are likely to attract new researchers over time. Metrics based on changes in researcher population, however, are best viewed as a long-term measure because the researcher population (monitored by departmental affiliation or field of training) will shift slowly over time due to the length of time required for training to undertake advanced research.
Student population	Yes	Indicators based on student population are valid measures of research trends. Fields of research with growing student numbers at the graduate level are likely areas of growing research interest. As in the case of researcher population, these indicators are best viewed as long-term measures.
Deliberative approaches	Yes	Deliberative approaches are valid for assessment of research trends, and critical in ensuring information gained from quantitative metrics is correctly interpreted in light of the research context for a field.

Indicators based on changes in the pattern of grant applications by research topic provide useful and timely benchmarks, and are directly relevant for funding organizations and programs making research funding decisions. Granularity of the metrics may be limited depending on the application or field of research administered by the granting program. Indicators based on student and researcher populations also are highly relevant to tracking interest and activity in research areas over time. These, however, are best viewed as long-term indicators because of the time involved for an investigator to move into a new research field or to train new researchers. The usefulness of these indicators may also be limited to the extent that they are based on standard field or program classifications within university faculties.

6.2.1 Bibliometric Methods

Many bibliometric variables and methods provide valuable insight for monitoring and tracking research trends. There is an extensive scientometric literature exploring these techniques exists (e.g., Upham & Small, 2010; Guo *et al.*, 2011). Simple publication counts (by research field or topic) can provide a reliable, basic guide to changes in intensity of output associated with areas, indicating changing interest and potential new directions. Bibliometrics also offer a diversity of more complex analytical possibilities. Advanced visualization techniques (and the underlying analytics) can enhance the value of bibliometric indicators because they facilitate new ways of presenting and analyzing data to identify previously undetectable patterns, and promise faster and easier recognition of emerging research areas (e.g., Small, 2003; Börner *et al.*, 2003; Boyack & Börner, 2003). For many years, forms of quantitative analysis have been used to monitor bursts of publication activity or research fronts (Goffman, 1966; Goffman & Harmon, 1971). More recently, science maps have been constructed based on keyword analysis to assist visualization of trends. Similarly, co-authorship or co-citation analysis can be used to identify patterns in scientific collaboration, and emerging new clusters of knowledge and their evolution over time (Small, 2006; Upham & Small, 2010).

6.2.2 Trends in Grant Applications, Researcher Population, and Student Population

Trends and patterns in grant applications, researcher population, and student population provide highly relevant and more proximate information for tracking and monitoring of research interests and activity over time. If constructed with care, they provide useful and timely benchmarks, but their validity can be limited by data continuity and granularity for a longer time series. Administrative nomenclature and standard disciplinary classifications of granting agencies and educational programs may inhibit long-term analysis of the dynamics within and across research fields or new interdisciplinary areas. As a consequence, retrospective and prospective extrapolations by field are usually difficult, and interpretation requires contextual knowledge. For example, quantitative data do not capture societal, political, or economic factors in different jurisdictions but will invariably influence application rate, researcher mobility, or attractiveness of a research field. Country-specific institutional settings (e.g., creation or elimination of graduate programs or investment in infrastructure and facilities) also play an important role in shaping long-term trends. All these factors are not synonymous with actual changes in scientific activity or interest, but rather with broader research capacity (see Chapter 7).

6.2.3 Trends in Research Funding

Although trends in level of research funding in an area can be an important and informative indicator of overall level of resources supporting research in that area, care must be taken in interpreting this type of information, which can be misleading. One rationale for using funding data to assess research trends is that new or emerging research fields are likely to attract growing levels of investment. Though this may be generally true, research investment decisions often reflect external factors such as changes in funding policies and priorities or changes in global macro-economic conditions and markets. These variables affect both general availability of research funding and its allocation across fields. As a result, funding trends are better interpreted as a determinant of research capacity (see Chapter 7).

6.3 DELIBERATIVE APPROACHES TO ASSESSING RESEARCH TRENDS

Deliberative methods are used to track research trends as well, usually in the context of establishing research priorities. Typically, an expert panel, with depth of expertise in a specific field, provides analysis of the current state and direction of research within its area of expertise and, based on that review, highlights promising areas of future research. Decadal studies carried out by the National Research Council in the United States are a well-known example of this type of initiative. These studies rely on expert panels and extensive consultation with relevant communities of scientists to offer insight into current state and future prospects of research in particular fields (see Box 6.1). Such studies may also include other types of consultative processes. An example from Canada is the report prepared by the NSERC Long-Range Planning Committee on prospects for research in subatomic physics for the period 2006 to 2016. This report relied on a call for submissions from the community and a meeting of physicists active in the field (NSERC, 2006b). Other methods for assessing research trends that rely fundamentally on expert opinions or judgment may include formal surveys or questionnaires targeting experts in the field.

Box 6.1
National Research Council Decadal Surveys in the United States

An important example of an expert panel process used, at least in part, to assess research trends at the national level are "decadal surveys" undertaken by the National Research Council (NRC) in the United States. In the 1960s the NRC began conducting periodic reviews of the state of research in physics and astronomy. These reviews, completed at approximately 10-year intervals by panels appointed by the NRC, are typically carried out through intensive consultation with the relevant research community. In recent years, the decadal survey model has been adopted by other disciplines (see NRC, 2007a, 2007b, 2010).

A 2007 review of these initiatives found that the surveys accomplish a number of goals (NRC, 2007a). They provide an authoritative description of research accomplishments in a field, define a compelling research program for the future, explicitly identify research and funding priorities for policy-makers and governments, characterize the existing state of research infrastructure and identify areas where investment is needed, and assist the research community and research funders in making the difficult decisions about the relative merits of different research directions and priorities (NRC, 2007a).

In general, these surveys appear to be well regarded by researchers and government departments and agencies. But some challenges associated with the model have been noted, including potentially inaccurate cost projections and risk assessments, and the need for recommendations to be resilient in the face of unexpected changes in the budgetary or political environment (NRC, 2007a). The scale of the NRC's decadal survey model may be prohibitive for other countries, but its long and successful history provides a rich source of experience from which to draw.

6.4 CONCLUSIONS

There is no general consensus, either in the extensive literature on science indicators or from existing international practices, that any one type of quantitative indicator is clearly superior to the alternatives in assessing research trends. The indicators that are available address different underlying factors and are relevant to different timescales. As a result, the choice of indicators or methodologies is dependent on context: the objectives of the assessment process in question, including the unit to be analyzed, the need to delineate specific aspects of activities, and the assessment purpose (i.e., to monitor existing trends or to identify novel areas of research activity). In addition, all available quantitative indicators are subject to risk of interpretation error, and can lead to unintended and negative behavioural consequences if included formulaically in research funding allocation. The most prudent approach to tracking research trends and identifying emerging research areas to support research funding decisions is based on a combination of indicators and expert judgment.

7

Indicators for Assessing Research Capacity

7 Indicators for Assessing Research Capacity

Key Points

- Research capacity indicators can be grouped into five general categories: research funding, people (researchers and students), research infrastructure, patterns in collaboration, and field characteristics.
- Because of the heterogeneity and diversity of data sources, it is impractical to assess these indicators in the same way as for research quality and trends. The validity of any indicator that is selected will depend on both the context of the intended use and the quality of underlying data.
- The most promising approaches to assessing research capacity at the level of a field in the natural sciences and engineering include quantitative indicators, but also expert opinion and deliberation.

In addition to measures of research quality and research trends, a comprehensive understanding of research capacity is relevant for informed field-level funding allocation decisions. Without this information, funding agencies risk investing in fields that have an insufficient capacity to support expansion or continuation of ongoing research. This chapter provides an overview of indicators that may be used to assess national research capacity at the field level. Given the heterogeneity of these indicators and their data sources (which differ substantially by country), it is impractical to evaluate the validity of research capacity indicators as was done for indicators of research quality and trends. Key features of these types of indicators, however, are highlighted below.

7.1 A TAXONOMY OF INDICATORS FOR ASSESSING RESEARCH CAPACITY

To enable high-quality research in existing and newly emerging NSE fields, funding decisions need to take into account research capacity, which can be described in terms of past achievements, or results, and of the mechanisms employed in their development (i.e., how support for research is provided). Research capacity evaluated with a comprehensive suite of measures helps to understand how the research environment enables both creativity and quality research. Because indicators to assess research results (e.g., the measures identified for research quality and trends) usually require a longer timeframe, measures of capacity provide helpful insight about effectiveness of the process for providing support. This is important because both the underlying policy and the final funding decisions are part of that process.

The dynamic process of developing research capacity and the environment within which NSE discovery research occurs can be measured using a broad range of indicators. Studies have identified a multitude of variables affecting research capacity, some focusing on macro-level determinants of capacity (i.e., the entire research system) and others on the micro level (i.e., characteristics of highly performing teams).[21] The macro-level aspects of human resource capacity, infrastructure, funding, and collaborations are important in field-level allocation decisions. Inherent characteristics of individual fields also dictate underlying capacity requirements and will nuance the measurement. A taxonomy of indicators is shown in Figure 7.1 and presents examples of indicators within five broad categories: research funding, people (researchers and students), research infrastructure, patterns in collaboration, and field characteristics.

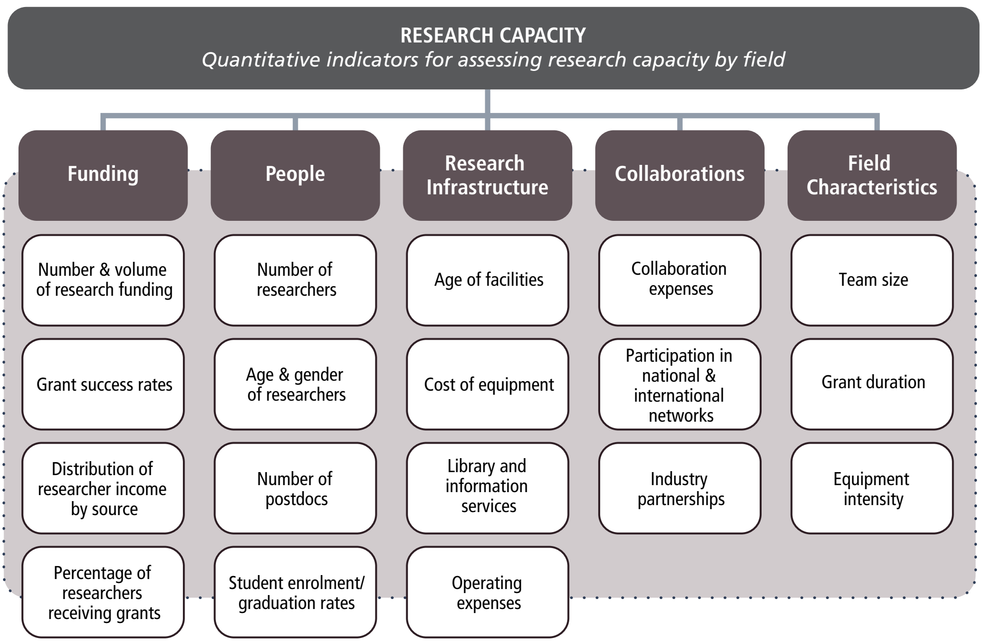


Figure 7.1

Taxonomy of indicators for assessing research capacity

This figure shows a range of quantitative research indicators that may be considered in assessment of research capacity. These indicators can be divided into five major types: i) funding; ii) people; iii) infrastructure; iv) collaborations; and v) field characteristics.

21 Jordan *et al.* (2001) provide a useful review of relevant literature.

7.2 INDICATORS OF RESEARCH CAPACITY

In general, a set of indicators for assessing research capacity would not only capture information across all five categories identified in the taxonomy (see Figure 7.1), but would also reflect significant interactions between and among them. International factors may need to be considered to anchor any indicators used as benchmarks. The discussion that follows is intended to inform choices of these indicator types by highlighting key features.

Indicators of research capacity should address critical questions related to operational capacity, as well as to global competitiveness, because high-impact research that aims to change the world depends on both (Kanter, 1988). Policy-makers may find it helpful to select indicators illustrating opportunities for leverage of a country's investment and its contributions to global research capacity. Although data on global trends are widely accessible, the use of data sources invariably requires caution and a thorough understanding of the way they have been collected.

7.2.1 Funding

For funding allocation decisions, funding data remain the foremost source of information for policy-makers because not only are the data relatively easy to comprehend and obtain in a timely manner (typical budget data may lag a year or two), but, more importantly, they can also be directly connected to a specific funding agency or program. The data also enable relatively easy comparisons across agencies, nations, and budgetary categories of spending. Thus, in the simplest form, spending data describe the volume of input into the research system. They can also be a powerful conveyor of information about important characteristics of research performers, potential opportunities, and the nature of relationships within the system — for example, geographic location, demographics of recipients, and networks of partnerships and collaborations (both national and global). And when combined in various measures of return on investment, funding data can be used to represent the perceived impact of the investment (as expressed in the selection of variables accounted for in the calculations).

Though largely retrospective, funding data continue to be one of the most significant international benchmarking information sources on competitiveness. As a result, these data have a direct impact on future funding, both private and public. The current state of data and tools used for measuring the returns on competitive research grants at higher levels of aggregation, however, does not support their use in program-level analysis and allocation decisions (Koizumi, 2011). Therefore, trends in research and development investments and funding data require diligent and expert consideration when used as an information source for field-level allocation decisions. Australia is an interesting example of the use of research funding data within the context of a research assessment process based on expert review (see Box 7.1).

Policy-makers need to consider both the overall level of funding available for a specific research field and the stability and diversity of that funding. While evaluations of individual research fields often examine the adequacy of funding within a field, comparing the cost of research across fields is a difficult task because there is no universally agreed upon approach. The cost of equipment and facilities has been deemed the most significant determinant of such differences, and these costs may be used as the basis of some comparisons. Many funding organizations, however, generally believe that historical funding trends and current funding requests reflect the perceived cost of research within a field; therefore, additional cross-field comparisons of the cost of research variables are rarely undertaken to support field-level funding allocation decisions.

Box 7.1
Indicators based on Research Funding in Australia's ERA

The Excellence in Research for Australia (ERA) initiative provides an interesting example of the use of external research funding indicators. Data on research income from selected sources are used as measures of research activity (ARC, 2010). These metrics are based on four types of external research income:

- funding from a specific national granting program (i.e., Australian Competitive Grants);
- other public research funding sources;
- industry research funding and funding from international sources; and
- funding associated with the Australian Cooperative Research Centre program.

Indicators are compiled based on income data submitted by universities. The overall ERA assessment process validates these metrics through expert review. Higher levels of external funding are interpreted as a sign of greater research volume and productivity.

For further information, see the case study in Appendix A.

7.2.2 People

The long-term capacity in scientific research is highly dependent on people: devoted students and researchers, and skilled research support staff (Bland & Ruffin, 1992). Long-term sustainability thus depends on training programs that build research skills and enable translations and dissemination of knowledge (across fields and globally). Investing in people — the training of students and the development of world-class researchers — is frequently an explicit criterion of public research funding, and key statistics are readily available. Canada operates a number of world-class programs that provide funding to top researchers. For example, the Canada Excellence Research Chair and Canada Research Chair programs target leading researchers from around the world. Both programs recognize that maintaining Canada's leadership and competitiveness in scientific research depends on the development of world-class researchers and the research networks they attract and build (CRC, 2011).

With the great variety of indicators used to monitor human resource capacity in national research systems, it is important to select a basket of measures that capture information about the various performers of discovery research e.g., researchers, students, support staff, and management capacity at funding agencies (ESF, 2009). Population-based metrics provide valuable insights into the location, geography, and competitiveness of research (to the degree they capture information about the ability to attract high-quality resources) (OECD, 2010). Comparisons across fields, however, can be problematic because the capacity to produce new graduates and researchers depends on many factors that mix aspects of a field with other general (though often significant) events in a country or region. For example, evolving industrial and regional economic needs may cause short-term spikes in the demand for skills, thus prompting higher education student enrolment rates.

7.2.3 Infrastructure

Adequate infrastructure, equipment, and facilities, as well as maintenance provisions, have long been seen as one of the critical determinants of research capacity. For scientific research, knowledge infrastructure is especially important; that means access to world-class collaborations and various other types of initiatives (networks, partnerships, or open platform projects) that can provide researchers with opportunities to gain experience in large-scale studies, leading-edge technologies, and interdisciplinary approaches, while deepening their field-specific knowledge and skills. The significance of research infrastructure, which differs across fields, may affect the development of human resources and collaborative networks, and thus the long-term capacity of a field. The state of infrastructure is therefore a key focus of broader research investment policies and priorities (regional, national, and global) (NRC, 1996b). For example, a 1985 comparison of the scientific performance of high-energy physics in Eastern and Western Bloc nations attributed the smaller output in the Eastern Bloc to limited resources, and inferior research facilities and scientific instruments (and management approach) (Irvine & Martin, 1985). Cyberinfrastrucuture is also an increasingly important determinant of R&D capacity in nearly all fields of research (see Box 7.2).

Box 7.2
Measuring Cyberinfrastructure

Many types of research now require access to high-powered computing facilities and high-throughput networking capacity. Accessibility of sufficient computational resources is often a key constraint in fields undertaking research involving sophisticated computer modelling or other computationally intensive processes (e.g., ecology, epidemiology, proteomics and genomics, ocean and climate science). Scientific work that involves remote instrument operation or necessitates large amounts of data transfer may also require advanced computing facilities and substantial networking capacity.

Fortunately, this is one area of research capacity where quantitative metrics are relatively easy to come by. In recognition of the growing importance of cyberinfrastructure to research capacity, some research funders and related organizations are taking additional steps to monitor and track these measures. A recent edition of the U.S. National Science Foundation's *Science and Engineering Indicators* provides an instructive example (National Science Board, 2012). The report includes data for U.S. research institutions on external and internal network bandwidth, number of internet connections, and the extent of wireless network coverage on university campuses. As the use of advanced computational facilities and networking technologies continues to permeate scientific research, these types of metrics will become increasingly important in the future.

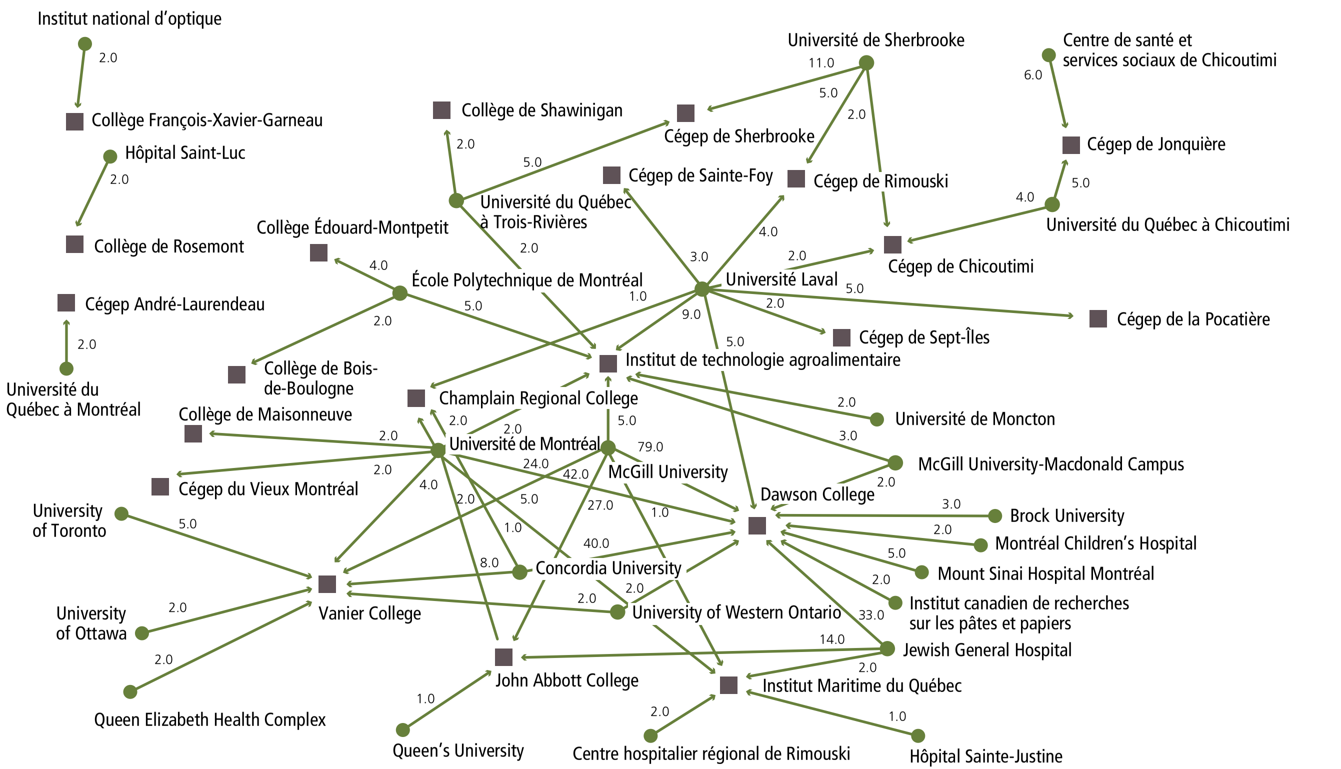
7.2.4 Collaborations

In recent years, a surge in various types of partnerships and collaborative efforts has taken over the research "towers" of traditional disciplines and institutional arrangements, thus extending global research capacity (National Science Board, 2010). Collaborations and interdisciplinary teams are now more frequently required or strongly suggested for research proposals. With collaborations commonplace, information exchanges transcend institutional, organizational, and national boundaries, and provide access to shared technology platforms with robust data. Although collaborations have been shown to boost research output (based on publications), there is no easy way to measure and report achieved capacity.

In Canada, patterns in research collaborations can be observed both nationally and globally. Figure 7.2a illustrates collaborations between colleges and universities in Quebec, while Figure 7.2b illustrates collaboration between Canadian universities.

When understood as a social system, contemporary science presents a number of paradoxes. This is evident when trying to assess and measure interdisciplinary research — research that "brings together multiple disciplinary perspectives or develops new interdisciplinary fields" (Powell *et al.*, 2011). Field-level funding allocation decisions address the choices of supporting novelty and innovations in research or strengthening what is already there. The paradox lies in the fact that strong fields are the foundation of creativity and potential synergies harvested by interdisciplinary initiatives (see Figure 7.3), which often emerge to address broader societal problems, such as climate change, sustainable development, or international security (Jacobs & Frickel, 2009). Thus the scientific merit of interdisciplinary work is often captured in relation to the epistemic cultures of established disciplines (Mansilla, 2006).

The increasing prevalence of large teams of collaborators presents new opportunities for development of interdisciplinary research platforms and networks including shared knowledge, research methods, and tools. The measurement of interdisciplinary research, however, remains a challenge. Recent initiatives have tried to address this knowledge gap (for example, see Paletz *et al.*, 2010), perceived by some as temporary and generated by anecdotal evidence (Porter & Rafols, 2009). Well-developed measures for interdisciplinary research are needed to improve understanding of its role and contributions, potentially illuminating needed adjustments in policy instruments and funding programs.

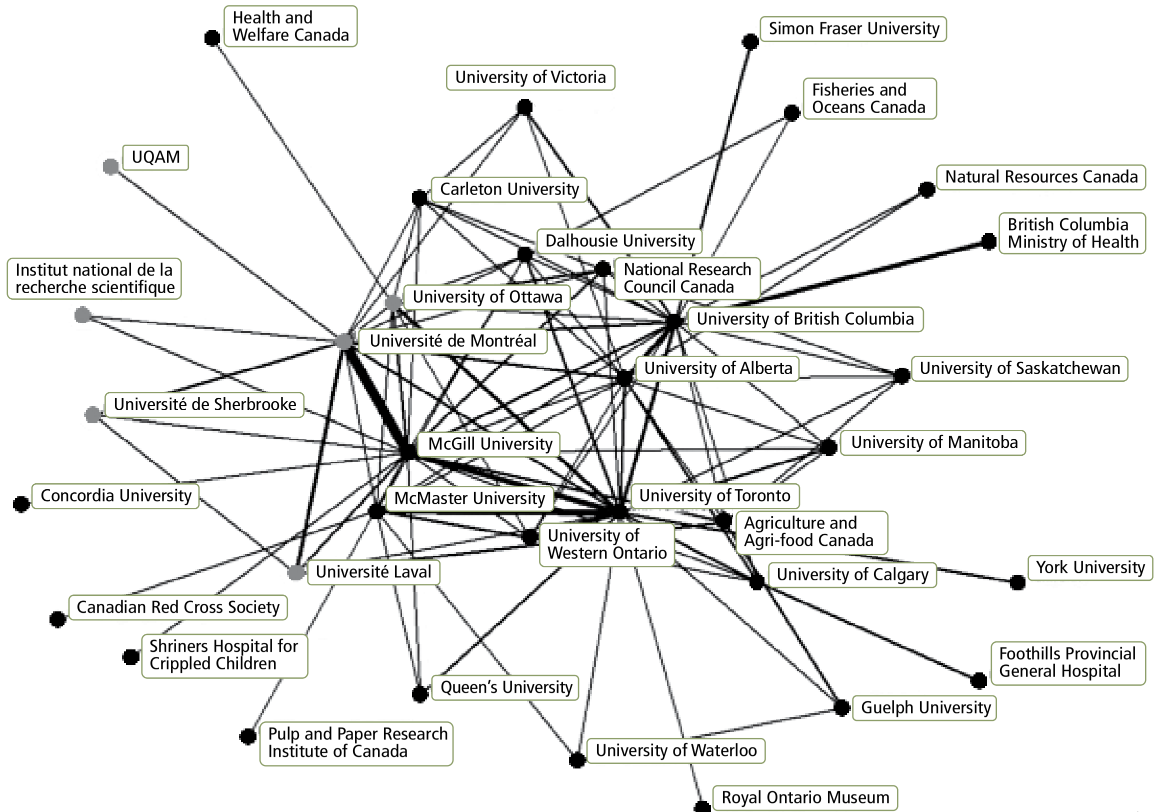


(Observatoire des sciences et des technologies, as cited in Piché & Lapostolle, 2009)

Figure 7.2a

University-College research collaborations in Quebec, 1972–2005

This figure is a visualization of research collaborations between and among universities and colleges in Quebec (where at least one of the collaborating institutions is based in the province of Quebec). The lines between institutions signify the number of research collaborations. The green circles indicate a university or a teaching hospital and the purple squares indicate a college or Cégep.

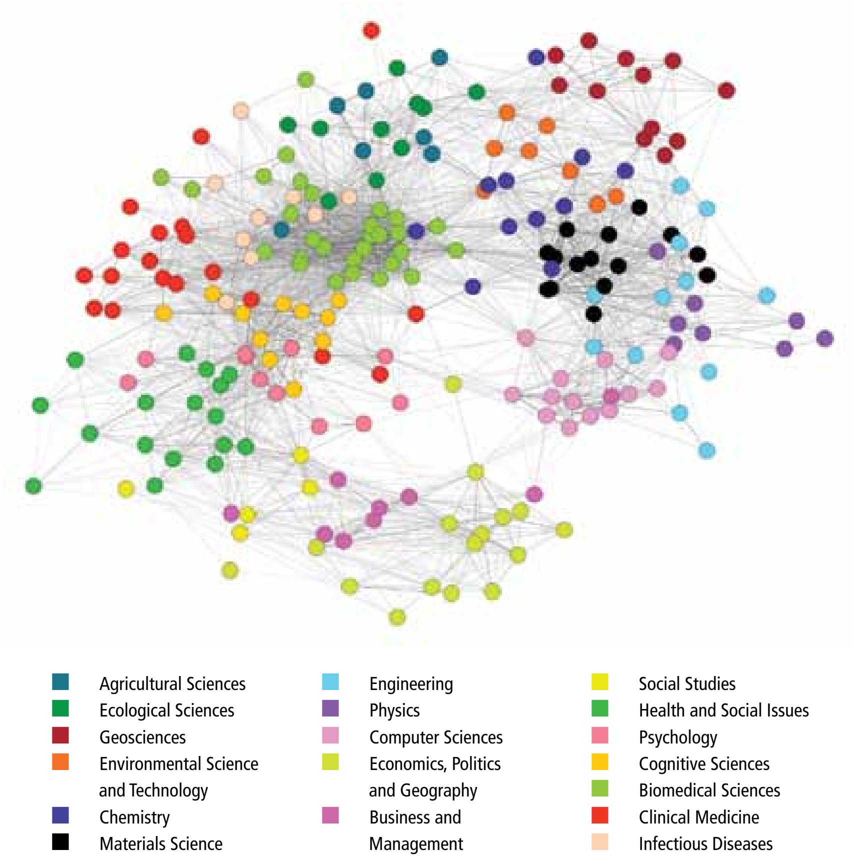


(Larivière *et al.*, 2006)

Figure 7.2b

Inter-institutional collaborative activities of Canadian researchers

This figure is a visualization of research collaborations between and among Canadian universities, colleges, and research institutions as measured by co-authored scientific publications. The lines between institutions signifies that there were 250 or more joint publications between researches in the NSE from 1990-2002. Black nodes represent Anglophone and/or bilingual institutions and the grey nodes represent Francophone institutions.



(Rafols *et al.*, 2010)

Figure 7.3

A global map of science by field

This figure shows the cross-disciplinary complexity of current scientific research, depicting connections between fields based on citations in journals in 2007.

7.2.5 Field Characteristics

There are also a number of possible indicators related to basic characteristics of research activity by field, such as average team size, average grant duration, and equipment intensity (i.e., the cost of research equipment and facilities relative to the total cost of research). These indicators, which highlight important differences in the nature of research activity across fields, are often related to specific funding programs and calculated using information from grant applications. Funding agencies routinely use these types of indicators to monitor the general evolution of research funding programs. These data, however, can also be used to inform more nuanced analyses of research performance. For example, one study of researchers funded by grants from the U.S. National Institutes of Health found that mid-sized labs (defined as those with annual funding in the range of US$750,000) showed the highest levels of research productivity and impact, as measured by article output and average impact factors (Wadman, 2010). In general, these types of characteristics reveal useful information about changes in research capacity across fields over time.

7.3 CONCLUSIONS

Similar to the assessment of research quality and research trends, there is no single indicator capable of capturing the relevant aspects and dimensions of research capacity. Nonetheless, there is general agreement that the relevant dimensions of capacity include human resources, funding, collaborations, infrastructure, and key characteristics of research activity in a given field. As with quantitative indicators discussed for research trends, using large pools of data creates a risk of misinterpretation and misuse if considered without appropriate contextual knowledge or field-specific expertise. In many cases, there is no straightforward answer to the question of how to select indicators of research capacity. Many publicly available data sources, however, provide cost-effective access to a broad set of standardized quantitative indicators that allow for meaningful comparisons across fields and for international benchmarking. Once again, deliberative processes can provide effective and efficient strategies for validation of data selection and consideration (for an example, see Harnad, 2008).

Indicators of research capacity serve a dual purpose in science budget allocation decisions. First, because public funding of basic research aims to develop long-term sustainable capacity across a diverse spectrum of existing and emerging research fields, the extent to which this is happening needs to be periodically assessed. Second, because world-class capacity is essential for the quality of science in existing and emerging fields, an appropriately selected basket of diverse indicators can help identify current or future areas of difficulties. In doing so it can help orient budget allocation decisions towards targeted capacity development within, but also across, fields (e.g., financing state-of-the-art computational equipment and facilities, maintaining field-level funding when alternative sources of funds are not currently available). The use of these indicators must be tempered by differences across fields. Since the cost of research capacity differs by field, the cost of maintaining a level of capacity relative to desired quality also differs by field.

8

Conclusions

- Responding to the Charge
- Responding to the Sub-questions
- Final Conclusions

8 Conclusions

Key Points

- Many quantitative science indicators are sufficiently robust to provide meaningful information about research at the level of nationally aggregated research fields. In almost all contexts, however, multiple indicators should be used to capture information on different aspects of research performance in the NSE.
- Effective indicator use is context dependent. International best practices therefore offer limited insight with respect to use of science indicators and assessment strategies. Whether an indicator is reliable or informative often depends as much on the evaluation context as on the construction of the indicator.
- Quantitative indicators should be used to inform rather than replace expert judgment. With respect to national research assessment in the NSE in the context of funding allocation, the weight of the evidence suggests that the best approach relies on a balanced combination of quantitative data and expert judgment.
- Mapping funding allocation decisions directly to indicators is far too simplistic, and is not a realistic strategy. Indicators may reveal useful information but funding allocation decisions are complex. As a result, any indicator or assessment process, no matter how robust, does not obviate the need for careful, strategic planning and judgment on the part of research funding agencies.

The deliberations undertaken by the Panel in response to its charge were wide ranging. The Panel considered the context of research funding allocation processes and how that context affects choices about indicators and assessment strategies. It also reviewed existing science indicators and assessment options, along with recent international experience with science indicators in support of research funding allocations. This chapter synthesizes and reiterates the main findings that emerged from the Panel's deliberations.

8.1 RESPONDING TO THE CHARGE

MAIN QUESTION

What do the scientific evidence and the approaches used by other funding agencies globally have to offer, in terms of performance indicators and related best practices in the context of research in the natural sciences and engineering, carried out at universities, colleges, and polytechnics?

The main question posed to the Panel has two parts. The first part asked for a consideration of what the existing scientific evidence has to offer in terms of using science indicators in support of research funding allocation for the NSE. The second part asked what insights can be gained from an examination of the approaches of funding agencies around the world in the use of these indicators and assessment methods.

In response to the first part of the charge, the Panel undertook a comprehensive review of the relevant scientific literature, which encompasses a wide-ranging and diverse body of evidence. Although the conclusions reached by researchers are not always in agreement, several clear findings emerged.

Many science indicators and assessment approaches are sufficiently robust to be used to assess science performance in the NSE at the level of nationally aggregated fields.

For example, bibliometric indicators based on weighted publication counts and citation-based indicators — when appropriately normalized by the field of research and based on a sufficiently long citation window — can be useful metrics in assessing the overall scientific impact of research in a given field at the national level. Many other types of quantitative indicators, such as those based on student or researcher population, research funding levels, and the state and quality of available scientific infrastructure and equipment, can be useful in characterizing research trends or national research capacity in certain assessment contexts. The distinction regarding the level of aggregation, however, is crucial. Many bibliometric indicators are subject to substantial margins of error when applied at the level of individual scientists. As a result, the application of bibliometric (and many other scientometric) indicators to individual researchers or small research teams in an evaluative context (especially when tied to research funding allocations) must be undertaken with a high level of caution. Many of these methodological issues prove much less acute, however, when indicators are aggregated to the level of national research fields.

Quantitative indicators should be used to inform rather than replace expert judgment in the context of science assessment for research funding allocation.

Although many types of quantitative indicators can be reliable and informative in science assessments at the national field level, these indicators should not be used to support research funding allocation without expert judgment. There are several reasons for this. First, some dimensions of research are not readily amenable to quantification e.g., the originality or creativity of a research program, the quality of the methods employed. Detailed contextual knowledge of the research environment in an area or recent developments in a field may also be required to interpret information emerging from quantitative indicators accurately. In addition, the formulaic use of quantitative science indicators in an assessment context, without the potentially moderating addition of expert judgment, increases the risks of negative behavioural consequences in the research community. For example, publication-based indicators may provide incentives for researchers to increase their output of publications at the expense of the quality of those publications.

In the past, academic literature on the use of quantitative indicators in science assessments often positioned the decision as one relying exclusively on either quantitative indicators or traditional, deliberative methods based on peer or expert review. Increasingly, this is no longer the case. The body of evidence now available recognizes that the most promising strategies, particularly with respect to higher levels of aggregation, rely on a balanced use of quantitative indicators and expert judgment (e.g., Moed, 2007; Butler, 2007).

Mapping funding allocation decisions directly to indicators is far too simplistic, and is not a realistic strategy.

The course of research developments is fundamentally unpredictable, and consequently past performance may sometimes be a poor predictor of future performance in discovery research. In most areas of scientific work, there is no compelling reason for certainty that past successes will lead to future successes or past failures to future failures. As a result, science indicators — essentially a measure of past performance — may not provide a reliable guide to future prospects. Research funding agencies, therefore, should exercise caution when using these indicators in research funding allocations. How information from science indicators should be utilized in support of a research funding allocation process is far from straightforward. In most respects, neither the existing body of evidence nor the experience of international research funders justifies a singular funding allocation or response to the results from a particular indicator. For example, funding agencies may choose to increase the allocation of resources to an area of research weakness (as identified by certain indicators) to bolster performance, or, alternatively, direct resources away from areas of research weakness and towards strengths. These choices are driven by the strategy of a funding agency and program, and the priorities adopted by policy-makers and funders in a specific research funding context. Overall, the Panel found no evidence of a single correct funding response to assessment results.

In response to the second part of the charge, the survey of the practices and experience of funding agencies around the world lends additional support to the key findings outlined above. The Panel undertook detailed case studies of national research assessment practices in 10 countries, and further considered a wide range of evidence from the experiences of other countries in its deliberations. Many countries show an evolution towards increased reliance on both deliberative methods and quantitative indicators in national research assessment exercises. Australia recently adopted a national research assessment system based on expert review informed by quantitative indicators. The United Kingdom explored the possibility of relying solely on quantitative indicators with its planned replacement for the RAE, but ultimately retained a system based on expert judgment and peer review, though one with an expanded role for metrics. While the United States does not have an analogous national research assessment process, the U.S. National Research Council has undertaken international benchmarking exercises that rely on expert review of quantitative data to inform assessment of national research performance in particular fields, and judged this a promising approach. Finland relies on both expert review and quantitative data in reviews of its scientific performance. Other countries, such as the Netherlands, rely on other models where institutions or research teams submit reports with quantitative data, which are then used in the process of supporting a review by selected experts. In all these cases, national governments and research funders prioritize the use of a combination of indicators and expert judgment in undertaking national research assessments.

The main finding, therefore, of this Panel's research is clear. Both the current state of scientific evidence and a review of current international practices support the conclusion that there are many ways in which science indicators can be informative and useful in aiding research funding allocation decisions at the national and field level. These indicators, however, should always be accompanied by expert judgment — both to ensure that indicator-based information is correctly interpreted and that research funders carefully consider the appropriate funding response to any information they may convey.

8.2 RESPONDING TO THE SUB-QUESTIONS

Sub-question #1

What existing qualitative and quantitative indicators and metrics are relevant to budget allocation in the context of support for research in the natural sciences and engineering, and how can they be categorized (e.g., shelf life; cross-disciplinary and international comparability; relevance to interdisciplinary vs. focused disciplinary areas; and applicability to emerging vs. established research areas)?

Existing science indicators and assessment strategies can be categorized in many different ways. This report has distinguished between those based on deliberative methods, such as peer or expert review, and those based on quantitative indicators, including publication and citation counts, numbers of researchers or students, research funding amounts, and grant applications. The Panel categorized indicators and assessment strategies by their intended function, and focused on three general assessment objectives: research quality, research trends, and research capacity.

For each assessment objective, the Panel developed a taxonomy of potential methodologies and indicators (Chapters 5 through 7), and assessed the validity of these indicators with respect to the objective. All types of indicators discussed in relation to these objectives can potentially be used to provide valuable information for determining research priorities and making resource funding allocation decisions in specific contexts.

Although the Panel recognized the existence and use of indicators and assessment methodologies related to assessing the socio-economic impacts of research, exploring and characterizing these types of methodologies in any depth were not within the scope of this study.

Sub-question #2

What are international best practices in the construction, methodological review, and use of quantitative and qualitative indicators for research evaluation and budget allocation in support of basic research in the natural sciences and engineering?

In the Panel's view, international "best practices" offer limited insight with respect to science indicator use and assessment strategies. The construction and application of indicators are context dependent. Whether an indicator is informative or reliable depends as much on the specific context of its use as on the nature and construction of the indicator. Research assessment practices at the field level occur in many different contexts internationally: they are carried out with respect to different funding programs or policy priorities, different types of research institutions and different funding systems, and by different types of research funders. As a result, *no single indicator, set of indicators, or assessment strategy offers an ideal solution in research assessment contexts for NSE discovery research*. The individual circumstances of the assessment and the research funding context must be considered.

There are however some general methodological guidelines that can be identified. Science indicators should never be used in isolation. When it comes to science assessment, there are always multiple dimensions of measurement at stake, calling for balanced sets of multiple indicators. The most reliable approach to science assessment incorporates these indicators into a model of informed, expert review. Some general methodological guidelines for developing an approach to science assessments are presented below, focusing on three types of science assessment relevant to informing NSE research funding allocations:

- **Research quality:** Weighted publication counts and certain citation-based indicators can provide useful information to research funders for assessing research quality. For the assessment of the scientific impact of research at the national field level, indicators based on relative, field-normalized citations (e.g., average relative citations) are the best available quantitative metrics. At this level of aggregation, when appropriately normalized by field and based on a sufficiently long citation window, these measures provide a defensible and informative assessment of the impacts of past research in the NSE.

- **Research trends:** Many indicators provide valid and useful information for assessing research trends. The best approach relies on a combination of indicators including those based on trends in grant applications by research topic and in the student population, and on advanced bibliometric variables or techniques building on trends in publications and citations.
- **Research capacity:** Many variables may be both relevant and helpful in assessing research capacity. Given the diversity of these variables, and the varying quality of their underlying data sources, it is impossible to define which ones are best suited in a given situation. Here research funders may consider a combination of indicators from five main categories: funding, infrastructure, student and researcher populations, collaborations and networks, and field characteristics such as the cost of research.

Sub-question #3

Considering the foregoing, and in light of the Government of Canada Science and Technology Strategy and NSERC's objectives for the support of research, what key considerations (e.g., risks, advantages/disadvantages, behavioural and institutional consequences) and principles emerge in determining defensible use and balance/weighting of performance indicators/metrics for budget allocation?

It was not the Panel's mandate to provide policy recommendations for national NSE assessment strategies. Rather, it was asked to review international practices and the available body of evidence on the use of science indicators in support of research funding decisions. Commenting on the implications of Canada's S&T strategy, or on the funding objectives of NSERC for the choice of assessment indicators and methods in relation to specific funding programs, is therefore beyond its remit. Any use of science assessment indicators by a funding agency would necessarily take into account national science and technology priorities as well as the mandate of the funding agency and the objectives of the funding program directly linked to the assessment. Fundamentally, any selection of indicators for evaluative purposes — whether scientific or otherwise — should be done with careful consideration of the proximal and final objectives of the funding program and agency in question.

The methodological guidelines for science indicator use described above are broadly applicable, and should help mitigate the risks associated with undertaking these types of assessment of scientific work. The risks, advantages, and disadvantages of various assessment approaches and indicators have been discussed throughout the report. It would not be prudent for the Panel to go beyond those guidelines, given that one of the report's key findings is that appropriate indicator use is context dependent. In the process of its deliberations, however, the Panel developed some guiding principles for developing a process for NSE assessment in the context of informing research funding allocation:

- **Context matters:** Effective use of science indicators or assessment strategies, as applied to research fields in the NSE, is context dependent. Thus any approach should take into account national science and technology objectives as well as the goals and priorities of the organization and funding program.
- **Do no harm:** Attempts to link funding allocation directly to specific indicators have the potential to lead to unintended consequences with negative impact on the research community. Promising strategies identified by the Panel for mitigating this risk include a balanced set of indicators and expert judgment in the assessment process.
- **Transparency is critical:** Assessment methods and indicators are most effective when fully transparent to the scientific community. Such transparency should include both the assessment methods or indicators (e.g., indicator construction and validation, data sources, criteria, procedures for selecting expert reviewers) and the method or process by which the indicators or assessments inform or influence funding decisions.
- **The judgment of scientific experts remains invaluable:** Many quantitative science indicators are capable of providing reliable and useful information in the assessment of discovery research at the national and field level. In the context of informing research funding decisions, however, quantitative indicators are best interpreted by experts with detailed knowledge and experience in the relevant fields of research.

8.3 FINAL CONCLUSIONS

The Discovery Grants Program has been a critical component of Canada's research funding landscape for several decades. Discovery research funded through this program has led to numerous, world-leading scientific advances, and past evaluations of the DGP have found the program to be extraordinarily effective in meeting its stated goals. The DGP has a high probability of continuing to meet these objectives in the future, given a prudent balance between informing funding allocation through expert judgment and available quantitative indicators.

Fundamentally, the principal message of this Panel is two-fold. First, many science indicators are sufficiently robust to provide useful, reliable information on various characteristics of science performance and the research environment. Although these indicators and methodologies are not without limitations, they are sufficiently well developed to be informative at the level of nationally aggregated research fields. Second, deciding how to use indicator-based information in the context of making research funding decisions is far from straightforward. The evidence suggests a direct mapping of allocation decisions to indicators is far too simplistic, and is not a realistic strategy.

As a result, quantitative indicators are far from obviating the need for human expertise and judgment in the research funding allocation decision process. Indicators should be used to inform rather than replace expert judgment. Given the inherent uncertainty and complexity of science funding decisions, these choices are best left in the hands of well-informed experts with a deep and nuanced understanding of the research funding contexts in question, and the scientific issues, problems, questions, and opportunities at stake.

References

References

Académie des sciences. (2011). Du Bon Usage de la Bibliometrie pour l'Évaluation Individuelle des Chercheurs. Paris, France: Académie des sciences.

Academy of Finland. (2010). The State and Quality of Scientific Research in Finland 2009. In P. Löppönen, A. Lehvo, K. Vaahtera & A. Nuutinen (Eds.), (pp. 288). Helsinki, Finland: Academy of Finland.

Academy of Finland. (2011). Chemistry Research in Finland 2005-2009: Evaluation Report (pp. 137). Helskinki, Finland: Academy of Finland.

Adams, J., & Gurney, K. (2010). Funding Selectivity, Concentration and Excellence – How Good is the UK's Research? Oxford, England: Higher Education Policy Institute.

Aksnes, D. W. (2009). *The Use of Bibliometric Indicators in Research Evaluations in Norway*. Paper presented at the 14th Nordic Workshop on Bibliometrics and Research Policy, Stockholm, Sweden.

ARC (Australian Research Council). (2008). ERA Indicator Descriptors. Canberra, Australia: ARC.

ARC (Australian Research Council). (2010). ERA 2010 Evaluation Guidelines. Canberra, Australia: ARC.

ARC (Australian Research Council). (2011). ARC Profile. Retrieved May 9, 2011, from http://www.arc.gov.au/about_arc/arc_profile.htm

Archambault, É., Campbell, D., Gingras, Y., & Larivière , V. (2009). Comparing bibliometric statistics obtainted from the Web of Science and Scopus. *Journal of the American Society for Information Science and Technology, 60*(7), 1320-1326.

Archambault, É., Vignola-Gagne, É., Côté, G., Larivière, V., & Gingras, Y. (2006). Benchmarking scientific output in the social sciences and humanities: The limits of existing databases. *Scientometrics, 68*(3), 329-342.

Australian Government. (2011). Improvements to Excellence in Research for Australia Retrieved December 15, 2011, from http://minister.innovation.gov.au/carr/mediareleases/pages/improvementstoexcellenceinresearchforaustralia.aspx

Bernanke, B. S. (2011, May 16). *Promoting Research and Development: The Government's Role*. Paper presented at the New Building Blocks for Jobs and Economic Growth, Washington (DC).

Bland, C. J., & Ruffin, M. T. (1992). Characteristics of a productive research environment:Literature review. *Academic Medicine, 67*(6), 385-397.

Bonaccorsi, A., Daraio, C., Lepori, B., & Slipersæter, S. (2007). Indicators on individual higher education institutions: addressing data problems and comparability issues *Research Evaluation 16*(2), 66-78.

Börner, K., Chen, C., & Boyack, K. (2003). Visualizing knowledge domains. *Annual Review of Information Science and Technology, 37*(1), 179-255.

Boyack, K., & Börner, K. (2003). Indicator-assisted evaluation and funding of research: Visualizing the influence of grants on the number and citation counts of research papers. *Journal of the American Society for Information Science and Technology, 54*(5), 447-461.

Bozeman, B., & Sarewitz, D. (2011). Public value mapping and science policy evaluation. *Minerva, 49*(1), 1-23.

Bush, V. (1945). Science The Endless Frontier. Washington, (DC): US Office of Scientific Research and Development.

Butler, L. (2003). Explaining Australia's increased share of ISI publications – The effects of a funding formula based on publication counts. *Research Policy, 32*(1), 143-155.

Butler, L. (2007). Assessing university research: A plea for a balanced approach. *Science and Public Policy, 34*(8), 565-574.

Butler, L., & Visser, M. (2006). Extending citation analysis to non-source items. *Scientometrics, 66*(2), 327-343.

CAHS (Canadian Academy of Health Sciences). (2009). Making an Impact: A Preferred Framework and Indicators to Measure Returns on Investment in Health Research. Ottawa (ON): CAHS.

CCA (Council of Canadian Academies). (2006). The State of Science and Technology in Canada. Ottawa (ON): CCA.

CCA (Council of Canadian Academies). (2009). Innovation and Business Strategy: Why Canada Falls Short. Ottawa (ON): CCA.

Chubin, D. E., & Hackett, E. J. (1990). *Peerless Science: Peer Review and U.S. Science Policy*. Albany (NY): State University of New York Press.

Cooper, R. G., Edgett, S. J., & Kleinschmidt, E. J. (2012). Portfolio Management: Fundamental for New Product Success. *Product Innovation Best Practices Series*: Stage-Gate International and Product Development Institute Inc.

Cozzens, S. E. (1999). Are new accountability rules bad for science? *Issues in Science and Technology, 16*(4), 59-66.

CRC (Canada Research Chairs). (2011). Canada Research Chairs Retrieved December 16, 2011, from http://www.chairs-chaires.gc.ca/home-accueil-eng.aspx

De Bellis, N. (2009). *Bibliometrics and Citation Analysis: From the Science Citation Index to Cybermetrics* (Illustrated ed.). Lanham (MD): Scarecrow Press.

DIISR (Department of Innovation Industry Science and Research). (2011a). Research Block Grants Retrieved May 9, 2011, from http://www.innovation.gov.au/Research/ResearchBlockGrants/Pages/default.aspx

DIISR (Department of Innovation Industry Science and Research). (2011b). The Process for Determining Joint Research Engagement (JRE) Scheme Grant Amounts. Canberra, Australia: DIISR.

Donovan, C., & Butler, L. (2007). Testing novel quantitative indicators of research 'quality', esteem and 'user engagement': An economics pilot study. *Research Evaluation, 16*(4), 231-242.

Dutton, J. A., & Crowe, L. (1988). Setting priorities among scientific initiatives. *American Scientist, 76*(6), 599-603.

EPSRC (Engineering and Physical Sciences Research Council). (2011). International Review Reports Retrieved March 7, 2012, from http://www.epsrc.ac.uk/newsevents/pubs/reports/Pages/internationalreviews.aspx

ESF (European Science Foundation). (2009). Evaluation in National Research Funding Agencies: Approaches, Experiences and Case Studies. A report of the ESF Member Organisation Forum on Ex-Post Evaluation of Funding Schemes and Research Programmes (pp. 92). Strasbourg, France: ESF.

European Commission. (2010). Assessing Europe's University-Based Research. Brussels, Belgium: European Commission.

Evaluation Associates Ltd. (1999). Interdisciplinary Research and the Research Assessment Exercise. Bristol, UK: HEFCE.

Fealing, K. H., Lane, J. I., Marburger III, J. H., & Shipp, S. S. (2011). *The Science of Science Policy: A Handbook*. Stanford (CA): Stanford Business Books.

Freeman, R., & Van Reenen, J. (2009). What if congress doubled R&D spending on the physical sciences? *Innovation Policy and the Economy, 9*, 1-38.

Garfield, E. (1979). Is citation analysis a legitimate evaluation tool? *Scientometrics, 1*(4), 359-375.

Georghiou, L., & Roessner, D. (2000). Evaluating technology programs: Tools and methods. *Research Policy 29*, 657-678.

Geuna, A., & Martin, B. R. (2003). University research evaluation and funding: An international comparison. *Minerva, 41*(4), 277-304.

Gingras, Y. (1987). La réception des rayon X au Québec: Radiographie des pratiques scientifiques. In M. G. Fournier, Gingras, Y.; Keel, O.; (Ed.), *Sciences et Médecine au Québec, Perspectives Sociohistoriques*. Québec: Institut Québécois de Recherche sur la Culture.

Gingras, Y. (2008). Du mauvais usage de faux indicateurs. *Revue d'Histoire Moderne et Contemporaine, 5*(4), 67-79.

Glänzel, W., Debackere, K., Thijs, B., & Schubert, A. (2006). A concise review on the role of author self-citations in information science, bibliometrics and science policy. *Scientometrics, 67*(2), 263-277.

Godin, B. (2002). Outline for a history of science measurement. *Science, Technology, and Human Values, 27*(1), 3-27.

Goffman, W. (1966). Mathematical approach to the spread of scientific ideas – the history of mast cell research. *Nature, 212*(5061), 449-452.

Goffman, W., & Harmon, G. (1971). Mathematical approach to prediction of scientific discovery. *Nature, 229*(5280), 103-104.

Government of the United Kingdom. (2006). Science and Innovation Investment Framework 2004-2014. London, United Kingdom: Government of the United Kingdom.

Government of the United Kingdom. (2010). Government Response to the House of Lords Science and Technology Select Committee Report: "Setting Priorities for Publicly Funded Research". London, UK: Coalition Government.

Gray, G. W. (2009). Introduction and Historical Development. In D. Demus; Gray, G.W.; Spiess, H.-W.; Vill, V.; (Ed.), *Physical Properties of Liquid Crystals* (1 ed., pp. 522). New York (NY): John Wiley and Sons.

Guo, H., Weingart, S., & Börner, K. (2011). Mixed-indicators model for identifying emerging research areas. *Scientometrics, 89*(1), 421-435.

Harnad, S. (2008). Validating research performance metrics against peer rankings. *Ethics in Science and Environmental Politics, 8*, 103-107.

HEFCE (Higher Education Funding Council of England). (1997). The Impact of the 1992 Research Assessment Exercise on Higher Education Institutions in England. London, UK: HEFCE.

HEFCE (Higher Education Funding Council of England). (2009). Report on the Pilot Exercise to Develop Bibliometric Indicators for the Research Excellence Framework. London, UK: HEFCE.

HEFCE (Higher Education Funding Council of England). (2010). Research Excellence Framework Impact Pilot Exercise: Findings of the Expert Panels. London, UK: HEFCE.

HEFCE (Higher Education Funding Council of England). (2011a). Assessment Framework and Guidance on Submissions. London, U.K.: HEFCE.

HEFCE (Higher Education Funding Council of England). (2011b). REF2014: Research Excellence Framework. Retrieved April 13, 2011, from http://www.hefce.ac.uk/research/ref/

Herbertz, H., & Müller-Hill, B. (1995). Quality and efficiency of basic research in molecular biology: A bibliometric analysis of thirteen excellent research institutes. *Research Policy, 24*(6), 959-979.

Hicks, D. M. (2004). The Four Literatures of Social Science. In H. Moed (Ed.), *Handbook of Quantitative Science and Technology Studies*. The Netherlands: Kluwer Academic.

Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America, 102*(46), 16569-16572.

Industry Canada. (2007). Mobilizing Science and Technology to Canada's Advantage. Ottawa (ON): Industry Canada.

Ingwersen, P. (1998). The calculation of web impact factors. *Journal of Documentation, 54*(2), 236-243.

Irvine, J., & Martin, B. R. (1985). Basic research in the East & West: A comparison of the scientific performance of high-energy physics accelerators. *Social Studies of Science, 15*(2), 293-341.

ISQ (Institut de la statistique du Québec). (2011). Compendium d'Indicateurs de l'Activité Scientifique et Technologique du Québec *Science, Technologie et Innovation* (2011 ed.). Montreal (QC): Institut de la Statistique du Québec.

Jacobs, J. A., & Frickel, S. (2009). Interdisciplinarity: A critical assessment. *Annual Review of Sociology, 35*, 43-65.

Jordan, G. B. (2010). A theory-based logic model for innovation policy and evaluation. *Research Evaluation, 19*(4), 263-273.

Jordan, G. B., & Malone, E. L. (2001). Performance Assessment. In US Department of Energy Office of Science and Office of Planning and Analysis (Ed.), *Management Benchmark Study*. Washington (DC): US Department of Energy.

Jordan, G. B., Streit, L. D., & Matiasek, J. (2001). Attributes in the Research Environment that Foster Excellent Research: An Annotated Bibliography. Albuquerque, (NM); Livermore (CA): Sandia National Laboratories.

Kanter, R. M. (1988). When a thousand flowers bloom: Structural, collective, and social conditions for innovation in organizations. *Research in Organizational Behavior, 10*, 169-211.

King, D. A. (2004). The scientific impact of nations: What different countries get for their research spending. *Nature, 430*(6997), 311-316.

Klavans, R., & Boyack, K. W. (2010). Toward an objective, reliable and accurate method for measuring research leadership. *Scientometrics, 82*(3), 539-553.

Koizumi, K. (2011). Science Policy: A Federal Budgeting View. In K. H. Fealing; J. I. Lane;, J. H. Marburger III; & S. S. Shipp; (Eds.), *The Science of Science Policy: A Handbook* (pp. 289-301). Stanford (CA): Stanford Business Books.

Langfeldt, L. (2001). The decision-making constraints and processes of grant peer review, and their effects on the review outcome. *Social Studies of Science, 31*(6), 820-841.

Larivière, V., & Gingras, Y. (2010). On the relationship between interdisciplinarity and scientific impact. *Journal of the American Society for Information Science and Technology, 61*(1), 126-131.

Larivière, V., Gingras, Y., & Archambault, É. (2006). Canadian collaboration networks: A comparative analysis of the natural sciences, social sciences and the humanities. *Scientometrics, 68*(3), 519-533.

Laudel, G. (2006). Conclave in the Tower of Babel: How peers review interdisciplinary research proposals. *Research Evaluation, 15*(1), 57-68.

Leydesdorff, L., & Opthof, T. (2010). Scopus's source normalized impact per paper (SNIP) versus the journal impact factor based on fractional counting of citations. *Journal of the American Society for Information Science and Technology, 61*(11), 2365-2369.

Lisée, C., Larivière, V., & Archambault, É. (2008). Conference proceedings as a source of scientific information: a bibliometric analysis. *Journal of the American Society for Information Science and Technology, 59*(11), 1776-1784.

Magoun, A. B. (2007). *Television: The life story of a technology*: Greenwood Publishing Group.

Mansilla, V. B. (2006). Symptoms of quality: Assessing expert interdisciplinary work at the frontier. *Research Evaluation, 15*(1), 17-29.

Marburger III, J. H. (2005). Wanted: Better benchmarks. [Editorial]. *Science 308*(5725), 1087.

Martin, B. R. (1996). The use of multiple indicators in the assessment of basic research. *Scientometrics, 36*(3), 343-362.

Martin, B. R., & Irvine, J. (1983). Assessing basic research: Some partial indicators of scientific progress in radio astronomy. *Research Policy, 12*(2), 61-93.

McNay, I. (1998). The RAE and after: 'You never know how it will all turn out'. *Perspectives: Policy and Practice in Higher Education, 2*(1), 19-22.

Meek, V. L., & van der Lee, J. J. (2005). Performance Indicators for Assessing and Benchmarking Research Capacities in Universities *Bangkok Occasional Paper Series* (Vol. 2). Bangkok, Thailand: APEID, UNESCO.

Meho, L. I., & Yang, K. (2007). Impact of data sources on citation counts and rankings of LIS faculty: Web of Science versus Scopus and Google Scholar. *Journal of the American Society for Information Science and Technology, 58*(13), 2105-2125.

Meyer, R. (2011). The public values failures of climate science in the US. *Minerva, 49*(1), 47-70.

Ministry of Employment and the Economy. (2009). Evaluation of the Finnish National Innovation System – Full Report. Helsinki, Finland: Ministry of Employment and the Economy.

Moed, H. F. (2005). *Citation Analysis in Research Evaluation* (Vol. 9). Dordrecht, The Netherlands: Springer.

Moed, H. F. (2007). The future of research evaluation rests with an intelligent combination of advanced metrics and transparent peer review. *Science and Public Policy, 34*(8), 575-583.

Moed, H. F. (2008). UK research assessment exercises: Informed judgments on research quality or quantity? *Scientometrics, 74*(1), 153-161.

Moed, H. F. (2010). Measuring contextual citation impact of scientific journals. *Journal of Informetrics, 4*(3), 265-277.

Moed, H. F., Burger, W. J. M., Frankfort, J. G., & Van Raan, A. F. J. (1985). The application of bibliometric indicators: Important field- and time-dependent factors to be considered. *Scientometrics, 8*(3-4), 177-203.

Morgan, M. G. (2011). Technically Focused Policy Analysis. In K. H. Fealing, J. I. Lane, J. H. Marburger III & S. S. Shipp (Eds.), *The Science of Science Policy: A Handbook* (pp. 120-130). Stanford (CA): Stanford Business Books.

Narin, F. (1976). *Evaluative Bibliometrics: The Use of Publication and Citation Analysis in the Evaluation of Scientific Activity*. Cherry Hill (NJ): Computer Horizons Inc.

National Science Board. (2010). Globalization of Science and Engineering Research: A Companion to Science and Engineering Indicators 2010. Arlington (VA): National Science Foundation.

National Science Board. (2012). Science and Engineering Indicators 2012. Arlington (VA): National Science Foundation.

Nature Editorial. (2010). Assessing assessment. *Nature, 465*(7300), 845.

Nederhof, A. J., & Van Raan, A. F. J. (1987). Peer review and bibliometric indicators of scientific performance: A comparison of cum laude doctorates with ordinary doctorates in physics. *Scientometrics, 11*(5-6), 333-350.

Nightingale, P., & Scott, A. (2007). Peer review and the relevance gap: Ten suggestions for policy-makers. *Science and Public Policy, 34*(8), 543-553.

Nobelprize.org. (2012). Liquid Crystals Retrieved February 27, 2012, from http://www.nobelprize.org/educational/physics/liquid_crystals/history/index.html

NOWT (Netherlands Observatory of Science and Technology). (2010). Science and Technology Indicators 2010: A Summary Report on the Netherlands (pp. 44). Hague, The Netherlands: NOWT.

NRC (National Research Council). (1996a). Understanding Risk: Informing Decisions in a Democratic Society. In P. C. Stern & H. V. Fineberg (Eds.). Washington (DC): NRC.

NRC (National Research Council). (1996b). World Class Research and Development: Characteristics for an Army Research, Development, and Engineering Organization (pp. 72). Washington (DC): NRC.

NRC (National Research Council). (2000). Experiments in International Benchmarking of U.S. Research Fields. Washington (DC): NRC.

NRC (National Research Council). (2006). A Strategy for Assessing Science: Behavioral and Social Research on Aging. In I. Feller & P. C. Stern (Eds.), (pp. 176). Washington (DC): NRC.

NRC (National Research Council). (2007a). Decadal Science Strategy Surveys: Report of a Workshop. Washington (DC): NRC.

NRC (National Research Council). (2007b). Earth Science and Applications from Space: National Imperatives for the Next Decade and Beyond (pp. 456). Washington (DC): NRC.

NRC (National Research Council). (2010). New Worlds, New Horizons in Astronomy and Astrophysics. Washington (DC): NRC.

NSERC (Natural Sciences and Engineering Research Council). (2006a). Final Report of the Evaluation of the Discovery Grants Reallocations Exercise. Ottawa (ON): NSERC.

NSERC (Natural Sciences and Engineering Research Council). (2006b). Perspectives on Subatomic Physics in Canada for 2006-2016. Ottawa (ON): NSERC Long-Range Planning Committee.

NSERC (Natural Sciences and Engineering Research Council). (2008). The Report of the International Review of the Discovery Grants Program. Ottawa (ON): NSERC.

NSERC (Natural Sciences and Engineering Research Council). (2011a). Report on Plans and Priorities 2011-12 Estimates. Ottawa (ON): NSERC.

NSERC (Natural Sciences and Engineering Research Council). (2011b). Discovery Grants Program Retrieved January 15, 2012, from http://www.nserc-crsng.gc.ca/professors-professeurs/grants-subs/dgigp-psigp_eng.asp

NSERC (Natural Sciences and Engineering Research Council). (2012). NSERC – Facts and Figures Retrieved March 7, 2012, from http://www.nserc-crsng.gc.ca/_doc/FactsFigures-TableauxDetailles/2010-2011Tables_e.pdf

OECD (Organization for Economic Co-operation and Development). (1997). *Policy Evaluation in Innovation and Technology: Towards Best Practices.* Paper presented at the Policy Evaluation in Innovation and Technology, Paris, France.

OECD (Organization for Economic Co-operation and Development). (2002). Frascati Manual 2002: Proposed Standard Practice for Surveys on Research and Experimental Development (6th ed., pp. 266). Paris, France: OECD.

OECD (Organization for Economic Co-operation and Development). (2008). The Role of Expert Review in the Evaluation of Science and Technology: Issues and Suggestions for Advanced Practices. Paris, France: OECD.

OECD (Organization for Economic Co-operation and Development). (2010). Performance-based Funding for Public Research in Tertiary Education Institutions: Workshop Proceedings (OECD, Trans.). Paris, France: OECD

OECD (Organization for Economic Co-operation and Development). (2011). OECD Science, Technology and Industry Scoreboard 2011. Paris, France: OECD.

Paletz, S., Smith-Doerr, L., & Vardi, I. (2010). National Science Foundation Workshop Report: Interdisciplinary Collaboration in Innovative Science and Engineering Fields *NSF Workshop on Interdisciplinary Collaboration in Innovative Science and Engineering Fields*. Boston, MA: National Science Foundation.

Piché, S. & Lapostolle, L. (2009). Moving on to Research 2.0! *Pédagogie collégiale, 22*(4), p.1-5.

Pontille, D. (2004). La Signature Scientifique: Une Sociologie Pragmatique de l'Attribution. Paris, France: Le Centre National de la Recherche Scientifique.

Popper, S. W., Wagner, C. S., Fossum, D. L., & Stiles, W. S. (2000). Setting Priorities and Coordinating Federal R&D Across Fields of Science: A Literature Review (pp. 30). Washington, D.C. : RAND Science and Policy Institute.

Porter, A. L., & Rafols, I. (2009). Is science becoming more interdisciplinary? Measuring and mapping six research fields over time *Scientometrics, 81*(3), 719-745.

Porter, A. L., & Rossini, F. A. (1985). Peer review of interdisciplinary research proposals. *Science, Technology, & Human Values, 10*(3), 33-38.

Powell, W. W., Owen-Smith, J., & Smith-Doerr, L. (2011). Sociology and the Science of Science Policy. In K. H. Fealing, J. I. Lane, J. H. Marburger III & S. S. Shipp (Eds.), *The Science of Science Policy: A Handbook* (pp. 56-84). Stanford (CA): Stanford Business Books.

Press, F. (1988). The dilemma of the golden age. *Science, Technology, & Human Values, 13*(3-4), 224-231.

Qiu, J. (2010). Publish or perish in China. *Nature, 463*, 142-143.

Rafols, I., Leydesdorff, L., O'Hare, A., Nightingale, P., & Stirling, A. (2012). How journal rankings can suppress interdisciplinarity. A comparison of innovation studies in business and management. *Research Policy (in press)*.

Rafols, I., Porter, A.L. and Leydesdorff, L. (2010) Science overlay maps: A new tool for research policy and library management. Journal of the American Society for Information Science and Technology. 61(9), 871–1887.

Rayleigh, L. R. J. S. (1942). *The life of Sir J.J. Thomson*. Cambridge: Cambridge University Press.

Reinhart, M. (2010). Peer review practices: a content analysis of external reviews in science funding. *Research Evaluation, 19*(5), 317-331.

REPP (Research Evaluation and Policy Project). (2005). Quantitative Indicators for Research Assessment – A Literature Review. Canberra, Australia: Australian National University.

Research Council U.K. (2006). Report of the Research Councils U.K. Efficiency and Effectiveness of Peer Review Project.

Rigden, J. S. (2000). *Rabi: Scientist and Citizen*. Boston (MA): Harvard University Press.

RIN (Research Information Network). (2009). Communicating Knowledge: How and Why Researchers Publish and Disseminate Their Findings. London, UK: RIN.

Rip, A. (2000). *Societal Challenges to Evaluation*. Paper presented at the U.S. – European Workshop on S&T Policy Evaluation, Bad Herrenalb, Germany.

Roberts, G. (2003). Review of Research Assessment. London UK: Report to the UK Funding Bodies.

Royal Society. (2010). The Scientific Century: Securing our Future Prosperity. London, UK: The Royal Society.

Royal Society. (2011). Knowledge, Networks and Nations. London, UK: The Royal Society.

Sarewitz, D. (2011). Institutional Ecology and the Social Outcomes of Scientific Research. In K. H. Fealing, J. I. Lane, J. H. Marburger III & S. S. Shipp (Eds.), *The Science of Science Policy: A Handbook* (pp. 337-348). Stanford (CA): Stanford Business Books.

Schneider, J. W. (2009). An Outline of the Bibliometric Indicator Used for Performance-Based Funding of Research Institutions in Norway. *European Political Science, 8*(3), 364-378.

Shao, J., & Shen, H. (2011). The outflow of academic papers from China: Why is it happening and can it be stemmed? *Learned Publishing, 24*(2), 95-97.

Sivertsen, G. (2009). A Bibliometric Funding Model Based on a National Research Information System Retrieved March 15, 2012, from http://www.issi2009.org/agendas/issiprogram/activity.php?lang=en&id=108

Slade, C. P. (2011). Public Value Mapping of Equity in Emerging Nanomedicine. *Minerva, 49*(1), 71-86.

Small, H. (2003). Paradigms, citations, and maps of science: a personal history. *Journal of the American Society for Information Science and Technology, 54*(5), 394-399.

Small, H. (2006). Tracking and predicting growth areas in science. *Scientometrics, 68*(3), 595-610.

SSHRC (Social Sciences and Humanities Research Council). (2008). Promoting Excellence in Research: An International Blue Ribbon Panel Assessment of Peer Review Practices at the Social Sciences and Humanitites Research Council of Canada: SSHRC.

Stewart, J. (1995). Models of priority-setting for public sector research. *Research Policy, 24*(1), 115-126.

STIC (Science Technology and Innovation Council). (2009). State of the Nation 2008: Canada's Science, Technology and Innovation System. Ottawa (ON): STIC.

STIC (Science Technology and Innovation Council). (2011). State of the Nation 2011 – Canada's Science, Technology and Innovation System: Imagination to Innovation – Building Canadian Paths & Prosperity (pp. 76). Ottawa (ON): STIC.

Sullivan, P. J. (2011). Seeing the bones of things: A scan of x-rays' early history. *Canadian Family Physician, 57*(10), 1174-1175.

Thelwall, M. (2008). Bibliometrics to webometrics. *Journal of Information Science, 34*(4), 605-621.

Thelwall, M. (2009). *Introduction to Webometrics: Quantitative Web Research for the Social Sciences* (Vol. 1). San Rafael (CA): Morgan and Claypool

Thomas, J., & Mohrman, S. A. (2011). A Vision of Data and Analytics for the Science of Science Policy. In K. H. Fealing, J. I. Lane, J. H. Marburger III & S. S. Shipp (Eds.), *The Science of Science Policy: A Handbook* (pp. 258-281). Stanford (CA): Stanford Business Books.

Travis, G. D. L., & Collins, H. M. (1991). New light on old boys: cognitive and institutional particularism in the peer review system. *Science, Technology, & Human Values, 16*(3), 322-341.

Tukey, J. W. (1962). The future of data analysis. *Annals of Mathematical Statistics 33*(1), 1-67.

Upham, S. P., & Small, H. (2010). Emerging research fronts in science and technology: patterns of new knowledge development. *Scientometrics, 83*(1), 15-38.

Van Leeuwen, T. (2008). Testing the validity of the Hirsch-index for research assessment purposes. *Research Evaluation, 17*(2), 157-160.

Van Leeuwen, T. N., Visser, M. S., Moed, H. F., Nederhof, T. J., & Van Raan, A. F. J. (2003). The Holy Grail of science policy: Exploring and combining bibliometric tools in search of scientific excellence. *Scientometrics, 57*(2), 257-280.

Van Noorden, R. (2010). Metrics: A profusion of measures. *Nature, 465*(7300), 864-866.

Van Raan, A.F.J., Moed, H.F., & van Leeuwen, T.N. (2007). Scoping Study on the Use of Bibliometric Analysis to Measure the Quality in the UK Higher Education Institutions. London, UK: Report to the HEFCE by the Centre for Science and Technology Studies, Leiden University

Van Raan, A. (2004). Measuring Science. In H. F. Moed, W. Glänzel & U. Schmoch (Eds.), *Handbook of Quantitative Science and Technology Research: The Use of Publication and Patent Statistics in Studies of S&T Systems* (pp. 19-50). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Van Raan, A. F. J. (1993). Advanced bibliometric methods to assess research performance and scientific development: Basic principles and recent practical applications. *Research Evaluation, 3*(3), 151-166.

Van Raan, A. F. J. (1996). Advanced bibliometric methods as quantitative core of peer review based evaluation and foresight exercises. *Scientometrics, 36*(3), 397-420.

Wadman, M. (2010). Study says middle sized labs do best. *Nature, 468*, 356-357.

Weinberg, A. M. (1963). Criteria for scientific choice. *Minerva, 1*, 159-171.

Werner, B. M., & Souder, W. E. (1997). Measuring R&D performance – state of the art. *Research-Technology Management, 40*(2), 34-42.

Wooding, S., Hanney, S., Pollitt, A., Buxton, M., Grant, J., & Project Retrosight Team. (2011). Project Retrosight: Understanding the Returns from Cardiovascular and Stroke Research. Cambridge, UK: RAND Europe.

WR (German Council of Science and Humanities). (2004). Recommendations for Rankings in the System of Higher Education and Research; Part 1: Research. (pp. 64). Hamburg, Germany: WR.

WR (German Council of Science and Humanities). (2008a). Research Performance of German Universities and Non-University Institutions in Chemistry. Cologne, Germany: German Council of Science and Humanities.

WR (German Council of Science and Humanities). (2008b). Steering Group Report on the Pilot Study Research Rating in Chemistry and Sociology (pp. 48). Cologne, Germany: German Council of Science and Humanities.

Appendix A

International Case Studies: Case Summaries

- **Appendix A** contains summaries of the international case studies developed by the Expert Panel on Science Performance and Research Funding. **Appendix B** presents the full versions of the case studies and can be found at www.scienceadvice.ca. **Appendix C,** which contains the Panel's review of quantitative indicators for research quality and trends, can also be found at scienceadvice.ca.

Australia

Key Statistics

• Gross Expenditures on R&D (GERD), 2007 (PPP$*):	$15.3 billion
• GERD as a % of GDP:	2.06%
• Total Researchers (Full-Time Equivalent, 2007):	87,140
• # of Researchers per Million People (2007):	4,224
• # of Scientific Publications (2008):	28,313

Source: UNESCO Science Report, 2010

CASE SUMMARY

Australia has substantial experience in national research evaluation and use of quantitative indicators in research funding allocation. Like many countries, Australia has a dual-support research funding system. Institutional research support is provided to universities through a series of block grants administered by the Department of Innovation, Industry, Science and Research (DIISR).[1] The funding allocation formulas of several of these grants have "performance-based" components, and incorporate indicators such as numbers of students and research publications. Project-based research funding is provided through grants from the Australia Research Council (ARC), and grant applications are evaluated on the basis of peer review.

Australia's early use of publication-based indicators in the distribution of institutional research funding (block grants) appears to have led to a substantial boost in the total output of research papers from Australian universities. That increase, however, may have come at the expense of a decrease in research quality and impact (as captured by citations). In response to concerns that Australia's system of performance-based block grants was prioritizing quantity over quality, the government launched a new national research evaluation exercise in 2004. Abandoned in 2007 after a change of government, it was replaced by the Excellence in Research in Australia (ERA) initiative. The ERA is based on expert review,

* Figures are in US$ adjusted for purchasing power parity.

1 On December 15, 2011 the Department of Industry, Innovation, Science, Research and Tertiary Education was established, replacing DIISR.

but informed by a range of quantitative indicators. It uses both citation-based analysis and peer review depending on the discipline being evaluated. Final ERA assessments are based on a standardized five-point rating scale.

The first ERA round of assessment was carried out in 2010, and results were published in early 2011. While it remains too early to draw any significant lessons, early indications suggest that the Australian research community has generally accepted the ERA assessments. Two areas of ERA's experience may be useful for countries considering similar initiatives: (i) the differentiation in indicators used by research disciplines, and (ii) the use of a well-developed base of quantitative data to inform final assessments based on expert judgment.

China

Key Statistics	
• Gross Expenditures on R&D (GERD), 2007 (PPP$*):	$102.4 billion
• GERD as a % of GDP:	1.44%
• Total Researchers (Full-Time Equivalent, 2007):	1,423,380
• # of Researchers per Million People (2007):	1,071
• # of Scientific Publications (2008):	104,968
• World Share of Exports in High-Technology Products (2007):	18%

Source: UNESCO Science Report, 2010

CASE SUMMARY

China is now unequivocally one of the world's leading funders and performers of scientific research.[2] It currently accounts for approximately 10 per cent of the world's total R&D investment and 10 per cent of the world's annual output of scientific papers; it ranks second in the world, after the United States, on both these measures. Government spending on science has increased at an annual rate of over 20 per cent in recent years, reaching US$29.6 billion in 2011.

China's large-scale investments in R&D are guided by the government's Medium to Long Term Plan for the Development of Science and Technology (MLP) (2006–2020). The MLP identifies research priorities and major research projects to be funded. It also specifies several high-level policy goals, such as increasing the ratio of R&D spending to GDP to 2.5 per cent by 2020 and making China one of the top five countries in the world in terms of patent applications on paper citations. Developing the MLP was a massive exercise in research, analysis, and planning, with input from thousands of scientists, engineers, business leaders, and policy-makers.

* Figures are in US$ adjusted for purchasing power parity.

2 "China," as used here, refers to both The People's Republic of China (PRC) and the central government of the People's Republic of China, depending on the context.

Despite the current scale of investment in research, however, research evaluation functions appear to remain underdeveloped in China. There is no known national evaluation of research performance at the discipline level. A recent Organisation for Economic Co-operation and Development (OECD) review of the Chinese innovation system concluded that S&T evaluation mechanisms are poorly integrated into many S&T funding activities, and that China would benefit from more systematic, rigorous, and transparent research evaluation practices. This leaves a limited body of science evaluation experience from which to draw lessons.

Two aspects, however, of China's experience with research priority setting and science indicators are worth highlighting. First, the extensive process used to develop national research priorities identified in the MLP could be an informative model for other countries to study, though one difficult to emulate due to its scale. Second, many Chinese universities and research institutes provide financial incentives to researchers based on their publication output. While these policies may have contributed to the rapid increase in the number of scientific papers in the past decade, evidence suggests they have also resulted in practices such as research falsification, illicit online markets for plagiarized or fictional research, and relatively widespread researcher misconduct. These unintended effects on the research community, which have led to calls for policy reform in the higher education system in recent years, argue for caution in the direct application of financial awards to indicators based on research output.

Finland

Key Statistics

• Gross Expenditures on R&D (GERD), 2007 (PPP$*):	$6.7 billion
• GERD as a % of GDP:	3.46%
• Total Researchers (Full-Time Equivalent, 2007):	40,879
• # of Researchers per Million People (2007):	7,707
• # of Scientific Publications (2008):	8,328

Source: UNESCO Science Report, 2010

CASE SUMMARY

In Finland, quantitative indicators are systematically used to inform budget allocation decisions for institutional research funding, which is administered by the Ministry of Education. About one-third of university research funding depends on research performance as evaluated by indicators. Three-quarters of this amount is allocated based on indicators that measure the extent of activities at universities (e.g., numbers of doctoral degrees completed, teaching and research person-years), and one-quarter is allocated based on indicators of quality and effectiveness (e.g., amount of external funding, numbers of scientific publications). A 2009 international review of the Finnish innovation system recommended several changes to the institutional research funding system, including discipline-specific allocations and increased emphasis on research quality based on quantitative output measures and "light" international peer review. A working group set up by the Ministry of Education is currently discussing the extent to which these recommendations will be implemented, and the Ministry will make a final decision in 2012.

The Academy of Finland, the most important source of project-based research funding for Finnish universities, allocates funding on a competitive basis. Scientific quality is the most important criterion, and there is no pre-determined allocation for research fields. Once a pool of high-quality proposals has been identified, allocation decisions are indirectly informed by several factors, including two types

* Figures are in US$ adjusted for purchasing power parity.

of field-level assessments (conducted by the Academy of Finland). First, every three years, the Finnish research system is broadly evaluated, the latest evaluation occurred in 2009), and an extensive evaluation of research fields is informed by both expert opinion and bibliometric analysis. Second, more narrowly focused national field-level assessments are conducted based on the priorities of the respective Research Councils, and two to three assessments are usually published every year. International experts carry out the assessments, which are informed by collected data, self-assessments of relevant research units (reporting both quantitative and qualitative information), and in-person interviews with research units.

Germany

Key Statistics

• Gross Expenditures on R&D (GERD), 2007 (PPP$*):	$72.2 billion
• GERD as a % of GDP:	2.54%
• Total Researchers (Full-Time Equivalent, 2007):	290,853
• # of Researchers per Million People (2007):	3,532
• # of Scientific Publications (2008):	76,368
• World Share of Exports in High-Technology Products (2007):	9.1%

Source: UNESCO Science Report, 2010

CASE SUMMARY

As a federated country, Germany's 16 states are responsible for core, institutional funding of research at universities, resulting in a diversity of research funding procedures. Many states incorporate performance-based funding allocation (sometimes subject-specific), which is generally based on indicators of third-party funding and PhDs awarded. The main federal third-party funder of basic research is the German Research Foundation (DFG). The DFG's bottom-up competitive funding program, the Individual Grants Program, is comparable to the Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grants Program, including its preliminary allocation of funding across fields of research.

The dominant form of assessing research fields consists of subject-specific rankings and ratings designed to compare research performance of higher education institutions. The Centre for Higher Education Development (CHE), an independent think-tank, is responsible for the most prominent rankings. Government-funded bodies also conduct rankings, including the DFG with its Funding Ranking, and the German Council of Science and Humanities (WR) with its newly developed Research Ratings. WR ratings are developed through peer review and informed by quantitative and qualitative indicators, while DFG and CHE rankings are based solely on quantitative indicators. Although there are many potential lessons to be

* Figures are in US$ adjusted for purchasing power parity.

learned from these activities, especially through the work of the WR, which has reviewed best practices in this area, these assessments are not meant to directly inform funding decisions, and between-field comparisons are purposely avoided.

Two other field-specific evaluations take place in Germany. First, the WR carries out structural analyses of selected research fields in German academia, and may make recommendations to federal and state bodies and universities on the use of funds to further these fields. Indicators do not appear to play a major role here. Second, the state of Lower Saxony carries out field evaluations through informed peer review, partly to inform the development of criteria for its performance-based funding.

The Netherlands

Key Statistics

• Gross Expenditures on R&D (GERD), 2007 (PPP$*):	$11.0 billion
• GERD as a % of GDP:	1.63%
• Total Researchers (Full-Time Equivalent, 2007):	51,052
• # of Researchers per Million People (2007):	3,089
• # of Scientific Publications (2008):	22,945

Source: UNESCO Science Report, 2010

CASE SUMMARY

The Netherlands has two noteworthy national research assessment processes. First, the Netherlands Observatory of Science and Technology (NOWT) reviews the national S&T system semi-annually based on existing quantitative indicators of research inputs and outputs. Although the focus is on providing an overview of the national research system, these reports also include field- and institution-level assessments of research outputs (primarily based on bibliometrics). Second, the Netherlands has developed a unique system of research assessment in universities based on a combination of self-evaluation and peer review by external evaluation committees. The current framework for this process, which has been in development since the early 1990s, is captured in the Netherlands Standard Evaluation Protocol (SEP) 2009–2015, published by the Royal Netherlands Academy of Arts and Sciences (KNAW). Neither the NOWT assessments nor the SEP process are directly linked to specific research funding mechanisms; however, both can reasonably be expected to influence research priority setting and management at the national and institutional levels.

A relative lack of evidence on the impacts of the approach to research evaluation in the Netherlands limits the conclusions that can be drawn. The SEP process is now well entrenched and appears to be widely accepted by the research community; it has recently been renewed for another six years. There is no indication of current plans within the government to move towards a research assessment system with additional emphasis on quantitative metrics, or one with greater direct

* Figures are in US$ adjusted for purchasing power parity.

linkages between research evaluation outcomes and funding allocation decisions. The combination of a high-level national review of quantitative indicators (the NOWT reports) with more detailed, periodic research assessments based on self-evaluation and external expert review (the SEP assessment process) appears to provide researchers and policy-makers with a robust base of knowledge to facilitate strategic planning and research priority setting. This assumption, however, is based primarily on the lack of evidence of any widespread criticism of the current system.

Norway

Key Statistics

• Gross Expenditures on R&D (GERD), 2007 (PPP$*):	$4.5 billion
• GERD as a % of GDP:	1.62%
• Total Researchers (Full-Time Equivalent, 2007):	26,062
• # of Researchers per Million People (2007):	5,468
• # of Scientific Publications (2008):	6,958

Source: UNESCO Science Report, 2010

CASE SUMMARY

In Norway, quantitative indicators of research quantity and quality (e.g., quality-weighted publication output, number of PhD graduates, amount of external funding) are used to determine the performance-based portion (10 per cent) of institutional research funding for higher education institutions (HEIs). Institutional research funding represents approximately 70 per cent of total funding for HEIs in Norway while competitive, project-based research funding accounts for just 18 per cent. The absence of qualitative review (i.e., by expert or peer review) in determining the performance-based portion of institutional research funding has been a source of controversy and discussion within Norway. Most of the competitive funding at the Research Council of Norway (RCN), the main public funding agency for scientific research, is top-down, often in line with national strategies and priorities. There is a relatively small amount of user-directed project funding through the Independent Grants Program (FRIPRO). The FRIPRO budget is divided between seven broad fields based on historical allocations; any budget increase is allocated in proportion to the total amount of funding received by each field.

The RCN conducts comprehensive discipline evaluations that inform funding decisions at several levels, including within government, the RCN, universities, and university departments. The evaluations are similar to those conducted by the Academy of Finland: the RCN appoints international expert panels informed in

* Figures are in US$ adjusted for purchasing power parity.

their analysis by several qualitative and quantitative sources, including detailed self-assessments by relevant research groups, in-person interviews, and commissioned bibliometric analyses. The use of bibliometrics as a supplement to peer review appears to be accepted and appreciated by all parties involved in the evaluations.

Singapore

Key Statistics

• Gross Expenditures on R&D (GERD), 2007 (PPP$*):	$3.0 billion
• GERD as a % of GDP:	2.52%
• Total Researchers (Full-Time Equivalent, 2007):	27,301
• # of Researchers per Million People (2007):	6,088
• # of Scientific Publications (2008):	6,813

Source: UNESCO Science Report, 2010

CASE SUMMARY

Singapore, which is often referred to as an "Asian success story" considers the funding of S&T as a key pillar of its economy. Upon independence in 1965, Singapore was a Third World country with no mandatory education. In the past 50 years, Singapore has rapidly shifted from a labour-intensive economy towards a knowledge-intensive one, and is now a business-friendly, First World country with a globally oriented economy. Between 1990 and 2009, Singapore's GDP almost quadrupled (A*STAR, 2009). Despite its small size, it is a lead player in certain industries; for example, it has the largest share (eight per cent) of pharmaceutical R&D in the world, twice that of China, its nearest competitor (four per cent).

Although Singapore invests significantly in S&T, the insights it can provide on use of indicators to inform research funding budget allocations at the field level may be limited. To compete on the global market, Singapore has created relatively narrow pockets of excellence by focusing funding on top-down, directed research that aligns with national priorities. Although the Ministry of Education (through the Academic Research Fund) provides some support for basic research, it has not found a satisfactory way to measure research quality across disciplines and thus takes the egalitarian approach of ensuring equal success rates across discipline clusters. National foresighting activities identify new and emerging

* Figures are in US$ adjusted for purchasing power parity.

areas, but no comparably comprehensive national assessment of research fields exists. Quantitative indicators are used to measure progress and set S&T targets on a national and agency level, which may or may not inform future funding decisions. These reviews include some analysis at the field level, but are by no means comprehensive.

South Korea

Key Statistics

• Gross Expenditures on R&D (GERD), 2007 (PPP$*):	$41.3 billion
• GERD as a % of GDP:	3.21%
• Total Researchers (Full-Time Equivalent, 2007):	221,928
• # of Researchers per Million People (2007):	4,627
• # of Scientific Publications (2008):	32,781
• World Share of Exports in High-Technology Products (2007):	6.0%

Source: UNESCO Science Report, 2010

CASE SUMMARY

South Korea (the Republic of Korea) has long prioritized investment in S&T as a critical component of its national strategy for industrial and economic development. In the 1960s, South Korea was one of the world's poorest countries. Today, its economy is 15th largest in the world, and national investment in R&D (in relation to GDP) is fourth highest after Sweden, Finland, and Japan (OECD, 2010).

No systematic, national assessment of research fields directly informs funding allocation in South Korea. It does, however, engage in two noteworthy research assessment processes related to public R&D funding. First, the government undertakes regular, comprehensive performance reviews of its R&D programs. These reviews are based on an expert review model, in which an independent review committee scores each government R&D program on a standardized set of qualitative criteria. Completed program evaluations then directly feed into the government's budget allocation process for R&D programs and departments for that year. Second, South Korea has a tradition of undertaking richly detailed technology roadmaps and identifying industrial technology development priorities at a high level of specificity. These priorities inform both general S&T policy and specific R&D funding programs, and may also help identify areas of basic research important to realizing these technologies.

* Figures are in US$ adjusted for purchasing power parity.

It is difficult to draw concrete lessons from South Korean experience with research evaluation and priority setting at the level of research fields. Its public research funding system is currently undergoing a period of transition. Several R&D agencies combined in 2009 to form the new National Research Foundation of Korea (NRF). Since the country's basic research capacity has long been regarded as under-developed relative to its applied research capacity, the government is taking steps to bolster basic research and provide enhanced support for research at universities. These ongoing changes complicate any lessons the Korean experience may offer, particularly on evaluation of investments in basic research. Korean research evaluators, however, appear to regard both the expert panel (committee) model of regular R&D program reviews and the use of detailed technology development roadmaps and priorities as effective tools that have contributed to Korea's success in development of national R&D capacity

United Kingdom

Key Statistics

• Gross Expenditures on R&D (GERD), 2007 (PPP$*):	$41.0 billion
• GERD as a % of GDP:	1.88%
• Total Researchers (Full-Time Equivalent, 2007):	261,406
• # of Researchers per Million People (2007):	4,269
• # of Scientific Publications (2008):	71,302
• World Share of Exports in High-Technology Products (2007):	3.6%

Source: UNESCO Science Report, 2010

CASE SUMMARY

The United Kingdom is home to one of the world's longest standing and most extensively studied national research evaluation programs: the Research Assessment Exercise (RAE). The RAE is undertaken periodically to evaluate research performance at the field/institutional level in U.K. universities. Evaluation is based on an informed peer review model, with independent review panels adjudicating submissions from research groups at participating institutions. RAE quality assessments are directly tied to distribution of approximately £1 billion of quality-related (QR) institutional research funding provided by the Higher Education Funding Council of England (HEFCE), with funding preferentially allocated towards institutions and departments with world-leading research. The RAE, first undertaken in 1986 and most recently in 2008, will be replaced in 2014 by the Research Excellence Framework (REF), a similar, though modified, assessment exercise. In addition to the RAE, the United Kingdom conducts in-depth analyses of specific research fields through international review panels sponsored by its Research Councils.

Most studies have confirmed that the RAE has been broadly successful in meeting its objectives of encouraging and developing high-quality, world-leading U.K. research. Bibliometric evidence suggests that the RAE has increased both research productivity and research impact (as reflected by citations) in the United Kingdom.

* Figures are in US$ adjusted for purchasing power parity.

Although reviews of the assessment methodology have resulted in changes to the process over the years, they have repeatedly vindicated the model's core reliance on informed, peer review as the basis for national research assessment.

The RAE, however, has also had unintended impacts on researchers and universities. Universities increasingly compete for leading researchers in advance of assessments, and researcher autonomy, morale, and job satisfaction have suffered in some cases as a result. The RAE has also influenced the choice of research outlets and author publication patterns, and may prioritize funding for established research over more risky research. While the REF will, like the RAE, be based on a model of informed peer review, it will also include consideration of research impact through case studies and may include additional bibliometric analysis (at the review panels' discretion). Further insights related to research evaluation may emerge from ongoing studies commissioned by HEFCE in preparation for the launch of the REF in 2014.

United States

Key Statistics

• Gross Expenditures on R&D (GERD), 2007 (PPP$*):	$398.1 billion
• GERD as a % of GDP:	2.82%
• Total Researchers (Full-Time Equivalent, 2007):	1,425,550
• # of Researchers per Million People (2007):	4,663
• # of Scientific Publications (2008):	272,879
• World Share of Exports in High-Technology Products (2007):	13.2%

Source: UNESCO Science Report, 2010

CASE SUMMARY

The United States does not have a centralized, national assessment of research performance at the field level directly tied to research funding allocation decisions. The majority of federal funding for basic research in the natural sciences and engineering is provided by departments and agencies in the form of project-based funding in response to specific funding programs. These programs are, in turn, a combination of standing, open invitations for unsolicited proposals and calls for proposals. Budget appropriations for these departments are established annually by Congress and guided by the President's budget proposal. Once budget appropriations are received, departments have significant autonomy in establishing research funding priorities across disciplines, but are guided by the Obama administration's innovation strategy, which prioritizes several areas of basic research. Grant applications for funding through federal departments are most often evaluated on the basis of peer/merit review, which may explicitly include consideration of potential broader impacts of research.

While the federal government has repeatedly sought to develop criteria for allocating funds across fields of research, little empirical evidence exists that these efforts have substantively affected allocations within agencies, among programs, or across proposals. The National Science Foundation (NSF) publishes information on a range of science outputs by field (including HQP and publications) biennially in its

* Figures are in US$ adjusted for purchasing power parity.

well-known *Science and Engineering Indicators* series. At the request of Congress and federal agencies, the National Academies, along with the National Research Council (NRC), have also undertaken a series of international benchmarking studies of U.S. research fields, as well as “decadal surveys” of researchers to identify future research priorities by field. Federal departments and agencies themselves are subject to evaluation within the ambit of the *Government Performance and Results Act* of 1993, and guidelines have been developed to assist research funding departments in monitoring and reporting on performance of their activities and funding programs. Finally, the White House Office of Science and Technology Policy (OSTP) have been actively engaged in developing new approaches to science assessment. The government is now supporting several initiatives (most notably the STAR METRICS project) to develop more robust approaches to assessing impacts of federal research funding. In general, the dominant paradigm for evaluating and assessing federal science investment in the United States remains that of informed, expert/peer review.

Assessments of the Council of Canadian Academies

The assessment reports listed below are accessible through the Council's website (www.scienceadvice.ca):

- Integrating Emerging Technologies into Chemical Safety Assessment (2012)
- Healthy Animals, Healthy Canada (2011)
- Canadian Taxonomy: Exploring Biodiversity, Creating Opportunity (2010)
- Honesty, Accountability and Trust: Fostering Research Integrity in Canada (2010)
- Better Research for Better Business (2009)
- The Sustainable Management of Groundwater in Canada (2009)
- Innovation and Business Strategy: Why Canada Falls Short (2009)
- Vision for the Canadian Arctic Research Initiative: Assessing the Opportunities (2008)
- Energy from Gas Hydrates: Assessing the Opportunities and Challenges for Canada (2008)
- Small is Different: A Science Perspective on the Regulatory Challenges of the Nanoscale (2008)
- Influenza and the Role of Personal Protective Respiratory Equipment: An Assessment of the Evidence (2007)
- The State of Science and Technology in Canada (2006)

The assessments listed below are in the process of expert panel deliberation:

- Canadian Ocean Science
- Energy Prices – Impacts and Adaptation: Assessing Canada's Preparedness
- Food Security Research in Northern Canada
- Harnessing Science and Technology to Understand the Environmental Impacts of Shale Gas Extraction
- Medical and Physiological Impacts of Conducted Energy Weapons
- Socio-economic Impacts of Innovation Investments
- The Potential for New and Innovative Uses of Information and Communications Technologies (ICTs) for Greening Canada
- The State of Industrial Research and Development in Canada
- The State of Science and Technology in Canada
- The Sustainable Management of Water in the Agricultural Landscape of Canada
- Therapeutic Products for Children
- Women in University Research

Board of Governors of the Council of Canadian Academies*

* Affiliations as at June 2012

Scientific Advisory Committee of the Council of Canadian Academies*

Tom Brzustowski, O.C., FRSC, FCAE, Chair
RBC Financial Group Professor in the Commercialization of Innovation, Telfer School of Business, University of Ottawa (Ottawa, ON)

Susan A. McDaniel, FRSC, Vice-Chair
Professor of Sociology and Prentice Research Chair in Global Population and Economy, University of Lethbridge (Lethbridge, AB)

Michel G. Bergeron, C.M., O.Q., FCAHS
Director, Division of Microbiology and le Centre de recherché en infectologie, University of Laval (Québec City, QC)

Murray Campbell
Senior Manager, Business Analytics Research, IBM T.J., Watson Research Center Yorktown Heights, (New York, NY)

Margaret Conrad, O.C., FRSC
Professor Emerita, Honorary Research Professor, University of New Brunswick (Fredericton, NB)

Marcel Côté
Founding Partner, SECOR Inc. (Montréal, QC)

Jean Gray, C.M., FCAHS
Professor of Medicine (Emeritus), Dalhousie University (Halifax, NS)

Judith Hall, O.C., FRSC, FCAHS
Professor of Pediatrics and Medical Genetics, University of British Columbia (Vancouver, BC)

John Hepburn, FRSC
Vice-President of Research and International, University of British Columbia (Vancouver, BC)

* Affiliations as at June 2012

Daniel Krewski
Professor of Epidemiology and Community Medicine and Scientific Director, McLaughlin Centre for Population Health Risk Assessment, University of Ottawa (Ottawa, ON)

Norbert R. Morgenstern, C.M., FRSC, FCAE
Professor (Emeritus), Civil Engineering, University of Alberta (Edmonton, AB)

John P. Smol, FRSC
Co-Director of the Paleoecological Environmental Assessment and Research Laboratory, Queen's University (Kingston, ON)

Robert Watson
Chief Scientific Advisor, Department for Environment, Food and Rural Affairs (London, United Kingdom)

Joseph D. Wright, FCAE
Retired President and CEO, Pulp and Paper Research Institute (PAPRICAN), Gibsons (BC)