For Official Use **DSTI/STP/TIP(2009)5**

Organisation de Coopération et de Développement Économiques
Organisation for Economic Co-operation and Development

24-Jun-2009

English - Or. English

DSTI/STP/TIP(2009)5
For Official Use

DIRECTORATE FOR SCIENCE, TECHNOLOGY AND INDUSTRY
COMMITTEE FOR SCIENTIFIC AND TECHNOLOGICAL POLICY

Working Party on Innovation and Technology Policy

ENHANCING PUBLIC RESEARCH PERFORMANCE THROUGH EVALUATION, IMPACT ASSESSMENT AND PRIORITY SETTING

29 June - 1 July 2009

Delegates will find attached the draft final CSTP report on Evaluation and Impact Assessment for discussion under item 8 a) of the draft agenda with a view to recommendation for declassification by the CSTP.

Delegates should note that this publication comprises several papers that have previously been reviewed and discussed by TIP including work on peer review, evaluation case studies, impact assessment and priority setting and are therefore encouraged to focus the discussion on the main messages contained in the introductory chapter.

Contact: Mario CERVANTES, Tel: (+33-1) 45 024 94 31, E-mail: mario.cervantesàoecd.org
Ester BASRI, Tel:(+33-1) 45 24 96 24, E-mail: ester.basri@oecd.org

JT03267199

English - Or. English

TABLE OF CONTENTS

Tables

Figures

Boxes

INTRODUCTION

1. Evaluation of publicly funded research has become a central concern of policy makers for two main reasons. Firstly, there is growing demand for evidence-based policies and for evaluation of the results of public investments. More precisely, governments increasingly seek to determine how much they should invest in science and technology (S&T) research and development (R&D) and innovation, identify where to invest and know what society gets in return. Ideally, evaluation should help determine both the economic effects of public investment in R&D and innovation, such as the contribution to growth, and the social impacts, such as better health outcomes. Moreover, policy makers also increasingly want public investment to help meet global challenges, such as energy, security and climate change.

2. Secondly, the demand for evaluation has expanded because OECD countries have substantially increased public investment in R&D despite budget constraints. Governments not only finance R&D in various sectors of performance such as the business sector or the higher education sector, they also fund the performance of R&D on their own behalf. Government budget appropriations or outlays for R&D (GBAORD) measures the funds committed by federal/central governments for R&D. In aggregate, this has been climbing faster than GDP across the OECD in recent years. Since 2001 GBAORD grew by 6.4% annually across the OECD. In addition to direct support, governments also finance business R&D indirectly through the use of tax incentives, an alternative to direct spending for achieving government policy objectives. The costs of these tax credits, in terms of foregone revenue, do not usually appear as R&D support in government budgets, although they may be significant.

3. The demand for effective evaluation tools to inform decisions on research funding and impacts will continue to increase in line with public investments in R&D and innovation as countries try to enhance competitiveness and improve innovation capacity but also due to greater accountability and demands to improve the effectiveness of public support. Evaluation assists governments in their decisions to prioritise R&D and innovation resources, and can help them design research programmes. Moreover, evaluation enhances public accountability, creates a better informed society and raises awareness of the contribution of research to a country's economic and social development. The demand for evaluation is also changing with increased interest in evaluating entire research systems and research portfolios. However the increased emphasis on evaluation has raised a number of important conceptual and methodological challenges.

4. This report reviews recent approaches to evaluation and aims to help countries develop a better understanding of the issues in designing and implementing their own evaluation practices and methods. It brings together work carried out by the Committee for Scientific and Technological Policy (CSTP) Working Party on Innovation and Technology Policy (TIP) in co-operation with the Working Party on Research Institutions and Human Resources (RIHR) including the results from two workshops. The publication also builds on TIP work on Assessing the Socio-Economic Impacts of Public R&D and work from the Working Party of National Experts on Science and Technology Indicators (NESTI) on output indicators for S&T. It is presented in five chapters.

5. Chapter 1 of the report presents key issues in the evaluation of S&T and sets the stage for the remainder of the report which addresses specific topics in more detail including the architecture of evaluation systems insofar as they exist, peer review as a tool for evaluation, and the interplay between

evaluation and priority setting. Chapter 1 begins by recalling the emergence of evaluation of public research, from its initial focus on improving the quality of research by evaluating discrete outcomes of research (*e.g.* publications) to more complex attempts to evaluate and assess economic and non-economic social effects of research as well as environment/ecology/systemic-wide effects. It discusses the diverse approaches to evaluation and notes that most work encompasses economic frameworks. While having many advantages the chapter highlights the limitations of economic-based approaches, particularly in relation to the impact of research on social change. It considers evaluation from different perspectives including governments, civil society, research institutions, funding agencies, evaluators and social scientists. Finally, it draws on the recent history in assessing the social impacts of research and presents new theoretical and empirical efforts (*e.g.* the science of science policy). While this work is still being developed it offers new lenses for analysis.

6. Chapter 2 presents case studies of the systems for evaluating S&T in a select number of countries (Austria, China, Finland, Japan and Korea). The case studies presented illustrate variety both with regard to the comprehensiveness of the evaluation systems, the objectives of the evaluations as well as the way in which they are conducted and used for policy making. In particular, the case studies map the institutional frameworks, actors, regulations and practices for designing and implementing evaluations (ex post or ex ante) in the area of S&T. Together these different elements that support evaluation form the "overall architecture" for evaluation at the level of an institution (e.g. public research institution or funding agency) or in some countries where these different elements operate in a more integrated and systemic manner, a "national system" for evaluation of S&T policies. For all countries studied, the rationale for evaluation is to improve their R&D performance and to ensure efficient use of the resources invested, yet the roles of the ministries, differ among the countries studied. The mechanisms and tools for evaluation also differ depending on the level of evaluation and country specific factors. Nevertheless, addressing social and economic impacts remains a challenge for most countries, mainly due to the difficulty of attributing impacts to a particular research programme, particularly after a lapse of several years. Scientific impact is easier to measure than social impact, and impact of applied research is more easily measured than basic research. This represents a considerable challenge. The outcomes of these evaluations appear to have the greatest influence on improving the design, implementation and effectiveness of programmes and a more limited impact on the policy decision for funding, with the exception of evaluations of basic research programmes.

7. Chapter 3 of the report assesses the role of "expert reviews" - a broader type of "peer review" - from its role in shaping research funding decisions (e.g. what, who and where to fund research) to its role in evaluating the output and impacts of public investment in R&D and innovation. Expert review is one of the most common methods used to evaluate public funding of S&T. While expert review has many merits including its low cost, it is fast to apply, it is widely accepted as an evaluation tool and it promotes mutual learning, it is currently facing its strongest challenges for several decades. Externally, there is some evidence of discontent among political decision-makers about the ability of peer review to reflect socioeconomic and political priorities. Internally, there is a hollowing out as increasing pressure on researchers' time makes it more difficult to find experts willing to undertake reviews. The chapter begins by describing the definitions, process and use of expert review especially at programme and policy level and then it summarises the challenges. Finally, the chapter concludes by identifying some solutions and good practices for improving the relevance and usefulness of expert review.

8. Chapter 4 presents findings of the TIP Impact Assessment project that relate to the rationale, methods and tools for evaluating public R&D at country or economy wide levels but also at the institutional and programme levels. It discusses key methodological challenges and suggests ways for improving impact assessments based on country experiences.

9. Against the background of increasing support for research and innovation as well as increasing demands for accountability and socio-economic impacts, the question of how ex ante and ex post evaluations relate to the selection of priorities in research and innovation has become central to policy makers. The issues of priority setting and evaluation are two distinct issues with their own dimensions. Priority setting is the conscious selection of activities at the expense of others with an impact on resource allocation in contrast with the type of priority setting that takes place in a self-organising system. Priority setting is concerned with questions such as shall we invest more in basic research or innovation? What are the technologies that have greater private and social returns? Or, in the case of an institution, shall we invest in earth observation platform or a particle collider? The priority setting process has several dimensions which vary with emphasis over time. Historically, thematic priorities like technological priorities have dominated. Then came along mission-oriented priorities to respond to societal demands; more recently, other types of priorities have come to the attention of policy makers such as "functional" priorities or those that affect the functioning of the system (e.g. shall we focus on policy strands instead of technologies?).

10. Chapter 5 analyses priority setting and the role of evaluation in policy feed-back and design. In particular it examines the links between ex post evaluation and ex ante evaluation in priority setting and decision making and seeks to identify best practices for improving both the quality of ex ante evaluations and their usefulness in the policy making process, in particular priority setting. It also assesses the process of priority setting in S&T itself and identifies structural weaknesses as well as best practice solutions.

11. The key messages from this report can be summarised as follows:

- ***There are growing demands on evaluation of S&T***. Demand for evaluation is increasing and changing: from evaluating the quality of research (via peer review) to assessing the outcomes, outputs and impacts of public R&D. There is also increasing interest in evaluating entire research systems and research portfolios and enhancing the role of evaluation as a tool for priority setting and decision making;
- ***Evaluation capacity remains weak and fragmented in countries.*** Evaluation is functionally fragmented in many countries, but elements of a system appear in place. These elements include a clear role for outside evaluators and stakeholders outside research establishment; the importance of setting standards in evaluation and the role of self-evaluation. A system of evaluation also requires establishing follow-up processes and consequences to evaluations;
- ***The peer review process remains a fundamental mechanism for research planning, implementation and evaluation***. Peer review is under stress but solutions to problems exist (*e.g.* extended peer review processes involving non-scientific stakeholders; combing metrics and indicators). There is a need to facilitate and improve the internationalisation of peer review because of increased international collaboration;
- ***Evaluation of social/economic impacts requires new metrics and approaches.*** In addition to new indicators and methodologies, assessing social economic impacts requires stakeholder involvement as well as new communication channels (to decision makers, to agents, stakeholders). Impacts assessment must also balance the tension between (scientific) relevance and social /economic impacts;
- ***Evaluation can no longer be done solely in relation to the national situation.*** The use of international comparisons in national policy analysis is increasingly (*e.g.* FP7 participation rankings, Shanghai university rankings). The evaluation of public policies, especially in those areas where countries are competing, cannot be done without reference to international benchmarks.

- ***Usefulness of Evaluation for policy making varies but is often limited.*** There are different uses of evaluations such as for improving project management as well as strategy. The feed-back /use of evaluations can be constrained by lack of data on negative findings as well as lack of political buy-in from leadership.
- ***Priority setting and evaluation interact but remain two distinct dimensions of policy making.*** Priority setting is a more complex exercise, involving a broader range of actors (*e.g.* funding agencies) and relys on different approaches and methodologies. This increases the need for mechanisms to ensure greater coherence between priority setting and policy making.
- ***The conceptual underpinning of priority setting remains rather weak and expert opinion continues to predominate.*** Improving the process of priority setting through the use of ex ante evaluations requires political "buy-in", commitment to invest in resources and develop skills as well as the creation of indicators and data to monitor policy or programme effectiveness.
- ***Impact Assessment is part and parcel of evaluation*** but the rationale for impact assessment is expanding as are the policy needs including for assessing the impacts of strategic research that falls outside the context of mainstream public research done in universities and government labs.
- ***Measuring impacts is neither straightforward nor an easy task especially as regards demonstrating causality.*** Many of the dimensions through which S&T impact upon society (*e.g.* societal, cultural or environmental impacts) are not easily captured by existing national statistical frameworks, as they are less tangible and therefore difficult to measure or evaluate and therefore difficult to link to policy interventions. Nevertheless, measuring economic impact should not be ignored due to methodological difficulties.
- ***The choice of methodology is not universal but context specific.*** Impact assessment methodologies are not universally applicable, but rather context specific depending on the objective of the impact assessment exercise, the timing of the exercise (ex ante and/or ex post); and the scope and nature of R&D.
- ***New and useful practices are being developed but methodological issues as well as (international) comparability remain challenges.*** These various methodologies are still evolving but until now, none of the available techniques has been able to capture the full range of impacts of public R&D on society, although they have opened new and encouraging lines of investigation.
- ***The international community and the OECD in particular can play a role in diffusing good practices and stimulating policy learning.*** The challenge of evaluation is rising not only in terms of the technical and methodological issues but also in terms of the political challenges. International co-operation can help improve the ability of countries to effectively develop and foster the use of ex ante and ex post evaluation for research and innovation policy.

CHAPTER 1

PUBLIC VALUE AND RESEARCH EVALUATION[1]

This chapter of the report discusses the recent history in the evaluation of public research and new theoretical and methodological efforts including efforts to assess both the social and economic impacts of science. In particular it highlights current approaches to evaluating public values and social impacts of research.

1. Introduction

12. Despite three decades' progress in the ability to conceptualize, measure, and evaluate research impacts, a gaping hold remains in research evaluation methods and technique- the ability to evaluate the *social and public value impacts* of research. Professional researchers have developed powerful economic tools to measure economic impacts of research, sophisticated bibliometric tools to measure validly the impacts of research outputs on scientific fields and the course of science and technology, and peer review techniques continued to be refined and employed in assessing projects, programs and proposals. But there has been remarkably little progress in the ability to measure directly, systematically, and validly the impacts of research on social change has progressed little. Many scientists (e.g. Ziman, 1968) have waxed eloquently about the communal and cultural value of scientific knowledge. However, possible cultural-aesthetics notwithstanding, the value of science inheres in its application. Without rejecting compelling arguments for the intrinsic value of research in intellectual, cultural and aesthetic terms (Fischer, 1997), most policy-makers and citizens seem to agree that the chief purpose for public funding of research is to improve the quality of life (Johnson, 1965).

13. The critical problem for understanding the social impacts of science is that we have no satisfactory analytical tools for understanding the causal impact and magnitude of effects of research on social change. This gap is not surprising when one considers the difficulty of the task and the adolescent stage in the development of research evaluation. This chapter highlights current approaches to evaluating public values and social impacts of research. The chapter considers: (1) why such approaches are needed and how they relate to extant approaches to research evaluation; (2) special difficulties or challenges of developing such approaches; (3) possible methods to be employed, including current "public value mapping of science outcomes" work of the author and research colleagues as part of the U.S. National

1. *Barry Bozeman* is Regents' Professor and Ander Crenshaw Professor of Public Policy, University of Georgia, and Distinguished Researcher and Professor of Public Affairs, Consortium for Science Policy and Outcomes and School of Public Affairs, Arizona State University. Contact information: *bbozeman@uga.edu.* The author acknowledges the support of the U.S. National Science Foundation's "Science of Science Policy" program (award number 0738203**,** Arizona State University, *"Public Value Mapping: Developing a Non-Economic Model of the Social Value of Science and Innovation Policy").* The author is grateful for the assistance and ideas of the members of the "Public Value Mapping" project, including Dan Sarewitz (principle investigator), David Guston, Catherine Slade, Ryan Meyer, Erik Fisher, Genevieve Maricle, Walter Valdivia, and Stephanie Moulton. The comments are solely attributable to the author and do not necessarily reflect the views of the U.S. National Science Foundation, the OECD, Arizona State University or the University of Georgia.

Science Foundation program the 'science of science policy.' Before turning to these questions, we consider very briefly the recent history of analytical methods for research evaluation.

2. Historical Roots of Research Evaluation

14. As late as the 1980's, research evaluation field was a field with few practitioners, mostly focused on economic evaluation of industrial firms' internal rate of return.[2] Whereas the Canadian government[3] and some European nations[4] had begun systematic evaluation of publicly funded research, in the United States and many other nations evaluation of public research impacts was not a field at all, but rather an agglomeration of fragmented, largely isolated works, many unpublished.

15. To understand the roots of research evaluation one can consider the state-of-the-art as reported in one of the earliest reviews focusing specifically on studies of the evaluation of publicly funded research. Salasin, Hattery and Ramsay's *The Evaluation of Federal Research Programs* (1980) stated intention was to "identify useful approaches for evaluating R&D programs conducted and sponsored by the federal government" (p. 1) and in pursuit of that objective they interviewed more than two hundred experts in evaluation generally or research evaluation specifically, most of them based in industry. The resulting monograph cited 49 papers, including exactly one journal article (Rubenstein, 1976) and one book (Andrews, 1979) focusing explicitly on systematic evaluation of *government-sponsored* R&D. The monograph identified four problems endemic to evaluating government R&D impacts, including (1) lack of a straightforward definition of effectiveness; (2) multiple and competing objectives; (3) problems in aggregating products and results, especially across programs, and (4) reconciling political and scientific measures of success- a list that would work just as well today. The monograph concluded with a problem identified by a great many of the more than 200 experts consulted: "It is not clear that it is possible to assess research quality based on the immediate outputs of a research project (e.g. reports or journal publications)" (Salasin, Hattery and Ramsay, 1980: p. 62). Today, studies and methods of R&D evaluation have greatly proliferated. But most of the problems identified nearly three decades ago in the Salasin, Hattery and Ramsay's pioneering monograph still exist, particularly the problems associated with a focus on discrete R&D outputs.

16. In 1993, Bozeman and Melkers edited *Evaluating R&D Impacts: Methods and Practice*, an R&D evaluation primer with contributions by leading authorities on such topics as case studies of R&D projects, rate of return on R&D investments, and operations research approaches, among others. This book, which was intended for a technical audience, received a surprising amount of attention from policy makers with a new agenda for assessing impacts of government-funded research. About the same time, the Critical Technologies Institute of the RAND Corporation published a report prepared for the Office of Science and Technology Policy reviewed methods available for evaluating fundamental science (Cozzens, et al., 1994) and this effort, too, received a good deal of attention. One of the earliest OECD research evaluation monographs was produced in the mid-1990s (OECD, 1995) and it too received a good deal of attention. In

2. During the history of modern science and technology policy and research evaluation, the most prominent approach to assessment has been peer review. While recognizing that peer review is crucially important, the present chapter focuses on systematic and potentially quantitative or mixed-method approaches and, thus, does not discuss peer review approaches to research evaluation. Similarly, this chapter does not deal with the many and increasingly useful bibliometic approaches to research evaluation. Since the primary concern here is with social impacts, measures related more to scientific impacts and scientific changes are beyond the study's scope.

3. For a history of government mandated research evaluation in Canada, including research evaluation, see Auditor General (1993). For a history of research evaluation activities in Canada see Barbarie (1993).

4. Several publications provide synoptic reviews of the history and methods of research evaluation in European nations; see for example Luukkonen (2002); Callon, Laredo and Mustar (1997).

short, as of the mid-1990's interest in research evaluation reached in inflection point and its literature, professionalization, dedicated journal pages, and policy applications have increased substantially ever since.

3. Economic Based for Research Impacts Evaluation

17. Each of the three works identified immediately above provided diverse approaches to evaluation but most of them falling squarely within an economic framework for analysis. Economic assessments of research and technology generally fall into two related[5] basic categories, one of which is most relevant to practical research evaluation. Less relevant are to practical evaluation, but well known and influential to broad economic development policy-making, are aggregate-level production function analyses (e.g. Solow, 1957), typically focusing on the contribution of technology to national or regional economic. More useful for research evaluation are those economic studies seeking social rates of return (e.g. Jones and Williams, 1998). While social rate of return approaches can be used in a wide variety of contexts, with respect to research and technology these studies attempt to estimate the social benefits that accrue from changes in technology and relate the value of these benefits to the cost of the investments that produced the changes of interest. Among social rate of return approaches, benefit-cost analysis has been most common and most prominent in project and program-level evaluations of research (see for example Link, 1996a; 1996b; Ruegg, 1996; Audretsch, et al. 2002; Saavedra and Bozeman, 2004).

4. Summary of Advantages of Economics-Based Research Evaluation

18. Economic approaches to research evaluation and, especially, cost-benefit based approaches, have a strong appeal, focusing as they do on discrete science and technology outputs such as the number of articles or patents produced in R&D projects, jobs created by technology transfer programs, and contributions of technology-based economic development programs to regional economies. The utility of these approaches is obvious even to skeptics. While the benefits of economics based approaches to evaluation are explored in more detail elsewhere (Link, 1996b), we can for present purposes summarize those benefits as follows:

1. Evaluation rooted in neoclassical economics seems to hold forth promise of "*harder" more rigorous analysis* and, thus, matches well the policymaker's need for justification of expenditures. Typically, these approaches yield numerical assessments, useful in a public policy domain increasingly dominated by "metrics" (Kostoff, 2001).
2. Whereas most approaches to research evaluation are either atheoretical and exclusively tool-oriented, or based on poorly developed theory, economics approaches can draw from decades of development of relatively strong (for the social sciences) theories of the firm, rational choice and economic growth.
3. While economists recognize that there are values that cannot be well accounted for by monetized units, many have been quite creative in developing quasi-economic techniques based on preference functions and units that mimic rational economic choice (e.g. contingent value analysis [Cummings and Taylor, 1999]).
4. Clearly, economic develop and growth is a driving impetus throughout the world and research policy pursues economic growth. It is not surprising that economics-based approaches to research evaluation underpin economics-rationalized research policy.

5. These approaches are related in the sense that they draw from the same base in economic theory, but for somewhat different applications.

5. Even if one's chief focus, as here, is the social impact of research, it is demonstrably the case that economic change affects social change. If one assumes (as we do *not* here) that economic growth and development generally lead directly to desirable social outcomes, then economic measures become satisfactory indices for social progress. Arguably, the linear model equating the development of basic science-to- technology development-to-economic growth-to-positive social impact is the predominant perspective in many nations (but for an opposing view see Bozeman and Sarewitz, 2005).

5. Limitations of Economics-Based Research Evaluation

19. Despite their many advantages, economics-based approaches to research evaluation have many limitations, especially one interested in gauging the impact of research on social change.

1. First, most economic approaches to research evaluation focus on the discrete products of research. If one if interested in the capacity created by research or in the transformational impacts of research, then cost-benefit or rate of return approaches provide only limited insights.
2. Related, a focus on particular products and projects works best there are crisp boundaries (e.g., a single R&D project) and this provides a limitation inasmuch as most social objectives are highly over-determined.
3. Third, despite efforts to consider implications of future streams of benefit, economics based evaluations tend to be static. They rarely take into consideration the mutability of the "products" evaluated, much less the changes in the persons and institutions producing them.
4. Product-oriented and output-focused evaluations tend to give short shrift to the generation of capacity in science and technology, and to the ability to produce sustained knowledge and innovations.
5. Most important for present purposes, many social benefits and costs of research, science and technology are not well or even validly accounted for in monetary units.

20. It is this latter limitation, the inadequacy of economics-based approaches for measuring and providing understanding about the social impacts of research that is the chief concern. To be sure, economics approaches are not unique in their inadequacy for this task. Currently, no satisfactory method (except, perhaps, case studies) has been developed to validly assess the impacts of research on social change.

6. Social Impacts of Research: Framing the Problem

21. It is perhaps not obvious that new approaches are needed for assessing the social and public impacts of research. Indeed, it may not be obvious what one means by the term "social and public impacts of research." First, let us be clear that in a broader sense of the term (not the concept used here) almost all research is focused on social impact and almost all measures are focused on impact. All research is socially embedded and the development and transmission of the results of research are social processes. In that sense, they inevitably have "social impact." Similarly, since all economic impacts are, first and foremost, social impacts (inasmuch as economies are embedded in societies), then these, too, are social impacts. But the present use of the term can be thought of as focusing on the *end- state* social outcomes, the final results. Thus, societies are concerned about economic development not chiefly because having a funds surplus is inherently desirable and not because a 10% gain in GDP is inherently more desirable than a 9% gain but because of the social impacts that are assumed to accompany these intermediate goals. If research leads to an extra percentage point gain in GDP and at the same time a reduction in public health, environmental quality or personal security then the benefits of the gain are not patent. With some

conspicuous exceptions (e.g. funding for astronomy) public investments in research are invariably rationalized by expected improvements in the social well-being of the citizens providing revenue to support the research. Thus, nations fund work on the human genome because there is an expectation that such funding will ultimately lead to improvements in the health of citizens. Similarly, support for developing technologies is redeemed by the expectation that resultant technologies will demonstrably make people better off. To be sure, the logic is sometimes indirect. Thus, one might assume that support for research leads to new technology which leads to economic growth which, in turn, will provide more disposable income to citizens which they can themselves employ to improve their conditions in way they deem fit. In all likelihood, the linear model does, at least in some circumstances, lead to desirable, even optimal social outcomes, especially in those instances (if any such exist) where the distribution of economic outcomes from research are entirely equitable.

22. Our concern is with more directly observable social and public impacts and with monitoring the direct effects of research on these impacts. Rather than a linear model, this is more a "churn model" (Bozeman and Rogers, 2002), with research leading variously and through often circuitous routes to dead ends, to positive social outcomes, to negative outcomes and, often, to both positive and negative outcomes. The question, then, is "to what extent does research contribute, either positively or negatively, to desired social change and to public values?"

7. Social Impacts of Research: Challenges to Theory and Method

23. The difficulties entailed in validly assessing the end-state social impacts of research are easily identified and impossible completely to resolve. The most important factor is that research is often only one factor in determining social outcomes and is rarely the most important one. Research contributes to such outcomes as, say, reductions in disease, but in most cases factors such as life style choices, economic opportunity, and environmental conditions play roles just as important. Moreover, disease-reduction social benefits from research are often a "best case" in terms of mapping the impacts of research to social impacts. Such desirable social outcomes as poverty alleviation, public education, improvements in housing and protection of public safety generally are highly-overdetermined. When research plays any significant role it is in concert with a great many other social, economic and natural determinants. In such circumstances it is virtually impossible to parse out the contribution of research. Whether one employs standard economics-based approaches such as cost-benefit analysis, social indicators monitoring and social accounting, or even in-depth case studies, causal attribution for complex social impacts is always fraught with great difficulty.

24. A related problem pertains to the "dependent variables." Determining causation is difficult enough, but often the effects are themselves interwoven in ways that are not obvious and difficult to unravel. Social outcomes occur in clusters. Some obvious examples come to mind: new birth control technology reduces unwanted pregnancy *and* gives rise to promiscuity and socially transmitted disease; smoking cessation programs reduce tobacco-related cancer *and* lead to increased obesity rates; technological innovations lead to increased wealth *and* sometimes to greater inequities. In short, in modeling social outcomes from research one has difficulty not only tracing cause to effect but also setting boundaries on effects.

8. Getting on With It: A Sober and Humble Rationale for Evaluation of Social Impacts of Research

25. The foregoing section identifies formidable obstacles to assessing the social impacts of research and, unfortunately, the list above is not exhaustive (for a more detailed discussion of problems in tracing social and public value impacts see Bozeman, 2007). It is easy to see why one might retreat from the task. However, it is important to work toward approaches to assessing the social impacts of research, even while

being realistic about the challenge. Here is the reason: policy makers will continue to make choices about research funding and these choices will continue to be premised on a causal logic. In making decisions about investments in research, policy makers make assumptions about the effects on those investments on such social outcomes as public health, transportation systems, education, and wealth creation. In most instances those choices will, perforce, be based on limited information. The evidence that can be brought to bear on those choices, even when obviously flawed, is in some cases likely an improvement over intuition, habit, rough-hewn ideology, political self-interest, and other such bases that guide investments in research.

9. The "Public Value Mapping of Science" Project: Recent Developments in Assessing Social and Public Impacts of Research

26. The idea of "public value mapping of science" was initially developed by researchers at the Consortium for Science, Policy and Outcomes (CSPO) as part of a grant from the Rockefeller Foundation. Current work is under development by a team of researchers at CSPO, now a part of Arizona State University. The new research project is funded by the U.S. National Science Foundation's "Science of Science Policy" program. The new project, whose full title is **"**Public Value Mapping: Developing a Non-Economic Model of the Social Value of Science and Innovation Policy," began in 2007 and terminates in 2010.

27. The primary rationales for the public value mapping of science (PVM) are that (1) the focus of science policy should be on end-state social goals and public values, and (2) current research evaluation and science policy analysis methods and techniques, while useful in many important respects are not entirely sufficient for such analysis of the impacts of research on public values. *PVM assumptions are provided in more detail in Table One. An especially important assumption to bear in mind is that PVM is not and does not aspire to be a unified method; rather it is an approach, or set of approaches, public values of science outcomes.*

Box 1. Core Assumptions of Public Value Mapping

1. PVM is either prospective (analyzing planned or projected research activities), "formative" (analyzing such activities as they are occurring), or "summative" (evaluating activities and their impacts after they have occurred).
2. It seeks to take into account the highest order impacts of activities (i.e. broad social aggregates) and, thus, focuses on social indices and social indicators.
3. It is multi-level in its analysis, seeking to show linkages among particular program activities of an agency or institution, activities of other agencies or institutions, relationships- either intended or not- among various institutional actors and their activities. Related,
4. PVM is concerned with understanding the environmental context for research and related programmatic activities, locating the activities and their institutional actors in terms of other actors in the environment, the constraints, opportunities and resources presented in the environment.
5. Research in any field by any method is embedded in a social context; in PVM analysis of the social context of the research (i.e. characteristics of research performers, their attributes and social relations) is a part of the analysis.
6. PVM is guided by a "public value model of science outcomes" rather than a market-based or market failure model. PVM explicitly rejects evaluation and assessment based on commodification of research values and outcomes. Market prices are viewed as weak partial indicators of the social value of research and research outcomes. Even as a partial indicator, market value is considered in terms of not only magnitude but also distribution and equity criteria.
7. Since market value is eschewed in PVM and since social values are not interpersonally transmissible, PVM anchors its outcomes values in a wide range of criteria derived from diverse sources including:[1] official, legitimated statements of policy goals; [2] goals implicit in poorly articulated policy statements; [3] government agencies' goal statements in strategic plans; [4] aggregated statements of value represented in opinion polls; [5] official policy statements by government actors; [6] official policy statements by relevant NGOs.
8. PVM analyzes (maps) the causal logic relating goals statements (any of the above) to science and research activities, impacts and outcomes, both measured and hypothesized. When possible, this analysis begins with the causal logic articulated by responsible officials. The causal logics, explicit or implicit, that are the basis of science and research activities are then considered in relation to various plausible alternative hypotheses and alternative causal logics invented by the analyst.
9. PVM is not an analytical technique or even a set of analytical techniques, but a model that includes a guiding theoretical framework (public value theory), a set of assumptions and procedures. Research techniques employed in PVM depend upon the needs and possibilities afforded by the context of its application. The only technical approach used in all applications of PVM is the case study method.
10. After gathering data to test hypotheses about causal logics and outcomes, appropriate analysis (selected depending upon specific analytical techniques used), is employed to test hypotheses and, at the same time, measure impacts and outcomes. Results of analysis focus on interrelationships among the causal logic, the environmental context and measured impacts and outcomes.
11. PVM concludes with a linkage of impact and outcome measures back to aggregate social indicators or other appropriately broad-based, trans-institutional, trans-research program measures of social well being.
12. PVM concludes with analysis and recommendations focusing on possible changes (in research or program activity, causal logic, implementation) that seem likely to lead to improved social outcomes.

Source: Bozeman, B. (2003). *Public Value Mapping of Science Outcomes: Theory and Method.* In D. Sarewitz, et al., Knowledge Flows & Knowledge Collectives: Understanding the Role of Science & Technology Policies in Development, 2(1).

28. As part of the Rockefeller project, two "beta tests" were developed as initial PVM applications were developed, one focusing on breast cancer research (Gaughan and Bozeman, 2002) and the other on genetically modified crops (Gupta, 2003). In both cases, the analyses proceeded by developing public policy statements about research and innovation goals (as surrogate indicators of public value) and developing indicators to determine the degree to which public value and social outcomes match those goals. Subsequently, most of the work developing PVM has been aimed at theory building and, more recently (i.e. since the outset of the NSF-funded project), development of multi-method analytical tools. Two aspects of PVM theory building are particularly relevant. First, a theory of innovation has been developed that matches the aims of PVM, the "churn theory of innovation" (Bozeman and Rogers, 2002). Second, a number of efforts have been employed to develop a theory of public value (Bozeman, 2007; Jorgensen and Bozeman, 2007; Bozeman and Moulton, 2008) and to apply the PVM theory in case contexts, including influenza vaccine research, development and commercialization, genetic suppression technology for seeds, and climate change (Bozeman and Sarewitz, 2005; Feeney and Bozeman, 2008; Bozeman, 2007).

10. PVM Theory "Building Blocks"

A. New Ways of Thinking about the Social Value of Knowledge.

29. A PVM taproot is new theoretical thinking about the value of knowledge and its assessment. Philosopher Elizabeth Anderson (1993) presents an especially interesting analysis of economic value and value theory as it pertains to economics. Anderson's position, one that would perhaps seem radical to many social scientists, is that economic values are inherently monistic. Because of the fundamental structure of assumptions built in to economic values, they cannot accommodate more pluralistic approaches to values. More troublesome, according to Anderson, is the fact that economic analysis of values, by insisting upon monistic interpretations of value, actively undermines richer and generally more useful pluralistic analyses. To put it another way, an analysis valuing exchanges, commodities, and services on the basis of market standards pre-empts simultaneous, comparable reference to other standards (see Marmolo, 1999; Anderson, 1993). These assertions have direct implications for models of innovation and the impacts of scientific and technical knowledge. In connection with Anderson's argument, let us consider a prominent example of a value in use that has long troubled economists seeking a useful index. Economists have never made much headway valuing scientific knowledge (see Machlup, 1962) and, thus, it provides an excellent case in point for understanding the relationships among intrinsic value, economic value and public value.

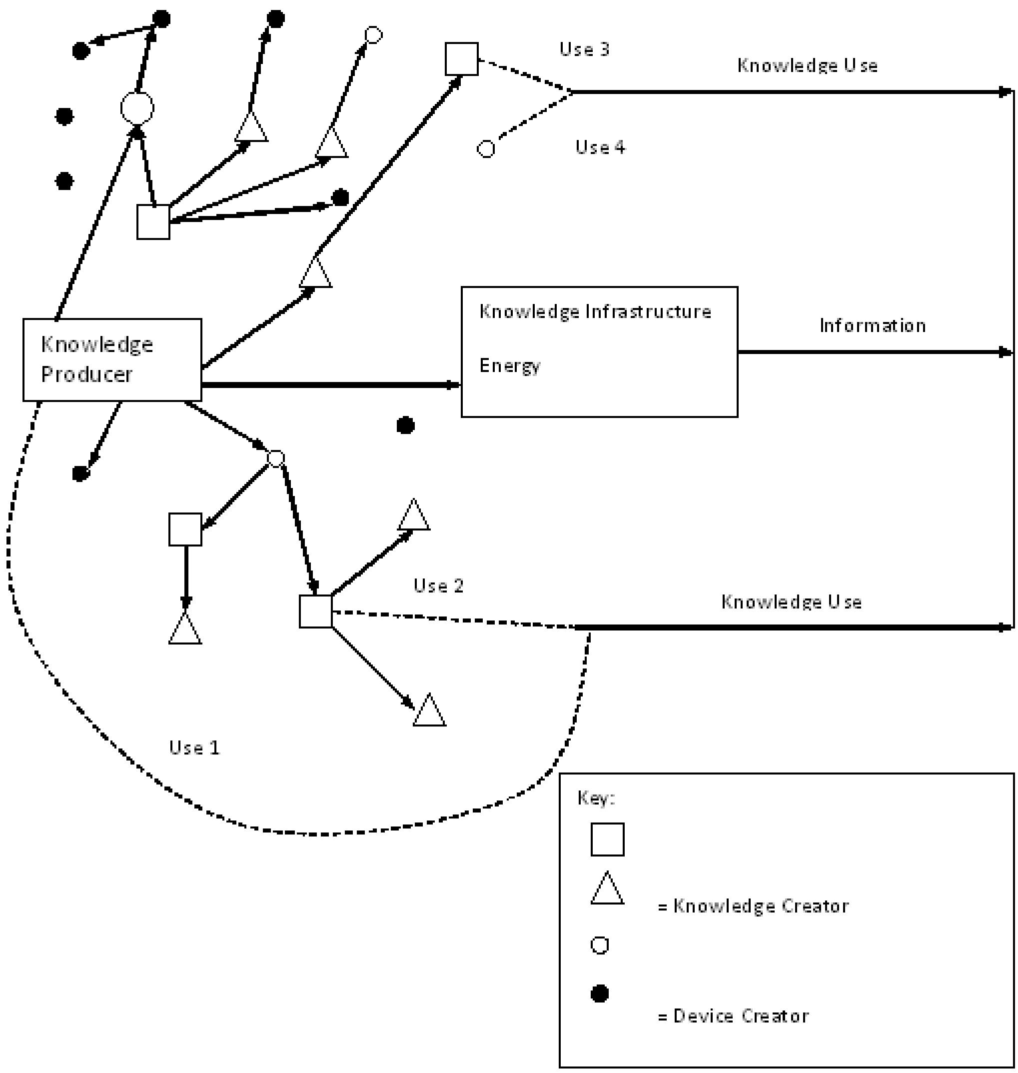
30. Bozeman and Rogers (2002) argue that economic currency is a poor surrogate for valuing *scientific knowledge*. This is chiefly because scientific knowledge simultaneously presents many types of values. The history of fundamental scientific research shows (Jewkes, Sawers, and Stillerman, 1958) that creating and deploying knowledge relates only obliquely to the creation of commodities and, from the standpoint of valuation, is as not much akin to technology creation and use (Nelson, 1959; Stephan, 1996). True, basic research (much like music and sculpture) can be priced, but price is a less exacting measure than repeated, broad-domain use. Economists refer to this application robustness as the "public goods characteristics of knowledge" and cast it as a as market failure. However, the problem is more fundamental and has less to do with failures of markets than with failures of market-based theories to account for problems pertaining to social choice (e.g. Arrow, 1958; Sen, 1982).

31. Economic assessments of scientific knowledge, whether grounded in cost-benefit reasoning, production function analysis or political economy theory, begin with one fundamental, generally unexamined assumption: the standard for knowledge valuation is price in an open market. To be sure, economists labour mightily to cope with widely recognized problems related to the economic valuing of knowledge, including, most conspicuously, the spill-over and free-rider problems occurring as a result of

the joint consumption properties of knowledge (one person's use generally does not diminish its availability and, often, its value to others). But these practical limitations of economic valuation tend to be viewed not so much as a limitation but a spur to developing allocation theories that take them into account. The analytical difficulties the nature of the "commodity" (scientific knowledge) sets for economic measure and valuation theory are acknowledged by all, but rarely is there much discussion of the difficulties economic valuation sets for the commodity and its translations.

32. Figure 1 below provides a simple depiction of the Churn model. It show sthat information (as we use the term above) is created from knowledge and that information can result, via use, in new knowledge or can lie fallow, depending on whether it is used. The figure also indicates the possibility of information put to multiple uses, in each instance creating value-as-knowledge.

Figure 1. Churn Model of Knowledge Use and Transformation



B. New Ways of Thinking about Public Values.

33. Certainly there is no single, canonical definition of public value, but one definition in use is closely related to PVM:

> Public values: *A society's "public values" are those providing normative consensus about (1) the rights, benefits, and prerogatives to which citizens should (and should not) be entitled; (2) the obligations of citizens to society, the state and one another; (3) and the principles on which governments and policies should be based.* (Bozeman, 2007)

34. Public values may be viewed as criteria by which to judge institutional arrangements for goods and services but should not be confused with them. Thus, public values neither support government action nor abjure markets. Criteria based on market failure or economic valuations often miss this fundamental point, one critical to assessing research and innovation.

35. An important challenge for any analytical approach to assessing public value (of research or other social goods) is identification of *particular* public values. To say that public values are held in common does not mean that they are universally embraced or that people agree on the exact nature or content of public values. Where does one look for public values? A nation's more fundamental laws and, if there is one, its constitution provide good starting points for identifying public values, though public law is best viewed as reflecting public values not as establishing them. Public values can be found in the fundamental myths of nations. Such myths as "the land of opportunity" often contain several broad public values. Public values can be found in the authoritative statements of duly authorized and legitimate policy-makers. Indeed, the early beta tests of PVM examined the concrete goal statements in the Department of Health and Human Services' *Healthy People 2000.*

36. The chief point is that it is less vital to agree on an exact approach to identifying public values than to agree that it is useful to evaluate research impacts from the standpoint of public values (as opposed to conventional analysis of economic impacts, intermediate-goals, or inputs). So long as the analysts is clear about the public values' definition-in-use, research evaluation can proceed and, if transparent, can promote dialogue about the social impacts of research. Thus, the answer to the question "Where does one find public values?" is that one finds public values in a great many places, nearly everywhere- formal scholarly literature, cultural artifacts and traditions, government documents, even some opinion polls (ones receiving valid and representative responses to questions about core values). Civil societies are necessarily permeated by public values since it is these that provide much of the structure of civil societies. The problem is not finding public values but understanding them in some analytically useful form. Arguably, it may prove useful in some cases simply to posit public values. If one's interest is in gauging the extent to which some value has or has not been obtained, then positing it as a public value may be unobjectionable. Thus, "improvements in public health and longevity" would seem to entail only minimal controversy as would "decreased infant deaths" or "cleaner air." Even the latter example is unlikely to stir much controversy *as a public value.*

C. A Public Value Mapping Criterion Model.

37. Disagreement about public values and their measurement proves less troubling in those instances when one has at his or her disposal public value *criteria.* Even when debates rage about choices of public value, concepts of public value, and the relevance of public values to particular states-of-affairs, one has hope of making headway if there are recognized public value criteria structuring arguments. The criteria presented in the PVM criterion model are not, then, public values themselves, but, rather, a set of diagnostics. The PVM model provides quite general criteria but these are applicable to questions science policy and research evaluation (see Bozeman and Sarewitz, 2005).

38. In some ways homologous to the market failure model and related concepts, the PVM approach seeks to identify public values failure. Public values failure occurs when neither the market nor public sector provides goods and services required to achieve public values. PVM criteria change the discussion of public policy and management by assuming that government (and market organizations as well) need be more than a means of ensuring market successes and technical efficiency in pricing structures. A fundamental assumption of the PVM model is that market failure actually tells us little about whether government should "intervene." With PVM, the key policy question becomes: "Whether or not the market is efficient is there nonetheless a failure to provide an essential public value?" The PVM criterion model is not a decision-making tool (a la benefit-cost analysis), but a framework to promote deliberation about public value (and its relation to economic value). Its primary use is for policy deliberation and promoting public dialog.

39. The PVM criterion model is presented in Table 2 (adapted from Bozeman, 2007 and Bozeman and Sarewitz, 2005.)

Table 1. Public failure and public policy: a general diagnostic model

Public Failure Criterion	Failure Definition	Science Policy Example
Mechanisms for values articulation and aggregation	Political processes and social cohesion insufficient to ensure effective communication and processing of public values	Peer review, the favored means of making decisions of individual-level projects, is appropriated for decisions about huge scientific programs, resulting in the displacement of social goals for more easily resolved technical goals
Imperfect monopolies	Private provision of goods and services permitted even though Government monopoly deemed in the public interest	When public authorities abrogate their responsibility for overseeing public safety in clinical trials for medical research, there is potential for violation of public trust andpublic value
Scarcity of providers	Despite the recognition of a public value and agreement on the public provision of goods and services, they are not provided because of the unavailability of providers	The premature privatization of the Landsat program shows that a scarcity of providers can create a public failure potentially remediable by government action
Short time horizon	A short-term time horizon is employed when a longer term view shows that a set of actions is counter to public value	Policy for energy R&D, by considering the short term, fails to fully capture the costs of global climate change on future generations
Substitutability vs. conservation of resources	Policies focus on either substitutability or indemnification even in cases when there is no satisfactory substitute	No-net-loss' policies fail to take into account the nonsubstitutability of many natural organisms ranging from wetlands protection to prohibiting the sale of human organs on the open market
Benefit hoarding	Public commodities and services have been captured by individuals or groups, limiting distribution to the population	A prime technical success of genetic engineering, the 'terminator gene,' proves an excellent means of enhancing the efficiency of agricultural markets, potentially to the detriment of millions of subsistence farmers throughout the world

11. Developing and Applying the PVM Model

40. While considerable conceptual work has already been undertaken to provide building blocks for PVM, these are not yet integrated to create a viable model that can generate practical analytical tools. That is the purpose of the work currently underway.

41. All PVM approaches begin as case study analyses. The current PVM cases include the following analytical approaches:

1. A search for "public values" pertaining to the case: We have discussed several approaches to identifying public values, including (a) surrogate public values (government mission statements, strategic plans, and broad policies, statutes); (b) distillation of public values from relevant academic literatures; (c) public values as expressed in public opinion polls and public statements. In addition, the Public Value Failure criteria will be used to guide the analysis of possible public value deficits or public value failures.
2. The application of the Public Values Grid: After developing information about putative public values and gathering data about the social impacts of SIPs and consequent STEM research and innovation, it will be possible to map historical and real-time cases on an improved version of the prototype Public Values Grid (Figure 1, above). *A key aspect of the proposed study will be to improve and further specify both the public value failure criteria model and the Public Values Grid, including further extending their direct relevance to SIPs.*
3. Developing value analysis chains: Among the many reasons why public value analysis of SIP has made little headway is that values analysis itself is remarkably underdeveloped. One of the difficulties of values analysis (Gaus, 1990) is that analysts sometimes fail to consider interrelationships among values, including such features as values hierarchies, conditional relations among values, logical structures of multiple and related values, and ends-means relations (Braybrooke and Lindblom, 1963). One of the key objectives of the proposed research is thus to develop the ability to clarify relationships among values.

42. The analytical lenses for the cases can be thought of as essentially master hypotheses about possible determinants of the public value outcomes for the cases. The cases proceed on the basis of four important contextual factors that affect the social impacts of research and science and technology policy.

a. *Characteristics of the knowledge that the research produces*. In some instances knowledge creation processes, innovation, and, ultimately, social impacts are very much governed by inherent characteristics of the science or technology (e.g., "technology push").

b. *Institutional arrangements and management affecting knowledge production and use.* "Institutional arrangements" pertain to the configuration of producers and users of scientific and technical knowledge, knowledge, the ways in which they interact, their internal and network management.

c. *Policy and political domains of knowledge production and use.* This analytical lens examines the political, legal, public policy and normative factors that determine research choices, utilization and impact (e.g. characteristics of intellectual property policy or structures of budgets for research).

d. *Market settings for knowledge production and use.* Public value may be achieved (or thwarted) by markets, quasi-markets, or government entities. In some instances, much can be understood about public values by considering such market features as the relative scarcity of resources, market actors controlling resources, market segmentation, extent and nature of competition.

12. The PVM Cases

43. Several cases are currently underway and aiming to apply and to further develop PVM approaches. These cases focus not only on research developments and the evaluation of research but also on the use and social impacts of knowledge produced within the domains of the cases. Cases at various stages of development include (1) nanotechnology-based water filtration; (2) alternative fuels; (3) climate change; (4) inequities in cancer research and treatment; (5) public value impacts of technology transfer.

44. These cases will not only perform the traditional role of "thick description" case studies but will also (1) provide a context for the application of a variety of analytical approaches including logic models and value chain analysis; (2) help determine the extent to which it is possible to distill public values in a satisfactory manner; (3) extent the theories upon which PVM is premised; (4) point the way for further development of analytical tools.

13. Summary

45. The assessment of the impacts of research on end-state social and public values remains at a quite primitive level, with few tools having been developed and many of the "borrowed" tools having important limitations. However, by simultaneously working to develop theory, concepts, cases and early form analytical approaches it is perhaps possible to develop approaches that can supplement the rational choice, cost-benefit, bibliometric and other extend research evaluation approaches that have been developed chiefly for purposes other than assessing public value impacts. This chapter has reviewed some of the motivations for a public value-oriented approach to assessment, some of the problems in developing such an approach and a few of the early steps that have been taken in fulfilling the perceived need for research evaluation approaches that are more sensitive to discerning social impacts.

REFERENCES

Andrews, F. (1979) Scientific productivity, the effectiveness of research groups in six countries. Ann Arbor, Mich.: University of Michigan Press.

Aubert, J.-E. (1997). On the OECD Experience of Country Reviews. In OECD (Ed.), Policy Evaluation in Innovation and Technology Towards Best Practices (Chapter 21). Paris: OECD. *http://www.oecd.org/dsti/sti/stat-ana/prod/evaluation.htm*

Auditor General (1993). Program Evaluation in the Federal Government. The Case for Program Evaluation. Treasury Board of Canada.

Averch, Harvey (1991). The practice of research evaluation in the United States. Research Evaluation, 1(3), 130-136.

Audretsch, D. B. Bozeman, B. Combs, K. L. Feldman, M. Link, A. N.; Siegel, D. S. Stephan, P. Tassey, G. Wessner, C. (2002) "The Economics of Science and Technology," 27, 2, 155-203.

Barbarie, Alain (1993). Evaluating Federal R&D in Canada. In Barry Bozeman & Julia

Melkers (Eds.), Evaluating R&D Impacts: Methods and Practice (155-162). Norwell, MA: Kluwer Academic Publishers.

Bozeman, B. (2007) *Public Values and Public Interest: Counter-balancing Economic Individualism*. Washington, D.C.: Georgetown University Press.

(2002) "Public Value Failure and Market Failure." Lead Article, *Public Administration Review*. 62(2): 145-161.

(2003) "Public Value Mapping of Science Outcomes: Theory and Method." In D. Sarewitz, et. al. *Knowledge Flows & Knowledge Collectives: Understanding the Role of Science & Technology Policies in Development*. 2(1).

Julia Melkers (eds.) (1993) *Evaluating R&D Impacts: Methods and Practice*. Boston: Kluwer Academic Publishers.

J.R. Rogers (2002) "A Churn Model of Scientific Knowledge Value: Internet Researchers as a Knowledge Value Collective," *Research Policy,* 31, 5, 769-794.

D. Sarewitz (2005) "Public Values and Public Failure in U.S. Science Policy," *Science and Public Policy*, 32, 2, 119-136.

Callon, M., P. Laredo, P. Mustar (1997) *The Strategic Management of Research and Technology: Evaluation of Programmes*. Paris: Economica International.

Cozzens, S., S. Popper, J. Bonomo, K. Koizumi, A. Flanagan 1994. Methods for Evaluating Fundamental Science. Report prepared for the Office of Science and Technology Policy. Washington: RAND Corporation, Critical Technologies Institute.

Cummings, R. and L. Taylor (1999) "Unbiased Value Estimates for Environmental Goods: A Cheap Talk Design for the Contingent Valuation Method," *American Economic Review*, 89, 3, 649-665.

Fischer, E.P. (1997) *Beauty and the Beast: The Aesthetic Moment in Science*, trans. Elizabeth Oehlkers. New York: Plenum Trade.

Johnson, H. (1965) "Federal Support of Basic Research: Some Economic Issues," *Minerva*, 3, 4, 500-514.

Jones, C.I. and J.C. Williams (1998) "Measuring the Social Return to R&D," *Quarterly Journal of Economics*, 113, 4, 1119-1135.

Kostoff, R. (2001) "The Metrics of Science and Technology," *Scientometrics*, 50, 2, 353-361.

LaFollette (1994). Measuring Equity: The U.S. General Accounting Office Study of Peer

Review. Science Communication, 16(2), 211-220.

Link, Albert N. (1996a). Economic Performance Measures for Evaluating Government Sponsored Research. Scientometrics, 36(3), 325-342.

Link, A. 1996b. Evaluating Public Sector Research & Development. New York: Greenwood.

Luukkonen T. (2002) "Research evaluation in Europe: state of the art," 11, 2, 81-84.

National Academy of Sciences (1996). Beyond Discovery: The Path from Research to Human Benefit. Washington, D. C.: National Academy Press. *http://www4.nas.edu/beyond/beyonddiscovery.nsf*

OECD (1995). Cost/Benefit Analysis of Large S&T Projects: Some Methodological Issues. Paris: OCDE/GD(95)57. *http://www.oecd.org/dsti/sti/s_t/ms/prod/e_95-57.pdf*

OECD (1997). The Evaluation of Scientific Research: selected experiences. Paris: OECD,

Committee for Scientific and Technological Policy. Document OECD/GD(97)194.

http://www.oecd.org/dsti/sti/s_t/scs/prod/e_97-194.htm

OECD (1997). Policy Evaluation in Innovation and Technology: Towards Best Practices.

Paris: OECD.

http://www.oecd.org/dsti/sti/stat-ana/prod/evaluation.htm

Rubenstein, A. (1976) "Effectiveness of Federal Civilian-Oriented R&D Programs," *Policy Studies Journal*, 5, 2, 217-227.

Ruegg, R. (1996), "Guidelines for Economic Evaluation of the Advanced Technology Program", NIST Internal Report 5896.

Saavedra, P. and B. Bozeman (2004) “The ‘Gradient Effect’ in Federal Laboratory-Industry Technology Transfer Partnerships,” *Policy Studies Journal*, 32, 2, 235-252.

Salasin, J., L. Hattery, and R. Thomas (1980)”The Evaluation of Federal Research Programs”, MITRE Technical Report MTR-80W123, June 1980.

Sen, A. (1982) *Choice, Welfare and Measurement*. Cambridge, MA: MIT Press.

Solow**,** Robert M. (1957) *Technical change and the aggregate* production function. Review of Economics and Statistics 39(3): 312-320. ...

U.S. General Accounting Office (1999). Federal Research: Peer Review Practices at Federal

Science Agencies Vary. GAO/RCED-99-99.

http://www.gao.gov/AIndexFY99/abstracts/rc99099.htm

CHAPTER 2

ARCHITECTURE OF EVALUATION SYSTEMS

This chapter presents case studies that outline the institutional and organisational structures in evaluation mechanisms and practice operate at national level in a select number of OECD and non-member economies.

1. Introduction

46. The overall framework, or architecture, for the evaluation of science and technology is a critical element of national innovation systems. As part of the CSTP project on evaluation, the TIP has undertaken a series of country case studies[6] that examine the elements of evaluation systems, the different actors, and their linkages and the relationship between evaluation pillars and policy design and implementation, in particular at programme, policy and institutional levels.

47. The case studies presented in this document illustrate variety both with regard to the comprehensiveness of the evaluation systems, the objectives of the evaluations and the way in which they are conducted and used for policy making. Of the six countries in this report, Finland is the country which is the closest to having a comprehensive system for evaluation. By comprehensive, it is understood to mean that evaluation is everywhere, from the evaluation of the country's performance as a system to the evaluation of programmes. In Finland all publicly funded research is evaluated, whether it takes place in universities or other research organisations. The evaluations are conducted by the Academy of Sciences and TEKES. Finland plans to conduct an impact assessment of the entire national innovation system in the near future.

48. In the other countries the responsibility for evaluation, the organisation and subsequent execution follow clearly delineated lines according to the focus of research or the funding streams. In Japan, for example, the Ministry of Economy Trade and Industry (METI) evaluates the competitive research funding to industry while responsibility for evaluating non-competitive research subsidies falls on the performing agencies, such as the New Energy and Industrial Technology Development Organization (NEDO), which has responsibility for evaluating its own (applied research) projects (which are subsidised by METI), and the National Institute of Advanced Industrial Science and Technology (AIST). In China, the Ministry of Science and Technology has responsibility for evaluating "national research programmes" (*e.g.* National Hi-tech R&D Programme or the National Key Basic Research Development Programme) which have served to concentrate and allocate public resources on priority areas for the social and economic development of China. In Norway, in contrast, fields of basic research in universities are evaluated by the Norwegian Research Council.

49. For all countries, the rationale for evaluation is to improve their R&D performance and to ensure efficient use of the resources invested. The roles of the ministries, however, differ among the countries. While the ministries play an active role in Japan and China, in other countries, the evaluation is delegated

6. The countries participating in the case studies on the architecture of evaluation systems are Austria, China, Finland, Israel, Japan, Korea and Norway.

to research councils or other funding bodies. Evaluation is also grounded in the legal framework and regulations whether at the level of the national government (Finland government resolution 2005) or the funding agency or research council. In China, until recently, there were no laws or regulations for R&D evaluation. The regulations aim at promoting the evaluation of R&D. Clear established guidelines such as those set out by METI's newly created Technology Evaluation and Research Division are also important to institutionalise and standardise evaluations practices.

50. It appears however, that the goal of using the results of evaluation for funding decisions, in particular budgeting decisions, is not explicit in the countries under study. This is case in China where evaluation is used more for improving existing programmes than for supporting resource allocation or priority setting. The objectives of the evaluations differ and depend on the policies/programme being evaluated, whether the programme focus on co-operation between universities and businesses as in the case of Israel and Japan or on the quality of research programmes (*e.g.* in China). In other cases, assessing the social and economic impacts is an explicit goal of the evaluations such as in Finland and Israel. In Norway, given the focus on funding basic research, the evaluations aim to improve quality.

51. The mechanisms and tools for evaluation differ depending on the level of evaluation and countries. While in Norway and Finland the evaluations are based mainly on peer review, the other countries used expert teams. In China, evaluations of large national programme involve several tools, including consultation of peers in the process. In Japan, peer review and expert review are both used as are interviews.

52. In all those countries evaluated, researchers, programme managers or firms, play an important role in the process. A common challenge seems to be availability of data on research. While data and bibliometric methods were used in Norway and Finland, and a range of other indicators are used in Japan, qualitative information such as that collected from interviews are also very important.

53. Addressing social and economic impacts remains a challenge for most countries, mainly due to the difficulty of attributing impact variable to a particular research programme, particularly after a lapse of several years. Scientific impact is easier to measure than social impact, and impact of applied research is more easily measured than basic research. This represents a considerable challenge. Finland is actively undertaking foresight studies with the aim to combine these with impact evaluations of R&D. In the case of Japan, one of the vital roles of the follow-up evaluation is to identify *negative* impacts and their linkages to research programmes.

54. What are the consequences of evaluations? From the case studies in this volume, the outcomes of evaluations appear to have the greatest influence on improving the design, implementation and effectiveness of programmes. Evaluations have a more limited impact on the policy decision for funding, with the exception of evaluations of basic research programmes (*e.g.* the Norwegian Research Council). Lastly, in the countries surveyed, evaluations are made public with the exception of China. However, the diffusion and dissemination of evaluations can help improve policy learning both with regard to the quality and process of evaluations as well as with regard to the results of the evaluations themselves.

CASE STUDY 1: THE EVALUATION OF GOVERNMENT FUNDING IN RTDI FROM A SYSTEMS PERSPECTIVE IN AUSTRIA[7]

1. Introduction

55. The Austrian Innovation System has by and large worked quite well in the past. Together with other favourable political and economic conditions, it helped Austria's income and productivity catch up with the most advanced countries by the nineteen seventies. It was instrumental in the following decades as Austria forged ahead relative to the average of the European Union. As a consequence Austria is now one of the top five countries in the EU as measured by income per capita and is ranked among the top ten industrialized countries worldwide.

56. Complacency is, however, the greatest danger to future prospects. Several strains are now noticeable in the Austrian Science and Technology System which make it necessary to increase innovation efforts, to boost efficiency and to foster radical changes in the innovation system. Challenges come from new global framework conditions (globalisation, EU enlargement, internationalisation of research; see *Jürgen Janger*). Austria is confronted with intensive competition both from neighbours and Asian countries. Radical change is urgent, specifically as a result of past success; a high-income country has to compete in sophisticated markets and products. Other countries are now moving into Austria's position as medium-tech specialists, deriving their competitive edge by adapting technologies imported from abroad and producing at somewhat lower labour costs. In addition, we see that higher innovation inputs in Austria have not been met by higher market shares and exports specifically in the highest quality segment in fast growing sophisticated industries. The number of firms innovating continuously remains small. Business research expenditure is highly concentrated in a small number of firms. Twice as much business research, as compared to the EU average, is financed by foreign resources in Austria, and multinational firms increasingly source research facilities and capacities at the low end of the spectrum from the globalizing world. Maybe the largest challenge to the system stems from an internal weakness: the Austrian innovation system is only loosely interlinked with, and insufficiently supported by, the education system. The gap between human capital available and the demand of firms is increasing at least with respect to the highest education level. More generally and most importantly, innovation and education are separated too much at all levels.

2. A radical strategic shift in six dimensions

57. A radical strategic shift in innovation policy is needed in the following six respects:

- From an **innovation policy in the narrow sense** to a **comprehensive innovation policy.** The latter is interlinked with education policy and includes improvements of the framework

7. This case study is based on the report on the evaluation of government funding in RTDI from a systems perspective in Austria by Karl Aiginger, Rahel Falk, Andreas Reinstaller. It is one of several reports prepared for the report "Reaching out to the future needs radical change: Towards a new policy for innovation, science, and technology in Austria". It has been commissioned by Bundesministerium für Wirtschaft, Jugend und Familie und Bundesministerium für Verkehr, Innovation und Technologie. It is submitted to the OECD Secretariat as the Austrian contribution to the case studies.

conditions (e.g. competition, international openness, mobility); while the former only concentrates on the measures and institutions directly involved in science and technology;

- from an **imitation strategy** to a **frontrunner strategy**. In a frontrunner strategy firms and researchers strive for excellence and market dominance in niches and high quality segments, increasing market shares in sophisticated industries and technology fields, and in areas or missions of particular relevance to society;
- from *fragmented public interventions* to *coordinated* and *consistent interventions derived from a vision* which specifies economic objectives, external and internal challenges and the market failures which call for public intervention;
- from a *multitude of narrowly defined financial programs to a flexible, dynamic policy defining broader tasks and priorities.* Some broad technology and research fields important for society (missions) should be defined top down in the vision, but clusters and centres of excellence will grow bottom up, and should be funded sufficiently so as to attract international leadership;
- from a *blurred division of responsibilities* between and within ministries (and other "players") to *well defined responsibilities***.** Ministries devise sub-strategies for their area of responsibility from the top-level vision, are coordinated on the government level by a high level commission and monitored by a Council of Science, Technology and Research;
- from managing public *intervention by bureaucratic procedures to modern public management techniques.* Goals are pursued either by internal competence centres in ministries or by delegation to outside agencies (agencification). Agencies are free to choose instruments and are controlled according to pre-defined output criteria without micro-interventions.

58. The report summarizes about fifty recommendations, which will enable the strategy shift as well as bringing about major or minor improvements of the Austrian System of Science, Research and Innovation as requested.

3. The reference point: an overarching vision

59. The preconditions of a new innovation policy are:

1. a commonly shared belief that research, technology and innovation are crucial for the welfare, growth and competitiveness of the Austrian economy and
2. a consensus on the policy changes necessary to increase the effectiveness of Austria's innovation policy to its highest level.

60. Therefore it is necessary to develop a general strategic "vision", which will define the mission and the goals of the Austrian System of Science, Technology and Innovation, including its relationship to the education system, and to societal and economic goals and the framework conditions needed for innovation.

61. Developing such a strategic "vision" at the highest level of government is needed. It will serve as a reference point for all sub-strategies of ministries, regions, institutions and agencies and thus form a blueprint for a new science, technology and innovation policy ("New STI"). The vision should be prepared by a team of national and international experts, but finalized and "owned" by the government. It should be put into legislation by parliament and monitored by a "Council of Research, Science and Technology" (see *convelop/Gerhardter*) as an external control. The "vision" has to define the mission and goals of the innovation system, its interaction with the education system, but also to other societal and economic goals. It is the base for all sub-strategies of ministries, regions, institutions and agencies.

4. Coordination and monitoring: reformed institutions

62. Implementation of a New STI policy does not end with the creation of a vision but needs consistent coordination between the policy strands defining the new comprehensive innovation policy. We propose to set up a "high level coordination commission" on research and technology with the Chancellor and the Deputy Chancellor as heads and the ministries responsible for innovation and education as members. It should meet about twice a year to promote the implementation of the vision. The government should be accountable to a new permanent "parliamentary committee for science and technology" (a merger of two existing committees; see *convelop/Gerhardter*). The parliamentary committee should also discuss an annual report of the "Council of Research, Science and Technology" on the progress of the vision.

5. Better governance: new role of ministries

63. The change in strategy calls for a new and better defined role of the ministries in charge of innovation policy. They will devise sub-strategies from the overall vision to implement the vision in their respective areas of responsibility, focusing on the frontrunner position and on links with other policies. They decide which part of the sub-strategy has to be fulfilled "internally" (e.g. (i) linking the innovation system and the education system, (ii) improving the framework conditions), and which part has to be delegated to agencies or institutions. Each ministry should be responsible for the implementation of well defined parts of the new strategy. Their activities should be coordinated by the "high level coordination commission" on research and technology that defines also the goals and milestones for each ministry.

6. Better governance: increased autonomy of agencies

64. The autonomy of the agencies in a process of agencification (*e.g.* according to the concept of earned autonomy; see *Sabine Mayer*) should be increased. This will require new governance procedures. At the administrative level we need to systematically build up competency to actually manage the agencies and to coordinate the intra-ministerial processes of policy development. Processes should be implemented that coordinate policy development activities across departments in order to avoid overlaps and conflicting assignments to the agencies. Broad tasks should be delegated to the agencies instead of narrowly defined programs and the delegated tasks should be monitored according to output goals whenever feasible. For these tasks actual goals and outcomes are specified. With these in mind the agencies themselves should develop suitable program (or sub-strategies) which also fit into their overall portfolio. If actual programs rather than tasks are still delegated, they should be much broader and should undergo a strict need-based test. However we need to move away from a culture where programs run forever to one where they can end in the wake of positive or negative evaluation programs. For this process to be efficient new and compatible reporting systems across ministries are needed, as well as within and across agencies. Micro-management and -intervention should be abolished. Control should happen exclusively on the basis of ex post assessments of outcomes. The extension of programs should be decided on the basis of these assessments.

7. Changing the track: Switch to a frontrunner strategy

65. The New STI policy for Austria should be a frontrunner strategy. A frontrunner strategy aims at supporting Austrian firms to achieve and sustain economic leadership through product innovation and productivity growth in niche markets. This requires an increasing number of Austrian companies to build up a winning margin in technological and market competencies over their principal competitors. This can only be achieved through more and more ambitious research and development in the business sector, more and better qualified people, and leading edge scientific research. Education is the driver which enables change in firms and institutions.

8. Government commitment: ambitious goals for 2020

66. The Austrian government has set the goal to increase research expenditures to 4% and expenditures for tertiary education institutions to 2% of GDP by 2020 (the two numbers should not be added up, since part of the second is included in the first). Europe is trailing the USA and Japan in research, and has set the 3%-of-GDP goal for 2010 without any chance of reaching it soon. Austria as a high-income country should be more ambitious. The Austrian government should take the necessary steps to make available sufficient financial means in the public budgets to finance the tax credit and direct support for R&D, to sufficiently finance university research, and to improve the quality of education and the number of graduates from higher education institutions. Economic growth and competitiveness of a country on the technological frontier according to the EU Commission is defined by:

A high level of expenditure for innovation and education,

their respective efficiency and;

Intensive synergies between higher spending and higher efficiency.

67. Since the 2% and the 4% of GDP are only input goals, complementary output goals would be necessary to track efficiency.

9. In face of crisis: keeping up dynamics

68. Keeping the dynamics of research expenditures is absolutely necessary. This had been the case for Austria over the past 10 to 15 years, but it is very much in danger today. Private investment in research will be curtailed in the crisis and this will happen over-proportionally in multinational firms. Empirical evidence shows that research expenditures are highly pro-cyclical; and even more so in Austria. The elections and the deferrals in the budgeting process for 2009 have already delayed spending by public funds and institutions (e.g. FWF, FFG, AWS). Other sources are drying up as the crisis deepens, and public money will be scarce in the further course of the crisis and thereafter. It would, however, be extremely important to keep the dynamics of research expenditures, since this is essential for a frontrunner strategy and for competitiveness in a tough environment. The current expenditure path for the next year is definitely lower than planned and as necessary to arrive at the 3%-of-GDP target in 2010.

10. A new simple tax credit: broadening the base and shifting the level

69. A frontrunner strategy needs a broader base. This means a higher number of research active firms, more firms innovating regularly, more innovative business start-ups, new research departments in existing firms, a larger number of firms choosing locations in Austria for research facilities, and more firms cooperating with universities. As a driving force for broadening and shifting the level of innovation expenditures we propose using a single tax incentive, namely an extended tax credit of 12%. This "new tax credit" should replace all existing schemes for expenditures based on patents or a certified "importance for the Austrian economy", all tax allowances, and the scheme for incremental innovation (see *Rahel Falk*). It is more generous for all firms (with the exception of firms liable to income tax), it is visible and provides incentives for shifts from physical investment into intangible investments within firms, for creating research departments, or e.g. for shifting research towards an Austrian location within a multinational firm. A strong driver for more investment into R&D is necessary if the 4%-of-GDP goal is to be reached in the foreseeable future, especially considering the economic constraints in the crisis since multinationals make a high financial contribution to R&D expenditures in Austria today. Total investment in R&D must increase by approximately 8% between 2008 and 2020 (in nominal terms) to reach the target of 4% of GDP by 2020. Tax incentives are an important condition for continuous investment and location decisions

(the "necessary" condition). Equally or even more important is the knowledge base, the research capacity of universities, research labs and human capital (the "sufficient" condition).

11. Deepening and changing the track: the complementary role of direct support

70. A frontrunner strategy must not only have a solid base but also has to support the peaks of excellence. Structural shifts, top level research and high efficiency are critical for the strategy. However these are much more difficult to bring about. Direct support should therefore be aimed in particular at quality and high risk projects. It should focus on firms with high innovation and knowledge intensity, technological excellence, quality in niches, hightech start-ups, programs of technological excellence and thematic priority. Excellent projects, important thematic areas and activities advantageous to innovation should be stimulated. Direct support is therefore an appropriate instrument to promote the deepening of innovation, to support risky projects and activities whose benefits are external to the firms in the pre-market phase and in institutions providing them. Direct support furthermore enables a learning process, provides information and a certain degree of consulting. It is therefore important for firms starting or upgrading innovation ("changing the track"). Direct support is therefore no substitute for tax incentives, but has a complementary function and should be made more complementary.

12. Direct support and tax credit: no trade off

71. To increase the effectiveness of direct support we recommend (1) reducing the number of programs (not the money spent) and to allow agencies more discretion in the choice of instruments; (2) defining output goals for agencies rather than input goals; (3) that thematic programs do not define narrow sub-fields, but allow these to develop and cluster bottom up; (4) basic, open programs to increase and promote the quality component, to enforce clustering and cooperation with universities, (5) science programs to support thematic fields if defined in the vision, and to foster cooperation, competence centres, excellence programs in a bottom up process, if the chance for excellence exists. Given the overall goals of 4% and 2% of GDP, it is necessary to increase funds for direct support at a rate exceeding GDP growth by far.

72. There is however no trade off between direct support of business research, support for scientific projects and tax incentives. If any of these drivers for increasing research efforts is not growing faster than the total economy, it would be more honest to give up the expenditure targets and the ambition to become a frontrunner.

13. Higher education: new funding rules and additional research money

73. The quality of universities, Universities of Applied Sciences and non-university research institutions is a crucial determinant of a frontrunner position. Quality is related to the financial means of higher education institutions and proper incentives. Currently these are not funded sufficiently to ensure a high quality of research or teaching. Incentives do not lead to excellence centres.

74. We therefore recommend increasing spending for tertiary education to the level recommended by the European Commission (2% of GDP). The current lack of tertiary graduates especially in the field of science and technology is an important bottleneck for industry and academia.

75. To increase the efficiency of tertiary institutions expenditure for research and teaching should be separated, funding per student should be based for universities on the model currently applied for the Universities of Applied Sciences, and additional research money (that should not reduce the funds for the Austrian Science Fund, FWF) should be allocated to universities on the basis of performance criteria (these criteria should also include cooperation with firms).

76. Money should be distributed within universities in a more competitive manner (inter alia to persons, and specifically young scientists, not to institutes). A new tenure track system based on international best practice should be envisaged and career steps in universities should depend on international experience and be under a competitive framework.

77. The budget of the Science Fund should be partly used to finance thematic programs (if defined by the vision). Research infrastructure should be supported e.g. by increasing the overhead costs covered in FWF projects from 20% to 50% (case and performance dependant).

78. Block grants to non-university research institutions should depend on the existence of a mission as well as defined milestones in academic research and infrastructure.

79. A career path from apprenticeship to the Universities of Applied Sciences should be created, marketed and promoted by organisational instruments (modules) as well as financial support.

80. R&D cooperation between university and industry should be stimulated since radical innovations often arise from academic research and scientific discoveries.

14. Guiding principles of a new policy: non-exclusivity, learning and mobility

81. The new strategy should be built on the principles of openness, non-exclusivity and mobility between firms and institutions. Openness for change and drawing knowledge from external sources should be overarching principles in education. Funding and policy decisions should be less influenced by the weight of interested parties, insider knowledge, and entropy (defined as the system's non-permeability of the system to information from external sources; see *convelop/Gerhardter*). The system should be open for experiments (e.g. pilot calls; see *Sabine Mayer*), and a culture moving from program based to task based intervention should be established which would make it easier to end programs. Continuous assessment and external evaluations (by international teams) should ensure that if the economic environment changes the system changes with it. The insiders and users of the current system are not overly critical of it as it is. They complain about administrative costs, but emphasize that they are guided well within the system, probably because direct funding did not dramatically affect their decisions. In comparison the new strategy allows new techniques to be learned, provides information and control and helps with planning. The strategy has to be implemented top down, information has to be gathered bottom up.

15. Regional, national, European: coordination and agenda setting on all levels

82. The Austrian research promotion policy should be redesigned and anchored within a multi-level system between the European Union, the federal and regional level. Deficits in coordination and specifically in agenda setting as well as the problem of cross-policies should be tackled. Demand for action exists specifically at the interface to the European level: while reflux of funds is working excellently (Austrian firms and institutions get more money back than government pays), there is not strategic, active co-design of the STI policy in the European Commission by Austrian authorities.

83. At the federal level we recommend the integration of further policies like educational, health, and environmental policies.

84. At the regional level a reorganisation of the one-way communication from the federal level is required. A two-way exchange of information and combined learning as well as possible support from cross-region activities should characterise the new system.

16. The case for radical change: a task beyond policy borders

85. Change is necessary, not because the Science, Research and Innovation System in the narrow sense has not worked, but because of new challenges and the new position of Austria as a high-income country. A successful innovation policy for a frontrunner has to be much more comprehensive and needs to interlink with other policies. The system should react to external as well as internal challenges, and to economic as well as societal trends (see *Michael Astor*). The changes needed are not minor changes. They need the attention of the top political level and an overhaul of current management and monitoring techniques; they rely on human capital, and build on the quality of the education system.

86. In addition, we now enter into a critical period in which firms, specifically multinational ones, will reduce research expenditures because of the crisis. At the same time, competitiveness of firms will depend even more on education and innovation in the crisis and thereafter, so that switching from an imitation strategy to a frontrunner position is absolutely necessary.

ANNEX

The Summary Report is based on nine special reports

Institute Author or Co-author

WIFO: Karl Aiginger, Michael Böheim, Rahel Falk, Michael Peneder,

Andreas Reinstaller, Susanne Sieber

OeNB: Jürgen Janger

KMU Forschung Austria: Iris Fischl, Sabine Mayer, Sascha Ruhland, Sonja Sheikh,

Jürgen Streicher

prognos: Michael Astor, Ulf Glöckner, Stephan Heinrich, Georg Klose,

Daniel Riesenberg

convelop: Gabriele Gerhardter, Markus Gruber, Simon Pohn-Weidinger,

Gabriel Wagner

Internal referees (WIFO):

Martin Falk, Klaus Friesenbichler, Werner Hölzl, Margit Schratzenstaller, Fabian Unterlass

External experts:

David Campbell (IFF Vienna), Heinz Hollenstein (KOF-ETH Zürich), Gernot Hutschenreiter (OECD), Barbara Kehm (University of Kassel), Hannes Leo (Leoon.at), Andrea Löther (GESIS Bonn), Herwig W. Schneider (iwi Vienna), Dorothea Sturn (University of Vienna), Gunther Tichy (ÖAW, WIFO)

Research assistance:

Julia Borrmann, Simon Braker, Dagmar Guttmann, Nadine Grieger, Elisabeth Neppl-Oswald, Michaela Reischitz, Karolina Trebicka, Ursula Weixlbaumer

Survey:

Gerhard Schwarz (WIFO), Roald Steiner (Salzburg University of Applied Sciences)

Steering Board

Federal Ministry of Transport, Innovation and Technology (BMVIT)
Federal Ministry of Economics, Family and Youth (BMWFJ)
Federal Ministry of Finance (BMF)
Federal Ministry of Science and Research (BMWF)
Austrian Council for Research and Technology Development

The Advisory Panel

All members of the Steering Board, and

The Austrian Federal Economic Chamber (WKO)
Federation of Austrian Industries (IV)
Federal Chamber of Labour (AK)
Austrian Trade Union Federation (ÖGB)
Austria Investment Agency and Promotional Bank (AWS)
Austrian Research Promotion Agency (FFG)
Austrian Science Fund (FWF)

OVERVIEW OF SPECIAL REPORTS

Report	***Work packages***	***Title***	***Authors***	***Institute***
1	WP 1	Framework Conditions	Jürgen Janger Michael Böheim Nadine Grieger	OeNB and WIFO
2	WP 2	Strategic Governance	Gabriele Gerhardter Markus Gruber Simon Pohn-Weidinger Gabriel Wagner	convelop
3	WP 3	Governance in RTDI – Relation between Ministries and Agencies	Sabine Mayer Iris Fischl Sascha Ruhland Sonja Sheikh	KMFA
4	WP 4	Tax Incentive Schemes for R&D	Rahel Falk	WIFO
5	WP 5	Direct Public Funding of RTDI	Sabine Mayer Iris Fischl Sascha Ruhland	KMFA
6	WP 6, 7	Effects of Block Grants on Research Institutes and Universities	Michael Astor Ulf Glöckner Stephan Heinrich Georg Klose Daniel Riesenberg	prognos
7	WP 8, 9	Public RTDI Funding - The Users Perspective	Sabine Mayer Sonja Sheikh Jürgen Streicher	KMFA
8	WP 12	Coherence of the Instrument Mix	Rahel Falk	WIFO
9	WP 10, 11, 13	Intervention Logic - Interaction between Institutions and Actors	Michael Astor Stephan Heinrich Georg Klose	prognos
Synthesis Report		Reaching out to the Future needs Radical Change	Karl Aiginger Rahel Falk Andreas Reinstaller	WIFO

REFERENCES

Aiginger, K., Competitiveness: From a Dangerous Obsession to a Welfare Creating Ability with Positive Externalities, Special Issue on Competitiveness, Journal of Industry, Competition and Trade, Vol. 6, No 2, June, 2006, pp. 161-177.

Aiginger, K., Walterskirchen, E., Tichy, G., WIFO-Weißbuch: Mehr Beschäftigung durch Wachstum auf Basis von Innovation und Qualifikation, WIFO, 2006.

Hölzl, W. (Koord.), Huber, P., Kaniovski, S., Peneder, M., Teilstudie 20: Neugründung und Entwicklung von Unternehmen, in: Karl Aiginger, Gunther Tichy, Ewald Walterskirchen (Projektleitung und Koordination), WIFO-Weißbuch: Mehr Beschäftigung durch Wachstum auf Basis von Innovation und Qualifikation, WIFO, 2006.

Hanisch, W., Turnheim, G., Einfluss der Forschung und Entwicklung von Auslandsunternehmen in Österreich auf die Erreichung der Lissabon-Zielsetzung 3,0% im Jahre 2010, Studie im Auftrag des Rats für Forschung und Entwicklung, April 2007.

Hutschenreiter, G., Innovation Policy and Performance in Austria, Chapter 2 in Assessing Innovation Policy and Performance – A Cross-Country Comparison, OECD, Paris.

Leo H. (Koord.), Falk, R., Friesenbichler, K.S., Hölzl, W., Teilstudie 8: Forschung und Innovation als Motor des Wachstums, in: Aiginger, K., Tichy, G., Walterskirchen, E. (Projektleitung und Koordination), WIFO-Weißbuch: Mehr Beschäftigung durch Wachstum auf Basis von Innovation und Qualifikation, WIFO, 2006.

Reinstaller, A., Unterlass, F., Forschungs- und Entwicklungsintensität im österreichischen Unternehmenssektor. Entwicklung und Struktur zwischen 1998 und 2004 im Vergleich mit anderen OECD-Ländern, WIFO-Monatsberichte, 2/2008, S. 133-147.

Tichy, G., Was ist das "Österreichische" an der österreichischen FTI-Politik?, erscheint in J. Fröhlich et al. Hg, Innovationsforschung in Österreich: Neue Konzepte und Herausforderungen, Innsbruck: Studien Verlag, 255-272, 2009.

CASE STUDY 2: GOVERNMENT-FUNDED R&D EVALUATION IN CHINA WITH SPECIAL REFERENCE TO THE EVALUATION OF NATIONAL R&D PROGRAMMES[8]

87. The need for an effective government-funded R&D evaluation system has become a high priority for the Chinese government. It is also attracting growing public attention. However, since evaluation is a relatively new concept in China, the institutional framework for R&D evaluation has not yet been established and the international evaluation community knows little about R&D evaluation in China. As one of the volunteer countries, China has participated in the effort to map the institutional frameworks, actors, regulations and practices of public R&D evaluation in different countries. This section presents[9] the findings of the Chinese case study.

1. China's evolving public R&D system

88. As a basis for understanding the R&D evaluation system in China, this section describes the key milestones in China's R&D system in the last 30 years and the major funding organisations at the central level.

1.1 Key milestones in the history of China's R&D system

89. China's R&D system[10] can be considered to have been created in the 1950s, when an R&D system was established, basically following the Soviet Union model. However the focus here is its evolution from the late 1970s.[11] The year 1978 is a milestone in the history of China's R&D system. The 1978 National Conference on Science and Technology was the starting point of a major shift in China's R&D system. Since then, the Chinese government has continually taken steps to move from a planned to a market-oriented economy.

90. From the mid-1980s to the early 2000s, national government-funded R&D programmes were included in the sixth, seventh, ninth and tenth five-year plans. These programmes addressed the priorities in each five-year plan period and were also the Chinese government's policy tools to promote innovation.

91. The year 2006 can be regarded as another milestone in the history of China's R&D system. At the beginning of 2006, China initiated the National Medium- and Long-term Science and Technology Development Plan (2006-2020) (the MLP). According to the MLP, China will invest 2.5% of GDP in R&D by 2020, up from 1.3% in 2005; raise the contributions of technological advances to economic growth to more than 60%; reduce its dependence on imported technology from 50% to less than 30%.[12]

8. Prepared by CHEN Zhaoying, HAN Jun and SHI Xiaoyong, National Center for Science and Technology Evaluation of China.

9. The opinions, findings and observations are those of the authors and do not necessarily reflect the views of the National Center for Science and Technology Evaluation (NCSTE) of China.

10. The discussion only covers civilian R&D.

11. Most studies of China's R&D system take their point of departure in the late 1970s.

12. These goals are part of the MLP documents. However the authors think that the measurement of these indicators, the policy implications and the utilisation made of them require further study.

The MLP also calls for China to become one of the world's top five countries in terms of number of invention patents granted to Chinese citizens, and for Chinese-authored scientific papers to become among the most cited in the world. For many observers inside and outside China, the MLP can be viewed as an important effort to shift China's current growth model to a more sustainable one and build an innovation-based economy by fostering indigenous innovation capability. It has attracted great attention from the international community.

1.2 Governance at the central level

92. The State Council Steering Group for Science, Technology and Education is a top-level co-ordination mechanism for dealing with strategic issues. The Ministry of Science and Technology (MOST), the National Natural Science Foundation of China (NNSFC), and the Chinese Academy of Sciences (CAS) are the main organisations that administer civilian R&D in China (Figure 2.1).

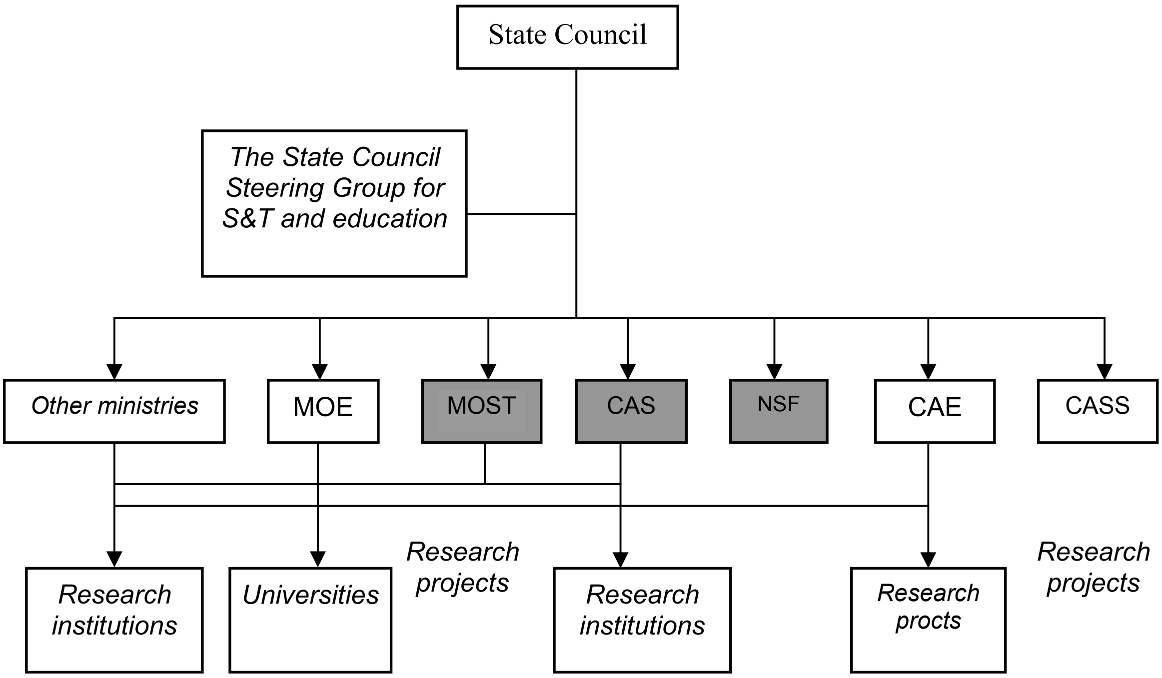
93. The main missions of MOST, under the direction of the State Council, are the formulation of innovation strategy and policies, promotion of the national innovation system, identification of R&D priorities, and the design and implementation of government-funded R&D programmes at national level.

94. The NNSFC aims at promoting and financing basic research in China. It mainly funds research in the natural sciences, such as physics, mathematics, chemistry and life sciences on the basis of proposals which are subjected to peer review. The principal recipients are Chinese universities and CAS research institutes.

95. CAS is essentially an R&D complex composed of about 120 institutes located around China. Major R&D funding for CAS comes from a line item in the government budget, projects supported by the national R&D programme, and some funding from the NNSFC.

96. A number of line ministries such as the Ministry of Agriculture (MOA), the Ministry of Education (MOE), the Ministry of Health (MOH) and the State Forestry Administration (SFA) also have R&D operations under their direct management.

Figure 2. Governance of public R&D at the central level



2. The Chinese case study

97. This section briefly describes the case study, including background on the evaluation of government-funded R&D in China, the reasons for the focus on the evaluation of the national R&D programme and the methods used.

2.1 The development of and demand for evaluation of government-funded R&D in China

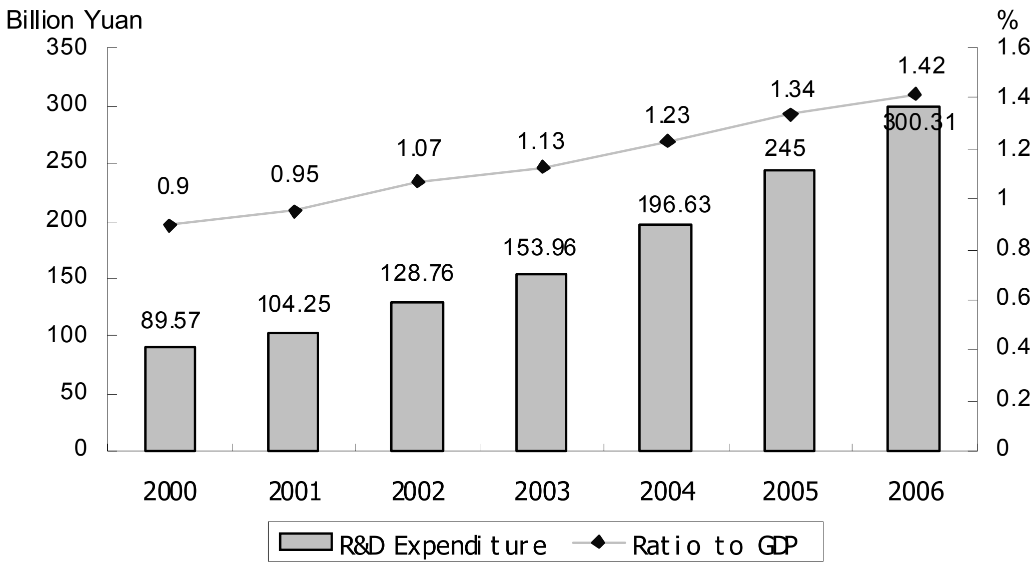
98. Evaluation of government-funded R&D in China can be said to have started in the 1990s, in particular with the programme evaluation carried out by MOST in 1994.[13] Largely through the initiative of MOST, some regulations on evaluation were released and the importance of evaluation was increasingly recognised. In MOST and a few other ministries, internal bodies responsible for managing evaluation were set up or specific staff were put officially in charge of evaluation. Since then, MOST has carried out evaluations of several national R&D programmes. A few line ministries have also carried out evaluation activities, primarily at the project level. Since 2005, China's leaders have called for government departments to be accountable for the results of public expenditure,[14] and new requirements for the evaluation of government-funded R&D are being established.

13. Some activities, such as policy analysis, management studies, surveys, programme reviews, were sometimes broadly termed "evaluation", but they differ a great deal from evaluation as understood by the international community in terms of design, implementation and the report presented.

14. Premier Wen Jiabao and the President Hu Jintao have called many times for government departments to be accountable for the results of public expenditure. The country's leaders require a government performance

99. The significant expansion of government funding of R&D has drawn more public attention to performance. Since 1999, China's spending on R&D has increased more than 20% annually. In 2006, it reached RMB 300.3 billion and 1.42% of gross domestic product (GDP), of which RMB 71.6 billion from the central government. Figure 3 shows the dramatic increase in R&D expenditure in China. This has raised concerns about the performance of R&D funding, and the government has come under pressure to establish an effective evaluation system for public R&D.

Figure 3. China's R&D expenditure and intensity 2000-06



Source: The Department of Development Planning, MOST, *Science and Technology Statistics 2007*, February 2008.

100. In its eleventh five-year plan the Chinese government proposes to improve administrative procedures and the management of public expenditure. While China's S&T policy already seeks to enforce accountability and improve the management of government-funded R&D through the introduction of an evaluation system, the implementation of the MLP will increase the pressure to evaluate the performance of government-funded R&D.

101. Since its accession to the WTO, China has become more active in bilateral and multilateral R&D programmes and projects and engages in a certain amount of joint activity. Examples include Galileo, ITER and the Framework Programme of the EU. Under the EU Sixth Framework Programme, China was the second largest country in terms of the number of projects in which it participated. As China's gradually becomes a major player in global innovation, its R&D system will become more global. This also requires Chinese government departments and relevant institutions to engage in the evaluation of international co-operation programmes or projects. However, the evaluation system and the culture of evaluation are relevantly weak at present in China, which is usually not sufficiently prepared to meet the evaluation requirements set by its co-operation partners. This has already drawn the attention of the MOST.

evaluation system to be set up to provide objective assessments of the government's policies, projects and programmes based on professional methods of evaluation and to improve administrative efficiency and ensure sound decision making.

2.2 Reasons for focusing on the evaluation of national R&D programmes

102. There are several reasons for focusing on the evaluation of national R&D programmes when introducing China's system for evaluating R&D, and specifically on the evaluation of government-funded R&D by MOST. Owing to the different duties and responsibilities of MOST, CAS and NNSFC, the type and focus of their evaluation processes have evolved quite differently. MOST is responsible for the evaluation of government-funded R&D. It also establishes relevant policies and manages evaluation activities, as set out in its mission statement approved by the State Council. The main focus of evaluation at NNSFC is on project selection (*ex ante* evaluation on the basis of peer reviews). At CAS, evaluations mainly involve internal R&D activities, primarily of the R&D labs, key projects and the research institutes. Therefore, a study of government-funded evaluation of R&D in China logically starts with MOST.

103. Second, among the evaluations carried out by MOST, the focus here is on the evaluation of national R&D programmes. It is widely recognised that national R&D programmes play a significant role in China's R&D system. These programmes are the country's most important policy tools for innovation. They allocate public resources to national priorities identified by the government and determine the most important government-funded R&D activities. For example, the National Hi-tech R&D Programme (the 863 Programme) and the National Key Basic Research Development Programme (the 973 Programme) have been the most important means of concentrating public resources on priority areas for S&T development to meet China's social and economic development needs.

104. Finally, the focus is on formal evaluations which rely largely on evidence and systematic design and implementation tools. Such evaluations have been relatively rare in China, and tend to concentrate on the national R&D programmes. While some cases of evaluations of R&D programme can be compared with those of other countries, they are relatively few in number. The evaluations described here are recognised as good practice in China. For policy evaluation, China is still at a preliminary stage and it is not easy to find suitable examples. The subject of e*x ante* evaluation (proposal review for project selection on the basis of peer review) is not addressed.

2.3 Methods for the case study

105. The case study follows the general TIP guidelines and the analytical framework, so that the results can be compared with those of other countries. Starting from the general guidelines and with a focus on the key questions to be addressed by all of the case studies, the study team based the study on the actual situation in China and established an analytical framework and a list of key issues.

106. The case study is based on desk study and expert review. Existing information, including relevant policy documents, statistics and evaluation reports, was examined. Interviews and discussions were held with a number of government officials, R&D researchers and evaluators. Specific evaluations, institutes and events with a strong influence on the development of the evaluation of R&D in China were also reviewed. Some of these are described below in boxes. For some issues mentioned in the general guidelines (mainly about goals, strategy and planning for evaluation), it is difficult to make a detailed statements at present, as most evaluations have not addressed them. The team can therefore only mention them briefly.

3. Institutional framework for the evaluation of R&D at MOST

107. R&D evaluation is a new concept in China, and prior to 2000, there were no relevant laws or regulations. Through an initiative of MOST, regulations on R&D evaluation were released in 2000 and

2002.[15] They aim at promoting the development of R&D evaluation. They do not deal with strategy, planning, implementation or budgeting of the evaluations.

108. In December 2007, the People's Congress revised the Law on Science and Technology Advancement which entered into force on 1 July 2008. The new law stipulates that the state will establish and improve a science and technology evaluation system which favours indigenous innovation. It requires the evaluation of investments in science and technology. While the principle has been established, the follow-up policies and regulations have yet to be determined.

109. This section focuses on the institutional framework for the evaluation of R&D programmes at MOST and briefly discusses strategy and planning for national R&D programme evaluation.

3.1 Evaluation requirements in regulations on the management of national R&D programmes

110. Recently, in order to improve the management of national R&D programmes and projects and in response to the general public's demand for accountability, the ministers and officials of MOST responsible for the management of R&D programmes have accorded more importance to evaluation. For example, evaluation requirements have been included in regulations on the management of national R&D programmes. During the eleventh five-year period (2006-10), the regulations for three major national R&D programmes– the National Hi-tech R&D Programme (863 Programme), the National Key Basic Research Development Programme (973 Programme) and the National Key Technologies R&D Programme – stipulate the evaluation of projects and sectors. However, they do not mention programme evaluation, so regular evaluations are not planned and programme evaluation is conducted on a case-by-case basis. Table 2 summarises the evaluation requirements for projects and sectors.[16] It is not clear which programmes are to be evaluated, when evaluations should be done and who is responsible.

15. They include the Provisional Regulation of S&T Evaluation Management (MOST, 2000); the Regulation of S&T Evaluation (MOST, CEPD, CET and MOF, 2002); the Policy Statement of Improving Activities for S&T Evaluation (MOST, MOE, CAS, CAE and NSFCC, 2002).

16. A national R&D programme is often structured with three levels: programme, sector and project. For example, the 863 Programme has ten sectors and each sector has its goals and implementation plan. The goals of the sector are to be achieved through projects.

Table 2. Evaluation requirements in management regulations of three major national R&D programmes

Programme name	Appraisal	Mid-term evaluation	Completion evaluation	Performance evaluation
863 Programme	All projects to be appraised through peer review or expert panel review.	Mid-term evaluation conducted by a professional evaluation organisation to assess the implementation of megaprojects or sectors.	All projects to be assessed when completed. This is organised by the MOST programme management offices.	Performance evaluation of megaprojects and sectors to be conducted after completion.
973 Programme	All projects to be appraised through peer review or expert panel review.	All projects to be evaluated after two years by an expert consultant group, with a focus on the project's status and prospect.	Project completion evaluation to be conducted by an expert panel commissioned by MOST. Evaluation of sub-projects to be conducted by the chief scientist of the project. The focus is on the achievement of objectives, effectiveness and nurturing of talent.	No requirement.
National Key Technologies R&D Programme	All projects to be appraised through peer review or expert panel review.	All projects of more than 3 years of duration to be evaluated at mid-term, organised by MOST.	Evaluation of completed projects to be organised by MOST	Performance evaluation to should be conducted for all projects. Can be combined with mid-term evaluation or completion evaluation.

3.2 Management structure for the evaluation of national R&D programmes

111. At MOST, the main actors for the evaluation of national R&D programmes are the Division for S&T Evaluation and Statistics of the Department of Development and Planning (DDP), different operational departments (ODs) and the National Centre for Science and Technology Evaluation (NCSTE).

112. The DDP is responsible for co-ordinating national R&D programmes and acts as manager of R&D evaluations. Its functions relating to evaluation are: *i)* to conduct research on theories, norms and standards of R&D evaluation; *ii)* to put forward regulations for managing evaluations; and *iii)* to organise and co-ordinate the evaluation of R&D policies, R&D strategies, R&D programmes and megaprojects.

113. Different ODs are responsible for the preparation and organisation of R&D programmes. For example, the Department of Basic Research is responsible for the National Key Basic Research Programme (973 Programme). Therefore, its evaluation is usually commissioned by the DDP and the relevant ODs. In some cases, MOST leadership will commission an evaluation to address issues of concern

to the leadership. The commissioners of the evaluations generally state the purpose of the evaluation and identify issues to be addressed.

114. Since its establishment in 1997, the NCSTE, which is affiliated to MOST, has the main responsibility for implementing evaluations of national R&D programmes and projects and of science and technology policies. As a professional evaluation agency, the NCSTE provides solid evidence for R&D decision making and makes recommendations for improving the management of R&D through objective, impartial and independent evaluations. Box 2.1 specifies its functions.

Box 2. The Chinese National Centre for Science and Technology Evaluation

Founded in 1997 with the approval of the Ministry of Science and Technology, the NCSTE is one of the leading organisations in the field of evaluations in China. NCSTE is responsible for implementing evaluations of R&D programmes, policies, institutes, as well as megaprojects. The NCSTE's responsibilities are as follows:

13. evaluation of major science and technology (S&T) development strategy;
14. evaluation of various national R&D programmes;
15. performance evaluation of national R&D institutes;
16. national R&D project and/or programme budget appraisal;
17. R&D human resource evaluation;
18. regional innovation capacity evaluation;
19. provision of technical support and evaluation quality control for other MOST agencies;
20. research on R&D evaluation norms, standards and methodologies;
21. international co-operation on R&D evaluation.

The NCSTE's human resources include: *i)* 25 permanent staff specialised in management consulting, public R&D policy research, technological and economic analysis, R&D evaluation and development evaluation, etc.; *ii)* about 25 contracted senior experts and advisors who are either senior specialists or retired senior officials in various fields; and *iii)* an expert database with approximately 3 000 registered experts.

In order to learn about international practice and experience with evaluation, the NCSTE has established close links with partners in various countries and international organisations, such as the United States, Canada, Japan, Korea, the United Kingdom, Germany, Denmark, the Netherlands, UNDP, the OECD, and the World Bank. In the past 10 years, NCSTE has held a number of evaluation workshops, seminars and training courses in Beijing in co-operation with international organisations and foreign ministries.

3.3 Standards for R&D evaluation in China

115. Mandated by the MOST, NCSTE has drafted the Uniform Standards for Science & Technology Evaluation[17] (the standards) which was published in 2001. In the same year, MOST made this the reference document for government regulation on evaluation management.

116. As the first and only R&D evaluation standards in China, the standards have three objectives. First, they give guidance on ethical conduct for evaluators and other actors in S&T evaluation. Second, they provide standards for the professional practice of various S&T evaluation activities, mainly programme and project evaluation, to enhance the quality of evaluation processes and to improve the

17. In China, especially at MOST, the term S&T evaluation is commonly used. However, judging from the purpose, content and implementation of evaluation, evaluation of R&D activities would be more appropriate. This is therefore the term used here except when referring to published documents or reports, when the term S&T evaluation is used as in the documents or reports themselves. Therefore, when introducing the standards, the term S&T evaluation is used.

utility of evaluation results. Third, they can be used as fundamental materials for training for S&T evaluation in China.

117. In order to achieve these objectives, the standards are divided into two parts: core content and reference content (Box 2.2). The former include principles on ethical conduct in evaluations and the standards for professional practices. The latter include further explanations of the core content and discussions of typical evaluations and can be revised as evaluation practices evolve.

Box 3. Structure and content of the standards

The standards are divided into core content and reference content, each of which is subdivided into two sections.

Core content

Section1. Guiding principles on ethical conduct of evaluation

It covers rules of behaviour for evaluators and evaluators' relationship with evaluated bodies, clients and end users, with the focus on the evaluator's behaviour. It also provides the terminology on science and technology evaluation used in the standards.

Section 2. Standards for professional practices

It addresses technical issues regarding evaluations, including the major evaluation procedures, the roles of the actors in evaluations, key steps and critical issues in the design and implementation of evaluations, detailed requirements for evaluation reports, and commonly used methodology and tools.

Reference content

Section 3. Explanations of the core content of the standards

It gives further clarification, explanations and supplements to the core content. It is mainly targeted at the key issues and issues that are easy to misunderstand or about which there is some debate.

Section 4. Discussions on typical cases

On the basis of current S&T evaluation practices in China, this section provides ten cases of evaluation design and implementation as a reference to facilitate the readers' better understanding of the core content.

118. Since the document was issued, the standards have been used in different types of R&D evaluations nationwide. They also provide a basis for evaluators from various regions and institutions to share R&D evaluation experience, and for evaluators to standardise their work. Moreover, it constitutes the basic material for training for science and technology evaluation. More than 800 evaluators from 70 evaluation institutions across China participated in S&T evaluation training courses organised by NCSTE between 2001 and 2003.

3.4 Strategy and planning for national R&D programme evaluation

119. Although MOST increasingly recognises the importance of evaluation, it does not at present have either a strategy or a schedule for evaluations of national R&D programmes, which are conducted on a case-by-case basis. There is no regular budget allocation and no timing requirement.

120. However, the purpose of individual evaluations is clear, although it varies to some extent depending on the evaluation. On the whole, the purpose of evaluations of national R&D programmes can be summarised as follows: *i)* to assess objectively the appropriateness of the goals, implementation and management, and the effectiveness and impacts of the programme; *ii)* to learn from past experience and

practices and identify weaknesses in the programme in order to improve its management; and *iii)* to provide evidence to the programme management and leadership of MOST for decision-making purposes.

4. Design and implementation of an evaluation of national R&D programmes

121. This section describes the design and implementation of an evaluation of national R&D programmes, including the methodology used. It also discusses the role and activities of decision makers, programme managers, external experts, and other stakeholders.

4.1 Design and organisation

122. When the NCSTE is asked to carry out an evaluation, it organises an evaluation team. The team is composed of two types of experts: evaluators from NCSTE and external experts from specific technological fields who provide their opinions. An NCSTE evaluator acts as team leader. This ensures that the evaluation respects the evaluation standards and fully considers the characteristics of the specific R&D programme. In the case of a comprehensive programme evaluation, the evaluation team is divided into groups to address different themes.

123. The organisation of an evaluation generally includes a steering committee (SC) in addition to the evaluation team. The SC is composed of officials from the commissioners, programme managers and staff from the implementing agencies of national R&D programme and, sometimes, representatives from other stakeholders. The SC is responsible for co-ordination and takes decisions on major issues that arise during the evaluation. The SC holds periodic meetings during the evaluation process to learn about its progress.

124. The design of an evaluation of a national R&D programme is mainly the responsibility of the evaluation team but it is generally decided jointly by the commissioners and the evaluation team. The evaluation team prepares a draft design for the evaluation based on the commissioner's needs and desk study of related programme documents which is then submitted to the commissioner. The draft document covers the objective, scope methodology, process and work plan of the evaluation. The commissioner discusses the document with the evaluation team and makes some suggestions. On the basis of the discussion, the evaluation team revises the document and resubmits it to the commissioner. In general, this document is confirmed as final by the commissioner and the evaluation team. Once the evaluation has been completed, the team leader, usually with a couple of core members, is responsible for preparing the evaluation report. The final evaluation report is submitted in the name of NCSTE which is responsible for the evaluation results.

4.2 Methodology

125. The NCSTE has developed a relatively mature evaluation framework which covers programme goals and objectives, programme management and implementation, programme effectiveness and impacts. Each dimension is examined with the use of some key questions (Table 3). The evaluation team can specify the questions to develop indicators that reflect the features of a given programme. For example, in the evaluation of the National Key Basic Research Programme (973 Programme), the evaluation of the programme's effectiveness and impact took account of major national development needs, progress of pioneering basic research, nurturing of human resources, academic communication and co-operation, and building of the research base.

Table 3. Framework for national R&D programme evaluation

Criteria	Key questions
Goals and objectives	• Are the programme's goals and objectives clear? • Is the programme designed so that it is not redundant or duplicative of any other programmes? • Is the programme planned in an efficient and effective way to achieve its goals and objectives?
Management and implementation	• Is the management model suitable for the implementation of the programme? • Are the programme and its projects organised and implemented in an effective and efficient way? • Is the distribution of programme funds appropriate? • Are the project implementers competent to carry out the projects?
Effectiveness and impacts	• What are the outputs of the programme, such as papers, patents, etc.? • Has the programme nurtured qualified human resources for R&D? • Has the programme enhanced the research infrastructure in its field? • Has the programme addressed bottlenecks of technology development in its field? • Has the programme facilitated the development of high-technology industry? • Has the programme promoted co-operation between industry, university and research?

126. Evaluations make use of both quantitative and qualitative indicators, with more qualitative than quantitative indicators. Because the information management system (IMS) for national R&D programmes is not good, the evaluation cannot obtain some key data, such as data on the effectiveness and socioeconomic impact of the programme. The quantitative indicators are based on data provided by the IMS or by questionnaire surveys, while qualitative indicators are obtained from evidence collected at workshops, interviews and questionnaire survey.

127. Methods used in evaluations are desk study, questionnaire surveys, field visits, focus group meetings, statistics, and cross-cutting analysis. The desk study reviews programme and project documents and related documents to understand the nature of programme and to collect key evidence needed for the evaluation. The self-administered questionnaire survey usually covers three types of actors involved in the programme: management experts, scientists and organisations.[18]. Focus group meetings are an important way to collect information on the programme and the opinions of the three types of stakeholders.

128. Peer review is used in different types of evaluations. At MOST, peer review is used in project appraisal and largely determines the approval of project. In evaluation of national R&D programmes, peer review is usually in the form of peer panel review, mainly for consultation about technical issues. During an evaluation, the evaluation team organises several workshops and invites a group of experts in certain S&T fields to participate. In the workshops, the evaluation team consults the experts about the quality of R&D results produced by the programme. If their judgments concur, their view is adopted in the evaluation; when their judgments differ, all are presented in the evaluation report.

18. Organisation includes the universities, research institutes and enterprises for which the scientists work.

4.3 Process

129. When the design of evaluation has been finalised, the evaluation is carried out. It should rigorously follow the procedure set by NCSTE. General speaking, there are five steps in the process of evaluating national R&D programmes: evidence collection and analysis, thematic evaluation, cross-cutting evaluation, report preparation and interaction with commissioners, and finalisation of the report.

130. The evidence collected and analysed covers R&D project data, programme statistics, programme-related documents such as project appraisal documents, mid-term reviews and final reports, related policy documents, and opinions of stakeholders. Evidence is collected from statistics, questionnaire surveys, field visits, focus group meetings and interviews.

131. It is difficult to collect evidence owing to the poor information management system for national R&D programmes. In 2006, MOST introduced a uniform IMS for national R&D programmes and projects. At present, data on project appraisals have been computerised, while data on project implementation and completion have only been partly entered into the database and the quality of some data is not good. Information on project effectiveness and the application of results is particularly weak. This has made it difficult to evaluate national R&D programmes.

132. Thematic evaluations are conducted to address some leading issues, such as participation by enterprises in national R&D programmes, industrialisation of R&D results, structure of R&D project implementers, etc. These issues are usually proposed by programme management and sometimes by the MOST leadership.

133. Based on this analysis, a cross-cutting evaluation is carried out to draw the conclusions of the evaluation of the national R&D programme. The conclusions are classified according to the evaluation framework described above: programme goals and objectives, programme management and implementation, and effectiveness and impacts of programme. Once the draft evaluation report has been prepared, the evaluation team consults with the commissioner and then amends and finalises the report.

134. During the evaluation process, the evaluation team interacts with the evaluatees. These are the officials and staff of operation departments, implementers of R&D projects, and the programme's management experts. The evaluators and evaluatees mainly interact is through focus group workshops. The evaluators present the purpose and design of evaluation to evaluatees to help them understand the evaluation. The evaluators also inform them on the progress of the evaluation. The evaluatees give their opinions and recommendations on the evaluation as well as on the R&D programme. This interaction helps both sides better understand each other and makes the evaluation progress more smoothly.

4.4 Role of stakeholders in the implementation of evaluation

135. While the evaluation is being carried out, the DDP gives the evaluation all necessary support, such as co-ordination of the stakeholders. Operational departments usually participate in discussing the evaluation design and are responsible for providing the necessary data and material on the national R&D programme. The project implementers under a programme are interviewed by the evaluation team and provide information and their opinion on their project.

136. External experts mainly consult on technical aspects of R&D projects and give their opinions on the goal, priority setting, management and implementation of the programme; this is important information for the evaluation. They also fill in the self-administered questionnaire provided by evaluators. Other stakeholders fill in the self-administered questionnaire, participate in focus group workshops and are interviewed by evaluators.

Box 4. The evaluation of China's National Key Basic Research Development Programme (973 Programme)

In March 1997, The National Key Basic Research Development Programme (973 Programme) was introduced to strengthen basic research in line with national strategic targets. The 973 Programme covers six sectors: agriculture, energy sources, information, resources and environment, population and health, materials, plus cross-disciplines and frontier sciences. Projects in the 973 Programme generally have a five-year implementation period.

The main tasks of the 973 Programme are to strengthen and support research on a number of major scientific issues of importance to national socio-economic development, to consolidate a highly qualified contingent for basic research and cultivate personnel with innovative capabilities, and to improve and perfect programme management to create a sound environment for innovation.

From April 2005 to March 2006, the NCSTE conducted the first evaluation of the 973 Programme. The evaluation adopted a framework based on NCSTE's past experience and practice, which covers goals and arrangement, management and implementation, and effectiveness and impacts. The methods used included policy analysis, statistics analysis, a questionnaire survey, case studies, field visits and focus group meetings. It further aimed to promote lessons learned and to provide evidence for decision making.

The main findings from the evaluation were as follows. First, the 973 Programme was launched in the 1990s by the Chinese government to enhance basis research and to improve China's indigenous innovation capacity, At the time, there was a lack of national needs-oriented basic research; and the programme has significantly improved China's basic research system and the integration of basic research and national needs. Second, the management mode of the 973 Programme is generally suitable for the programme's characteristics and its implementation. It works efficiently towards achieving the programme's goals, but needs to be improved in some respects. Third, owing to the relatively short period of implementation, the overall effectiveness and impact of the programme have not been fully demonstrated and breakthroughs to resolve important national needs or to reach the scientific frontier are still on the horizon. The evaluation recommended the establishment of a stable funding mechanism for the 973 Programme so as to increase total funding for the whole programme and the intensity of project funding.

Source: Chinese National Centre for Science and Technology Evaluation, Evaluation Report of National Key Basic Research Development Programme (973 Programme), March 2006.

Box 5. The evaluations of China's National Hi-tech R&D Programme (863 Programme)

The Hi-tech Research & Development Programme (863 Programme) is China's largest R&D programme. It is committed to addressing strategic, advanced and forward-looking high-technology issues that are crucial to the nation's future development and security. It plays a leading role in the future development of emerging industries by developing, integrating and applying proprietary high technologies.

So far, three evaluations of the 863 Programme have been conducted, in September 1995, August 2000 and May 2006. All three have adopted a combination of comprehensive and thematic evaluations. The comprehensive evaluation covers programme goals, programme management, and effectiveness and impacts. The thematic evaluation covers that are relevant at the time. For example, the first two evaluations mainly focused on the adjustment of programme goals, the management model of the programme, accountability in the programme management, and the impacts of the programme on high-technology industrialisation. The third emphasised issues such as the organisation and implementation of megaprojects, the participation if enterprises in the programme, patents produced by the programme, etc.

The usual methods were followed: desk studies, field studies, surveys using questionnaires, and data and information from the 863 Programme management, and information collected directly during the evaluation. Probably the most striking and original aspect of this evaluation was the "stakeholder dialogue approach", with the organisation of several roundtable workshops. These included programme managers or persons with a direct interest in the programme such as project managers, conductors of 863 projects and S&T experts not participating directly in 863 projects. Debates were led by the NCSTE professional evaluation staff, which afterwards summarised them in the form of reports.

Although the evaluation results have not yet directly led to decision making, some received attention from the 863 Programme management and the leadership of MOST, and some follow-up measure were taken. For example, the third evaluation found that during the tenth five-year period (2001-05), enterprises became major project implementers. They took charge of or participated in 50% of projects and received 60% of central funds, and the number of patents produced was 3.8 times that of the preceding 15 years. Yet much information about the operation and financial status of these enterprises was lacking or of poor quality, which made it difficult to appraise and control risks that could emerge during project implementation. Thus the evaluation recommended that the DDP improve its MIS and supplement its information on the operation and financial data of enterprises. The DDP and the programme management recognised the importance of these findings and took measures to track and improve data about enterprise development and patents.

Source: Chinese National Centre for Science and Technology Evaluation, Evaluation Report of the Tenth Five-year National Hi-tech Research & Development Programme (863 Programme), November 2007; Chinese National Centre for Science and Technology Evaluation, Evaluation Report of National Hi-tech Research & Development Programme (1986-2000), March 2001; Chinese National Centre for Science and Technology Evaluation, Evaluation Report of National Hi-tech Research & Development Programme (1986-95), January 1996.

5. Utilisation of national R&D programme evaluations

137. This section discusses the utilisation of the evaluation by programme management and the response to the evaluation results by the leadership. At MOST, feedback from evaluation results is weak and unstable owing to the lack of any institutional mechanism. Evaluations have had little effect on priority setting and on the budget co-ordination and allocation process, but they have a certain indirect influence on programme management. There is much room for improving the use of evaluations.

5.1 Circulation of evaluation reports

138. The evaluation reports on national R&D programme have not all been made public. This is mainly because MOST considers the reports as being for internal use. Summaries of the early programme evaluations were published in some newspapers, but not of later ones. The reports are, however, circulated within MOST and occasionally to some external stakeholders. The commissioner of the evaluation determines what is circulated. In sum, there is no standard procedure for circulating evaluation reports and practices differ.

5.2 Utilisation of evaluations by programme management

139. The primary users of the evaluation findings are the commissioners, DDP and operational departments. They are informed about the implementation and effectiveness of the programmes and thus take measures to improve the management of the programme. In some cases, the evaluation reports are also circulated to implementing agencies of the programme at MOST.

140. More generally, the evaluation's role in project management takes place during the evaluation process. The evaluation team informs programme management about problems identified in the management process and discusses the causes and possible remedies with the management. This allows for resolving some problems before the evaluation is completed. The evaluations also make recommendations for improvements which help managers enhance their management and planning capacities.

5.3 Leadership's response to the evaluation results

141. The leadership's response to evaluation results can affect the importance accorded to evaluations. Early evaluations received more attention from the leadership than more recent ones. Currently, when an evaluation is completed, the evaluation team briefs the minister responsible for that programme. Sometimes the evaluation team or the DDP also briefs the leadership on the findings, in which case the leadership attaches more importance to the evaluation and the issues. Usually, the leadership responds and

indicates which departments and agencies should address the relevant issues. In such cases, an evaluation may influence decision making and lead to policy improvements.

6. Observation and key challenges

142. This section offers some observations about the present status and the key challenges for the evaluation of R&D in China.

6.1 Current status of R&D evaluation

143. China's R&D system has developed strongly and the country occupies a special position among developing countries in terms of R&D. However, the evaluation of R&D does not seem to be keeping pace with other aspect of the system. At present, China's R&D evaluation system is at an early stage of development.

144. Although the terms "evaluation" and "assessment" are broadly used for R&D management in China, both within government departments and among experts, a shared understanding of the function, implementation and utilisation of evaluation is lacking. China's R&D evaluation system can be characterised as follows.

145. ***Weak institutionalisation****:* Compared to other aspects of the R&D system, the institutional context is weak. Over the past decade, government departments have issued regulations on the regular evaluation of the government's programmes and key projects. However the regulations have not clearly stipulated which programme should be evaluated, when and how. There is no annual planning of work and budget for evaluation activities, and there are no specific implementation guidelines for different kinds of evaluation. In consequence, the regulations have not truly been implemented. The evaluation of R&D has not yet been incorporated into the management and decision-making process and is not a regular and obligatory activity.

146. ***Imbalance in evaluation capacity****:* Attention is paid to the capacity to undertake an evaluation but little attention is paid to the capacity to use evaluation results. However while the gap between China and the developed countries is reflected in the lack of professional personnel to design and implement evaluations, it is reflected even more in the lack of competent officials to manage and co-ordinate evaluation activities. There is already a group of professional evaluators in China, although they are few in number. Yet within government departments, only a few officials understand why the government needs evaluation and know how to ensure that evaluation findings are used appropriately.

147. ***Some good programme evaluation practices****:* China has already completed some programme evaluations, such as those of the 863 Programme and 973 Programme. Judging from the evaluation framework and methods adopted in these cases, the methods are not very different from those used elsewhere in the world. The Country-led Joint Evaluation Dutch ORET/MILIEV Programme in China,[19] completed in 2006, is regarded as a success by the international evaluation community. However, there are few such cases.

---

19. The evaluation was jointly launched and completed by IOB (the Policy and Operations Evaluation Department of the Netherlands Ministry of Foreign Affairs) and NCSTE (National Centre for Science and Technology Evaluation of China). NCSTE and IOB made a joint presentation on the evaluation at the Sixth Meeting of the OECD DAC Network on Development Evaluation, 27-28 June 2007, Paris. The report of the joint evaluation has been placed on the DAC website.

148. ***Use of evaluations depends on top leaders****:* The influence of evaluation results on decision making depends mostly on top leaders in government departments. Even if two programme evaluations are basically on the same level in terms of design, information collection and analysis, the influence of the evaluation on decision making varies greatly, owing to different officials' understanding of evaluations. In sum, what are lacking are the institutional arrangements and the feedback and learning loops that involve policy makers and programme managers.

6.2 Building China's government-funded R&D evaluation system

149. The government-funded R&D evaluation system should cover R&D at different levels, starting from projects and moving through programmes and ministries to the system as a whole. The question is, based on the Chinese profile, whether to start building at the project, programme, ministry or national level. China may not yet be in a position to adopt a strategy at the ministry or country level; a mixed approach may be a more realistic choice at present. In terms of what has already been and is being done, the point of departure is usually pilot work for core national R&D programmes and key projects supported by the programmes.

150. In terms of evaluation priorities, outcomes should perhaps be the first priority. The Chinese government is under increasing pressure to produce and to demonstrate the results of R&D. MOST is held accountable for the national R&D programmes and is required to report on performance. That is, the main initial focus of evaluations should be the identifiable changes resulting from the programmes and the key project deliverables.

151. It is more difficult to measure the impact of R&D than the outcomes. The impacts are usually generated not simply by a single programme/project but depend upon numerous factors which are very difficult to attribute directly to the programme/project because they are also affected by exogenous factors. The goal of measuring project impacts might be addressed at a subsequent stage.

6.3 Peer review in the programme evaluation

152. Peer review plays different roles in *ex ante* and *ex post* evaluations, so that the problems which arise for peer review as an evaluation tool also differ and must be addressed differently. In China, criticism or debate relating to peer review mainly focuses on its role in *ex ante* evaluation. As a fundamental mechanism for project selection, peer review is used by almost all government departments. Table 2.1 shows that, for the three major S&T programmes, all projects are to be appraised through peer review or expert panel review. Peer review is an important element in the allocation of R&D resources, and the most serious criticism of peer review at present concerns its fairness.

153. Government departments try to improve the peer review process for project selection in various ways: by setting up a pool of reviewers pre-approved by relevant management departments, by randomly selecting reviewers from the pool and obliging reviewers to declare in advance any possible conflict of interest; by sending feedback of the panel summary to the applicant, etc. Transparency is the most important way to improve peer review and it is increasing in China.

154. Much less attention is currently given to peer review in *ex post* evaluations. As noted above, peer review usually takes the form of expert panel review in evaluations of national R&D programmes. This primarily involves providing opinions on technical issues to the evaluation team, such as the quality of the programme outputs. In programme evaluations, the evaluation team's judgement is based on careful analysis; the expert panels do not make judgements themselves.

155. Guidelines can help peer review play a better role in programme evaluation. Before inviting experts to participate in the programme evaluation, the evaluation team usually prepares guidelines, including background information on the programme being evaluated, key issues that experts need to address, the consultation procedures, requirements for the reviewers' report, and so on. Practice shows that the quality of peer reviews is enhanced when guidelines exist and are used.

6.4 Programme evaluation and priority setting

156. As described above, evaluations are used more for improving existing programmes than for supporting resource allocation. There are no data to show that the results of evaluation affect the setting of priorities for the following reasons:

- First, when the evaluation task is assigned, the proposed objectives of evaluation usually do not include priority setting.

- Second, in China, the key moment for priority setting is the first year of every five-year plan. Evaluation does not fit into programme cycles because of the absence of any institutional arrangements for programme evaluation.

- Third, in terms of timing, evaluations are often conducted when the programme has achieved a certain level of results. Thus, the evaluation of a programme usually starts in the last year of the five-year plan. Moreover, since an efficient monitoring system which could facilitate evaluations is lacking, it takes quite some time to collect the necessary data and information. As a result, for priority-setting decisions are often made before programme evaluations have been finalised.

6.5 The challenge of learning from international experience

157. Government officials and experts in China have tried to draw lessons from current international experience. However, they must consider the extent to which that experience suits the Chinese context, which elements can be especially valuable for establishing a Chinese R&D evaluation system, and which may not be suitable. There are important challenges that must be faced for learning from international experiences. The Chinese government will need to pay attention at least to the following issues.

158. The challenges of transferring policy or management tools from developed to developing countries are well known. Since evaluation is a tool from a "Western" management tradition, it may be asked whether the lessons are applicable or applicable to the same extent in China.

159. Since China and "Western" countries are not at the same stage in the development of the economy, society and R&D system, some earlier evaluation practices in Western countries may better fit China's current situation and some current Western practices might be better kept for the future. For example, the evaluation of the impact of R&D programmes has recently attracted much attention in the evaluation community, especially in OECD countries, but it is less a focus of attention in China.

160. In terms of design and implementation of evaluations, learning from international experience should be comparatively easier, as Chinese evaluators face similar challenges. For example, an indicator framework to guide performance evaluation of R&D programmes has been an important concern for policy makers, programme managers and evaluators in China. However, no agreement has been reached on the kind of indicator framework which would be suitable and how to develop it.

REFERENCES

Chinese National Center for Science and Technology Evaluation, Evaluation Report of National Key Basic Research Development Programme (973 Programme), March 2006.

Chinese National Center for Science and Technology Evaluation, Evaluation Report of the Tenth Five-Year National Hi-tech Research & Development Programme (863 Programme), November 2007.

Chinese National Center for Science and Technology Evaluation, Evaluation Report of National Hi-tech Research & Development Programme (1986-2000), March 2001.

Chinese National Center for Science and Technology Evaluation, Evaluation Report of National Hi-tech Research & Development Programme (1986-1995), January 1996.

The Department of Development Planning, MOST, Science and Technology Statistics 2007, February 2008.

Chinese National Center for Science and Technology Evaluation, Uniform Standards for Science and Technology Evaluation, Price Press, China, 2001

Ministry of Science and Technology, Provisional Regulation on the Science and Technology Evaluation (Rg.No.2000-588), December 2000

Ministry of Science and Technology and Ministry of Finance, Management Regulation of National Hi-tech R&D Programme (863 Programme) (Rg.No.2006-329), July 2006

Ministry of Science and Technology and Ministry of Finance, Management Regulation of National Key Basic R&D Programme (973 Programme) (Rg.No.2006-330), July 2006

Ministry of Science and Technology and Ministry of Finance, Provisional Management Regulation of National Key Technologies R&D Programme (Rg.No.2006-331), July 2006

Ministry of Science and Technology, the Handbook of the Ministry of Science and Technology of the People's Republic of China, 2008

OECD/DSTI/STP, the OECD Review of Chinese National Innovation System and Policy – Synthesis Report, 25 October, 2007, OECD, Paris.

EVALUATION, IMPACT ASSESSMENT AND FORESIGHT IN FINLAND[20]

1. Introduction

161. This section provides a comprehensive stocktaking of the current status of evaluation, impact assessment and foresight in Finland, and an overview of the main results of evaluations carried out especially since 2005. Furthermore, it can be seen as an example of how evaluation is treated in the Finnish policy regime, especially in light of the current process of transition in science, technology and innovation policy (STI) and the entire innovation system.

162. It relates to the public statement made by the Finnish Science and Technology Policy Council (STPC) on August 2007.[21] It addresses both the emerging need to evaluate the Finnish innovation system (NIS) as a whole and the need to enhance existing evaluation practices. It also follows to some extent the guidelines set for the OECD project on evaluation systems.

163. The discussion also touches upon the general framework for Finnish evaluation activities. However, Finland has no evaluation *system* as such, although evaluation is deeply rooted in the Finnish innovation policy system. It concerns particularly *ex post* evaluation carried out by domestic specialists, often with foreign experts. In this sense, almost everything is evaluated, from projects and single policy measures to programmes and institutions (see Table 2.3). Even the national innovation support system was evaluated a few years ago. However, a comprehensive evaluation at the level of the innovation system has not been carried out since the mid-1980s, when it was done by the OECD (*Review of National Science and Technology Policy: Finland*, published in 1987).

2. Recent evaluation-related developments in STI policy

164. Investments in research and innovation have been growing steadily in Finland, particularly since the latter part of the 1990s. Both public and private parties, sometimes in partnership, have increased their efforts to create and exploit new knowledge, technology and skills. At the same time, the evaluation of the impact of policy measures and funding has become a focus of attention. Evaluation has gradually become an increasingly systematic activity, yielding more reliable and useful results. However, evaluation and impact assessment, as well as foresight, are very challenging areas. They require highly developed methods, a range of measures for development, visionary outlook and insight. Hence, there is still a lot to do to develop evaluation and foresight as such and to bridge the gap between results derived from these activities and (political) decision making.

---

20. Prepared by Mr. Kai Husso, Chief Planning Officer, Science and Technology Policy Council of Finland and Mr. Esko-Olavi Seppälä, Secretary General, Science and Technology Policy Council of Finland.

21. Much of this material comes from a background document for the STPC statement (24.8.2007) on the development of impact assessment and foresight in Finland. While preparing the text for the OECD project on evaluation, the STPC secretariat was greatly aided by the Finnish OECD/TIP delegate, Mr. Pentti Vuorinen, from the Ministry of Employment and the Economy.

165. Since 2005, dozens of evaluation and foresight reports concerning research, development and innovation (RDI), their funding and policy measures have been published in Finland. The increase in the cost of, and in investments in, R&D, changes in the nature and execution of R&D, and the globalisation of the operating environment have all led to greater need for evaluation and foresight as well as higher requirements for such activities. The Academy of Finland[22] and Tekes[23] have been quite active in evaluation and foresight, but reports have also been produced by the Research Institute of the Finnish Economy (ETLA), the Finnish Innovation Fund (Sitra), the Government Institute for Economic Research (VATT), among other important R&D organisations. The key results of these publications are described here.

166. The most important STI policy document in recent years, the so-called Government Resolution (2005[24]), points out that the structural development of the public research system is needed in order to raise the quality and relevance of R&D. Development measures should be targeted at boosting the prioritisation of activities, the international and national profiling of research organisations, and selective decision making based on foresight. Successful implementation requires accurate assessment, high-quality foresight and a sufficient information base. Along the same lines, the STPC's latest STI policy review (2006) suggests that the promotion of innovation dynamics, the development of S&T policy tools, and improvements in the productivity and impact of the related measures all require support in terms of the enhancement of competencies related to evaluation and foresight. The role of the Academy and Tekes and the specific objectives of their partnership are to develop evaluation activities, methods and data and to investigate the impact of measures on the economy and society, in general, and the (structural) development of public RDI system, in particular (see Figure 4 and Table 2.3).

167. Besides being connected to the development of research and technology, evaluation and foresight are also linked to societal development in general. In particular, the impact assessment of policy measures and the evaluation of the standard of legislative drafting have been strongly emphasised, particularly in the field of sectoral research (*i.e.* research mostly done by government R&D institutes).

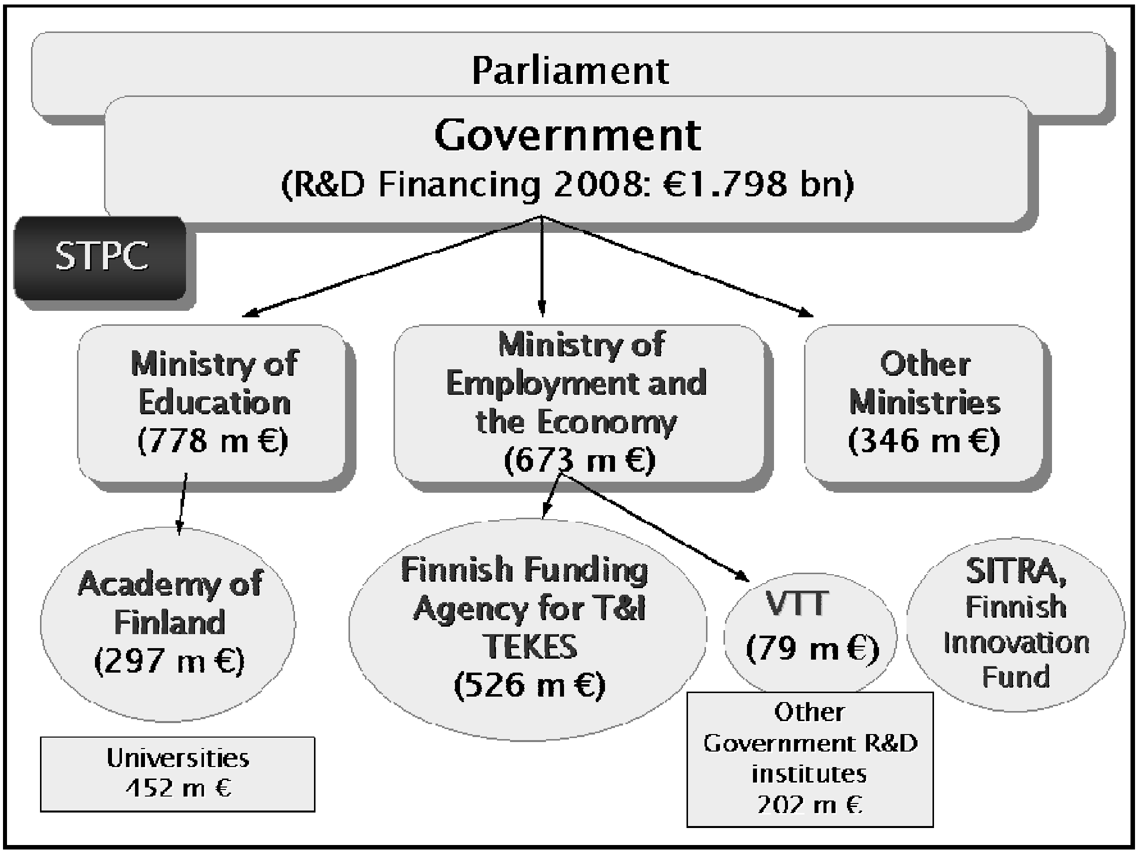
---

22. The Academy of Finland's mission is to finance high-quality scientific research, act as a science policy expert, and strengthen the position of science and research in the society. The Academy's operations cover the full spectrum of scientific disciplines. It supports PhD training and careers in research, internationalisation and the practical application of R&D results. In 2008, the Academy had a total budget of EUR 297million. It operates under the auspices of the Ministry of Education, which steers its operations by following the method of management by objectives and results.

23. Tekes, the Finnish Funding Agency for Technology and Innovation, is the largest public R&D financing and expert organisation in Finland. It finances industrial R&D projects but also projects in universities and R&D institutes. It promotes innovative, risk-intensive projects. In 2008, Tekes had a total budget of EUR 526 million. Together with the Academy, it accounts for some 46% of total government R&D investments. Tekes operates under the auspices of the Ministry of Employment and the Economy. The ministry steers the operations of Tekes through management by objectives and results and uses a set of measurement indicators in this regard.

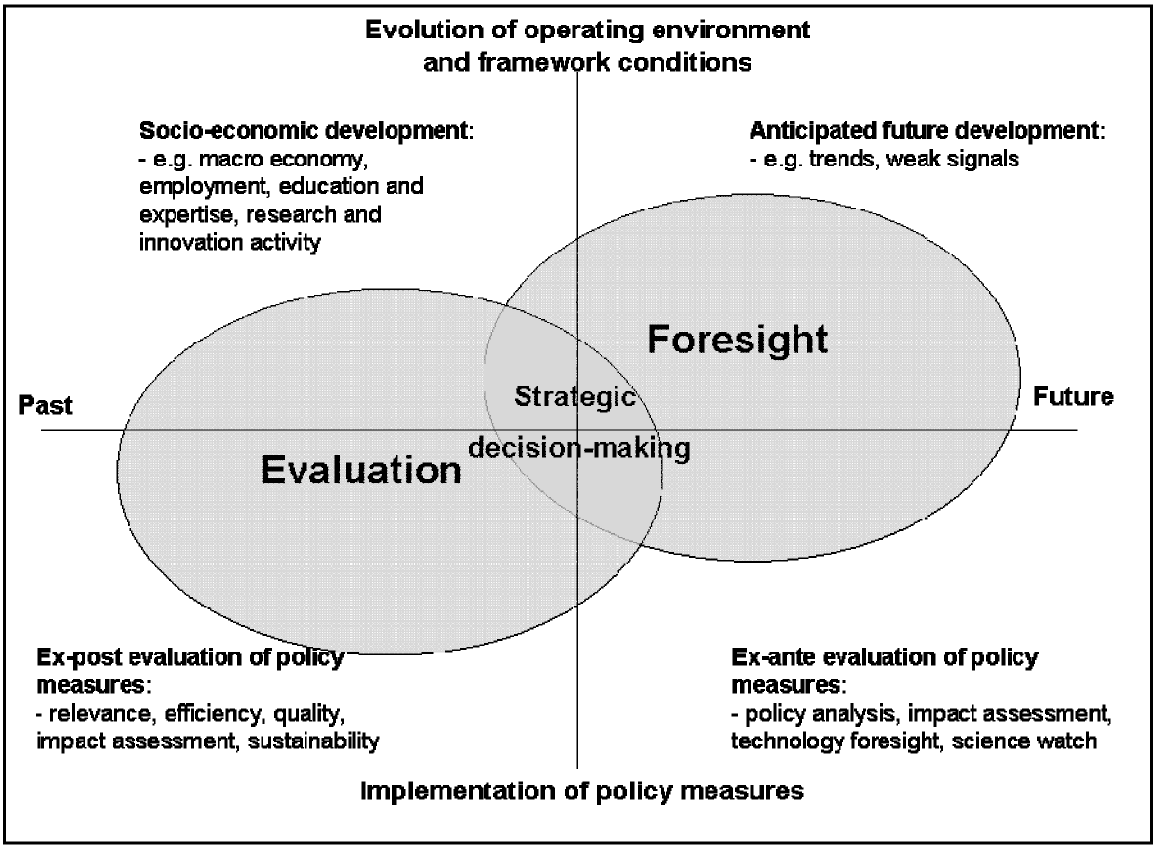
24. The Resolution on the Structural Development of the Public Research System of 7 April 2005. The resolution includes the general guidelines and basic principles for structural development measures. Based on the resolution, numerous policy measures (which combine funding and structures) has been initiated and implemented. The resolution emerged from an evaluation project launched in 2003 by the STPC. To a large extent, the resolution is related to a number of separate studies, which shed light on the structural issues of universities, ministries and public research institutes, as well as intermediaries operating at the interface between the public research system and the rest of society.

Figure 4. Structure of Finland's public RDI system



168. There is a fundamental link between evaluation and foresight: evaluation results are used for foresight purposes and *vice versa*. Accordingly, development measures should target both areas (Figure 5). Evaluation and foresight support the continuous development of research and innovation policy and strategic decision making. As such, they are instruments of common learning, understanding and exploitation. Therefore, it would be valuable to pay special attention to the successful organisation of evaluation and foresight and the division of responsibilities for the activities involved.

Figure 5. The relationship between evaluation and foresight and their relation to strategic decision making and the development of policy measures



Source: Based on Valovirta and Hjelt (2005).

Box 6. The general framework for evaluation in Finland

Finland has a long tradition of performance and impact evaluations and of assessments of policy instruments, measures and S&T organisations. This evaluation culture started to emerge in the latter half of the 1980s. It has remained strong and has even strengthened in the past decade. Today, efforts are being designed and implemented in order to fortify connections between the findings of evaluation reports and the development of policy operations. More attention is also being paid to the different needs of evaluation-based information in various stages of the policy cycle. The challenge that remains is to build a firm, interactive connection between evaluation and impact assessment activities and selective decision making.

Owing to the fragmentary and loosely co-ordinated development of evaluation within the innovation system over the years, there are no cross-sectoral or horizontal strategies or common evaluation practices evaluations in place today. It can also be argued that the role of evaluation in relation to other practices of strategic intelligence has not been sufficiently considered. The policy expert and funding agencies have their own evaluation objectives, practices and procedures. For example, Tekes has often been presented as an example of the Finnish evaluation model. It is indeed a good example, but it cannot be generalised to the whole Finnish policy system.

Responsibility for improving evaluation and foresight is shared by all key stakeholders of the national innovation system: producers of new knowledge and technology, financiers of research and innovation, and the actors benefiting from the results. A particular responsibility for initiating further development measures is carried by public expert and funding organisations (Academy of Finland, Tekes, Sitra) and the providers of innovation support services. This institutional set-up with divided responsibilities quite naturally emphasises the need to pay increasing attention to co-ordination and steering issues.

Since the early 1990s, all the stakeholders of the Finnish innovation system – ministries, funding organisations, universities, government R&D institutes, public and private intermediaries etc. – have been evaluated, some of them many times. This fact describes reasonably well the intensity and the depth of the evaluation culture and its evident link to policy making at the different levels of NIS. Evaluations have usually been linked with methods, procedures and objectives to:

a. Improve steering of RDI organisations by using the method called "management by objectives and results" or "results-based management and steering". This method is a key element of the NIS and has been effective in fortifying evaluation activities and culture without the need to establish a (centralised) evaluation system based on legislation or any other procedures or methods steered by law or strict regulations.
b. Increase the volume of public (competitive) R&D funding (channelled through Tekes and the Academy of Finland) and the impact of financing and other means of support.
c. Ensure that public funding is used effectively and efficiently.
d. Develop the entire NIS, not only individual organisations, fields of research, technology and research programmes, and procedures.

However, it should be emphasised that political and other decisions are almost never directly based on the results of evaluations and impact assessments. Instead, the results have often provided a basis for (public) debates on development needs and other issues. Evaluations are expected to give new perspectives and a positive push in an organisation's development or mode of operation.

There are two principal clients or commissioners of evaluations and impact assessments: the Ministry of Employment and the Economy and the Ministry of Education. Furthermore, as mentioned in the government decree on the STPC, the Council (chaired by Prime Minister and comprising seven other ministers relevant to STI policy and ten members well versed in STPC matters) was established to handle important matters concerning research, technology and their utilisation. The STPC also addresses issues relating to the development and exploitation of evaluation and impact analysis.

In addition, other ministries (*i.e.* those with R&D institutes of their own) have carried out evaluations as well. These evaluations have used rather different methods. To date, evaluations of R&D institutes have largely focused on their scientific impact and relatively little on their wider social impacts.

The two main actors and implementing organisations in evaluation and foresight are the Academy of Finland and Tekes. In addition, universities, polytechnics and government R&D institutes have carried out numerous self-assessment and external expert evaluations on their own initiative since the 1990s. The results of these evaluations and assessments are sometimes very intensively used to develop strategies and operations at ministries, RDI policy expert and funding organisations, and within the STPC. However, there is still a lot to do to improve and intensify the exploitation of evaluation findings.

Evaluation and impact assessment at the Academy of Finland focus on:

1) The research system:

a. Reviews of the state and quality of scientific research in Finland: an extensive review of the state and quality of science and research in Finland will be done once during the three-year term of its Research Councils. The next review (the fifth) will be completed in 2009. To date, they have mostly been done by Academy staff together with the Research Councils. The methods used have changed from one review to another (formula, methodology, data).
b. Evaluations of (the impact of) disciplines and fields of research: the aim is to provide information on the societal, technical and economic impacts of research in a given field. Evaluation criteria may include scientific visibility, researcher training, postdoctoral research careers, international recognition of researchers, societal impacts, *e.g.* contribution to popularisation of science. The impetus for the evaluation may come from the research community, the Research Councils, other science funding agencies or the authorities. Evaluations are done by external expert panels.

2) The Academy's own actions and processes

a. Research programmes: the evaluations consider whether the goals set for the programmes have been achieved, their success in generating new knowledge and the value added produced by the programmes. One of the main components of every research programme is societal impact. Evaluations are done by external expert groups.
b. Centre of excellence programmes: the main focus is on their scientific impact, and secondarily their societal impact. The evaluations of the CoEs 2000-05 and 2002-07 was scheduled to be completed at the end of 2008 by an external expert group.
c. Reviews of funding application: 4 000 a year.

3) The organisation of the Academy: an international expert evaluation of the Academy was carried out in 1992

and 2004.

The major features of evaluation and impact assessment at Tekes are as follows:

Input and output results (balanced score card) on the operational level (by Tekes itself).

22. Evaluations of individual instruments and programmes (commissioned by Tekes, usually done by international experts).
23. Evaluations, assessments and studies on outcome level (done by universities, R&D institutes or consultants independently or commissioned by Tekes).
24. National and international indicators and benchmarking on impact level (collecting data for indicators by Tekes).
25. System-level evaluations (commissioned by the STPC and ministries and often organised by Tekes).

3. Impacts and mechanisms of RDI

169. Impact is the last phase in a process that begins by organising the input factors for achieving a goal. In a research project, input factors include the necessary material and intangible resources as well as the manner of organising the work process. The further the distance in time and space from the immediate quantitative results of the activity (*e.g.* formed partnerships, use of resources, publications, patents, new models and algorithms, process and product prototypes), the more difficult the evaluation process. At the macro level, impact refers to a change that has resulted from the perspective of the goals set by society. While impact areas can be defined, examining the nature of specific factors and their role in the overall change is far more difficult.

170. The impact assessments of scientific research implemented by the Academy of Finland (2006) approach this problem by dividing "impact" into two parts: scientific and social impacts. This is justified for the evaluation of basic research. The actual time span from the original research results to the actual far-reaching social impacts may be decades and thus makes it difficult to identify and measure impacts. Impacts are the result of complex cause-and-effect chains over long periods of time. Furthermore, as concluded in the Academy's review of methods for evaluating the impact of basic research funding, there is no best method for this kind of impact assessment.

171. Research and its impact manifest themselves as the generation of new information and the growth of information capital. This is more widely reflected in the formation of world views, in social debate, and in greater understanding of various phenomena and issues. The output and impacts of research are mutually supportive: they partially overlap and interact. For instance, PhD training, the growing volume of knowledge, and the mobility of experts form, in all, a process that can be considered a result of research and a diffusion mechanism for R&D results and competencies, as well as an impact of research.

172. The impact of research depends on the degree of applicability of research results and the extent to which a society can exploit information and skills and transform them into innovations. A large volume of reports published recently in Finland indicate that research may have the following kinds of impact:

- ***Scientific**:* the accumulation and renewal of the knowledge base, including tacit knowledge.
- ***Technical:*** technological solutions; new products and processes; patents; new research equipment; methodological and technical skills.
- ***Economic:*** new (R&D- and technology-based) entrepreneurial activity; technological and social innovations that promote, for instance, production and foreign trade; the development of the competitiveness, productivity and efficiency of enterprises.
- ***Social:***

a) societal and cultural impacts, including education (lifelong learning, adult education and in-service training among others);

b) impacts associated with health and welfare, including safety;

c) impacts on working life, including new skilled workforce, employment;

d) impacts on nature and the environment;

e) political impacts, *e.g.* exploitation of research results and use of the expertise of researchers in decision making;

f) administrative and organisational impacts, *e.g.* introduction and application of new modes of operation and structures, regulatory environment.

173. The mechanisms for exploiting and distributing the results of research and innovation include:

- partnerships between companies, research organisations and the administration as well as measures promoting networking (research and technology programmes, common research centres, cluster projects);
- mobility of researchers and experts;
- new R&D- and technology-based enterprises, including academic entrepreneurship, spin-off companies and activities supporting these (intermediaries, technology centres and science parks, R&D and innovation agents, business incubators);
- transfer of R&D results and expertise from research organisations to those introducing, applying and commercialising these (including patenting, licensing, university companies, IPR transfers against share ownership);
- publications, technical reports and other documents;
- dissemination and application of tacit knowledge through channels including official, unofficial and personal contacts; education; meetings, workshops and conferences; expert services.

4. Education, research and innovation and the impact (of their financing)

4.1 Scientific and societal impact of (basic) research

174. The evaluation of scientific impacts and research standards are a more familiar subject than social impact and its assessment. An example is the Academy of Finland report (2006), which uses bibliometric methods to examine Finnish science and its international status. Given the method used, the results show that (internationally speaking) the quality of Finnish basic research and its visibility are relatively high. During the period 1995-2004, Finnish scientific publications received 13% more citations than those of other OECD countries on average.

175. Some critical comments were also heard during the ensuing debate on the report. According to professor Raimo Väyrynen, then President of the Academy, the basic questions can be summarised as follows*: i)* Is technological research capable of producing sufficient scientific results relative to the public resources allocated for the purpose? *ii)* Can the biosciences generate sufficient social and industrial benefits relative to the significant research resources allocated for the purpose? *iii)* How can the social sciences and humanities justify their social importance in comparison with the relatively modest scientific results they have achieved internationally?

176. However, bibliometric methods alone cannot describe even the impact of basic research. Other measurement methods and qualitative assessments are also needed. Furthermore, bibliometrics is not neutral in regard to fields of science. The Academy's Research Council for Culture and Society pondered this question in its impact report, *Knowledge creation cannot be imported* (2006) and made the following proposals:

- It should be emphasised that the most important impact exercised by researchers on society comes through the education of a new generation of experts.
- We should monitor the extent to which researchers influence the development of their own disciplines on a global level.
- It should be accepted that, in most cases, social impacts of research cannot be seen for years, and often decades.
- Quantitative data should be collected on the concrete effects of research on innovation.
- Researchers should be encouraged to consider the various impacts of their research and to report on them.

177. In relation to social impacts, the Research Council for Culture and Society considers the researchers to have applied two strategies in the debate:

> "On the one hand, they have pointed out that not all values can be measured in money terms. The aim is to try and contain the impacts of research within the field of science. Knowledge is considered valuable in and of itself; it is not thought to be necessary separately to identify any instrumental values. The thinking is that civilised society should support and sponsor science in the same way as it supports and sponsors art and other forms of culture. Science is thus protected from the practices and principles of business. The concepts of accountability, benefits, efficiency, performance targets, competition, measurement and impact are rejected by reference to civilisation and scientific autonomy. Dependence on the outside society is a problem, because science has to be able to justify the value of civilisation and autonomy in order to get the funding it needs. Nonetheless, it is accepted that at least within the scientific community, it is necessary to have debate about what is good and worthwhile and what has a positive impact. The other strategy is to accept the challenge presented by society. Questions of impact and even benefits are discussed and debated, but not without weighing and defining the rules of discussion and debate and interpreting the concepts. The utility and benefits of all research must be open to discussion so long as those concepts are understood in broad enough terms."

178. The Research Council points out that even a strategy with a favourable perspective on the impact debate faces the problem that even a broader concept of benefits (when referring to the economic impact) would be too narrowly defined.

179. In recent years, the most prominent evaluations in institutions of higher education have been the comprehensive research evaluations implemented by certain universities.[25] With respect to the exploitation

25. For instance, the University of Helsinki, on its own initiative, has carried out two very extensive research assessment exercises (1999 and 2005). The evaluation in 2005 combined an external assessment by 21 international evaluation panels (a total of 148, mainly European, experts) with an internal self-assessment exercise. The assessment results revealed strengths, weaknesses and opportunities in the university's research activities. The University will spend a total of EUR 15 million of its own funds over the next six years to reward the units that were most successful in the evaluation.

of these evaluations, the critical issue is whether they can serve as a basis for the necessary strategic decision making.

4.2 The impact of research funding by the Academy of Finland

180. The Academy invited an expert panel to investigate the impact of Academy research funding; their report was published in 2006 (Publications of the Academy of Finland 11/2006). As far as impact assessment is concerned, the key findings are as follows: in practice, concrete and detailed impact mechanisms can be described by means of separate innovation-specific and project-specific case studies. "It would be impossible to perform a comprehensive analysis taking account of the cause-and-effect relationships involved in the impact of the Academy of Finland or any other actor in the national innovation system." The report contains some sharp observations on the funding operations of the Academy and the instruments used. However, as regards the key question, the panel has had to be satisfied with somewhat general statements:

> "With respect to the Finnish innovation system as a whole, it is important that the Academy of Finland take the goals and needs of the entire innovation system into consideration when setting goals and developing its research funding system.
>
> "The best possible assessment of the effects and impacts of the Academy can be achieved by scrutinising the Academy's operations from different perspectives, combining both quantitative and qualitative research methods."

4.3 What kind of social impact is scientific research expected to have?

181. The joint FinnSight2015[26] project of the Academy of Finland and Tekes identified several development trends with many expectations relating to research. The findings can be summarised as:

- increasing productivity is crucial to Finnish competitiveness;
- global risk management will assume ever greater importance;
- energy and environment issues are of critical importance globally;
- increasing emphasis on basic human and social knowledge and skills will raise the quality of future-oriented development;
- know-how facilitating scientific and technological breakthroughs must be fostered.

182. The list mainly contains factors which, in relation to research, can be linked to the word *relevance*. However, the reports of the Academy's research councils strongly underscore the scientific quality and impact of research as a prerequisite for social impacts. For instance, the Research Council for Health which, on the one hand, points out that the long-term social effects of research in the medical sciences can be associated with changes in health policy, health care and health behaviour (and reflect upon the health and welfare of the population) is, on the other hand, of the opinion that free project funding is probably the most effective way of raising scientific quality and the impact of research. The Research Council for Biosciences and Environment, however, estimates that promoting genuine demand for knowledge is of crucial importance. On the basis of an inquiry implemented by the council, it draws the conclusion that the focus of the current debate, which considers the structures of knowledge production

26. FinnSight 2015 – Science and Technology in Finland in the 2010s – was a joint foresight project of the Academy and Tekes carried out in 2005–06. The actual work was done in ten panels of ten experts. The panels included representatives from public- and private-sector organisations.

and know-how as the problem, is wrong. In reality, structures on the demand side for knowledge would seem to constitute the most significant structural problems.

183. A questionnaire addressed to the directors of environmental research projects financed by the Research Council for Biosciences and Environment asked about obstacles to the practical application of R&D results. In order of importance, these were: *i)* users do not understand the opportunities for application of the results (49% of answers); *ii)* lack of suitable actors to exploit the results; *iii)* legislative and political obstacles; *iv)* no opportunities to apply the results in sight; *v)* the development of applications would not be cost-effective; and *vi)* lack of willingness to use the results (5%).

184. The report of the Research Council for Natural Sciences and Engineering lists the channels through which technology generated in the course of research can be transferred for business use. In order of importance these are: *i)* personal contacts; *ii)* research co-operation; *iii)* dissertations; *iv)* scientific publications; *v)* conferences, seminars and lecture events; *vi)* mobility; *vii)* technical reports, documents; *viii)* consultation and expert services; *ix)* licensing, sale of patents; and *x)* businesses specialising in the transfer of technology. The council also points out that the EUR 1 million of funding provided by the Academy generates, on average, 31 peer-reviewed articles in international journals, 4.0 doctoral degrees, 1.9 licentiate's degrees and 4.8 master's degrees.

185. These results in addition to the discussions above provide a reason for seeking improved methodological knowledge and better comprehension of significant relationships between basic research and its scientific and social impact. This issue has also been raised by the Nordic research funding bodies. Studies conducted to date only show that the issue should be approached from various perspectives, applying both qualitative and quantitative analytical methods. Tekes and the Academy have therefore recently launched development projects aimed at establishing a databank on the impacts of RDI (including systematic collection of information), creating a fixed set of impact indicators, and introducing more detailed procedures for carrying out evaluation and impact assessments. Furthermore, in the final reports on projects funded by the Academy, concrete attention is now paid to collecting data on the impacts of research.

4.4 R&D and innovation and the impact of funding: Tekes' point of view

186. The assessment of Tekes' impacts and the generation of a database for this purpose is a continuous process. Impact assessment is implemented by means of project follow-up evaluations and programme and operation evaluations. For projects, impact information is collected in advance (output expectations), during the final project stages (what the project produced as compared to the launch situation), and three years after the project has ended, at which point a more comprehensive picture of the results is possible. Various parameters form a basis for an overall view of the impacts.

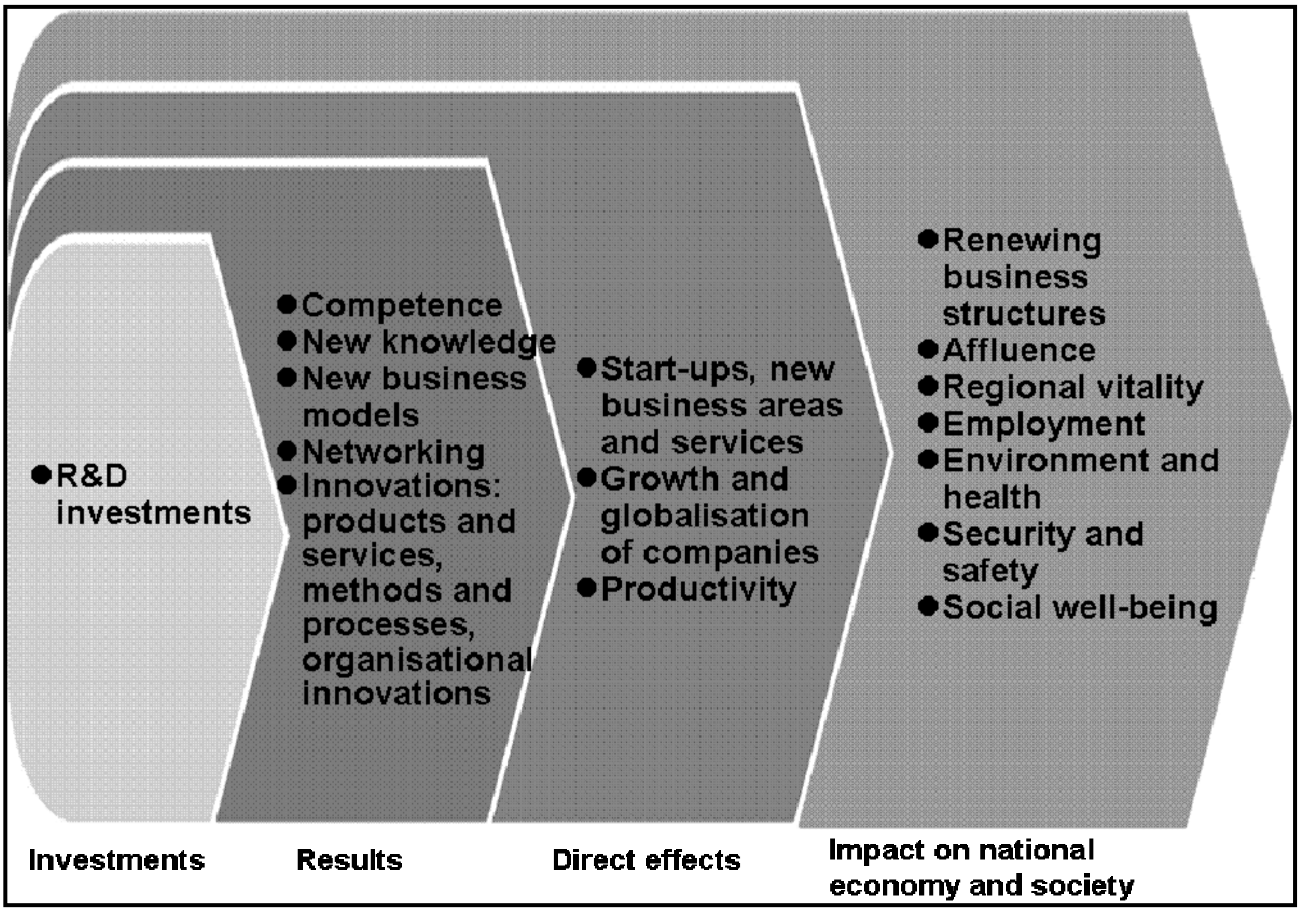
187. The model Tekes uses for impact analysis is a description of various effects of RDI projects based on cause-and-effect relationships. This model is based on the notion that R&D and the development of innovations generate results with immediate effect which, for their part, affect the national economy and society in a wider sense. *The four-phase model* (Figure 2.5) divides funding and results into separate phases, which together form a general view of the impacts of research and innovation activities. The approaches used in the impact model include the concepts of additionality (input, behavioural, output), knowledge spillovers and new growth theory.

188. The first phase in the model is RDI investments. Tekes gives its customers incentives (through funding) to increase their investment in RDI. The first level is in this sense is thus also an *input* additionality level. The second phase describes the results, *i.e. behavioural* changes which take place within firms, universities and research institutes as a result of public RDI funding and other activities of

Tekes. The third phase concerns innovative firms and renewal of business, *i.e.* whether public RDI funding increases company *outputs*, and if this in turn improves their overall performance. This level also involves industrial dynamics, such as the utilisation of information technology to enhance productivity and the efficiency of public services. The fourth and final phase analyses the impacts of RDI financing on the economy and society as a whole.

189. Tekes' recent evaluations (see references) show that R&D and innovation activities as well as their public funding are of great importance to the national economy and welfare. Public support significantly increases the competence and networking of enterprises, generating patents, publications, products, processes and services. Know-how and new modes of action are manifested in the form of companies' increased turnover and productivity and a higher employment rate.

Figure 6. Tekes' impact model: four-phase approach to effects and impacts



Source: *Tekes Technology Review* (2007) 203: 3.

190. Tekes' funding is targeted at the R&D of businesses and public research organisations. Its goal is to encourage customers to engage in more R&D than they otherwise would (*input additionality).* According to recent studies, EUR 1 of Tekes funding increases private R&D investment by EUR 0.4-1.0, for a total increase of EUR 1.4–2.0. Public support increases companies' R&D and innovation activity both nominally and in real terms. Thus, Tekes' funding has a very positive *impact on the company's own R&D investment*. This has been one of the key justifications for increasing public funding for RDI, as it does not seem to crowd out or replace private investments. Results have been similar elsewhere. Input additionality has been EUR 1.4–1.9, depending on the country and period in question.

191. The *impact* of Tekes on *output, commercial results and productivity:* its funding seems to boost growth, patenting activity and demand for labour in companies, and reduce the probability of business closures and mergers. Tekes' support has enhanced productivity growth and improved the employment rate in companies with R&D staff. The impact on productivity is highest in SMEs and companies operating close to the productivity frontier. The added value of public support is to facilitate implementation of projects and influence how they are executed. Funding has also increased co-operation and interaction between key stakeholders of the innovation system.

192. In addition to the financial impact, companies receiving funding have also often changed their modes of operation and their behaviour (*behavioural additionality)* in a manner that has strengthened the importance of education and research and the capability to exploit new information and competence. The *impact* of Tekes' support *on modes of operation and competence* (Table 4) include the following: *i)* public funding has a positive effect on the innovation activities of companies; *ii)* funding has encouraged companies to build closer partnerships and has promoted the establishment of co-operation networks; *iii)* funding has spurred industry into engaging in co-operation with several (service) companies of various kinds; *iv)* funding has raised the risk-taking level of projects and productivity, raised the level of competence in companies and impacted on long-term business strategies. The importance of public support is emphasised by the development of innovations with more challenging competence requirements. Approximately 60% of significant innovations developed in Finland in 1985–2000 received Tekes funding. In Tekes' technology programmes, the product development projects of small companies emphasise the importance of co-operation. In particular, low-technology sectors seem to benefit from the programmes. This underscores the importance of policies that support networking, even in such fields.

Table 4. The impacts of funding on companies receiving support from Tekes1

	Agree	Disagree
Enhanced the **know-how of personnel**	90%	10%
Enabled projects with **more risk and return**	86%	14%
Improved the **credibility** of the company	84%	16%
Helped the company improve sales	84%	16%
Helped the company improve sales	77%	23%
Helped build **partnership networks**	68%	32%
Improved the **profitability** of the company	64%	36%
Also aided other projects by improving the quality of **equipment**	63%	37%
Helped **internationalisation** of the company	**60%**	40%
Over time had impacts on **business strategy**	**56%**	44%
Increased the **publicity** of the company	**55%**	45%
Helped to find **potential investors**	14%	86%
Influenced the **location** of R&D facilities	10%	90%
Influenced the **location** of production facilities	7%	93%

1. Figures based on a 2003 Tekes survey on behavioural additionality, including company interviews (survey replies from 200 companies, of which 80 are Tekes customers). 'Cannot say' option and blanks are not included in these figures.

193. Programme activities have incontestably had positive effects. Public co-financed programmes have played a particularly important role in the creation of common strategies and visions. They have enhanced R&D activities in the following cases: *i)* the activities are related to a new area of research; *ii)* the research is justified by benefits to the national economy and business; *iii)* the theme is important from the viewpoint of political decision making; *iv)* the research is related to problems that are interesting from the viewpoint of enterprises; *v)* the combination of the information generated by different researcher communities supported produces something new and significant. Research usually brings added value in situations where there is a genuine will to merge the objectives of various organisations. According to studies, increasing the (public) resources of a single organisation in the name of co-operation but without seeking more extensive common goals and consolidation of the objectives of different partner organisations does not generate added value to the same extent.

194. The more comprehensive social effects of Tekes funding are relayed through various channels, as part of the dissemination of expertise and technology. For instance, in every second case, Tekes projects have also had a positive impact on the turnover of enterprises that are not direct Tekes customers. They have had a favourable impact on productivity and number of jobs in other enterprises in every fourth project. The impact of programmes is greater than that of separate projects, because information and know-how are transferred more effectively through programme activities and services.

4.5 The economic and social impact of RDI

195. The broader connections between R&D, productivity and economic growth have been analysed by VATT and ETLA, and partly financed by Tekes. They have studied the impact of R&D funding in companies and extended the scope of the impact analyses to cover various sectors and the national economy. In terms of average productivity and gross national product per capita, Finland is close to the OECD average. In comparison to the EU states, Finland's income level and average productivity are above average.

196. In addition to technological development, increased productivity requires organisational, administrative and structural reforms which also enhance cost-effectiveness and economies of scale. Productivity increases when companies introduce better technology and organise work more efficiently. In the long run, innovations – new products, production methods and services – are a key source of productivity and technological development. As far as the development of technology is concerned, the key observation is that the decisive factor is the use not of the newest but the most productive technology. Results also indicate that policies prioritising the generation of new technologies do not always have an optimal outcome. There is also the possibility of introducing technologies created elsewhere and putting them to as extensive a use as possible.

197. In Finland, productivity is among the highest in the world in key industrial export-based sectors. However, to strengthen productivity at the level of the national economy, the productivity of traditionally low-productivity sectors also needs to be raised. Consequently, productivity in these sectors has risen more rapidly in Finland than in other countries, even in the services sector. To promote productivity growth beyond the high performers, special attention must be paid to the comprehensive improvement of competence, the development of technology and conditions for their exploitation.

198. Recent reviews emphasise the simultaneity of the different phases of the innovation process and the fact that enterprises no longer perform all of their R&D themselves but outsource some of it. Thus, *open innovation* is gaining more ground. Another phenomenon often grouped with open innovation and which has become more and more widely recognised in technology and innovation policy is *user innovations.* Here a key part of the process involves end users improving on products and inventing new uses for them. In this way, the know-how of end users becomes part of the development process and the resulting innovations may become more widespread within society. End-user innovations draw attention to one of the key reasons for the failure of certain innovations which do not take the needs of consumers and contexts sufficiently into account. Integrating this aspect of R&D activity with policy measures is a challenge which requires a new kind of thinking and interaction.

199. Around ten out of more than 300 companies studied as part of the Innovative Growth Companies project were particularly rapidly growing "gazelles" (growth of turnover over 30% during each accounting period in 2001–04). These companies are of major financial importance. In spite of their small number, they were capable of compensating, almost by themselves, for the reduced turnover in the other company groups. Policy measures are particularly necessary for identifying and encouraging growth companies, a need which receives further emphasis from the fact that, according to studies, new companies apply new technologies intensively and their productivity grows faster than that of old ones. Information and

communications technology increases productivity, and it has been observed that the impact is greatest in new companies. In other words, the establishment of new, growing companies is important from the point of view both of the national economy and employment.

200. In research-based enterprises, innovations are based on the latest R&D from universities and research institutes. Typically, both the companies and the innovations take a long time to become profitable. A report commissioned by Sitra on the commercialisation of university research considers it a key problem that, even today, Finland does not have a clear policy on the role universities should play in the commercialisation process.[27] The resources and expertise of R&D service units have increased, but the challenge of clarifying their role in commercialisation activities and developing their management practices remains. Business operations are difficult to steer without clearly specified objectives. In addition, the evaluation of such activity is difficult. Even in this respect, much remains to be done in following up the exploitation of R&D.

201. According to studies, the practical implementation and commercialisation of R&D results are not as successful as they could be. In many reports, two factors emerge as obstacles to exploitation. The first is related to the initial phases of the process: domestic research produces promising early-stage business ideas, but there are problems identifying them and moving them ahead. The second factor is the lack of financing at the stage at which capital would be needed for long-term R&D and in situations in the later stages of product development, when the process is on the verge of production. Accordingly, it would be important to use actual needs as the foundation for the ongoing development of public innovation services. Even the current service structure fails to support the exploitation of the results of R&D and university-company co-operation in the best way.

202. The differences in the innovation activities of various sectors are significant. In certain sectors, new innovations are created every few months, whereas in others product cycles are significantly longer. Since service innovations have not been widely studied, their characteristics, impact and operative mechanisms remain largely unknown. For instance, R&D in the services sector and its impact are limited. Innovations in information-intensive sectors are often services containing a great deal of silent knowledge. Service concepts, distribution channels and organisational models are typical forms of service innovations which contain a great deal of such knowledge. The public sector, on the other hand, has not traditionally been considered innovative, but for social innovations, its image is quite different. Public-sector innovations often have a legislative background or are dictated by social needs and developed in collaboration with the private sector. All in all, the measurement and assessment of the impact and productivity of services is difficult. It would be useful to develop comprehensive, reliable methods and improve the accumulation of data for that purpose.

203. In addition to investments in material and R&D, innovations – and their creation and successful commercialisation – depend on immaterial investments and capital related to a competent workforce, patents, customer relationships, organisational structures, business models, design, digitised data and software. Since the impact of immaterial investments is very difficult to measure, their social and

27. Finnish universities produced, on average, 40 reported inventions per EUR 100 million invested in R&D. The corresponding figure in the United Kingdom was 28 and in the United States 87. Finnish universities were granted, on average, seven patents per EUR 100 million used for R&D; the corresponding figures in the United Kingdom and the United States were 15 and 11, respectively. In 2005, licences produced an average income of EUR 250 000 per EUR 100 million spent on R&D for universities. In the United Kingdom, the figure was EUR 1.6 million and in the United States EUR 7 million (in 2004). Finnish universities generated five spin-off companies per EUR 100 million invested in R&D, whereas the UK figure was six and the US figure only one. Taking into consideration all R&D-based companies established in Finland in 2000-05 (171 in all), the turnover of only one exceeded EUR 2 million in 2005. The growth of such companies is relatively slow.

economic importance is largely unexplored. The challenge of enhancing the quality of research and innovation and its impact is connected with the development and exploitation of competencies within these fields.

4.6 Ministries and research institutes

204. Increasing attention has also been paid to the evaluation of the effects and impact of research institutes. One example is the JYVA project (Assessment of Social Impact in Public Research Organisations), led by Technical Research Centre of Finland (VTT) and completed in 2006, which examined such issues in five organisations, the MTT Agrifood Research Finland, the Finnish Defence Forces Technical Research Centre, VTT and two polytechnics, Helia and Satakunta University of Applied Sciences. As a continuation of this study, four ministries [Ministry of Finance, Ministry of Agriculture and Forestry, Ministry of Trade and Industry MTI (currently Ministry of Employment and the Economy MEE) and Ministry of Social Affairs and Health] launched a joint project entitled VALO (Impact as Part of Results-based Management and Steering of Research Organisations) and implemented it in 2006-08. Among other issues, the project will investigate how the social impact of research organisations might be assessed systematically. This will lay foundations for the development of the impact assessment of sectoral research to a level corresponding to the activities of financier organisations.

4.7 Impact of education

205. The importance of a high level of education for the quality and impact of research and innovation as well as the favourable development of individuals and society as a whole is widely acknowledged. However, relatively few detailed analytical studies have been conducted on the role and impact of education, particularly with respect to research and innovation activity. VATT's research paper, "Finland 2025" (2005), points out that the rise in the education level of the population of active working age will continue almost automatically and the share of the highly educated will increase over the next 25 years. However, it is not evident that this will in itself generate growth. This serves as a reminder that productivity is among the key means of securing economic growth.

206. Nevertheless, in its macro level analysis, Sitra's *Economic Impact of Education* (2006) concludes that investments in education undeniably have a favourable effect on productivity. Still, the information on the primary mechanisms through which investments in education accelerate growth is insufficient. In any case, such investments have a positive effect on the labour market success of individual citizens, particularly their future income. From an individual citizen's point of view, in Finland a longer education is still a good investment. However, the salary benefits of those with an intermediate level of education over those who settle for a basic level of education appear to be negligible in terms both of real differences in salaries and of productivity rates derived from education.

5. Foresight, its development and exploitation

207. Foresight brings new information and insight to the creation of the research, technology, business and competencies which are of great importance to society (Figure 2.4). Finland's position at the top of international comparisons requires that, instead of following outside examples, it must be capable of developing future prospects independently and implementing experiments which open up new opportunities. Foresight brings different kinds of actors together and creates a common knowledge base and insight into the future and its needs. Furthermore, it supports strategy work and selective decision making (*e.g.* the need to strengthen and target funding and research; policy contents). The use of foresight when deepening co-operation or launching new projects has often been very fruitful. The goal is to develop practices and enhance competencies and networking by means of common foresight.

208. In recent years, foresight activities have increased significantly in Finland. Conscious efforts have been made to reduce the incoherence and lack of a systematic outlook in foresight activities. Key projects include the Government Foresight Network, MTI's Foresight Forum and Sitra's National Foresight Network. The Committee for the Future of the Finnish Parliament also partially operates in this arena. A special Technology Futures Forum has been launched under the auspices of the Technical Research Centre of Finland (VTT), and several other organisations are implementing their own foresight projects.

209. In the summer of 2006, *FinnSight 2015*, a joint foresight project of the Academy of Finland and Tekes, was completed. By convening ten expert panels, the project explored the future prospects of science, technology and society. The themes of the panels were: learning and renewing society through learning; services and service innovations; welfare and health; environment and energy; infrastructure and security; bio-competence and bio-society; information and communications; understanding and human interaction; materials; and the global economy. The panels identified over 80 areas of expertise, which could help Finland achieve breakthroughs in S&T and develop new innovations. In other words, the exercise produced a very extensive list of areas of expertise and specialisation. While this does not make prioritisation, choices and decision making easier, it clearly shows the diversity of alternative lines of development and a situation in which the development of the economy and society requires varied and broad-based skills.

210. Accordingly, efforts are being made to promote foresight, as presented in Figure 2.4, with the help of studies of development trends (*e.g.* population trends and the related migration patterns) and so-called weak signals. An example of the latter is the foresight project Signals implemented by Tekes as part of its strategy processes and the development of technology programmes in autumn 2006. Co-operation partners included MTI, Finnvera, Finpro, the Foundation for Finnish Inventions, Sitra, the Academy of Finland, the T&E centres and VTT. The goal was to identify so-called weak signals, identify promising fields and new opportunities for Finland, and recognise threats. More than 1 200 experts representing the companies and communities involved in research and innovation participated in the project. The respondents agreed almost unanimously on Finland's success factors: technology and its exploitation, high quality and comprehensive skills, and generation of innovations. For instance, the export of methods and techniques and environment-friendly technologies could open up increasing opportunities. International competition and co-operation were considered to be closely related. The responses emphasised the need for dynamism and courage.

211. The participants considered the exploitation of Finland's special status and competitive edge, its investments in business expertise, and the promotion of entrepreneurship to be important. Attention should also be paid to the creation of infrastructures and the reduction of bureaucracy. Factors viewed as threats included the sustainability of the economy; demographic factors; developments in neighbouring regions; the purchase of research-based innovations by foreigners; climate change; and the division of society into haves and have-nots. Otherwise the exercise produced wide-ranging results, leaving room for different interpretations. No synthesis report was produced.

212. Tekes is continuing its foresight activities as part of an ongoing international benchmarking activity. This will map out Finland's opportunities and threats with the help of foreign experts. Alongside other Finnish funding and expert organisations, Tekes has a host of international connections and co-operation partnerships in the area of foresight. For instance, Tekes and the Academy of Finland have a co-operation agreement concerning technology foresight and evaluation with NISTEP, which represents Japan's science and technology policy. Sitra, VTT and Helsinki University of Technology TKK have taken advantage of the co-operation opportunities offered by the agreement.

213. The Government Foresight Network (first term 2004-07; the second term started in 2008) has been a co-operation platform consisting of representatives of Finnish ministries. It has co-ordinated foresight by ministries and acted as a discussion forum. Its aim has been to develop foresight competence and foresight at the regional level and promote the exploitation of foresight results in decision making. It has been recognised that, in this field, development work should continue.

214. The actual foresight work of the Foresight Network was carried out in the ministries. In 2005, the network compiled an extensive report on the foresight information produced by various administrative sectors, which served as a basis for drafting the report of the ministries. The final report of the network considered the current status of foresight and its development needs. In all, foresight practices, competencies and the connection with planning and strategy processes vary greatly from one ministry to another. Foresight has been implemented separately by each ministry, taking the ministry's own strategic needs as the starting point. A foresight process that brings all administrative sectors together is necessary for the government's future report compilation process and, in this sense, the first foresight network fulfilled its task satisfactorily. In the future, closer commitment to networking by participants, sufficient resources for implementing common projects and, possibly, a separate support organisation/unit for co-ordinating and leading joint efforts will be needed. Accordingly, the government will probably set up a new foresight network to co-ordinate foresight in policy measures.

215. The purpose of Sitra's National Foresight Network has been to enhance interaction between experts and decision makers, and to promote the exploitation of foresight. The network has brought together key foresight actors and decision makers who use foresight in their organisations' strategy processes. Social challenges and opportunities have been identified as topics for consideration by decision makers and for public debate and as the targets of further studies and innovation. ***The network surveys new trends and signals, and organises workshops for those exploiting foresight information. I***n 2007, for instance, the focus was on the exploitation of the results of recently implemented foresight projects. This was considered necessary, since insufficient attention has been paid to this. The same applies to the exploitation of evaluation results.

6. Conclusions and the rationale for development measures

216. To date, evaluations of the direct impact of public financing of research and innovation on the operation of companies and research organisations have been mainly positive. They have also of course identified shortcomings and development needs. [28]

217. Although there are positive evaluations of the social impacts of RDI funding organisations, on the one hand, and of research and innovation activities, on the other, many reports fail to provide an organised and sufficiently detailed picture of their impacts. The reasons are well known: the multi-dimensional and indirect nature of the impacts; complex interaction with other factors in the environment; different exploitation mechanisms and dissemination channels; and delays in the emergence of the impacts. These factors have not yet been fully grasped, not because of the nature of domestic funding and research organisations but because impact assessment is difficult everywhere. Furthermore, it is particularly challenging to generalise the findings of detailed case studies.

218. Evaluation and its development is further impeded by the one-time nature of the activity. Remedying the lack of resources and shortcomings in the competence and information base require concrete financial investments and systematic, long-term development. The strengthening and

28. On the basis of the material presented here, the STPC raised many of these issues in drawing its conclusions on the need to develop evaluation, impact assessment and foresight in Finland.

systematisation of the information base is also a prerequisite for the development of evaluation, impact assessment and foresight in policy measures, and legislative drafting and legislation.

219. Tekes has made major investments in impact assessment and, recently, in the evaluation of social impacts on its business customers. For research organisations, the challenges relating to impact assessment still exist. In addition to the direct financial impact on business performance, enterprises receiving funding have often changed their modes of operation and their behaviour in a manner that has strengthened the importance of education and research and the capability to exploit new information and skills. Alongside the direct impact, the indirect impact (as information, know-how and expertise disseminate more extensively into the environment) raises the estimate of the overall impact of financing. In order to gain a better understanding of the indirect impact, the direct impact of support instruments and incentives should be studied in greater detail, along with the question of the added value generated and whether they produce the desired outcomes.

220. The evaluations conducted by the Academy indicate that the difficulty of assessing the impact of (basic) research includes the fact that the impact is indirect, complex and unpredictable and takes a long time to become apparent. In addition, the differences between the various fields of science are great in terms of the nature of research and the form and impact of the results and their dissemination mechanisms. In the future, more attention should be paid to the following areas:

- When should the assessment of the impact of (basic) research be made?
- What is the significance of the research itself in relation to the observed impact?
- Which parties benefit from funded research?
- By what mechanisms and processes is the impact generated?

221. In a more comprehensive perspective, the challenges associated with impact assessment and foresight activities include the following: *i)* the relatively limited exploitation of the results of impact assessment and foresight; ii) the lack of clarity in responsibilities for the development of such activities and the division of labour; *iii)* insufficient and fragmented competence. Shortcomings can also be observed in the development of methodology and the generation and strengthening of the information base. In many sectors, impact assessment and foresight are still new concepts and no established methods exist. It would also be worthwhile paying attention to the means by which evaluation and foresight can be more closely integrated with other development measures and strategic decision making. In addition, hitherto, evaluations have chiefly been implemented separately within each organisation or (policy) sector. Thus, the focus has been on the evaluation of a single actor, programme, project, or field of science. Furthermore, impact assessment is often implemented too soon after completion of the activity. The need for increased horizontality is evident. Internationalisation also emerges as a clear challenge.

222. Foresight has advanced from the earlier focus on technology foresight towards a more comprehensive, social perspective which places the anticipated development of technologies and research in a wider context. However, expertise in foresight is somewhat limited in Finland, including among those commissioning the studies. Efforts have been made to improve this situation through network-based co-operation.

223. Collective, broad-based foresight at the national level should be strengthened. Foresight activities remain fragmented. Various organisations exercise plenty of foresight, which leads to overlapping actions (*e.g.* the Employment and Economic Development Centres and regional councils). In fact, the volume of foresight actions at national level is comparatively small. Projects implemented until now have not usually been based on broader national starting points or strategic needs.

224. Foresight is receiving ever more emphasis for many reasons, particularly related to horizontal innovation policy and its development. In the future, the ethical, social and societal aspects of the results and impact of research and innovation will be emphasised in a manner that gives a crucial role to foresight.

225. In the area of STI policy, in terms of evaluation and foresight in general and impact assessments in particular, the following fields emerge as broader development areas:

- supply of intellectual resources; the renewal of the knowledge base;
- improving the conditions for the growth of productivity and GNP;
- internationalisation (associated with *e.g.* research, mobility and business activities);
- quality and productivity of services;
- entrepreneurship, establishment and growth of new companies (including availability of capital and risk funding; development of innovation services; FDIs);
- commercialisation of R&D results and their transfer from research organisations into practical use.

226. Based on the RDI evaluations discussed here, it has become apparent that more resources should be directed to comprehensive measures designed to improve innovation dynamics, develop the assessment of effectiveness and productivity of public services, and help to include the user perspective in policy measures. In this context, it is also essential to better link impact assessment and foresight with decision making on STI policy issues.

227. In relation to evaluation and foresight, the STPC has pointed out the increasing need to remedy methodological shortcomings, to bring together fragmented skills scattered throughout Finland, to increase training and to strengthen the knowledge base. It is also necessary to open up the access to, and enhance the application of, publicly funded data, statistics and registers. This challenge relates to micro-level data, in particular. The ministries should earmark sufficient resources and expand opportunities for joint evaluation and foresight efforts across administrative boundaries. Furthermore, there should be more multilateral international projects on impact assessment and foresight. Owing to the growing needs of decision makers and policy planners for more, and more accurate, information on the functioning of the Finnish RDI environment and the efficiency of the current STI policy measures, the impact assessment of the entire innovation system will be carried out by external experts in the near future.

REFERENCES

Tekes reports and reviews on the themes of R&D, innovation and impact assessment:

Autio, E., K. Miikkulainen and I. Sihvola (2007), "Innovatiiviset kasvuyritykset", *Tekes Technology Review* 201/2007.

Berghäll, E., T. Junka and J. Kiander (2006), "T&K, tuottavuus ja taloudellinen kasvu", Government Institute for Economic Research VATT, Research Reports 121.

Hjelt, M., M.-L. Niinikoski, M. Syrjänen, V. Valovirta and T. Törmälä (2006). "Julkisten tutkimustulosten kaupallinen hyödyntäminen", *Tekes Technology Review* 192/2006.

Hyvärinen, J, and A.-M. Rautiainen (eds.) (2006), "Innovaatiotoiminnalla kilpailukykyä ja kasvua: tutkimus- ja kehitystoiminnan vaikuttavuus yhteiskunnassa," *Tekes Technology Review* 188/2006.

Innovaatiotoiminta luo menestystä ja kasvua (2007), Tekes Report. Tekes, Helsinki.

Kanninen, S., P. Kutinlahti, T. Luukkonen, J. Oksanen and T. Lemola (2006). "Finnish national evaluation of EUREKA and COST," *Tekes Technology Review* 13/2006.

Rajahonka, M., S. Svahn, M. Tinnilä and M. Valtakari (2006), "Kohti verkostomaista liiketoimintaa. (2000-luvulla päättyneiden teknologiaohjelmien onnistuneisuus eritoten verkostomaisen liiketoiminnan kehittämisessä)," *Tekes Programme Report* 15/2006.

Rilla, N. and J. Saarinen (eds.) (2007), "Tutkimusmatka innovaatioihin," *Tekes Technology Review* 197/2007.

Rouvinen, P. (2007), "Yrityksen tutkimus- ja kehitystoiminnan ulkoisvaikutukset," *Tekes Technology Review* 209/2007.

Syrjänen, M., A. Pathan, S. Ahvenharju and M. Hjelt (2006), "Yhteistyöllä tietoa ja verkostoja. Innovaatiorahoittajien yhteisohjelmien arviointi," *Tekes Programme Report* 20/2006.

"Tekesin vaikuttavuus. Asiakaskyselyjen ja teknologiaohjelmien arviointien analysointi" (2007), *Tekes Technology Review* 203/2007.

"Tilastoanalyysi Tekesin vaikuttavuudesta" (2008), *Tekes Technology Review* 229/2008.

The Academy of Finland's SIGHT 2006 project produced evaluations and assessments of the state, quality and impact of Finnish scientific research. Seven publications were prepared. Unfortunately, a synthesis report was not made. The publications form a collection of research results, data and views concerning research done in Finland. The reports were as follows:

Suomen Akatemian rahoittama luonnontieteiden ja tekniikan alojen tutkimus: Arviointi hankkeiden vaikuttavuuksista (2006), Publications of the Academy of Finland 6/2006.

Tutkimuksen vaikuttavuus biotieteiden ja ympäristön tutkimuksen aloilla (2006), Publications of the Academy of Finland 7/2006.

Strategisella rahoituksella vaikuttavampaa tutkimusta? Kolme esimerkkiä vaikutusten ja vaikuttavuuden arvioinnista terveyden tutkimuksen alalta (2006), Publications of the Academy of Finland 8/2006.

Kanninen, S. and T. Lemola (2006), *Methods for Evaluating the Impact of Basic Research Funding: an Analysis of Recent International Evaluation Activity*, Publications of the Academy of Finland 9/2006.

Suomen Akatemian tutkimusrahoituksen vaikuttavuus: arviointiraportti (2006), Publications of the Academy of Finland 11/2006.

Lehvo, A. and A. Nuutinen (2006), *Finnish science in international comparison*, Publications of the Academy of Finland 15/2006.

Civilisation cannot be imported – Researcher commentary on the impact of cultural and social research (2007), Publications of the Academy of Finland 3/2007.

Other relevant publications:

Ali-Yrkkö, J. and C. Palmberg (eds.) (2006), “Suomen kilpailukyky t&k-toiminnan sijaintipaikkana,” *ETLA B* 218.

Asplund, R. and M. Maliranta (2006), “Koulutuksen taloudelliset vaikutukset,” *Sitra Reports* 60.

Berghäll, E. (2006), “R&D and productivity growth in Finnish ICT manufacturing,” Government Institute for Economic Research VATT, Discussion Papers 388.

Bibliometristen aineistojen käytettävyys yliopistojen julkaisujen laadun ja tuottavuuden arvioinnissa, Publications of the Ministry of Education 2007:2.

Ebersberger, B. (2005), “The impact of public R&D funding,” *VTT Publications* 588.

Finland’s response to the challenge of globalisation. Report by the secretariat of the Economic Council, Part 2 (2006), Prime Minister’s Office Publications 19/2006.

Globalisation challenges for Europe. Report by the Secretariat of the Economic Council, Part 1 (2006), Prime Minister’s Office Publications 18/2006. (The report comprises 13 articles by European top experts.)

Himanen, P. (2007), *Suomalainen unelma. Innovaatioraportti.* Teknologiateollisuuden 100-vuotissäätiö, Helsinki.

“Impact evaluation of research – from theory to practice” (12.10.2005), Material of the workshop (jointly organised by the Academy of Finland and Tekes) held in Helsinki on 12 October 2005, at *www.aka.fi/*.

Kankaala, K., P. Kutinlahti and T. Törmälä (2007), “Tutkimustulosten kaupallinen hyödyntäminen – kvantitatiivisia tuloksia,” *Sitra Reports* 72.

Maliranta, M. and P. Ylä-Anttila (2007), “Kilpailu, innovaatio ja tuottavuus,” *ETLA B* 228.

Pekkanen, J. and T. Riipinen with E. Ahola (2004), "Investing in innovation," Presentation at the meeting of the OECD Group on Behavioural Additionality, held in Manchester on 10.–11 May 2004.

Piekkola, H. (2005), "Public funding of R&D and growth. Firm-level evidence from Finland," *ETLA Discussion Papers* 996.

Piekkola, H. (2006), "Knowledge and innovation subsidies as engines for growth. The competitiveness of Finnish regions," *ETLA B* 216.

Piekkola, H. (2006), "Seutukuntien kilpailukyky. Osaamispääoma ja innovaatioiden tukeminen kasvun lähteenä," *ETLA B* 216.

B. Foresight

Publications on foresight are mainly produced in joint projects carried out by the central organisations of the national innovation system and by top experts in the field.

Building on innovation: Priorities for the Future. The focus areas in the Tekes strategy (2005), Tekes, Helsinki.

FinnSight 2015. The outlook for science, technology and society (2006), The Academy of Finland and Tekes, Helsinki. *www.finnsight2015.fi/.*

Hjerppe, R. and J. Honkatukia (eds.) (2005). *Suomi 2025 – kestävän kasvun haasteet*, Publications of the Government Institute for Economic Research 43.

"Hyvä huomen! Introduction of the Foresight Forum of the Ministry of Trade and Industry," www.ennakointifoorumi.fi/index.php?alasivu=etusivu.

Hämäläinen, T. (2006), *Kohti hyvinvoivaa ja kilpailukykyistä yhteiskuntaa: kansallisen ennakointiverkoston näkemyksiä Suomen tulevaisuudesta*, Sitra, Helsinki.

Signals 2006. Results of the foresight project.by Tekes, www.tekes.fi/TilastotJaVaikutukset/tulokset.html.

Ståhle, P. (ed.) (2007), "Five steps for Finland's future," *Tekes Technology Review* 202/2007.

Valovirta, V. and M. Hjelt (2005), "Onnistumista punniten, tulevaa luodaten," *Hallinnon Tutkimus* 2005: 3, 95–111.

Valtioneuvoston ennakointiverkosto ja ennakointi ministeriöissä 2005. Government Foresight Network, Report 1/2005.

CASE 3: EVALUATION OF ISRAEL'S R&D PLAN – MAGNETON AS A CASE STUDY[29]

1. General

257. The Government of Israel, in order to encourage investments in industrial R&D, enacted the R&D law (1984), whose purpose is to allocate funding to industrial companies that perform high-technology R&D. The goal was to make it easier for Israeli companies to invest in R&D because R&D was considered the best way to promote competition and could give companies relative advantages in international markets.

258. The rationale behind the law is participation in risks and in results. Therefore, a company which obtains an R&D grant and successfully sells the products based on the financed R&D should return the grant money as a percentage of yearly sales up to the total of the grant. The money received from companies is recycled into the budget of the Chief Scientist in the Ministry of Industry, Trade and Labour, who is the executor of this law, in order to finance new R&D activities. If the projects funded do not reach the sales phase or the sales do not yield a return sufficient to return the full grant, there is no requirement to return the balance of the financing.

259. During the 1990s, it was understood that it is necessary also to invest in the development of technologies in order to lay the technological foundation for additional innovative developments. In a state which lacks governmental research institutions that deal with applied research for industry, this task should be performed in research institutions and in industry, preferably in co-operation. Accordingly, from 1994 a separate activity called the Magnet Programme was established within the Chief Scientist's bureau to deal with the advancement of the technological infrastructure of industry, Its objective is to strengthen industrial companies' capacity to draw from a vast and varied pool of research and technology, giving them the capability to develop innovative, high value-added products with significant export potential.

260. The programme provides financial support to "pre-competitive" R&D projects developed jointly by enterprises and research institutes organised as consortia specifically dedicated to the project and governed by "collaborative agreements" among the respective parties. The intellectual property rights derived from technologies developed by a consortium belong to the members that developed it; however, each other member receives a licence to use the technology for further development of its own products free of charge.

261. As of 2000, the activity was broadened to Magneton and is the case study presented in this section. Its mission is to encourage and assist in the transfer of technologies from research bodies to industry through co-operation between an enterprise and a research group. The Magneton programme was created to make the achievements of Israel's scientific research in research bodies more readily available to Israeli industry for the benefit of the Israeli market. The goal is to encourage a process that contributes to technological innovativeness, a primary engine of growth in Israel's industry and economy. The time frame allocated for projects within the framework of the Magneton programme is usually two years, during which a technological feasibility study is performed for the product/technology. Upon completion the firm

29. Prepared by Dr. Daphna Getz, Mrs. Vered Segal and Mr. Ilan Peled.

can decide whether to move on to an independent full-scale product or industrial process development or to stop the project if it does not appear technologically feasible or lacks profitability.

262. The programme started with a pilot which lasted two years. The aim of the pilot was to test the programme's potential, characteristics and ability to enrich industry with advanced technologies. The pilot phase was constantly supervised and there was continuing dialogue with the research institutions, researchers and companies' representatives which took part in the pilot in order to study the main parameters. Although the pilot included 25 projects, it was not large enough to carry out a statistical/qualitative evaluation because of the intentionally large variety of the projects, which involved different types of companies (small, medium and large), different fields (communications, electronics, materials, biotechnology, etc.) and the types of research institutions (universities, medical centres, research institutions). The evaluation of the pilot found it successful, the programme's procedures were updated, and it went into operation immediately following the completion of the pilot.

2. Intermediate conclusions

263. The programme's administration required the evaluation of the programme during the pilot period to ensure that the programme can reach its main goal, that its definitions coincide with the needs of industry, and that the parameters for evaluating the projects reflect Israel's priorities.

264. The programme is not systematically evaluated each year, but since there are two rounds of submissions a year, data on past performance is made available to the selection committee as background material Usually these data are qualitative rather than quantitative. However, at the beginning of 2008 quantitative research began to be regularly performed. The findings are presented here.

265. Within the framework of the Magneton programme, up to 2008, 280 requests for financing of projects had been submitted, out of which 150 projects were approved. Currently there are 24 active projects and 16 additional projects have been approved but have not yet begun.

266. In 2004, the Samuel Neaman Institute for Advanced Studies in Science and Technology[30] in the Technion conducted, on its own initiative and financing, a first effort to evaluate the Magneton projects by interviewing ten institute researchers and eight industry managers who participated in Magneton projects. The Neaman research team also conducted a comprehensive literature review of the main goals and methods in use for evaluating R&D proposals, models for evaluating the main programmes, and methods for evaluating indirect results. Also, information on programmes similar to Magneton in other countries was accumulated. Based on the interviews and the literature review, research questionnaires for the research institutes and industry were constructed and transferred to participants in 50 projects (most projects examined in this preliminary research were performed while the programme was in its pilot phase). The survey findings, based on answers from 34 institute researchers and 20 industry participants, were summarised in a report sent to Mr. Ilan Peled, the manager of the Magnet programme. Based on this preliminary research and on the knowledge, infrastructure and experience accumulated by the research team at the Neaman Institute on evaluations and on policy issues regarding technology transfer from research bodies to industry, the manager of the programme decided to order an evaluation of the Magneton programme, to be performed by the Neaman Institute in 2008, with financing from both bodies. At the time

30. The Samuel Neaman Institute for Advanced Studies in Science and Technology is an institute for national policy research with emphasis on technology and science, industry and economics, higher education and quality of the environment. It is currently considered the largest and leading research centre in Israel on economic and social policy regarding science and technology. The Neaman Institute is an independent non0profit organisation working within the framework of the endowment made by Samuel Neaman for the advancement of technological and economic research of Israel.

of writing, parts of the evaluation project had been carried out, and it was to be completed by the end of 2008.

3. The goals of the evaluation of the Magneton programme

267. In the Magneton evaluation project it was decided to focus on the results, the outcomes and the impacts of the projects executed within the framework of the Magneton programme. A number of meetings were conducted between the Neaman evaluation research team and Mr. Ilan Peled, the manager of the programme, to discuss the goals of the evaluation and its scope. These concern the main outcomes of research-industry co-operation in the projects performed and their influence:

Construction of a database of all of the programme's projects to date.

Identification of the factors influencing the success of the programme and the projects from the point of view of the characteristics of the participants, the characteristics of the projects, and the working processes of the projects' partners.

Discussion of further issues which can influence the success of the programme. For example, would the projects have reached industry without participation in the programme? What are efficient ways to market the programme? How can these factors be identified at the stage of examination and approval of the projects.

Formulation of conclusions and recommendations on the basis of the research findings to the manager of the programme to help decision makers establish the programme's policies regarding future projects.

268. The conclusions of the evaluation will be aimed at policy makers and Magneton decision makers and will supply them with extensive information on the efficiency and effectiveness of the programme's efforts to transfer technology from research bodies to industry. This information can help in the management of the programme and the projects. Information on the characteristics of projects and participants with greater chances of success can help in the project selection process and increase the programme's effectiveness and success. Information on work processes and elements of the programme's management policies which contribute to its success can also help establish policy on links between research bodies and industry during the project.

269. When planning the current evaluation of the Magneton programme, the Neaman Institute's team considered other parties interested in the programme, such as: the Chief Scientist and R&D policy makers in the Ministry of Industry, Trade and Labour; the project selection team; the team of examiners from the Ministry of Industry, Trade and Labour which accompanies the projects from the examination phase, examines their compliance with milestones, attempts to assist in problem solving, and can authorise an additional year for projects on the basis of their advancement; the technology transfer offices (TTO), which represent the research bodies and are responsible for preparing the agreements between them and industry, for signing the agreements, for management, and for legal issues; and company managers or project heads who participated in Magneton projects and indeed researchers and project heads whose proposal was rejected.

270. Those interested parties (stakeholders) do not take part in the present evaluation programme and no effort was made to obtain information from them. In the future a broader scope may be considered, in order to collect and exchange information with such interested parties as a way to deepen understanding of the findings and the larger picture.

4. Evaluation methodology

271. The Neaman Institute research team submitted to the Magneton programme manager a detailed proposal for the evaluation which included the research goals, the indicators to be examined as part of the research, the research population, and the detailed research plan, timetable and budget. The proposal was updated on the basis of feedback from the programme manager, the project duration was set at 12 months and is being carried out according to the plan agreed between the two bodies, which will remain in contact throughout all its phases.

272. In choosing the methodology for the evaluation, the research team took into consideration the fact that the Magneton programme is very new, having commenced operations following the two-year pilot (2003-05). The short time since then makes it difficult to evaluate the long-term economic effects of the programme through measurement and analysis of economic parameters.

273. Therefore, it was decided to survey project participants. The advantage of a survey is that it can obtain information on the programme and participants in the early phases of the programme. Other reasons motivating the choice of a survey as an evaluation tool were:

- It is the quickest and most readily available method of reaching all programme participants (an Internet programme was used which helps construct and distribute the research questionnaires and collect answers).
- The survey makes it possible to build a database on the projects, which can be used in the future to continue and broaden research on programme evaluation in different ways and using different methods.
- The survey supplies aggregative information which can be analysed using statistical tools that enable examination of the connections between the different variables.
- The survey obtains both subjective information supplied by project participants on their background, the projects' characteristics, their rate of success and their opinion of the programme, as well as objective information regarding the outcomes of the programme.

274. In the future it will be possible to adopt additional evaluation methods, such as: economic analysis of the programme's influence and construction of an analytical model for understanding different aspects of the programme, and more.

275. The evaluation process is not meant to remain as an isolated event but to continue in the future and include additional interested parties and broaden the evaluation to areas and aspects not included in the present phase of the project and use additional evaluation methods which will throw light on other aspects of the programme.

276. The present evaluation focuses on the effects of participation in the Magneton programme and the characteristics of the projects (participants, fields, innovativeness, uncertainty, co-operation, etc.) which can be tied to measures of success and the influence the projects have.

277. The main indicators measured in the research questionnaire are:

1. Measurements for success of the project:

   Products resulting from, or expected to result from, the project (articles, patents, development of new product lines in enterprise, entrance of the company into new markets, etc.), direct and indirect benefits (spillovers) of the project, and the rate of success of the project according to different criteria (ability to continue to develop independently the product/technologies in the companies, realisation or expectation of realisation of the technology in the industry, broadening of employment opportunities in the company, etc.).

2. Characteristics of the projects and participants which influence the success of the project:

- Background data on programme participants: education, years of experience, function, previous experience in academic–industry co-operation, previous participation in R&D projects.
- Project characteristics: state of advancement of the knowledge prior to beginning the project, relevant infrastructure for the project such as laboratory, equipment, type of knowledge developed, rate of innovativeness, technological classification, level of uncertainty of the project.
- Characteristics at the start of the project: type of connection between partners prior to the project, expectations, calculations regarding entering the project.
- Characteristics of the co-operation: the pattern of the co-operation throughout the project, location of the work performed, factors contributing to co-operation, continuation of co-operation after the end of the project, etc.

278. The indicators are measured on a scale of 1 to 5, where 1 indicates low contribution or influence, and 5 indicates high contribution or influence.

279. The research questionnaire is mostly composed of multiple choice questions. Two different questionnaires were prepared, one for researchers and the other for heads of industry projects. Many of the questions were included in both questionnaires in order to examine the partners' differences in approaches to the different subjects. Short versions of both questionnaires were also prepared (in which the general questions were omitted) for participants who took part in more than one Magneton project. These were to be filled in for each additional project in which they participated. The research questionnaires were sent to the programme manager for feedback and subsequently modified on the basis of his remarks.

280. Most of the questions are subjective. Objective questions concern the products resulting from the project, subsequent decisions, etc. Respondents are asked to choose what they considered the most suitable responses on the basis of a number of alternatives, or to rate the strength of the factors or characteristics presented. Opinio software is used to prepare and conduct the surveys and to follow up respondents in real time, to analyse and present data and outcomes, and to export data into statistical software for detailed analysis.

281. After the data have been transferred to statistical software, a database will be compiled. Before analysing the data, their internal reliability will be examined by using Cronbach's α. The research findings will be presented using descriptive and inferential statistics, which will include the testing of the research assumptions by relevant statistical examinations. The descriptive analysis will present the distribution of replies in chosen variables through analysis of frequencies, charts, single dimensional tables and bi-dimensional tables (cross tabulations). The purpose of the inferential analysis is to identify the connections between key variables and to find differences between the two research populations (academic and industrial). The inferential analysis uses tests and models such as t-tests, linear and logistic regression models, statistical correlations, chi-square tests, as well as a-parametrical tests and models such as Mann-Whitney U, Kruskal-Wallis H test and the Spearman correlation coefficient.

283. Mr. Ilan Peled, the manager of the Magnet programme and the client of the evaluation project is a partner in the evaluation of the Magneton programme. Mr. Peled initiated the Magneton programme and took part in defining its objectives and directives. He is also responsible for implementation of the programme's policy and making policy adjustments and alterations throughout the programme.

284. After it was decided to evaluate the Magneton programme, with joint financing by the Magnet administration and the Neaman Institute, Mr. Peled, together with the team of the Neaman Institute, defined the objectives of the evaluation. A summary of the findings of the Magneton evaluation, which will be composed by the Neaman team, will be presented to Mr. Ilan Peled so that he can decide if changes in the programme's policy and directives are required. The research team of the Neaman Institute is responsible for constructing and carrying out the evaluation programme, constructing the database on the Magneton programme's projects, and submitting the summary report to the project manager. The team performs the evaluation independently and is responsible for the analysis of the research results and the conclusions drawn.

285. The research team in the Neaman Institute and the manager of the Magnet programme were in contact and co-operated closely during the performance of the evaluation project. The main phases in the research for evaluation of the Magneton programme were:

- In 2004 preliminary research was carried out by the Neaman Institute for the evaluation of the Magneton programme and the results were presented to the programme manger. This preliminary research formed the basis for the construction of the present evaluation programme.
- Several meetings were held between the research team from the Neaman Institute and the Magneton programme manager in order to agree on the targets of the evaluation. On that basis, the Neaman Institute formulated an offer for an evaluation programme, the scope of the evaluation, expectations, research population, research tools, budget milestones, etc. The programme was authorised by the manager of Magnet.
- 13 interviews were conducted with participants in Magneton projects who had finished their activities (six interviews with academic researchers and seven industry project managers) to learn whether changes had occurred in the programme after the preliminary research had been performed and in order to test the clarity of the questions in the research questionnaire and their suitability to the issues to be covered in the current project.
- A database of primary data on participants in the projects of the Magneton programme was constructed.
- The research questionnaire was updated and adjusted to the present evaluation goals, based on the interviews and feedback from the manager of the programme.
- The questionnaire was uploaded to Internet software and the e-mail addresses of the research population were updated in order to be able to conduct the survey.
- Performance of the survey: The research questionnaire was sent to the participants, followed by mail and phone reminders to ensure a high percentage of returns. The research team called all the participants who did not return the questionnaire and asked each of them personally to do so. For some participants the Neaman Institute team filled in the questionnaire during a telephone interview with the participant. Data was subsequently analysed and entered into the database.
- Analysis of the survey results was performed by the Neaman Institute team which also wrote a summary report which respects the anonymity of participants. It was presented to the Magneton programme manager and then published.

FOLLOW-UP EVALUATION TO ADVANCE R&D POLICY IN JAPAN: ASCERTAINING LONG-TERM IMPACTS[31]

1. Introduction

286. Japan's economy began to stagnate during the 1990s. One of the reasons was the mismatch between conventional technology development for industry and the emerging science-based industries, particularly in the fields of information technology and biotechnology.

287. In an effort to modernise Japan's science and technology system, the national Science and Technology Basic Plan was issued, followed shortly thereafter by innovation policy. The 3rd Basic Plan (for 2006-10), promotes innovation policy, itemising priority fields and indicating the required budget.

288. The evaluation process reviews the entire R&D phase, which is an important area of innovation policy. It seeks to streamline the budget process and reduce redundancy. In addition, with advances in science and technology, the evaluation process allows timely adjustments in innovation policy. However, modern S&T has become so specialised that decision makers for R&D find it difficult to fully evaluate its contents. This makes well-established methodologies and standard procedures for evaluation necessary.

289. The follow-up process identifies elements that accelerate, as well as impede, the innovation process. The two case studies of follow-up evaluation described here demonstrate the innovation processes. The results section indicates essential action plans to further promote R&D.

2. The evolution of evaluation in Japan

2.1 The history of evaluation in Japan's public entities

290. In 1995, the government enacted the Science and Technology Basic Law with a view to making Japan an S&T-based nation. In 1996, on the basis of this law, the government established the 1st Basic Plan for Science and Technology. Its purpose was to attain higher standards in S&T and encourage comprehensive and systematic S&T policies, while contributing to the economic development of Japan. The plan began in fiscal year 1996 and ran to 2000. It also included projections for the following ten-year period. It has been revised every five years.

291. In 1997, the General Guidelines for the Evaluation of Research and Development were issued to enhance the efficiency of R&D, and ministries began to evaluate R&D projects on the basis of these guidelines. This was the starting point of public evaluation in Japan. The evaluation process was part of the reform of the budget system from an internal decision process to a co-ordinated decision process involving the intervention of external experts. To date, the guidelines have been revised twice and are scheduled to be revised again. The evaluation system has evolved over the past ten years and will continue to do so in line with the changing world context. Also in 1997, METI's newly created Technology Evaluation and

31. Report prepared by Yasukuni OKUBO, Director for Technology Evaluation, Technology Evaluation and Research Division, Ministry of Economy, Trade and Industry.

Research Division issued the Guidelines for Technology Evaluation in which METI established the standard items and criteria for evaluation.

292. In January 2001, in a major restructuring of central government, 22 ministries and agencies were consolidated into the Cabinet Office and 12 ministries. In April 2001, to separate operational sectors from the central government, independent administrative institutions (IAI) were established. In 2004, 99 national universities were reorganised into the 89 IAI national universities. Prior to the restructuring, the number of central government staff was approximately 544 000. However, as a result of the restructuring, staff was reduced by approximately 25% (or 136 000 individuals).

293. The Cabinet Office was established in January 2001 as a new agency within the *Cabinet of Japan*. It is responsible for handling the day-to-day affairs of the *cabinet.* The position of cabinet secretary has traditionally been filled by a sitting cabinet minister. Also in 2001, the Council for Science and Technology Policy (CSTP) was established in the Cabinet Office, under the leadership of the prime minister. The Council serves as the headquarters for the promotion of S&T policy. The CSTP meets every month and works to oversee, formulate, and co-ordinate national S&T policy. CSTP examines its compatibility with the Basic Plan and sets priorities for each ministerial research project. This cross-cutting, inter-ministry and top-down decision maker is responsible for organising a new science and technology administration system in Japan.

2.2 METI's role in S&T administration

294. METI's mission is to secure Japan's sustainable economic growth, while ensuring stable growth of the international economy. This is achieved by enhancing Japan's competitiveness, creating new markets, and improving the socioeconomic infrastructure. METI covers six policy areas: *i)* manufacturing industry, economic and industrial policy; *ii)* foreign economic policy; *iii)* policy for information and communication and services; *iv)* policy for small and medium-sized enterprises and regional economy and industry; *v)* energy and environmental policy; and *vi)* nuclear and industrial safety policy. Under these policies, 34 programmes oversee approximately 120 R&D projects. These projects serve as the minimum unit of policy.

295. METI's total budget for R&D is approximately JPY 500 billion per year. The budget is mainly allocated as non-competitive subsidies to the New Energy and Industrial Technology Development Organization (NEDO) and to the National Institute of Advanced Industrial Science and Technology (AIST). NEDO evaluates its own projects (which are subsidised by METI). AIST evaluates its own performance. Excluding its non-competitive subsidies, METI mainly evaluates the competitive budget subsidies to industry. METI's Technology Evaluation and Research Division is responsible for evaluation of the R&D project and its budget and evaluates about 50 projects a year.

2.3 METI's R&D-related groups

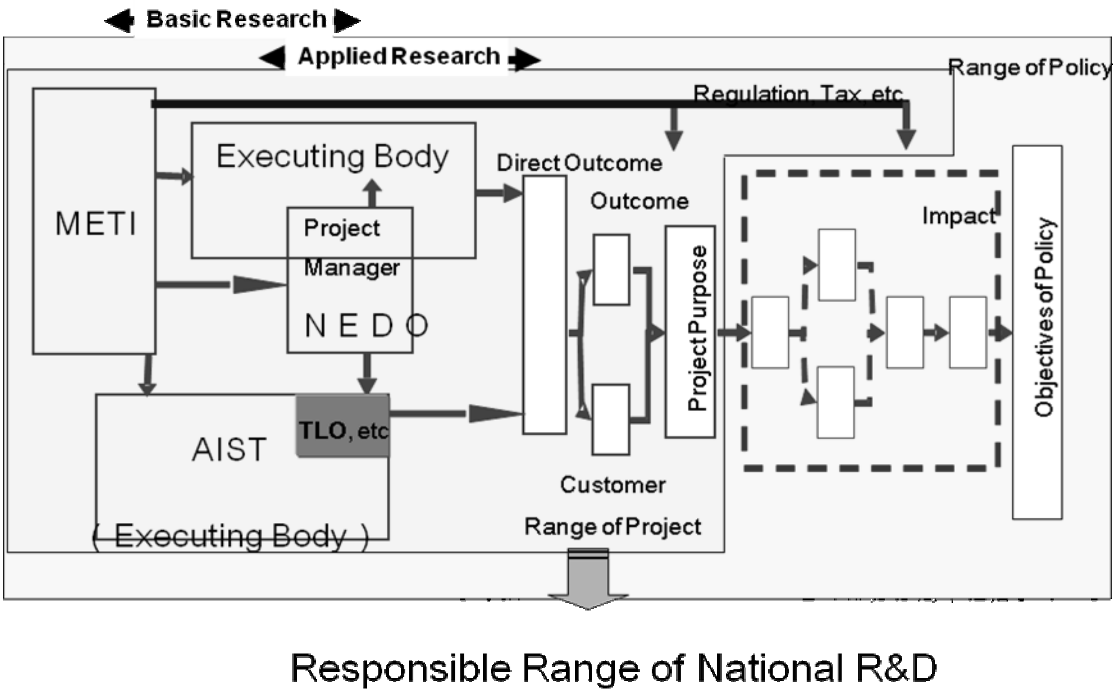
296. METI's R&D-related groups include; NEDO, AIST, private industry and universities. NEDO serves as a funding agency which co-operates with private bodies to conduct applied research. In some instances project managers from the executing body work directly with the NEDO office and its staff. This type of co-operation results in a strong sense of responsibility and accountability among all participants.

297. AIST is a research institute charged with conducting basic research; however, AIST currently extends its purview to resulting products as well. AIST has a technology licensing organisation and an innovation office which are responsible for ensuring the outcome of the research conducted at its facilities also contributions to society. METI endeavours to stimulate the international economy by promoting S&T

through direct support to the private sector, NEDO, and AIST. In addition, METI works with regulations, taxes, etc., to enhance the impacts to the innovation process.

298. The blue area of Figure 7 shows METI's range of responsibilities relating to national R&D.

Figure 7. Role of METI, AIST and NEDO in relation to Japan's R&D effort



299. Both METI and NEDO conduct follow-up evaluations, however, they differ in the range of their evaluations. NEDO's projects aim at short-term outcomes and reaching the project's goal. METI's follow-up evaluation takes into account a long-range outlook to achieve the policy objective. Hence, METI's evaluation incorporates NEDO's evaluations.

2.4 METI's evaluation process

300. The diffusion model of technology makes it possible to sketch the large range of outcomes from a project's outputs. In METI's current innovation policy, a key discussion point is whether and how R&D has contributed to or affected various outcomes. Collect a wide range of evidence to identify and understand outcomes requires a significant effort. It has been a challenge for METI to evaluate outcomes in the project evaluation phase, the impacts on industry, and the cost-benefit ratio of R&D.

301. METI remits interim, *ex post*, and follow-up evaluations to external committees, while the *ex ante* evaluation is accomplished via self-evaluation. The two types of external evaluation evaluate projects in terms of the output and the outcome. The first consists of a peer review and the second an expert review. The peer-review evaluation team of four to five members reviews specific R&D projects. Meanwhile, the expert-review evaluation team of ten members reviews all R&D projects. Follow-up evaluation methodologies used by METI include; indicator evaluation, quantitative evaluation and interview/questionnaires. Follow-up evaluations are not accomplished on all projects; rather, one or two projects are selected each year from the entire pool of projects completed five years earlier.

3. METI's follow-up evaluation

3.1 Key role of follow-up evaluation

302. The first item in the follow-up evaluation process is impact analysis. Under METI's guidelines this includes technology impact, improvement of R&D performance, economic benefit, social/life improvement, and feedback to policy. These items reveal the extent of progress achieved by the project. The concept of the impact on industry is always opaque and controversial in the evaluation.

303. A common question of reviewers is the definition of "benefit". It is defined here as the cost-benefit ratio. It is a useful measure because it is quantitative and provides a numerical indicator. However, collecting the relevant evidence usually requires a lot of effort.

304. As mentioned earlier, the follow-up evaluation is conducted on projects concluded five years earlier. The evaluation looks at the project from a "hindsight" or "after-the-fact" viewpoint in order to identify longer-term effects. Changes in points of view and circumstances following the project can throw further light on the earlier *ex post* evaluation and reveal the legitimacy of the follow-up actions. This is a way to learn about elements that accelerate and/or impede longer-term outcomes.

305. External, internal and political factors affecting the flow of research are collected and analysed by METI via interviews of the executing bodies. Examples of external factors include advances in competing technologies and changes in the marketplace. Examples of internal factor include maturity of a technology, justification of technology, possibility of fusion with another technology, justification of strategy for industrialisation, etc.

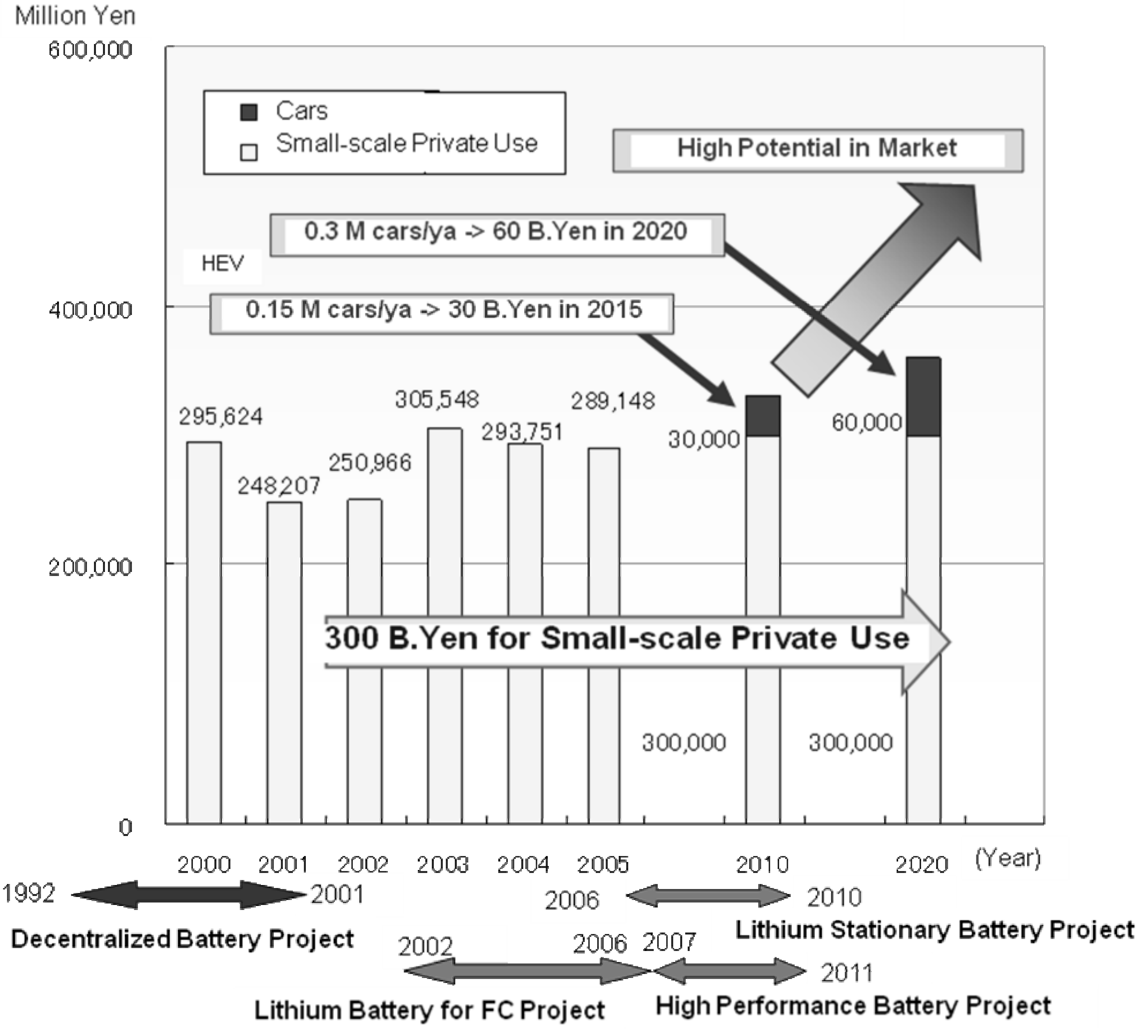
306. One of the vital roles of the follow-up evaluation is to locate or identify negative elements and their linkages. Positive elements are easily identified because they advance or accelerate a project; negative elements inhibit the progress of technologies or practices. Since the innovation process may be terminated in the latter cases, little relevant evidence may be available. Linkages also represent a problem, as missing linkages are hard to identify but can impede progress on technologies and practices. All these factors can help improve an R&D strategy and identify new R&D programmes, funding and regulations.

3.2 METI case study instituting follow-up evaluation

307. From 1992 through 2001, METI conducted a "decentralised battery project". This project resulted in the launching of three new projects; one of which is an ongoing stationary battery project; the other two targeted automobiles.

308. Figure 8 examines the prospective lithium battery market. The market includes small-scale personal computers and digital cameras for which demand remains constant and stable over a span of 15 years and is worth approximately JPY 300 billion. The know-how generated from the initial project and subsequent projects was responsible for producing medium- to large-scale batteries. Currently there is no demand for a lithium battery in automobiles, but it should emerge within the next two years and grow over the next ten years.

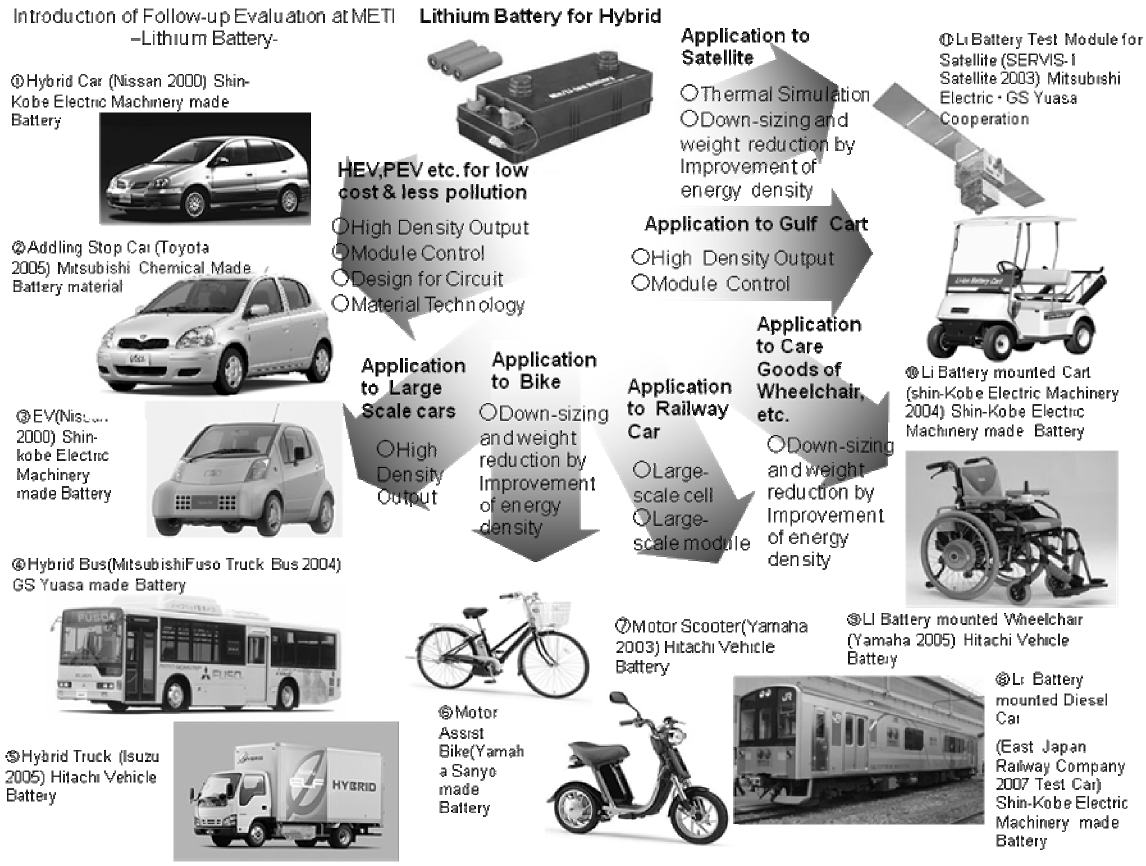
Figure 8. Prospective lithium battery market and subsequent national projects



309. The advantages of the lithium battery, compared to the lead-acid battery include its high density storage capability, enhanced module control, design for circuits, material technology, facilitation of downsizing, reduced weight, large-scale cell and module production. The lithium battery has been applied to hybrid systems for transport, including the delivery system in hybrid automobiles, electric vehicles, hybrid buses and trucks, motor assisted bicycles, motor scooters, diesel cars, wheelchairs, golf carts, as well as a satellite test module (Figure 9). High density storage helps to accelerate the development of a large-scale hybrid system.

310. A follow-up evaluation enabled METI to understand that the decline in the price and demand for electricity and regulation to maintain conventional use of electricity were factors that inhibited progress on the creation of new demand for battery storage systems. However, the stationary battery has the potential to stabilise power from wind power generator. Seizing the opportunity, NEDO began a lithium stationary battery project in 2006 (scheduled to run through 2010). The project aims to develop a battery capable of generating electricity at 40 000 kWh when combined with wind power generation. This is an example of policy intervention to assist the diffusion of technology by identifying impeding elements.

Figure 9. Diffusion of lithium battery for transportation



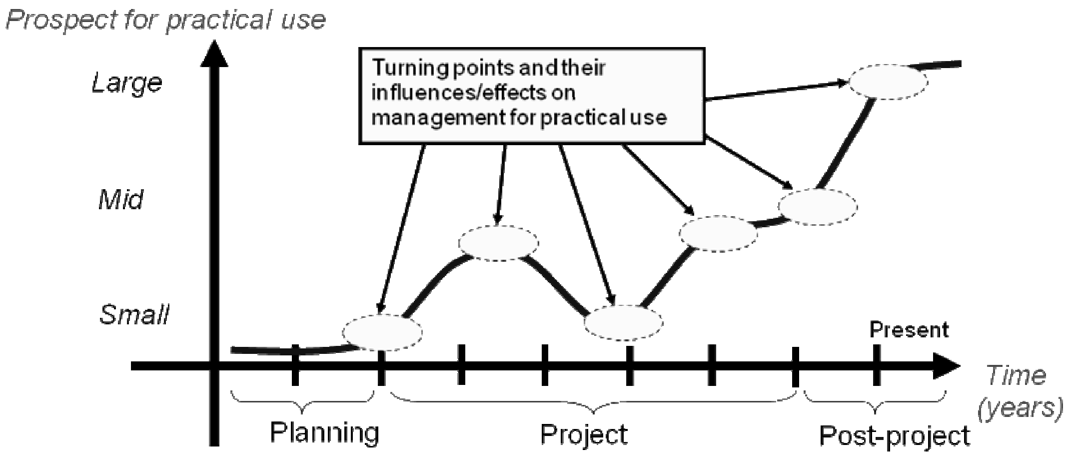
4. Follow-up evaluation at NEDO

4.1 Monitoring for follow-up evaluation

311. NEDO conducts follow-up monitoring to observe the direct outcome of its research. The monitoring of results ensures accountability to Japan's taxpayers, provides feedback for improving NEDO's R&D management, and contributes to the planning of future R&D strategies. NEDO monitors the activities of project participants annually, in particular the practical applications of R&D achievements, while collecting evidence related to accelerating and impeding elements. The collected evidence and the *ex post* evaluations allow NEDO to follow technology trends through post-project research activities.

312. In order to visualise its findings, NEDO creates a follow-up chart (Figure 10). The follow-up chart is a tool that identifies chronologically significant events in the project management flow. The main objective is to show "success" or "failure" scenarios in an executing body. The methodology used to collect evidence is interviews of the companies involved in the project. The interviewee is a person who understands the entire project and its evaluation results.

Figure 10. Follow-up chart



313. The horizontal axis of Figure 2.9 represents time, while the vertical axis signifies the prospect that the technology will have practical use in the view of the interviewees. The diagram reflects the change in prospect (practical use of the technology) over time and the turning points, as well as the influences/effects on management. To date, NEDO has conducted interviews of approximately 69 companies. Of these, 46 achieved commercialisation (*i.e.* proved successful); while 23 terminated their R&D on the basis of demonstrated failure.

314. Based on company interviews, NEDO defines four categories. Of the 46 success stories, 33 are categorised as Type A (Figure 2.10), in which the executing body steadily overcame technical problems with a clear vision of social/user needs. The remaining 13 were categorised as Type B (Figure 11), in which the view of technical problems during the early stages had been overly optimistic but R&D breakthrough did occur during the latter half of project.

315. Of the 23 failures, 15 were categorised as Type C (Figure 11), in which technical problems were steadily overcome but soon after the project was launched, there was decrease in the need for and/or the cost competitiveness of the intended products. The other 8 were categorised as Type D (Figure 12), in which the executing body had been too optimistic about its capacity to solve the technical problems and failed to attain R&D breakthrough.

Figure 11. Follow-up charts of Type A and B

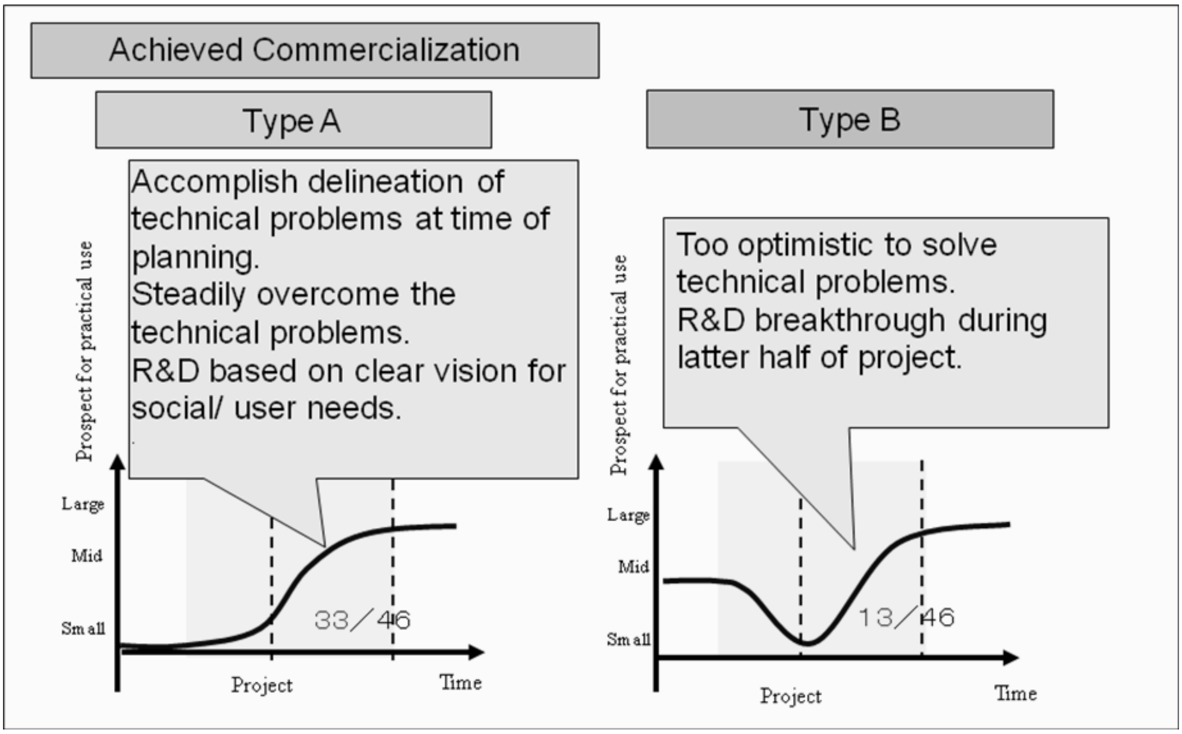
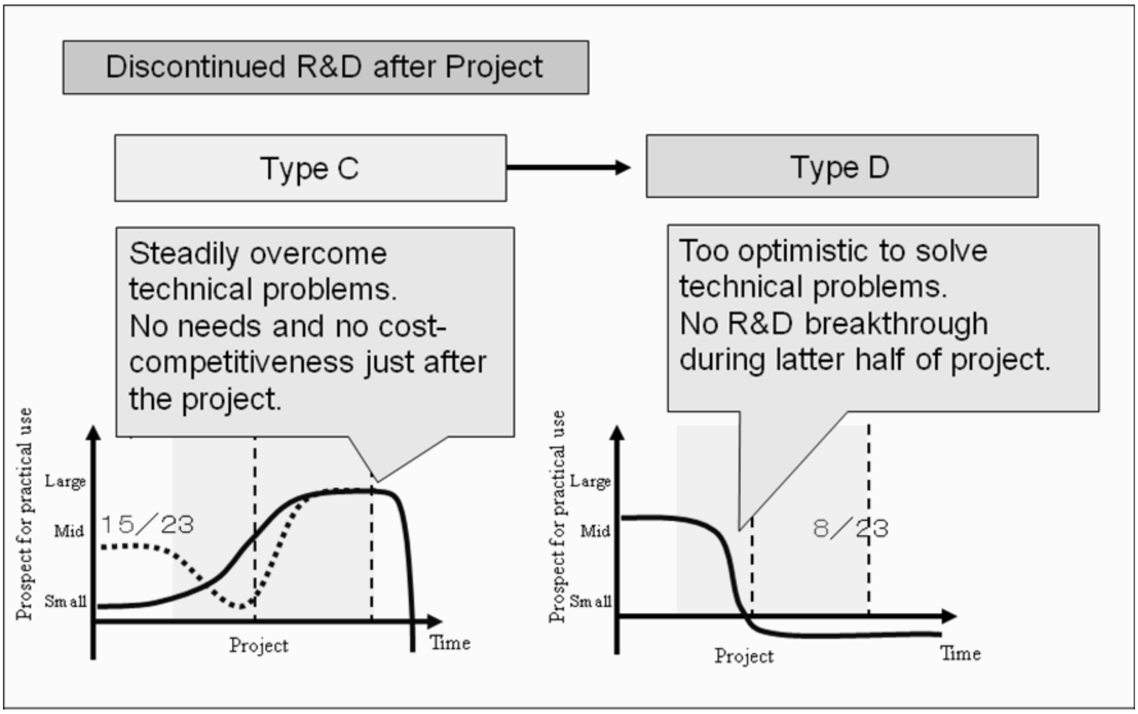


Figure 12. Follow-up charts of Type C and D



4.2 Advantages of follow-up monitoring

316. Advantages to follow-up monitoring include the fact that lessons learned from follow-up monitoring have been consolidated and documented in the NEDO R&D Project Management Guideline which is accessible through NEDO's intranet system and serves as a knowledge-sharing venue for companies participating in NEDO projects. NEDO also organises discussions on innovation policy with the companies. Ultimately, participants benefit through these types of activities and evaluation exercises. Also, while conducting follow-up monitoring, there are times when NEDO discovers that the technology developed has come to a dead end. In these cases, NEDO provides financial assistance aimed at removing obstacles, if necessary. NEDO submits policy proposals, including results from the follow-up monitoring to METI. An example of this is the proposal on reforms of legislation and national innovation plans.

5. Conclusions

317. By monitoring the diffusion of outputs, METI uses the follow-up evaluation to view R&D projects over the long term with respect to policy objectives and long-term outcomes. This type of observation reveals factors that accelerate and impede innovation; contributes to finding new R&D projects and ensures effective funding and regulations, while also being instrumental in justifying policy actions. The follow-up evaluation works well for adjusting innovation policy to the rapidly changing S&T environment. By reviewing the diffusion of outputs, R&D decision makers ensure the legitimacy of policy intervention. This enables a "policy mix" for research and innovation policies. The disadvantages of follow-up evaluation include difficulties for collecting historical information (the data are five years old) and the inability to capture intangible factors and their impact on the project.

318. Unlike METI, NEDO focuses on applied R&D. NEDO monitors projects in the follow-up evaluation by including the executing body and collecting direct evidence on project outcomes. The monitoring reveals a sequence of events which suggests many turning points and factors that accelerate or impede progress. Identification of the factors influencing the project leads to appropriate reforms in R&D strategy. The results are loaded on METI's open database, thus creating feedback for the participants.

319. The disadvantages to follow-up monitoring include the lack of timely data. Since the follow-up review is conducted on projects completed five years earlier, obtaining historical information is cumbersome. Another disadvantage with the follow-up monitoring process is that it does not capture intangible factors that may have affected the R&D project.

320. The monitoring after the end of project conducted by NEDO is a good way to refresh the data. The next step for METI is co-operation with NEDO for better follow-up evaluation.

REFERENCES

Arimoto, T. (2007), “Innovation Policy and Evaluation –Changing the national innovation system of Japan”, International Symposium on Innovation Policy and Evaluation, 19-20 November.

Nakamura, O. (2007), “Strategic Policy and R&I Management by METI: ‘Eco-innovation’ and ‘Strategic Technology Roadmap’”, International Symposium on Innovation Policy and Evaluation, 19-20 November.

Nakamura, O., M. Sawada, S. Kosaka, M. Koyanagi, I. Matsunaga, K. Mizuno and N. Kobayashi (2008), “Strategic evaluation of research and development in Japan’s public research institutes”, In C.L.S. Coryn and Scriven (eds.), *Reforming the Evaluation of Research, New Directions for Evaluation*, 118, pp. 25-36.

Ogasahara, K., O. Nakamura, K. Inahashi, C. Miyokawa, Y. Kimura and Y. Okubo (2007), “Analysis of Follow-up Evaluation Results of Research and Development (R&D) Projects Applying Logic Model to Elucidate the Process of Innovation”, American Evaluation Association Annual Conference in Baltimore, November 7.

Suzuki, J. and S. Yumitori (2006), "Behavioural Additionality of Public R&D Funding in Japan”, *Government R&D Funding and Company Behaviour,* pp. 151-165, OECD, Paris.

CASE 4: EVALUATING RESEARCH FIELDS/DISCIPLINES IN NORWAY[32]

1. Introduction

321. This particular case study was chosen because it represents the most systematic approach to evaluating research on a national level in Norway. It has become an important tool for the development of the research in terms of its organisation and for maintaining a more consistent focus on quality improvement. The evaluations of research fields/disciplines have been greatly encouraged by the Ministry of Education and Research and are mentioned in several political documents.

322. The evaluations focus on university research, but may also include research environments at university colleges or research institutes. They are used in priority setting in the research institutions and by the Research Council, and have produced important input for the development of new strategic measures (the Norwegian CoE scheme among others). The results are also useful to the Ministry of Education and Research and other ministries.

323. The following 16 evaluations of research fields/disciplines have been undertaken:

- Chemistry (1997).
- Earth sciences (1998).
- Biology, basic including biomedicine (2000).
- Physics (2000).
- Mathematics (2002).
- ICT (2002).
- Linguistics (2002).
- Political Science (2002).
- Medical and Health (2004).
- Pedagogy (2004).
- Technology and engineering sciences (2004).
- Nordic languages and literature (2005).
- Pharmaceutical research (2006).
- Development research (2007).
- Economic research (2007).
- Historical research (2008).

324. The evaluations are conducted as a combination of quantitative analyses, bibliometry and other performance measures, along with qualitative peer reviews (an international expert committee). They are considered part of a continual learning process for all stakeholders involved in the development of basic research/research within fields/disciplines.

325. Over the years this particular approach to evaluation has been improved to make it more standardised and more efficient. Methods are continually improved along with the development of more

32. Prepared by Gro E M Helgesen, Special Adviser, Research Council of Norway.

sophisticated research indicators and improved focus of the terms of reference. Recently a five-year plan for the evaluation of research fields/disciplines was adopted by the Division for Science in the Research Council. The plan has been presented to the Ministry of Education and Research.

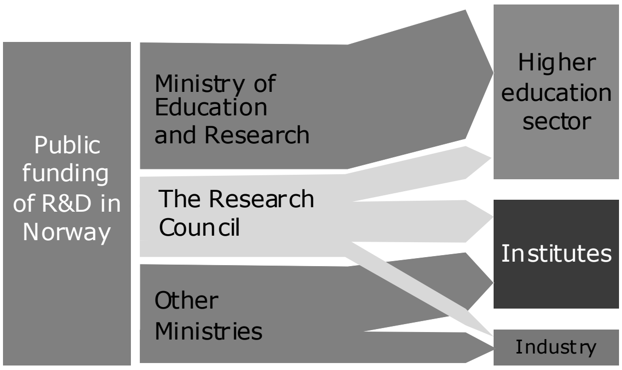
2. Environment

2.1 Resources for R&D in Norway

326. Norwegian expenditure on research and experimental development (R&D) in nominal terms amounted to NOK 33.9 billion in 2006. This represented 1.57% of the gross domestic product (GDP). Norway spent NOK 6 410 per capita on R&D in 2005; the OECD average was NOK 5 770 the same year.

327. The higher education sector accounted for NOK 9.1 billion of R&D in 2005. The equivalents for the industrial sector and the institute sector were NOK 13.6 billion and NOK 6.9 billion, respectively. In 2005 NOK 13.2 billion of Norwegian R&D expenditure was funded by industry, NOK 12.9 billion by the government and NOK 3.5 billion from other sources and abroad. Approximately 30% of all public expenditure on R&D is channelled through the Research Council of Norway (Figure 13).

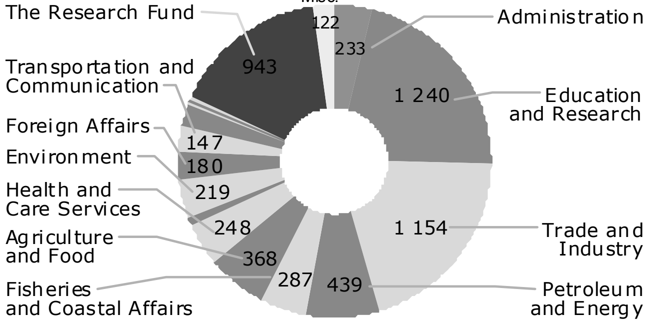
Figure 13. Public funding of R&D in Norway



2.2 The Research Council of Norway

328. The Research Council of Norway is a public administrative body with special powers of authority and is organised under the auspices of the Ministry of Education and Research. The Research Council serves as a national strategic and executive body for research. In 2008 the Research Council's funding amounted to NOK 5 691 million (including administration) or about EUR 820 million (Figure 14).

Figure 14. Research Council funding from all ministries, NOK 56 991 million in 2008



329. The Council is responsible for improving the general knowledge base and for helping to meet society's research needs by promoting basic and applied research as well as innovation. The Research Council also promotes international research co-operation, and serves as an advisory body to government authorities on matters concerning research policy. The research-oriented activities of the Council are organised into three divisions, with responsibility for the following areas:

- Academic research and disciplines (Division for Science).
- Innovation and user-initiated research (Division for Innovation).
- Strategic research in priority areas (Division for Strategic Priorities).

330. According to the Statutes of the Council, one of its main tasks is to "ensure the evaluation of Norwegian research activities". Among other more general evaluation tasks, evaluations of designated research fields/disciplines have been performed by the Council since 1996. The Division for Science has a particular responsibility for these evaluations.

2.3 Frameworks

331. Inspired by general evaluation trends and by experience with the evaluation of research disciplines in other countries, the Research Council has taken on the responsibility of assessing the quality and status of research disciplines on a national level, starting with the natural sciences. An increased focus on the role of these evaluations has been developed in co-operation with Ministry of Research. They are referred to in research white papers as well as in the annual budget allocation from the parliament.

332. Over the years a systematic approach to carrying out these evaluations has been developed, building on previous experience and expanding to all research fields. Recently a five-year plan was adopted by the Division for Science in the Research Council and presented to the ministry.

333. National research statistics serve as a basis for collecting information on the target groups for specific evaluations. Research publication indicators are also available as databases.

334. Information on the individual evaluations is communicated through the Research Council's website, and the reports are also made available there.

3. Goals/strategy

3.1 Purpose

335. Evaluating research fields/disciplines may be summarised as a way of exploring the "health" of Norwegian research. The aim is to critically assess research quality in a given research field or discipline in relation to international standards: its strengths and weaknesses, available infrastructure and organisational conditions, as well as training and recruitment of young researchers. The purpose is to:

- contribute to the dynamic development and improved quality of Norwegian research;
- provide insights on the strengths and weaknesses of the field in question for continued long-term development;
- expand and strengthen the knowledge base for the strategic and advisory tasks of the Research Council and other science actors.

336. The evaluations are not carried out for budgetary purposes, but to improve the quality and developmental potential of the field in question.

3.2 Scope

337. The evaluations mainly focus on fundamental or academic research carried out in universities. Sometimes research at "university colleges" and research institutes is included, depending on their relative strength in contributing to the general development of the particular research field. The evaluations are carried out on a research group or institute level.

338. The research institutions are involved in the planning process to some extent and in the delimitation of the research being evaluated. Communicating the multiple goals to the researchers and institutions involved is important.

3.3 Role

339. The financing of public universities and university colleges in Norway is based on three components:

- basic funding (around 60%);
- a teaching component (around 25%) based on teaching credits/exams and exchange students;
- a research component (around 15%), only part of which is performance based.

340. Research institutes usually have less public basic funding and obtain most of their income from contract research. From 2009 parts of the public basic funding of the research institutes will be based on performance measures.

341. There is an ongoing debate on the financial structure, in particular on how to improve transparency in the allocation of basic resources and on how to separate teaching and research allocations on the basis of improved indicators and evaluations. As an increasing share of university budgets is allocated according to performance measures these evaluations may come to have an even broader utilisation, and even be used for the allocation of funds in the future.

342. The Ministry of Research and Higher Education has actively encouraged these evaluations. With an increasing acceptance of the impact evaluations have on both national strategy development and the institutions' own development, the evaluations are increasingly referred to in general policy debates on a national level. Furthermore, the evaluations of research fields/disciplines are considered to be an essential part of the quality assurance of research in the higher education sector.

4. Planning

4.1 Budgets and resource allocation

343. The Research Council takes on the planning responsibility, and the Research Board (for Science) within the Council sanctions crucial aspects of the process such as the scope of the evaluation, terms of reference, committee members as well as the timeline for the whole process. The budget for each evaluation varies with the size of the field or discipline to be evaluated. In addition to human resources from the Research Council to organise the process, the budget covers "mapping" the research population to be included, commissioning bibliometric studies, fees and travel expenses for a peer review committee and secretary, and travel expenses for researchers involved. Another important element of the cost aspect is the time spent by the evaluatees in preparing self-evaluations and presentations that constitute the basis for the evaluation.

344. Usually at least two years must be allowed from the start of planning until the report is presented.

4.2 Designing the evaluation

345. As mentioned, these evaluations are conducted as a combination of quantitative analyses, bibliometry and other performance measures, along with qualitative peer reviews by an international expert committee. This means that the design is more or less given. A handbook for the administrative planning process has been developed to assist the officers involved, but differences in research cultures may still influence the approach. The collection of general background information is often commissioned to institutions working with research statistics and indicators, as are the bibliometric analyses. Most often the ISI database is used. Also a recently developed national system for field-specific publication indicators at universities and university colleges are used as input.

346. Every evaluation has it particulars when it comes to:

- delimitation of research groups to be included
- choosing methods and back-ground information.

347. Defining the population to be included in the evaluation is an important point of departure and involves meetings with research institutions and different national organisations before the final conclusion is made by the Research Council. These meetings involve discussions of the evaluation model, input for the terms of references as well as input on appropriate statistics and indicators for the research field in question. The dialogue with relevant research institutions has proved important for a successful implementation and smooth operation of the evaluation process.

4.3 Terms of reference

348. The terms of reference for each evaluation are more or less standardised. They state the objectives and organisation of the evaluations and also contain the mandate for the evaluation committee. The committees should focus on:

- scientific quality
- relevance
- collaborations (nationally and internationally)
- research organisation
- scientific leadership and management.

4.4 Methods and material

349. The evaluations combine quantitative and qualitative approaches based on self-evaluations, bibliometric analyses, SWOT (strengths, weaknesses, opportunities, threats) analyses and interviews.

350. Self-evaluation forms are part of the planning process, and build on previous experience as well as special aspects of the research in question. The Research Council is responsible for collecting these. The Research Council also commissions the bibliometric analyses and schedules the meetings between the committee and the evaluatees.

351. Site visits are very time-consuming for an evaluation committee. Unless there are crucial aspects connected to infrastructure assessments, the meetings between the committee and research institutions are set up as "hearings". This involves a fixed time schedule for the committee to meet with representatives from the research groups and institutions being evaluated. It usually includes a standardised presentation in the shape of SWOT analyses and designated time for questions and answers/dialogue.

352. An instrument has also been developed for the peer review. Based on pre-defined criteria it consists of a five-scale ranking system, using the categories *excellent – very good – good – fair – weak.* The instrument is applied in most evaluations.

353. A website is established for each evaluation to present information about the ongoing evaluation. This may include time schedules, terms of reference, committee members, self-evaluation forms, etc. Transparency in the processes is a major goal.

5. Implementation

354. A general procedure for undertaking these evaluations has been developed. The key steps are:

- communication with the research environment
- appointing committee members
- preparing an evaluation report
- follow-up.

5.1 Communication with the research environment

355. Usually a "start-up" meeting with representatives from the institutions involved is called to present the process, the purposes of the evaluation and follow-up plans. The Research Council is responsible for this. General, as well as more specific, information about the evaluation is also presented on the Research Council's website.

5.2 Appointing committee members

356. Considerable work is devoted to the selection of peers. Broad scientific experience, high international standing in their research fields and a high level of scientific integrity are important qualifications for committee members to ensure a legitimate evaluation. Evaluation experience and experience in research organisation and policy making are also desirable qualifications. There is also a sharp focus on possible conflicts of interest. The broader the research field, the more challenging the task of composing a well-qualified committee of workable size. The evaluatees and relevant institutions, including foreign research councils, are consulted with and asked to suggest peers. There is close contact between the Research Council and the committee chair before the evaluation work starts, in order to make sure that the purpose of the evaluation and the methods involved are clearly communicated.

5.3 The evaluation committee's work

357. The work of the evaluation committee is based on the following elements:

- the "mapping" of the research groups/institutes and a presentation of the structural framework for that particular area of research;
- a general presentation by the Research Council of that particular research field in the national setting;
- research groups' self-report and self-evaluation (standardised);
- bibliometric analyses;
- committee meets the evaluatees;
- committee's own overall assessments (peer review in light of international research).

358. A central part of the evaluation is the meeting of the committee with the evaluatees. These meetings are frequently conducted as "hearings". Representatives from the research groups being evaluated

are invited to meet with the committee at a set time, with a prescheduled agenda, including a short presentation and time for questions and answers. Evaluatees come from all over the country to the committee's meeting place, and the hearings last from two to five days. This has been shown to be very efficient in terms of the committee's time. Alternatively, site visits may be conducted, depending on the number of institutions involved and the relative importance of observing the home territory (laboratories, etc). It is the committee's responsibility to conduct the meetings and interviews. Usually the Research Council is represented as an observer in these hearings/site visits, but the Council's representative never attends the committee's discussions concerning their assessment of the evaluatees.

359. If the committee's work is well planned and organised, it may have a draft report at the end of the daily meetings with the evaluatees. The finalising of the report may then be carried out later through Internet contact.

360. Some committees, however, prefer to have a meeting or two to discuss their final report. Usually the parts of the report that describe facts about the different evaluatees are sent to the evaluatees and to the Research Council to check for corrections before the report is finalised.

5.4 Presenting the evaluation report

361. The evaluation report is the committee's final product. It is made public in printed and electronic versions by the Research Council. Often the release is combined with a seminar or a press conference. The Research Council distributes the report to all stakeholders, often inviting specific comments on the committee's findings. The Board of the Research Council has a formal discussion of the findings in a meeting. In most cases the Research Council will have a meeting with representatives of the evaluatees and stakeholders to discuss the findings. The evaluates are also invited to give written comments on the final report and its findings as input to the follow-up process.

6. Utilisation

362. The primary users of the evaluation findings are the research institutions that have been evaluated and the Research Council. Secondary users are the ministries, in particular the Ministry of Research, and other research stakeholders, including the media.

363. In 2004 the Ministry of Education and Research commissioned a meta-evaluation of the evaluations that had been carried out. Some general findings on Norwegian research quality gave input to policy discussions and also drew a positive conclusion about the evaluations as such. It also gave important input to further development of the processes involved in this kind of evaluation.

6.1 Follow-up processes

364. The Research Council initiates the follow-up of the evaluation. This usually involves appointing a working group in which major stakeholders are represented. This group comments on the findings and advice of the evaluation committee. Recommendations are classified according to possible actions and responsibilities and adapted to a realistic follow-up plan of action.

365. While the plan is discussed by stakeholders, there is no enforcement of follow-up. However, the Research Council has increasingly been allocating funds and inviting the evaluated institutions to apply for so-called strategic project funding.

366. The research institutions conduct their own discussions as to the consequences of the evaluation findings. The ministries, in particular the Ministry of Education and Research, assess the report internally

and use it as point of departure for policy discussions. The reports are frequently referenced in White Papers and other policy documents for higher education and research.

367. In line with general quality control procedures, the follow-up is also reported to the Office of the Auditor General in connection with its performance audits of Norwegian universities.

368. The media also frequently cover findings from the evaluations, and this often leads to high-profile discussions of the quality of Norwegian research.

369. Periodically the Research Council requests reports on the consequences and actual follow-up from the institutions, either written or as part of a workshop.

370. Evaluations have led to the reorganisation of institutions; to new collaborative initiatives; and to new financial schemes for the Research Council.

6.2 Consequences of undertaking evaluations of research fields/disciplines

371. Experience shows that acceptance of the findings is strongly related to how well the goal of the evaluation is communicated, and to the degree of involvement of the evaluatees in the planning process. The less control-oriented, and more learning-and-development the focus is, the better. This must also be communicated to the evaluation committee.

372. The findings from the evaluations constitute an important knowledge base for the ministries and for the Research Council the important of which usually exceeds the immediate distribution of follow-up grants. This information is used in strategy discussions in many contexts, including more general budget proposals from the Research Council as well as institutions. And it is used in discussions on improving policies and policy measures. One example is the development of the Norwegian CoE scheme. This came about as a result of several evaluations on variable research quality, research groups below optimal size, too little international collaboration, etc. There was an apparent national need to improve this situation.

373. The institutions also use the results actively, both in terms of organisational changes and the allocation of funds internally.

374. The meta-evaluation of these evaluations carried out in 2004 strongly supported the use of the evaluations, and also gave important input to the way they are implemented. Moreover, after each evaluation, the involved Research Council staff is encouraged to make a short report on the process for each particular evaluation. The aim is to improve the planning and methods of this type of evaluation.

7. Conclusion

375. Quality control of planning and implementation are very important to the Research Council in order to legitimise these evaluations. Ensuring quality is a continual process. Both the follow-up processes and improved utilisation of results are subjected to continuous monitoring and improvement.

376. This case study describes only one of several evaluation activities undertaken by the Research Council of Norway. The overall evaluation "system" may be characterised by two points: *i)* evaluation of research quality (projects, scientific research fields/disciplines, programmes, institutions); and *ii)* evaluation of the quality of policies (instruments and schemes, economic impact, societal and other non-economic impact). The five-year plan for evaluating research fields/disciplines that has been developed should eventually be integrated into a more general evaluation strategy for the Research Council as a whole.

REFERENCES

Research Council of Norway (2007), *Report on Science & Technology Indicators for Norway*.

Brofoss, Karl Erik (2004), "En gjennomgang av Forskningsrådets fagevalueringer", NifuStep Arbeidsnotat 7/2004 (Norwegian only).

CHAPTER 3

ASSESSING THE ROLE OF PEER REVIEW

This chapter reviews the role of expert review as a key tool for evaluation of public R&D. It explores the current challenges faced by the peer review process as well as broader process known as "expert review" as tools for ex post and ex ante evaluations of research policies, programmes and public research organisations (PROs). The purpose of this paper to provide a comprehensive assessment of expert review at the programme and policy level but also a summary of methodological issues and good practices that have emerged from the experience of OECD member countries. It does not however aim to be guide or handbook for programme managers in evaluation agencies.

1. Introduction

1.2 Aims and scope of this paper

377. This chapter on peer review is structured around seven parts. The main body of the report (Sections 2, 3, 4, 5) addresses definitions, procedures, the underlying issues and suggestions surrounding research policy & programme expert review. Section 2 summarises definitions, purpose and applications, and merits and limits of expert review for common understanding of readers. Section 3 describes in detail expert review process protocols based on a few good practices guidelines of member countries. Section 4 addresses methodological issues in expert review and suggest solutions based on the experience of countries and research. Section 5 summarises many of the principles and suggestions included in the main body of the present paper. And finally, the references and suggestions for further readings contain an extensive list of primary and related references to the expert review literature.

1.3 The problems in "expert review"

378. The 'expert review' process is perhaps one of the most dominant and common methods used in the evaluation of science and technology. It plays a significant role in many of the key stages associated with research. It is the main mechanism for deciding who gets funded and what type of science is funded; it determines who gets to publish in the scientific literature; and is used in the selection and promotion of individual within research institutions (Scott, 2006). It is also the core tool used in various R&D programmes and innovation policies.

379. Expert review has many merits. It is as a relatively quick, low-cost, fast-to-apply, well-known, widely accepted, and versatile evaluation method that can be used to answer a variety of questions throughout the project performance cycle, as well as in other applications. It also provides an opportunity for mutual learning. Expert review could very well be the best of all known methods of assessing R&D programmes and policies so long as it is *properly* managed.

380. There are, however, some concerns that the expert review system is under pressure and losing confidence among users because it depends on the professional but subjective decisions of individuals and it is increasingly time consuming and resource intensive. It is not an exaggeration to say that expert review is currently facing its strongest challenges in several decades. At a higher level, we see external and internal challenges. Externally there is some evidence of discontent among political decision-makers about

the ability of expert review to reflect socioeconomic and political priorities. Internally, a hollowing out is occurring as increasing pressure on researchers' time makes it more difficult to find experts willing to undertake reviews. From the perspective of the method of evaluation, it is therefore an appropriate time to assess the status of expert review and to identify the challenges and solutions.[33]

381. Among the key challenges issues that arise in expert review are:

- How to reflect socioeconomic and political priorities effectively and link these priorities to decision making in expert review processes?
- How can expert review be combined with other both quantitative and qualitative methods in order to improve evidence-based policy?
- How to enhance cost efficiencies at the various stages of the peer review process?
- How to develop an effective international frame of reference for expert review?
- How to manage the conflicts of interest in the expert review process?
- What opportunities does the internet offer for improving expert review?
- What type of expert review is fit for the evaluation of policy, programmes, or PROs?
- What are the key principles for a high-quality programme/policy expert review?

2. Definitions and applications

2.1 Definitions

382. There are several definitions on peer review. Hartmann & Neidhardt (1990) define peer review as various processes to evaluate the quality of research by peer scientists. Chapman & Farina define peer review as "a process of assessment on research proposal by peer scientists" (Chapman & Farina 1983). Kruytbosch (1989) also provides a simple definition of peer review in science as "advice about proposed actions solicited by decision makers from recognized experts in relevant technical areas." Chubin & Hackett(1990) say that peer review is an organized method for evaluating scientific research in order to enhance the exactitude of research process, evaluate the authenticity of results, and allocate scarce resources.[34] An OECD document provided a comprehensive definition of peer review as follows (Gibbons and Georghiou 1986):

> *Peer review is the name given to the judgement of scientific merit by other scientists working in, or close to the field in question. Peer review is premised upon the assumption that a judgement about certain aspects of science, for example its quality, is an expert decision capable of being made only by those who are sufficiently knowledgeable about the cognitive development of the field, its research agenda, and the practitioners within it.*

383. Peer review in this form is intrinsic to the practice of science, being used in publication, career and resource allocation decisions. It is widely used by industry, government, and academia. Increasingly it

---

33. For these reasons, the 2005 OECD-BMBF Conference on Evaluation and subsequent meetings have highlighted a number of issues in the domain of peer review of research. For example, see *http://www.internationales-buero.de/de/2193.php* and *http://www.pragueforscience.cz/Scientific-Programme.php*

34. Some people use the term "peer advice," "peer evaluation," "peer judgment," "quality control," "peer censorship," "merit review," "refereeing" as an equivalent.

is also being used as an instrument for ex-post evaluation. The model of peer review has also been extended to encompass additional criteria, notably socio-economic criteria and the potential to contribute to innovation as well as other considerations of merit beyond scientific quality. According to this trend, EERE's Peer Review Guide (2004) defines in-progress peer review as:

> *A rigorous, formal, and documented evaluation process using objective criteria and qualified and independent reviewers to make a judgment of the technical, scientific, and business merit, the actual or anticipated results, and the productivity and management effectiveness of programs and/or projects.*

384. As we can see above, there is no single definition of peer review used in the evaluation literature. However, all of definitions of peer review adhere to the fundamental concept of a review of scientific or technical merit and socio-economic impacts by individuals with professional competence and no unresolved conflicts of interest (GAO 1999; Guston 2001).

From Peer to Expert Review

385. Expert review however is a broader concept than peer review.[35] The classical definition of a peer is "A person who has equal standing with another." A peer review, then, could be defined as "A review of a person or persons by others of equal standing." The crucial issue then becomes how "equal standing" is defined. For example, although scientists who participate in an evaluation may be identified as the "peers" of the applicants when evaluating research proposals, in a programme evaluation, experts in other fields in addition to peer scientists should be included. The term "expert review" is therefore more appropriate than peer review for an evaluation of a programme. The term "expert review" could be defined as follows (Ruegg and Jordan 2007):

> *Qualitative review, opinion, and judgment from individuals with professional competence on the subject being evaluated, based on objective criteria.*

386. The best-known form of expert review is actually ***peer review***, developed from the premise that a scientist's or engineer's peers have the essential knowledge and perspective to judge the quality of research and are the best qualified people to do so. Peer review is commonly used to make many kinds of judgments: about the careers of individual researchers, about the value of their publications, about the standing of research institutions, and about the allocation of funds to individuals and to fields of research (COSEPUP, 1982). Some people therefore often use the term "peer review" instead of expert review.

387. In conclusion, expert/peer review has distinguishing characteristics such as being a *qualitative* method, judgement by qualified *individuals*, and based on *objective criteria*. Whichever definition one uses, the following three issues might be key for high-quality expert review or peer review. That is, who should be the evaluator? How to enhance the credibility of subjective opinions and judgements of individuals? How to develop and provide materials and criteria to the evaluators for objective evaluation?

---

35. According to COSEUP (1999), 'expert review' could be classified into three types: *i) peer review*, which is commonly used to make judgments about the careers of individual staff members, the value of publications, the standing of institutions, and the allocation of funds to individuals, organisations and fields of inquiry; *ii) relevance review*, which is used to judge whether an agency's programmes are relevant to its mission; and *iii) benchmarking*, which is used to evaluate the standing of an organisation, programme, or facility relative to another.

2.2 Purposes and applications

388. The evaluation of a policy/programme involves assessing one or more of five domains (Rossi 2004): *i)* the need for the policy/programme, *ii)* the policy/programme's design, *iii)* its implementation and service delivery, *iv)* its impact or outcomes, and *v)* its efficiency. The general goals of the evaluation relate mainly to programme improvement, enhancement of accountability, or knowledge generation (Chelimsky, 1997). Expert review is one method of evaluation. Therefore, in a basic way, expert review is used to help policy makers reach their goals.

389. According to the literature (Kostoff, 2004; Alassaf, 1996; Armstrong, 1997; Cram, 1992; Levine, 1988; Palli, 1993; Rainville, 1991; Ramsay, 1989; Stull, 1989; Wakefield, 1995; Wicks, 1992), expert reviews of projects and programmes serve a broad range of purposes:

- It serves as quality filter to conserve scarce resources;
- Papers published in peer-reviewed journals are assumed to be above a threshold of minimal quality, such that the reader can focus limited time resources on the highest quality documents assumed to be contained in journals;
- Projects and programmes selected for initiation or continuation by expert review are assumed to be above a threshold of minimal quality;
- Precious labour and hardware resources can be focused on these high quality tasks selected;
- Expert review has the potential to add value to, and improve the quality of, the manuscript or programme under review;
- Expert review can provide a mark of approval for legitimacy and competency to increase a programme's visibility and support;
- The objectives of expert review range from being an efficient resource allocation mechanism to a credible predictor of research impact; and
- A properly conducted expert review of a research programme can provide research sponsors with a credible indication of the programme's quality, relevance, management, and appropriateness of direction.

390. Policy makers and programme managers want to know through evaluation whether their research is being done right (*e.g.* has high quality and efficiency); whether the programme's R&D efforts are focused on the right areas; how programme-created knowledge finds varied applications that generate additional benefits to the nation; how collaborations and other activities stimulated by the program have affected the nation's R&D capabilities; and if their past efforts or new planned initiatives are worthwhile, and so forth. A good expert review should be able to provide programme managers and policy makers with answers to these questions. Ruegg and Jordan provide a good summary of uses of *programme* expert review as follows (Ruegg and Jordan 2007)[36]:

36. Policy makers and programme managers want to know through evaluation: If the research is being done in the right way (*e.g.* has high quality and efficiency); If the programme's R&D efforts are focused on the right areas; How programme-created knowledge can find various applications that generate additional benefits to the nation; How collaborations and other activities stimulated by the program have affected the nation's R&D capabilities; If their past efforts were worth it and if planned new initiatives will be worth it.

- To conduct in-progress reviews of scientific quality and productivity;
- To help answer questions about the relevance, timeliness, riskiness and management of existing programme research activities, and resource sufficiency of new programme initiatives;
- To score and rate projects under review to aid decisions to continue, discontinue, or modify existing or planned project, programmes, or programme initiatives;
- To help assess the appropriateness of programme mechanisms, processes, and activities and how they might be strengthened;
- To integrate across multiple evaluation results and render judgments about the overall success of a programme or programme initiative;
- To provide information to help programme managers make decisions to design or revise their programme, re-direct existing R&D funds, or allocate new funds.

2.3 Merits and limits

391. Like other methodologies used in evaluations, expert review has its own strengths and limitations. This section summarises the *merits* and *limitations* of expert review. The merits of expert review can be understood as follows:

- *Expert review is relatively fast and convenient*. Given that the most appropriate experts are selected, expert review, in any form, may be very time-efficient;
- *Expert review may be carried out in diverse situations*. It is also easy to persuade stakeholders for the following reasons. One often finds numerous experts on a given evaluation object; because these experts participate as a third-party, it is easy to persuade the evaluated and the stakeholders;
- *Expert review is relatively cheap*. Because it involves using existing knowledge of the experts, one may reduce the costs of additional analyses;[37]
- *Expert review provides opportunities for mutual learning to those involved*. There is much discussion and exchange of ideas through expert reviews and one may find intended and/or unintended benefits from such activities.

392. Despite the merits, there are many limitations to expert review, including:

- *Difficult to ensure accuracy and quality of the resultant evolutions* when expert review is applied to the impact assessments of research and development programmes. Therefore, its usefulness as a method to guarantee reliability and consistency (or repeatability) is limited.
- *The quality of a review is limited by the biases and conflicts of interests* of the reviewers. Although one may reduce biases and conflicts of interests through various measures, in principle, they may never be completely eliminated.
- *Expert reviews tend to perpetuate orthodox and conservative paradigms*, and tend to reject new paradigms that threaten the structure of the status quo.

---

37. However, there are considerable hidden indirect costs in expert review. Also, as regards programme evaluation, the actual cost may increase by a large amount due to additional resources needed to analyse the programme.

393. The second and third drawbacks mentioned above in particular pose challenges to the reliability of or confidence in expert review. These risks mainly concern the review of grant applications or scientific papers (i.e. project level expert review) and have been most frequently examined in the context of "peer review". While reviewers should be as objective as possible, in practice the peer judgment is affected by different factors (e.g. bias, favouritism, conservatism, discrimination, and so on) which have nothing to do with the subject of the evaluation, which raises the risk of a crisis of confidence

394. The first *bias* is known as the "Matthew effect". According to the "Matthew effect", the allocation of research funds is more likely to be skewed towards more famous and influential researchers, and researchers who received funds before have a higher propensity of getting funds repeatedly (Merton, 1973). Gustafson (1975) shows that 46% of the entire research funds is awarded to the top 10 research organisations in NIH, and the top one-third of the total funds goes to the top 20 organisations in the NSF. The "Matthew effect" can be a severe problem especially when not enough research funds are available, and such a problem is often pointed out more by those who are unsuccessful in their proposals (Pouris, 1988).

395. Peer review by definition is not immune to the risk of *cronyism* as the established scientists mutually support each other. *Informal cartels* or personal connections play an important role especially in the evaluation of a major project that may have a large impact on a researcher's reputation. The selection of panel members and their evaluation processes may also be influenced by *favouritism* and *discrimination*. For example, when a member holds a key post too long in the evaluation committee and that person may even appoint his/her successor personally, the evaluation committee may not represent the entire science community but reflect interests of only a certain group. Such a problem can lead to discrimination against certain groups, including women, young researchers, and researchers who work for less renown institutes and universities (Gustafson, 1975). It is therefore very important for an objective and fair evaluation to avoid the effects of the social replication or the so-called "Old-Boys-Network."

396. The *conservatism* of peer review has also been criticised by many. Peer review can be seen as supporting an orthodox and conservative paradigm whereby it is hard to accept a new and innovative idea that may threaten the stability of the present structure. Given that the goals of evaluation is to promote, support, and indemnify the new innovation and paradigm in science and technology, it is important to choose members of the review panel who value these goals in their evaluation. One of the weak points in peer review is that only specialists who know their own specific fields make up the review panel instead of experts who have a broader view. Established scholars may be satisfied with a present position in the science and technology community and may be against any new paradigm that may threaten the current paradigm. If the review panel is composed of the only mainstream researchers, they will be more concerned with questions such as "is this research successful?" as opposed to more fundamental questions like "is this research really needed?" Such a narrow evaluation will come to support only the views of the mainstream scholars (this is known as "Pied Piper Effect" in the literature (Kostoff, 1996).

397. Certain authors note that peer review is quite conservative in its analysis and therefore not able to acknowledge the scientific achievements of other fields (Bozeman and Melkers, 1993). The established fields may also have a better chance than new fields to obtain a grant because new fields are placed at a disadvantageous position in accessing the mass media and in lobbying (Pouris, 1988).

398. Ethical issues also threaten confidence in the peer review process. There are many ethical issues in the scientific community: fraud, plagiarism, fabrication, image manipulation, leakage of commercial confidentiality etc. (Campbell, 2006). For example, plagiarism and wilful delay in the evaluation by reviewers can damage researchers' interests. It is relatively easy for reviewers to appropriate or use the grant applicant's ideas by delaying the evaluation process intentionally especially when reviewers conduct research on the same topics as grant applicants. In addition, a leading scholar in a certain area may not

want to see another rival who might challenge his/her authority later so he/she would try to hold him back by criticising a new researcher's work inadequately (Pouris, 1988). Scientific misconduct like this has enormous impacts but is often hard to document and prove. In fact, much scientific misconduct in science and technology originates in the peer review process. The academic world is spending a great deal of effort to prevent such misconduct given that one of the purposes in the peer review is to protect the ethical values in the science and technology community (Goodstein, 1995). Recent work by the Global Science Forum addresses scientific misconduct in more detail (see*: http://www.oecd.org/dataoecd/37/17/40188303.pdf*).

3. Key processes of expert review

399. Even though the purpose of this paper is to highlight the challenges to expert review and present some emerging solutions, the key processes of expert review of programmes/policies deserve to be touched upon briefly because the evaluation of programmes and policies have a different focus and deal with different uses, different stakeholders and a different level of complications than evaluations of research projects. The section below covers the key processes of the expert review of programmes while highlighting the important aspects of these processes.

400. Good examples of the process for expert review at programme level are provided in several existing national or institutional guidelines (EERA, 2004; Kostoff, 2003; Kostoff, 2004; Rigby, 2002; The British Academy, 2007; EPA, 2000).[38] In relation to the review process, Kostoff suggests the following five phases: *i)* initiation of the review; *ii)* establishing the foundations for the review; 3) preparing for the review; 4) conducting the review; 5) post-review actions. EERE's guide describes four phases: *i)* Preparations;*ii)* Pre-Review; *iii)* Conduct of the Review; *iv)* Post-Review Activities. EPA's *Peer Review Handbook* (2000) describes three stages: *i)* Planning a peer review, *ii)* Conducting a peer review, *iii)* Completing a peer review. Rigby (2002) suggests twelve key steps as follows: Setting the Terms of Reference; Overall Time Available; Appointment of Panel Chair; Appointment of Panel Members; Appointment of the Panel Secretary or Scribe; Operating Procedure; Schedule of Work of the Panel; Links from Panel to Programme/Client and other Sub-contractors; Identifying the Requirement for External Support; Interim Reporting; Final Reporting; Dissemination. As above, although there are differences in the literature, expert review is generally understood to have the following three phases: Pre-review – Implementing review – Post-review phase. This section describes the main phases and steps of expert review with key steps and actions mainly based on the materials mentioned above.

3.1 Pre-review

401. The pre-review phase is as a preparation and planning stage and includes the following three activities: establishing the foundations of the review, selecting reviewers, preparing tools and materials.

3.1.1. Establishing the foundations of the review

402. ***Initiation of the review: Assigning the responsibilities***. A successful R&D programme expert review requires full participation by the unit undergoing a review. With few exceptions, no one likes or wants to be evaluated. How, then, can the stakeholders be motivated sufficiently to participate fully, and insure that the best review product will results? Motivation and participation derive from the actions of organisation's senior management at the initiation of the process. In this process, it is of utmost importance that a senior manager (that is, a senior decision-maker in evaluation agency) sends out an initial letter to all participants including: the purpose of the review and its importance; the goals, objectives, and scope of the review; the identity and responsibilities of the review manager(s), the general responsibilities of the

38. The EERE guide and Kostoff (2004) provide information and examples useful for planning, conducting, and utilizing expert reviews based on best practices in the US.

reviewers, and the responsibilities and reporting chain of the reviewers through all phases of the review process etc. (Kostoff, 2004).

403. ***Identifying the purpose and scope of the review***. Once the responsibilities have been assigned by the senior manager, the principles that govern the review must be established. The first step is to determine the purpose and scope of the review within the context of other review and management activities. Identifying clear the objectives of the review and the boundaries of the programme to be reviewed provides a framework for the remainder of the review. If the purpose is unclear and if the scope is too large, the evaluation gets confusing and the evaluation questions get nebulous. On the other hand, if the scope is too narrow, it is difficult to have a birds-eye view of the programme, and it is hard to draw conclusions on the redistribution of resources and modifications vis-a-vis other programmes. The smallest unit of review also should be determined at this stage. General speaking, at the project level, the review focuses on whether the "projects are being done right" and many of the reviewers have a high level of topical expertise. At the programme level, the focus is on whether the "right things are being done." Evaluation on R&D programme may include in-depth technological reviews of the accomplishments of S&T projects within the programme. It may also fix programmes as the review unit and assess the uniformity of the programme with the policy objectives, the relationship with other programmes, the relevance of the project portfolio, and the relationship with the external environment. Therefore, the review unit needs to be selected in advance according to the objectives and the uses of evaluation.

404. ***Identifying the evaluation criteria and review questions to be used***. Expert review requires pre-established evaluation criteria. Evaluation criteria should be identified and selected primarily by the mission and review objectives as well as the nature of programme and material being reviewed. The criteria and related standards for judging any aspect of the programme reflect the programme's definition of success and characteristics of the programme or projects. The criteria should focus on the right questions and the tough questions, the questions that most need to be discussed by an objective expert group. Criteria and associated questions need to be stated as clearly and succinctly as possible to reduce the likelihood that reviewers will use their own interpretation (EERE, 2004).

405. The fundamental evaluation criteria for a R&D programme are research quality, research relevance, and overall programme quality. For some evaluations, the fundamental evaluation criteria have been further subdivided into research merit, research approach/plan/focus/coordination, match between resources and objectives, quality of research performers, probability of achieving research objectives, programme productivity, potential impact on mission needs (research/technology/operations), probability of achieving potential impact on mission needs, potential for transition or utility, and overall programme evaluation (Kostoff, 1997a; Kostoff, 2004).

406. For example, there are a few criteria that are often recommended and used by the DOE, OMB, NAS, and others. Although programmes may choose to define additional criteria, at a minimum all EERE programmes are expected to use the following three criteria (referred to as "core criteria"). The three core criteria are the following: *1)* quality, productivity, and accomplishment, 2) relevance, 3) management. In addition to specific criteria, reviewers often could be asked to provide an overall assessment. The OMB R&D Scorecard of the US provides another example of criteria (US DOE FY 2002 R&D Scorecard): 1) accomplishments, 2) relevance, relevance of future research), 3) approach to performing, technology transfer/collaboration. Asking specific questions has an advantage that it becomes easier for the reviewer to do the job requested. Therefore, evaluation criteria are often presented to evaluators as questions tailored to the particularities of the evaluated project or the programme. Of course, these questions will not be applicable to all programmes.

407. ***Identifying information needed and data collection/analysis processes***. Once the purpose, scope, criteria, and questions of the review have been determined, attention turns to the review process itself: What type of review should take place? How should one collect and analyze necessary data and transmit them to the evaluators? How should one assemble evaluation results from the evaluators? The focus of information and analysis certainly depends on the particularities of programme/policy and as well as the objectives and uses of the programme/policy. For example, if the main objective of the evaluation is on the performance of a programme, the collected data is focused on the performance of the programme and analysis is focused on output, input and impact of the programme. On the other hand, if the objective was to modify a programme or to decide on the continuation of a programme, analysis on the relevance of the programme as well as its portfolio is important.

408. The data collected must be sufficient for reviewers to judge the set of activities against the standards that have been set by the definition of the criteria and the specific questions. The data includes material that is provided prior to the review and during the review. A balance must be struck between having too much data and not having enough data. To the extent possible, the burden on researchers should be minimized by using materials already developed or planned for other purposes, rather than developing new materials just for the peer review. Depending on the type of programme, data can include the following (EERE, 2004): information on the programme or project mission, goals, and targets and milestones including data on how funding is allocated across key activity areas; summary project reports, plans, and budgets; principal investigator or project manager presentations; lists of publications or patent applications and the results of citation analysis; customer surveys, available impact studies; various reports prepared by other external groups; and/or any additional data and information reviewers may request.

409. ***Identifying the types of review group and the audience.*** In programme/policy review, the competence of the review group might be more important than the individual reviewer's technical competence. The selection of the type of review group therefore is an important issue, and should be addressed at the initiation of the review process. Many types of groups are possible in order to achieve the aim of the review: For example, 1) an independent panel which is a group of experts independent of the agency, and typically funded under a contract; 2) external reviewers group which consists of experts individually contracted to the agency. Which type of review group to select depends on the objectives and the particularities of the programme or policy. Generally, in the case of expert review whose purpose is to assess the performance or the accountability of a programme, an independent panel is frequently used.

410. A programme review could provide an excellent forum for disseminating programme information and results to a wide audience. A determination therefore needs to be made early in planning about whether or not the public will be invited to be present or participate in the review sessions. Care should be taken to insure that the review audience includes: actual and potential customers, stakeholders and other oversight groups, co-sponsors, users, and other agency representatives (Kostoff, 2003 & 2004).

411. ***Establishing a timeline and determining logistics for the review***. Timing is an important factor because evaluation is not for academic research but for practical use. Consequently, after setting the date of the presentation of evaluation results which will serve as the basis for the timeline, major deadlines of the evaluation process should be clearly determined in advance. The primary intent of programme review is to provide information that assists programme managers and staffs in their efforts to improve programme performance. The timing of when the report becomes available to provide useful input is therefore also important. Of course, resources (time, money, people etc.) need to be considered to identify the logistics for the review. Although in theory, resources (time, money) are determined by programme size, objectives of evaluation etc, in practice these resources are scarce. Therefore, while respecting the definite timeline and the format of the evaluation, one need to take these limitations into account when determining specific logistics and concentrate the limited resources on key issues and fundamental processes.

3.1.2. Selecting and inviting the reviewers

412. ***Identifying criteria of selecting reviewers***. When seeking nominations, it is important that the criteria for selection of reviewers be clearly presented. The review manager, working with staff, the external steering group, if any, and others establish qualifying criteria that individuals should meet for selection to the peer review panel. These qualifying criteria include: 1) in-depth knowledge of the subject area for which he/she is being selected; 2) that reviewers have no real or perceived conflicts of interest.

413. ***Developing a list of possible reviewers and nominate***. Once the overall technical description of the programme is generated, and technical descriptions of the technical sub-areas are provided, the identification of the reviewer can be initiated. Sources of candidate reviewers can include: programme manager recommendations, membership lists of prestigious organisations, agency review boards, agency consultant pools, contributors to technical databases (such as journal article authors or technical report authors), and other similar lists. The review manager, working with the external steering group and/or others, develops an initial list of candidate chairpersons and reviewers according to like the following: 1) Arranging for several independent, external, and objective groups familiar with the programme to nominate candidates; 2) Identifying candidate chairpersons and reviewers from experts identified in a bibliometric search of the published literature on the topic, or from their roles in research or management institutions or professional societies; 3) Employing a co-nomination approach for identifying and nominating reviewers, where reviewers are selected from those nominated by more than one external expert in the relevant field.

414. ***Gathering background information and developing an initial selection list***. The review manager develops information on the candidate chairpersons and reviewers using approaches such as the following:

- Reviewing the performance of reviewers in past reviews, noting who did or did not meet selection criteria based on this experience.
- Contacting candidates to determine their general interest and availability; sending them project summary descriptions to further identify interests and possible conflicts; and requesting and reviewing self-assessment forms.
- Obtaining staff and/or public input, as appropriate, to identify candidates that may have known biases or other issues. Considerable care is needed here to prevent gathering of materials or other input that could unfairly or inappropriately characterise an individual and to make sure that privacy or other concerns are not raised.

415. ***Selecting the chairperson and reviewers from list of nominees***. The review manager should select the chairperson and reviewers from the list of nominees by working with the external steering group, the chairperson (after selection) and/or others, using processes such as the following:

- Arranging for independent, external, unbiased, objective university, professional society, or other groups familiar with the programme, as identified above, to select the chairperson and/or the reviewers from the nominees.
- Selecting from the nominees the review chairperson, who then chooses the rest of the reviewers.
- Identifying the chairperson and the reviewers based on a co-nomination process among the candidates, as described above.
- Using an independent, unbiased, objective contractor to select from the nominees either directly, or in collaboration with the steering group, independent, external, unbiased universities, professional societies, or others.

- The selection process should be carefully and fully documented to ensure transparency, as other aspects of the peer review process are, and included in the final peer review report.

3.1.3. Preparing Tools and Materials

416. ***Developing guidelines and tools for the review***. Both the review panel and the presenters should clearly understand the objectives and guidelines for the review as well as the specific evaluation criteria that will be addressed. The review leader and chairperson should determine how the projects/program would be rated and distributed to both reviewers and those being reviewed a written description (evaluation guidelines) of the evaluation method. These guidelines should describe the purpose and scope of the review, the evaluation criteria and questions, data to be presented, and how the data will be collected from reviewers, analyzed and reported.

417. Rating or scoring systems are often used to improve the effectiveness of the evaluation. In this case, clear standards should be provided. The comparability of ratings across peer reviewers and review groups requires that all reviewers use the rating scale in the same way. Thus, it is imperative that the scale be well defined so that all reviews are calibrated in the same way and an adjective or numerical rating will represent the same cognitive appraisal by different reviewers.

418. ***Developing the presentations***. Although in the case of research project review, presentations may be easily prepared by the project leaders, it is a lot more complicated to present the evaluation results of a programme as one need to take into account various socio-economic factors as well as the numerous components of the programme itself. Therefore, evaluation managers should provide appropriate guidelines on presentation to relevant managers.

419. ***Providing evaluation material***. Before embarking upon evaluation, one needs to provide the evaluators as well as those being evaluated (e.g. presenter, programme manager) clear instructions as to what materials are needed for the evaluation by when. This way, the evaluated can effectively prepare for the evaluation.

420. It is recommended that a variety of background material be supplied to the reviewers (and the invited audience) before the review. When the evaluated submits background materials and analysis results according to the guideline provided by the evaluation manager, these materials and analysis results must be distributed in a timely manner to reviewers with a guideline clarifying evaluation criteria, processes and indicators. It is important to provide sufficient time for the reviewing of these materials in order to ensure the quality of the evaluation.

421. Evaluation managers can provide documents containing programme accomplishments at this time. Although these documents may be provided during an evaluation, it is better to distribute them in advance. The reviewers may request additional materials in advance after having examined the initial materials.

422. ***Creating an expert review record.*** The peer review record is established at the beginning and maintained throughout the review process. The record should contain all the key documents of the review. This record is an important part of transparency of the process and will aid evaluation manager's efforts to continually improve its expert review process.

3.2 Conducting the Review

423. ***Providing on-site instructions to the reviewers***. Having provided reviewers with written direction prior to review, it is recommended that the review leader or chairperson reinforce guidelines orally at the opening of the review. This will ensure that the reviewers are clear on what is being asked of them and clarify the purpose of the particular peer review. This provides time to settle any outstanding reviewer concerns or questions before the review begins. And reviewers are instructed to keep all evaluations strictly confidential during and after the review.

424. The specifics of on-site instruction depend on choices made by the review leader, review chairperson, and/or group. However, in general, reviewers could be instructed to: *i)* read and understand the evaluation criteria and peer review procedures; *ii)* evaluate each programme element; *iii)* prepare preliminary comments on the merits of the project/program in accordance with the peer review evaluation criteria; *iv)* be prepared to discuss each project and/or the program at the meeting or assign a rating or ratings that reflect the reviewer's opinion of the merit of the project/program in accordance with the specific evaluation criteria, and; *v)* complete the post-review evaluation form.

425. ***Programme presentation and Q&A***. Given that expert review promotes new ideas through discussions between evaluators and the evaluated and provides mutual-learning opportunities, presentation is a crucial step in this process. Concerned parties from various levels – organisation unit head, programme manager, technical unit head – could give presentations; the content of the presentation depends on the presenter. For example, the broader technical portion of the presentations is initiated by the head of the organisational unit in which the program resides, and it includes the following informational material: the mission and objectives of organisational unit, a list of all programs in organisational unit, a description of objectives of each program, the funds and people associated with each program and with the program to be reviewed, an overview of the accomplishments and transitions of programs not being reviewed, and their relation to the accomplishments and transitions of the organisational unit's mission and potential national impact, etc. And the program manager(s) provides a more detailed overview of the program under review, including: objectives of program under review; requirements to be met and derived target capabilities for the S&T initiative.

426. ***Discussion and judgment.*** Reviewer-to-reviewer interaction, for example in a special closed session to discuss their preliminary rating and then finalise each of their individual ratings, can improve the quality of the review findings. This discussion can be useful for clearing up misconceptions or bringing in new information. Such interactions may be particularly important at the higher level program review in order to better understand the full range of issues. The review chairperson needs to ensure that no single reviewer dominates the ratings discussion and to make clear that consensus is not expected.

427. After the discussion among the reviewers, judgment takes place on the level of evaluation panel or evaluation committee. Sometimes individual opinions of the reviewers are merely accumulated and sometimes, a consensual judgment is reached based on the individual review results. On programme evaluation, the latter option is often preferred. What is important is that this choice of final judgment method must be determined in advance in the preparatory stage of the review.

428. ***Synthesising evaluation results from the reviewers***. After discussion and judgment, the evaluation results submitted by individual evaluators or by an evaluation committee are confirmed and synthesized for the final report. In the case of evaluations which determine priorities among different programmes, a rating or scoring system is often used. When this is the case, the type of rating or scoring system which will be in use should be determined in advance.

429. ***Developing review documents and report***. The peer review report provides managers with an independent assessment of the programme's productivity, relevance, and management. The report should include the following features when applicable:

- Programme/project identification, description, and budget;
- A narrative summarizing the salient features of the comments of the individual reviewers and their primary reasons for their judgments;
- Support of conclusions with specific observations;
- Summary of reviewers' rating or assessment on each individual criteria as well as the overall assessment;
- Actionable recommendations aimed at improving program performance, including areas where further study is desirable;
- As appropriate, comments on the status of recommendations made at prior reviews; and
- Appendices with the full text of reviewer input.

430. The review chairperson concurs and signs off on the report, which is often also sent to reviewers for review of the record of their own response. With the conclusion of this report, the "conducting review" phase comes to an end and the report is distributed to stakeholders such as the programme manager.

3.3 Post-Review Process

431. ***Integrating additional comments.*** Before the report is distributed publicly, the evaluated like programme manager develop and add their response to reviewers' comments and recommendations. And any additional comments about the review, either from the reviewers, the external audience, or senior management should be considered and integrated into the review report.

432. ***Drafting a final report***. In general, there are two forms of the final report, a long version and a short version. The long version includes all the written material that was generated during the course of the review. It provides an archival record of exactly what was done during the review. The short version would summarize the process details, and would focus on reviewer comments and other significant inputs, conclusions, and recommendations. And the final report should include the viewpoints of all the reviewers, with appropriate weightings given for judgment and expertise of specific contributors.

433. ***Make the report available to the public.*** When the final report is presented to policy-level decision-maker or higher-level committee and is recognized as official evaluation results, the report should be available to related parties as well as the general public through publications and the Internet.

434. ***Assigning action items and evaluating responses to action items.*** If internal management accepts the conclusions and recommendations of the report, action items should be assigned to the appropriate personnel for responding to problems identified in the report. There are many types of responses possible such as a corrective action or a rebuttal disagreeing with the conclusions and recommendations. The response therefore should be evaluated, and appropriate follow-up action taken. These action items, responses, and follow-up actions should be presented at the introduction of the next review.

435. ***Evaluate the expert review process itself, including the lessons learned***. This step is considered as a type of meta-evaluation. Expert review is used as valuable resource for improving future expert reviews by providing information on problems faced during the process, suggestions and requests by the stakeholders.

Table 5. Phases and Key Actions for the Expert Review

Phases	Key actions
Pre-Review	**Establishing the Foundations of the Review** • Initiation of the review: Assigning the responsibilities (K) • Identifying the purpose and scope of the review • Identifying information needed and data collection/analysis processes • Identifying the evaluation criteria and review questions to be used • Identifying the types of review group and the audience (K) • Establishing timeline and determining logistics for the review **Selecting and Inviting the Reviewers** • Identifying criteria of selecting reviewers • Developing a list of possible reviewers and nominate • Gathering background information and developing initial selection list • Selecting the chairperson and reviewers from list of nominees **Preparing Tools and Materials** • Developing guidelines and tools for the review • Developing the presentations • Providing evaluation materials • Creating the expert review record
Conducting Review	• Provide final instructions to the reviewers • Programme presentation and Q&A • Discussion and judgement • Synthesizing evaluation results from the reviewers • Developing review documents and report
Post-Review	• Integrating addition comments • Writing a final report • Make the report available to the public • Assigning action items and evaluating response to action items • Evaluate the expert review process itself, including lessons learned

Source: Adapted with changes form U.S. DOE EERE (2004), EERE Peer Review Guide: Based on a Survey of Best Practices for In-Progress Peer Review, August 2004; Kostoff, Ronald N. (2003), Science and Technology Peer Review: GPRA, Office of Naval Research. Kostoff, Ronald N. (2004), Research Program Peer Review: Purposes, Principles, Practices, Protocols, Office of Naval Research; Rigby, John (2002), Expert Panels and Peer Review," Fahrenkrog, Gustavo, Wolfgang Polt, Jaime Rojo, Alexander Tubke, and Klaus Zinocker eds., RTD Evaluation Toolbox: Assessing the Socio-Economic Impact of RTD-Policies. IPTS Technical Report Series, EUR 20382 EN.

4. Issues and suggested solutions

4.1 The changing context

436. There are a number of changes in the environment which affect how expert review operates. These changes offer new challenges and opportunities for expert review.

437. *Emphasis on performance*. There is more emphasis on the evaluation of the results and the performance of public policies and utilization of evaluation results as a result of 'new public administration' promoted since the 1990s in the UK, Australia, New Zealand and the United States. For example, in order to enhance the accountability of government programmes, the US'GPRA (Government Performance and Result Act) requires Performance-based Management, and Performance-based budgeting.

438. *Progress on international benchmarking and the internationalisation of evaluation.* In many OECD countries, there has been an increase in international benchmarking regarding policies on science and technology (OECD, 2007d). This could be seen as a continued effort to promote the quality and the objectivity of evaluation. As seen by European Science Foundation's member organisations, it is particularly the case among EU member states.

439. *Development of methodologies*. Recently, there have been more efforts to evaluate programmes/policies using quantitative indicators and these efforts have led to the development of new indicators. Also, various methods have been developed to measure socio-economic impacts of the programme. Therefore, there is growing interest in coming up with ways to effectively complement expert review with other evaluation methods.

440. *Requirements for greater transparency.* Given the limited resources for research and development, there is greater competition in priority-setting. This requires more transparency in priority-setting processes. Elimination of biases and conflicts of interest in the evaluation process also remains as a challenge.

441. *Development of information and communication technologies.* There is greater flexibility in expert review with the development of various communication tools such as phone-conference, video-conference and the Internet. Also, with the expansion of the Internet and the development of online databases, there are no time or spatial limits in accessing and exchanging information. Real time entry and reviewing of evaluation data/information became possible. These developments have contributed to the effectiveness as well as the quality of evaluation and allowed for network-centred expert review. Electronic communications now means that expert review can more easily be an international process, potentially widening the range and number of reviewers.

4.2 Methodological issues and solution based on country experience[39]

442. Over the years peer review has received much attention in the evaluation literature. Studies have suggested a number of challenges, solutions, and issues. And most of these are related to project level evaluation such as grant application, paper publication, and ex-post project evaluation. For example, Wood and Wessley (2007) covers issues in their recent systematic review which are mainly related to grant peer review as follows: Is peer review of grant application fair?; Are peer reviewers really peers?; Is there institutional bias?; Do reviewers help their friends?; Age and getting grants; Gender bias and grant peer review; Misuse of confidential information; Reliability of grant peer review; Does peer review of grant

39. What is covered under this section is centred around issues presented at the OECD workshop held in Paris in October 2007.

applications serve the best interests of science?; Is peer review of grant application cost effective?; Can peer review of grant applications be improved?; Should peer review of grant application be replaced?

443. And Kostoff (2004) describes the strength and weakness of major peer review components and issues, including: Objectives and purposes of peer review; Quality of peer review; Impact of peer review manager on quality; Selection of peer reviewers; Selection of evaluation criteria; Secrecy (reviewer and performer anonymity); Objectivity/bias/fairness of peer review; Normalization of peer review panel; Repeatability/reliability of peer review; Effectiveness/predictability of peer review; Global data awareness; Cost of performing a peer review; Ethical issues in peer review; Alternatives to peer review; and Recommendations for further research in peer review.

444. It is therefore impossible to cover all issues raised. The next section of this paper therefore focuses on a few issues for high-quality expert review in the evaluation of *policy*, *programmes*, and *institutions*. Although targeted toward research policy/programme expert review, most of the issues in this report apply to many kinds of expert review including project selection review.

Issue 1. *Consider socioeconomic factors in evaluation: How to effectively reflect socioeconomic and political priorities and link these priorities to decision making in the expert review processes?*

445. This issue may be one of most important ones in policy making and evaluation. Some decision-makers doubt the ability of expert review to reflect socioeconomic and political priorities. Expert review, in fact, is likely to ignore wider social and economic effects due to its self-oriented and highly scientific approach. Expert review panels are dependent on sound and detailed information on which to base their judgments about a programme's progress or impact, and they are vulnerable to poor and insufficient information. The type of data needed for retrospective impact assessment cannot be created in an expert review panel format. For this reason, expert review tends not to be appropriate for evaluating impacts of programmes (Ruegg and Jordan 2007).

446. How can we solve this problem? A couple of solutions could be suggested. To begin with, reviewers could be provided with a pre-analysis of socioeconomic needs and priorities. For example, the Korean government has been informing evaluators of the National Master Plan of Science and Technology about the results of technology foresight, expenditure priorities at a national level, and the status of public R&D expenditure, and analysis of programme's portfolio and performance during the R&D programme evaluation process (Oh and Kim 2006).

447. Diversifying the fields of experts could be the most common solution. While it is reasonable to compose a review panel of *peers* from the same field to assess the excellence of the research proposal and to judge whether or not to award a grant, it may be inappropriate to construct expert panels who only know about their own specialized field or technology, especially when evaluating a programme or a policy aimed at addressing general social and economic problems. To put it simply, it is important to seek *balance* in various aspects when selecting the review panel (See Box 7).

448. Research & development evaluators should not only have technical expertise but a perspective on the broader issues (for example, the impact of the research, mandate of the programme, economic utility, political and economic effects etc) (Klahr, 1985; Marshall, 1996). Although it would be ideal for an evaluation to have both of these qualities, different evaluators as a team can complement one another to provide the expertise and the broad perspective necessary. Some of experts should have a non-S&T background and have expertise in economics, business, accounting, public relations and policy, industrial policy, and other areas as well. Even in projects related to very specialised technological areas such as biotechnology or nanotechnology, the social and economic impact, needs, relevance, and value that those

projects can bring to society as a whole should be treated as important as the scientific merits of technological advances in a programme and policy level evaluation.

Box 7. Balancing Expertise on the Peer Review Panel

The most important aspect of "balance" is to include the range of respected intellectual perspectives in the scientific and technical community. Considerations in developing "balance" also include:

- Balance between technical specialists and multi-disciplinary types, while ensuring adequate coverage of critical technical disciplines for each project and the overall program;
- Balance between academic, industrial, national laboratory, governmental, and non-governmental organisation perspectives, as well as that of customers;
- Balance between "old hands" and "young bloods";
- Gender balance;
- Geographic balance, possibly including international expertise and perspectives;
- Balance across time, maintaining some continuity with prior peer reviews; and, in some cases, if appropriate,
- Balance across interest groups, including representation from environmental, labour, and other organisations, particularly for higher, program-level reviews.

Source: U.S. DOE EERE (2004), EERE Peer Review Guide: Based on a Survey of Best Practices for In-Progress Peer Review, August 2004 .

449. Recently, the question of how science can be made more relevant to the needs of society is increasingly central not only in science-policy debate but also in project selection (Scott, 2006; Nightingale and Scott , 2007). In Canada, there is an increased expectation to address political and socio-economic priorities. At NSERC, expert review is mostly used to evaluate applications for research grants. Generally, programme managers guide the work of peer panels regarding the goals, criteria and applicable polices, but they are not involved in the peer review. In some programmes, officers make a recommendation based on the peer review input and analysis of merit relative to the selection criteria. Most panels have members from industry, government and university sectors. There is also diversity in the panels including a mix of national and international geographic representation, stage of career, gender, language and size of the institution. The diversity on panels works well in 'problem/priority areas' (OECD, 2007d).

450. Another solution is to establish a dual review committee: One review group focuses on the scientific and technological excellence of the subject of the evaluation, while the other focuses on the relevance and socioeconomic priorities of research. For example, NIH has a 'Dual Review System' for grant applications (Scarpa 2006):

- The first level of review by a scientific review group (SRG) provides initial scientific merit, review of grant applications, rates applications and makes recommendations for the appropriate level of support and duration of the award.
- The second level of review by council makes a recommendation to institute staff on funding, evaluates programme priorities and relevance, and advises on policy.

451. Through this dual review system, it is possible to carry out a proper evaluation of the scientific and technical quality of the research as well as its socio-economic value and utility.

452. "Bicameral review"‒ a slight modification of expert review – may also be applied to the evaluation of programmes. According to bicameral review, research grants are assessed with two different, independent criteria. One is the past accomplishments of the researcher and the other is the proposed

research project. The former is assessed through peer review; the latter is assessed internally, based on the budget (Forsdyke, 1991; Forsdyke, 1993). These methods could be applied to evaluations with the purpose of setting priorities or the allocation of resources. In other words, it is possible to draw a final conclusion from two independent processes assessing past achievements as well as a country's strategic priorities and budgetary concerns.

453. It is also possible to use a *Delphi* method, frequently used in technology foresight. In Netherlands, under the assumption that "when evaluating scientific projects, the best standards come from outside the field," when evaluating a grant in the field of physics, researchers are selected from various fields such as physics, chemistry, mathematics, and astronomy as evaluators. This does not mean that the internal criteria (competence and experience) are not important but that 'relevance' should also be included in the set of criteria. The two-stage Delphi procedure has been adopted in order to preclude bias and misunderstandings. In the first-stage, for each research proposal, the grant applicant is given feedback consisting the questions, criticism and advantages/disadvantages brought forth by 4-6 evaluators. In the second stage, on the basis of the responses from the grant applicant, the evaluator assesses the applicant's abilities, objectives, methods and the general level of research; he/she then sets the priority accordingly (Pouris, 1988). This allows for improving the quality of the evaluation as the evaluators are aware of others' opinions. It also prevents unnecessary conflict and power struggle among the evaluators as it guarantees the anonymity of the evaluators. Most importantly, it promotes the relevance of projects in national research development policy. With increased emphasis on the issue of relevance, the application of this method in evaluating programme/policy would be very useful.

Issue 2. *Interface of expert review with other means of judgment: How to use objectives indicators or ranking tables effectively in order to enhance the objectivity of evaluation result? How to combine expert reviews with other both quantitative and qualitative methods for evidence-based policy?*

454. With the rise of indicator-driven judgements and ranking tables, the interface of expert review with these other means to judgement is also of interest. Policy makers and R&D programme managers have attempted to adopt a more quantitative indicator based evaluation system as a complement or substitute to expert review. The problem that occurs in expert review procedures is the identification of relevant performance indicators that are closely linked with the desired outcomes. Programme theory (or a logic model) is used often in order to develop the most suitable performance indicators in programme or policy evaluation. The level of a future target performance of the programme is often presented in advance by objective numbers as well. Developing the more quantitative indicator based on objective numbers rather than qualitative analysis based on subjective opinions by *experts* is becoming an urgent task.

455. On the other hand, it is also important to raise the accuracy and enhance the subjectivity of the expert evaluation by employing both qualitative and quantitative methodology properly. The qualitative and the quantitative method have their own advantages and disadvantages and taking advantage of the strong points of each can make up for the weak points in the current peer evaluation system.

456. In fact, there are few examples of combining peer review with other tools. For instance, various methods – surveys, case study, sociometric/social network analysis, bibliometrics, historical tracing – have been used with expert judgment in ATP programme evaluation (Ruegg and Feller 2003). Another case shows a similar situation. The technology development programmes in the US Department of Energy (DOE) extensively and successfully utilise expert review to evaluate research and development (R&D) activities at the project and programme levels. In addition to expert review, R&D programme managers in DOE are encouraged to use other evaluation methods in order to obtain information on programme effectiveness and the benefits generated that cannot be provided using the peer review method (Ruegg and Jordan 2007). Application of quantitative indicators should be based on the notion that expert reviewers need to be provided with condensed, systematic, verified, objective information on the research

performance of the groups to be evaluated, and that the grounds for their judgment, or the assumptions underlying it, should become more explicit, thus making the process more transparent (Moed, 2007).

Issue 3. *Cost efficiency of expert review: How to enhance the cost efficiencies of the various parts of the expert review process?*

457. It is also important to enhance the cost efficiency of the various parts of the expert review process including the administration. Given that the evaluation process aims at creating new value, the benefits of evaluation should outweigh the costs. Evaluation costs are easily underestimated in a real world because those costs usually incur as an implicit opportunity cost not an explicit payment. For instance, evaluators have to sacrifice their own working time and performance in order to spend their valuable time on evaluating projects assigned to them, but this kind of opportunity cost is often neglected in the benefit-cost analysis in evaluation due to its nature of implicit cost. According to research, indirect costs related to the time value of evaluators, presenters, staffs and visitors are more than ten times as high as the direct costs like travel expenses (Kostoff 1996). In particular, in the case of the panel evaluation done by renowned experts, the total costs (i.e. sum of implicit and explicit costs) become much higher than direct costs. In sum, given that costs of expert evaluation are not negligible, evaluators should do their best to achieve cost efficiency in the evaluation process.

458. Efforts should also be made to reduce expenses in each review process. The key process of expert review which we have seen in the previous section (Section 3) suggests the possible ways of minimizing evaluation costs in the process of application, selection of expert reviewers, and panel discussion of discipline committees. Often the size and scope of the programme or project determines the venue for the expert review. Scheduling the event using public facilities, meal planning, and audiovisual requirements, all should be completed well in advance of the actual meeting. Typically, meeting logistics are one of the major costs of an expert review. Expenditures vary depending on the number of projects reviewed, the number of reviewers, whether the meeting is open to the public, and the length of the review. Ways of controlling the cost of the review meeting include the following: Structuring the agenda carefully so that the agenda is focused and people's time is used efficiently; making maximum use of teleconferences, videoconferences, and other electronic media to prepare the review panel. This is particularly helpful when international reviewers are involved (EERE, 2004).

459. Building an appropriate database of evaluators will have long term consequences in reducing the cost associated with the selection of evaluators. Given today's internationalisation of science, there is much value in promoting international cooperation in building evaluator databases.

460. There are also many suggestions for reducing the costs in research project evaluation. For example, Klahr (1985) points out that NSF was able to reduce the number of final proposals they should evaluate by one-third using the screening method by comparing the results of "mail review" at the first stage evaluation with the ones of "panel review" at the second stage evaluation. In addition, NIH runs the "Center for Scientific Review (CSR)" to maximize the efficiency in the evaluation process, and CSR also operates "Streamlined Review Procedures (SRP)". SRP was able to save evaluation costs by concentrating on only the quality proposals that rank 50% and above (Lee, Om, and Ko 2000).

461. Various types of alternative methods can possibly be employed with the help of various tools supported by the Internet as it delivers real-time news and information and facilitates networking among the persons concerned. For instance, NSF is operating the "NSF Fast Lane System" for more effective, convenient, and faster administration. It has various applications to prepare, submit, and revise research proposals (*www.fastlane.nsf.gov*). NIH also announces recent policies regarding an assessment of the research proposals officially on the internet through SRP so that researchers can be well aware of the most recent evaluation criteria and policies (*www.drg.nih.gov/refrev.htm*).

Issue 4. *International frame of reference: How to develop an effective international frame of reference for expert review?*

462. There are basically two approaches in selecting a panel: intra-national and international. The intra-national panel is composed of local experts in the field, i.e. academics, professionals, policy makers, etc. This type of panel selection is useful in large countries where the possibilities for selection of experts are more numerous due to well-developed S&T systems, and where the possibility of subjective evaluation is minimal. The international panel is mainly composed of foreign and internationally recognised experts in the respective field. Both approaches can be criticized. The first for its inability to cope with local lobbying of interested groups within the scientific community, and the second for lack of knowledge of external evaluators on certain particularities of the respective country and the possibility that the expertise might be misleading due to the different scientific environments prevailing in different countries (OECD 1998).

463. An international frame of reference has been *increasingly* used as the standard for expert review, with the use of foreign experts being seen as the answer both to potential conflicts of interest in small communities and as a means of assuring stakeholders that the work stands up to global scrutiny. In Finland and Portugal, for example, proposals are submitted in English because it increases the number of international reviewers that can be used. As we can see from the examples of countries including Finland, the receptiveness of evaluation results increases when foreign experts are selected as evaluators (Pouris, 1988; OECD, 2007d). The internationalisation of expert panel is needed more in countries that have a small science and technology community. For example, the Korean government is aware that the Korean S&T society and expert pool is very limited. The Government thinks the internationalisation of evaluation would be a solution to enhance objectivity and reliability, and therefore, especially, tries to enlarge the expert pool including foreign experts. But, at the same time, the Korean policy-makers also know that it is very difficult for foreign experts to evaluate Korean R&D programmes because they should have sufficient knowledge of the Korean scientific community, the context of programmes and related policy, and national strategies (OECD 2007).

464. The internationalisation of science itself is increasingly important in evaluation at the national level. Research has been internationalising and thus requires international reference points in measuring outcomes. Evaluation needs criteria, standards, and benchmarks to assess the quality and achievements of policy, programme, project, or institutes. In a global innovation system national standards or approaches are limited, and therefore increasingly should be defined internationally. Not only should the performance of an institute be assessed in an international environment but also should the effectiveness of policies and programmes. For these reasons, international indicators, evaluation criteria or benchmarks are needed in expert reviews of research institutions, programmes and projects.

465. To be sure, the international frame of reference reflects the growing concerns about the role of science in competitiveness and competitiveness in science. However, caution is needed when policy is transferred across different cultures and contexts, particularly when understandings or policies are incomplete. Therefore, there is a need for *taxonomy* of the internationalisation of expert review.

***Issue 5**. Managing the conflicts of interest: How to manage the conflicts of interest in the expert review process?*

466. One of basic hypothesis of expert review is that expert's judgments are trustworthy and reliable, since the persons who do evaluation have judgment, experience, and a professional ethos. However, decisions made by evaluators are easy to be affected by personal relationships with others and this often potentially prevents the entire evaluation process from being impartial and objective. It is therefore to effectively manage the potential or existing conflicts of interest in the expert review process.

467. The United States Office of Management and Budget's (OMB) Peer Review Standards points out that factors relevant to whether an individual satisfies these criteria include whether the individual: *i)* has a financial interest in the matter at issue; *ii)* has, in recent years, advocated a position on the specific matter at issue; *iii)* is currently receiving or seeking substantial funding from the agency through a contract or research grant (either directly or indirectly through another entity, such as a university); or *iv)* has conducted multiple peer reviews for the same agency in recent years, or has conducted a peer review for the same agency on the same specific matter in recent years. (OMB's Draft Peer Review Standards for Regulatory Science under Executive Order 12866, August 29, 2003).

468. One of the direct ways to avoid such a problem is to exclude evaluators who might have interests with proposers in selecting reviewers. It is however nearly impossible to nominate experts to a review panel who have absolutely no interest. Besides the trade-off between choosing reviewers who are indeed peers and the resulting increased chance of a conflict of interest is one of fundamental dilemmas (Wood and Wessely 2007). Thus, it's better to conclude a mutual agreement out of conflicting views among evaluators in the panel review.

469. Conflicts of interests could also occur between an evaluation manager and a reviewer. Those conflicts are usually related to the questions such as "who is responsible for the evaluation results?" and "how deep should a manager and a reviewer be involved in decision making?" For example, both an expert reviewer and a manager may want to make the final decision on the proposals, not just supports the other's decision making. The severe disagreement regarding resource allocation decisions in the expert review often comes from the conflict of interests among parties involved in the evaluation process. Proper management of interests and dissolving conflicts among parties would thus enhance the receptivity among them, and it is important to construct peer review mechanism that follows objective evidence, not a personal interest, in overcoming such difficulties.

470. Concerning the potential conflict of interests, declaration of interests by the evaluators is proposed as a solution (Bozeman, 1993). The UK Research Assessment Exercise requires declaration of interests in order to avoid obvious or potential conflict of interests. It is even argued that the authors of papers should declare their financial interests (RAE 2001). The scientific journal *Nature* required authors of papers to declare their financial interests.

471. Another solution is to internationalise evaluators as foreign experts may have less biases and interests. The Academy of Finland invited a quartet of British, American, West German, and Swedish experts to evaluate the country's progress in Inorganic Chemistry. The assessment believes that 'it succeeded only because the panel came entirely from beyond the frontiers of Finland (Dixon, 1987; Pouris, 1988).

472. In many review practices, all reviewers must sign a Conflict-of-Interest form prior to the beginning of the review process. In addition, during the review, the reviewer should agree to disclose any actual or perceived conflicts of interest as soon as the reviewer is aware of the conflict.

473. It is advisable to limit the number of evaluations or the duration of evaluating activities for the participating experts. If an expert participates in too many evaluations, the expert might develop a relationship with the evaluated bodies and might be susceptible to lobbying from them. Moreover, trapped in their judgments from previous evaluations, they might not be able to take a fresh look at similar programmes. However, to limit the duration of evaluating activities excessively may be counter-productive. For example, it might decrease a sense of responsibility. If the expert participates in the evaluation only once, their responsibility might be lower than if they were to participate in future evaluations.

474. Lastly, it is advisable to prevent an expert who has expertise only in a particular domain from judging the quality or the value of what is being evaluated. However, it is advisable that his opinions are transmitted to other experts in the panel to be used in the joint-decision making. For example, an expert on biotechnology evaluating a biotechnological program might know more about the technical aspects than other experts but cannot judge correctly the socio-economic value of the evaluated program. Also, they might insist on higher allocation of resources to their area of expertise.

Issues 6*. Expert review in the Internet age: What opportunities does the Internet give us for improved and enhanced expert review? Could an Internet based "open evaluation" tool organized by the scientific community be an alternative to the classical approach? Can network-centred expert review replace classical review? Is evaluation possible without expert review panels?*

475. Technological progress like the Internet provides not only new means and modes of communication but also opportunities for advanced evaluation. Panel review and mail review are the most general type of peer review and both of them are not efficient regards to time and expenses. But using internet in constructing the panel and in evaluating proposals enhances efficiency. Most importantly, peer evaluation systems based on the Internet boost rationality and receptivity dramatically because the Internet conveys all kinds of useful information in real time. That is, all the information regarding text, speech, graphics, music, video, images, 3d-models, and raw data is digitalized and delivered to evaluators, and they can open up necessary information at anytime and anywhere. Besides, the Internet provides evaluators with search engines and alert systems as well as data analysis tools.

476. Internet could enable a new style of peer review. Whether it's a panel review or a mail review, traditional peer review is a "closed evaluation" by the nominated experts group. An Internet-based "open evaluation" tool organized by the scientific community can be an alternative to the classical approach because the internet can secure additional evaluators around the world without a boundary. For example, a project or an evaluation results can be reviewed by the numerous people once it is posted on the Internet. Also, open evaluation is found to be a very powerful tool to solve data fabrication, which is one of the hot issues in science these days. In fact, the people who first caught the data fabrication in "Hwang's affair" in South Korea were Internet users. Finding errors and data fabrication by researchers are almost impossible to be found during the normal panel review that should be done within a rather short time period, but it is difficult for researchers to deceive all potential reviewers on the Internet at once.

477. The publishing system of *Journal of Atmospheric Chemistry and Physics* gives a good example, Interactive Open Access Publishing (Mehlhorn 2006). Papers for *JACP* are handled in two phases:

- In a first phase, the author submits a paper to the editor. The paper is published in *Journal of Atmospheric Chemistry and Physics Discussion* as a paper for discussion. The paper is *openly* reviewed by the scientific community as well as appointed referees with reactions by the author.
- In the second phase, the author is required to submit a revised paper based on comments from referees and the scientific community. The editorial board makes a final decision to publish the final revised version of the paper or not.

478. The review and publishing system of *JACP* has many advantages. It provides authors, referees and readers with: free speech and rapid publication (authors & readers); direct feedback and public recognition for high quality papers (authors); prevention of hidden obstruction and plagiarism (authors); documentation of critical comments, controversial arguments, scientific flaws, and complementary information (referees and readers); deterrence of careless, useless, and false papers (referees and readers); public discussion and final revision (readers). In short, it could be said that JACP's publishing system provides maximum quality assurance of papers through public, interactive, and collaborative peer review.

479. Information technology, such as Groupware software, has the potential to significantly improve the efficiency and overall value of the expert review process. Information technology brings real time data entry, screen sharing, data manipulation, and statistical analysis capabilities to the expert review process. Individual reviewers can enter anonymous review and rating data, and the review manager can compute summary rating statistics to share with them in a timely manner. This increased information handling can free up time to permit additional time allocation for important reviewer-to-reviewer or reviewer-to-review manager interactions. Box 8 compares a network-centric expert review with the traditional review process.

Box 8. Comparing a groupware-based peer review with the traditional review process

Traditional peer review	**Network-centric peer review**
• Data input is via the evaluation form completed during the Q&A session or shortly thereafter. • Each reviewer completes their evaluation during the session, and the individual and summary result for the panels are computed at the end of each presentation day or after the review has concluded.	• All the members of the on-site audience are linked by Group-Ware information technology. All data input is mechanized, and instantly recorded. • Each reviewer completes their evaluation during the session using the groupware. During the presentations, the reviewers enter final ratings and any additional comments they believe are important based on last-minute observations or insights. Individual and summary results for the panel are made available in real-time and routed back to each individual for further discussion.
• Statistical analysis of reviewer comments (summary and integrative statistics, as well as aggregating comments) typically is not available instantly or in time for use in onsite panel discussion. • Reviewers could meet in closed session to discuss their preliminary reviews. However, during closed session discussion, reviewers often do not have access to the full statistical analysis of ratings for the panel.	• Statistical analysis of reviewer comments is completed onsite to provide useful performance data quickly. • To complement the groupware tool, reviewers could meet in closed session to discuss the preliminary reviews and once the interactive cycle is complete, they may make final changes to their individual review comments and ratings. The groupware technology would enable reviewers to have access to the full statistical analysis of ratings for the panel.

Sources: *www.inform.nu/Articles/Vol2/v2n1p11-18.pdf* and Ronald N. Kostoff (2001), *Network Centric Peer Review*, Office of Naval Research.

Issue 7. *Expert review for policy, programme and/or PROs: What type of expert review is fit for the evaluation of policy, programme, or PROs? Is expert review a relevant tool for evaluating research institutions?*

480. This issue is also important for programme level expert review. Because the peer review method is generally used at the project level, it is necessary to consider what type of the peer review (actually, expert review) is appropriate for an upper level decision making in programme, policy, or institution. The outcomes of the review can affect the decision making and those are useful as the reference data as well. For example, an author classified peer review into three categories as follows based on the level of its impact on the final decision making (Bozeman 1993): pre-emptive peer review, traditional peer review, and ancillary peer review. Pre-emptive peer review is the one that the final decision depends entirely on the results of the peer review and a programme manager has no right of judgement. Following the already determined format, scoring model or ranking model is employed in the pre-emptive peer review. The dual review system in NIH is an example.

481. In the traditional peer review, the decision is also influenced by other factors like the decision of a programme manager while the result of the peer review is still the important factor affecting the final decision. Along with the result of the peer review, an academic standard of the organisation a research proposer is affiliated with and a geographic area are also considered in the traditional peer review. NSF typically uses this method. Ancillary peer review can provide only the partial information out of all the crucial data and thus play a minor role in the decision making process. Assuming that different aspects of evaluation should be considered differently in search of the most suitable evaluation method, economic and political areas and the case of geographical distribution of scare resources are evaluated by different evaluation methodologies from the peer review while the peer review is a suitable method for the science and technology field. These are often used in the major programme evaluation or building up a science complex.[40] Because most programme evaluations have many policy issues to consider, among the three types of peer review mentioned above, the pre-emptive review is rarely used.

482. On the other hand, selection of the type of review group is a core issue, and should be addressed at the beginning of the review process. Although there are many types of external expert reviews, two types draw our special attention: the independent panel and the external reviewers group (EPA, 2000; Kostoff, 2003; Kostoff, 2004). The independent panel is a group of experts independent of the agency, and typically funded under a contract. The independent panel has a chairperson, attempts to reach consensus on issues, and generates a written report containing the results of the review and sometimes recommendations. The group of external reviewers consists of experts individually contracted to the agency. The reviewers report to the agency review manager.

483. In contrast, the external reviewers group does not have a chairperson; the review manager serves in this role. While the group may engage in technical discussions during the course of the review, it does not reach a consensus. While there may be individual written inputs from each group member, there is no group report. The review report is written by the agency review manager based on the individual written inputs plus other considerations. Because of the technical understanding required to write a credible report, as well as select the appropriate mix of reviewers, and conduct all aspects of the review, the review

40. Within the general category of expert review, there are a number of sub-types according to the level of specialisation and professionalisation (Gibbons and Georghiou, 1987; Rigby, 2002): traditional peer review (canonical academic review), direct peer review, modified direct peer review, pre-emptive peer review, indirect peer review, merit review (extended form of peer review), ancillary peer review, expert panels, panel review, professional evaluators.

manager should have a solid technical background and some understanding of the subject matter to be reviewed.

484. Each of the two review group approaches has value for specific applications. The group of external reviewers is less formal, and has fewer reviewer and audience restrictions. It is useful for internal reviews where structural program issues are paramount and need resolution or improvement, and where comparison with other programs is not the major focus. The independent panel is more formal. The independent reviewer panel has more specific reviewer, meeting, and audience selection constraints/requirements. From the agency's perspective, either group has very high utility for addressing the agency's programme improvement needs. From a perspective external to the agency, the independent panel has higher credibility because of its independent nature. For performance evaluation or evaluation for priority setting, the independent panel is more appropriate, because of its perceived independence.

485. It is also important to determine whether to make the expert review process, for example, the presentation or the contents of the discussions, publicly accessible. Those in favour of having open-to-the-public reviews (EERE, 2004) suggest that having the review meeting open can: help sharpen the questions raised; improve the transparency of the peer review process; help improve or legitimize the technical or management approach; strengthen integration networks for research, deployment delivery, or business management; broaden public learning by providing an opportunity for individuals to hear firsthand what others are accomplishing and how they manage their work; and encourage participants to improve performance due to the pressures of presenting publicly to their peers. It is generally believed that in the case of evaluations of programmes or institutions, making the evaluation process public has a more positive effect than in the case of evaluation for priority-setting.

5. Ways to High-quality Expert Review

5.1 Essential Requirements for Good Practice

486. What are the key factors for a high-quality expert review? Chubin (1994) suggests seven *requirements* order to enhance the quality and credibility of peer review as follows:

- *Effectiveness*. Peer review should be effective to allocate resources and to set research priorities.
- *Efficiency*. Resources including time, money in peer review should be used most efficiently.
- *Accountability*. Peer review should enhance the accountability of science to not only scientists but also the general public.
- *Responsiveness*. Peer review should be flexible mechanism which can lead the development of new fields and support policy maker's decision of new direction of innovation.
- *Rationality*. Peer review process should be transparent and rational.
- *Fairness*. Peer review should be impartial and observe a social norm that everybody is equal under the law.
- *Validity*. Peer review should obtain the same result from repeated assessment and remove the effect of contingency in review process.

487. In practice is but all but impossible for peer review to satisfy all requirements mentioned above. Some of these values involve a range of contradictions. Some experts consider that peer review embodies tensions between five 'value pairs' — desirable properties that are in tension with each other (Hackett 1997; Scott 2006):

- *Effectiveness and efficiency.* Increases in effectiveness will require more work on the part of peer reviewers and will therefore impose greater costs, while high levels of efficiency will usually come at the expense of thoroughness.
- *Autonomy and accountability*. Wider accountability might reduce autonomy; more autonomy implies less public accountability.
- *Responsiveness and inertia*. This is the tension between tradition and originality.
- *Meritocracy and fairness*. A poor paper by a respected academic may be published due to his or her reputation.
- *Reliability and validity*. Reliable criteria may be narrow and rigid, and thus may not produce the most valid results.

488. The most important point here is finding the *optimal* balance between the contradictory requirements. Trade-offs between desirable properties of peer review are inevitable: there is an ongoing challenge for research funding bodies to be able to determine what constitutes a defensible, appropriate, and workable balance (Wood and Wessley 2007). For instance, although the pursuit of greater effectiveness could enrich the exactitude of evaluation, it requires too much time and resources and reduces cost efficiency. On the other hand, focusing only on cost efficiency may lead to a superficial assessment. Evaluation designers should therefore consider all resources and conditions and choose a more optimal evaluation process and method.

489. Autonomy is one of key values in the professional community but it often conflicts with accountability. Scientists, as experts, would like to decide what and how to do research by themselves. However, the general public wants to see the output or performance of scientists whose work is supported by taxpayers. Furthermore the dissemination of performance-based budgeting has further increased the emphasis on the accountability of public research. It is very important for a successful peer review to find the optimal balance between the contradictory values in the evaluation process.

490. While expert review is one of most flexible methods for determining value, capable of application to a wide number of fields, in order to apply it, a number of critical pre-conditions must be met. Much of the literature addresses conditions or requirements of expert review (especially, focusing on peer review of project evaluation) for applications. Rigby (2002) suggested four essential pre-conditions for applying peer/expert review as follows:

- Experts with knowledge of a particular area must be available and be willing to participate. Because it can be difficult for government officials to identify the relevant peers as they not usually part of the social or professional networks of scientific peers, it is important for programme evaluators and responsible bodies to maintain access to such networks.
- The panel of experts cannot be expected to answer questions which are beyond the scope of the available knowledge. Terms of reference need therefore to be set with some sense of what it is possible for the experts themselves to know or to infer and to judge collectively from their specialist knowledge.
- The panel should only be asked to come to a judgment on a single area of knowledge or expertise rather than more than one as peer review is known to be weak where comparative judgments between different fields of expertise have to be made.
- While the costs of peer review are low, sufficient resources should be made available to facilitate the work of the panel. Some panel reviews are often supported by a secretariat.

491. By definition, a high quality peer review should provide an accurate picture of the intrinsic quality of the research being reviewed, irrespective of whether this intrinsic quality is high or low. The fundamental problem is the lack of absolute standards (analogous to physical standards for primary measurements such as time and length) for measuring research quality. Presently, evaluation of intrinsic research quality is a subjective process, depending on the reviewers' perspectives and past experiences. A high quality review under these imperfect circumstances, then, would occur when two generic conditions are fulfilled: 1) utilization of highly competent reviewers, and 2) no injection of additional distortions in the reviewers' evaluations as a result of biases, conflict, fraud, or insufficient work (Kostoff, 2004).

492. High quality expert review processes require as a minimum the conditions summarized by Ormala (Ormala 1989):

- The method, organisation, and criteria for an evaluation should be chosen and adjusted to the particular evaluation situation;
- Different evaluation levels require different evaluation methods;
- Program and project goals are an important consideration when an evaluation study is carried out;
- The basic motive behind an evaluation and the relationships between an evaluation and decision making should be openly communicated to all the parties involved;
- The aims of an evaluation should be explicitly formulated;
- The credibility of an evaluation should always be carefully established;
- The prerequisites for the effective utilization of evaluation results should be taken into consideration in evaluation design.

493. EERE's *Peer Review Guide* (2004) describes the minimum requirements to be prepared for expert reviews of EERE's R&D programmes as follows:

- **Scope of Review.** All EERE programs in both Technology Development and Business Administration offices and their key projects will be reviewed by qualified and objective peers on a regular basis. This should typically cover 80-90% of RD funding and supporting business analysis and management programs. Earmark projects will be included in the review and treated on the same basis as other activities.
- **Frequency of Review.** All EERE programs and their key projects will be reviewed, on average, every two years, depending on the characteristics of the program and needs for information.
- **Timely Preparation.** Preparation for a peer review will include designation of a review leader, determination of the purpose of the review and the review agenda, and communication of this information to reviewers and those being reviewed in time for them to prepare for the review.
- **Core Evaluation Criteria.** Clear standards for judging the program or projects will be defined prior to the review. This includes the criteria and the kinds of evidence (data) needed to judge those criteria. At a minimum, programs will be assessed on quality, productivity, and accomplishments; relevance of program success to EERE and programmatic goals; and management.
- **Reviewers.** There will be a minimum of three reviewers for each discrete program element or smallest unit that is assessed and reported on. Each reviewer will be independent, competent, and objective, selected by a transparent, credible process that involves external parties. Together the reviewers will cover the subject matter. Reviewers will sign Conflict of Interest forms prior to the review and Nondisclosure Agreements if/when proprietary information is presented or discussed.

- **Plan for Collecting Reviewer Data.** Review leaders will plan ahead for how review inputs will be documented, analyzed, and reported, as well as how individual reviewer comments will be tracked while maintaining their public anonymity. The review agenda will allow sufficient time for a rigorous Question & Answer period for reviewers. Reviewers will be encouraged to support their comments with citations or data wherever possible.
- **Producing the Peer Review Report.** The peer review report will reflect the full range of reviewer comments with high fidelity. The report should also include all individual inputs from the reviewers and will be reviewed by the panel chair and/or the review panel before release.
- **Program Manager Review and Response.** Before the report is finalized and goes to senior management, the program manager/office director will add written responses to peer reviewer findings and recommendations, including actions to be taken to improve the program.
- **Peer Review Report Distribution.** The final peer review report will be promptly communicated to senior management, associated staff and researchers involved with the R&D program or project, and all persons involved in the review, and the report will be made available publicly.
- **Peer Review Record and Ex-post Evaluation.** A peer review record will be established at the beginning of, and maintained throughout, the review process. The record should contain the final form of all the key documents of the review for all phases of the review. An evaluation of the peer review process is necessary to aid continuous process improvement.

494. Based on the variety of experiences, examining the peer review literatures, and managing hundreds of peer reviews, Kostoff (2004, 2003, 2001, 1997, 1995) concludes the followings as *the factors critical* to high-quality peer review for programme evaluation:

- **Senior management commitment**: Senior management's commitment is the most important factor in the quality of an organisation's S&T evaluations.
- **Evaluation manager motivation**: The second most important factor is the operational manager's motivation to perform a technically credible evaluation.
- **Statement of objectives**: The third most important is transmission of a clear and unambiguous statement of the review's objectives (and conduct) and its potential impact and consequences to all participants.
- **Competency of technical evaluators**: Fourth most important factor is the quality of the technical evaluators themselves, specifically their role, objectivity, and competency. This fourth factor consists of the evaluation *experts' competence and objectivity*.
- **Selection of evaluation criteria**: The fifth important factor is selection of evaluation criteria. These criteria will depend on the interests of the audience for the evaluation, the nature of the benefits and impacts, the availability and quality of the underlying data, the accuracy and quality of results desired, the complementary criteria available and suites of diagnostic techniques desired for the complete analysis, the status of algorithms and analysis techniques, and the capabilities of the evaluation team.
- **Relevance of evaluation criteria to future action**: Every S&T metric, and its associated data, should answer a question that contributes to forming the basis for a decision.
- **Reliability of evaluation**: The reliability and repeatability of an evaluation is also crucial. To minimize repeatability problems, a diverse and representative segment of the overall competent technical community should be involved in the construction and execution of the evaluation.

- **Evaluation integration**: A sound evaluation processes should in general be seamlessly integrated into the organisation's business operations. Evaluation processes should not be incorporated in the management tools as an afterthought (which is typical practice today), but should be part of the organisation's front-end design.
- **Global data awareness**: Data awareness is also important. Placing the technology of interest in the larger context of technology development and availability world-wide is absolutely necessary.
- **Normalisation across technical disciplines**: For evaluations that will be used as a basis for comparison of S&T programs or projects, the next most important factor is normalization and standardization across different S&T areas.
- **Secrecy**: Secrecy is as important as normalization: reviewer anonymity and reviewer non-anonymity. "Blind reviewing" has been used for the noble purposes of providing fairer.
- **Cost of S&T evaluations**: cost is also a critical factor for quality of S&T evaluation.
- **Maintenance of high ethical standards**: The final critical factor, and perhaps the foundational factor in any high quality S&T evaluation, is the maintenance of high ethical standards throughout the process.

5.2 Principles and suggestions for successful Expert Review

495. Some principles or policy recommendations have been suggested for successful expert review (Bozeman, 1993; Rigby, 2002; Ormala, 1989; EERE, 2004; Kostoff, 2004, 2003, 2001, 1997, 1995; Nightingale & Scott, 2007; Moed, 2007; Donovan, 2007; The British Academy, 2007; ESPRC, 2008; Noble, 1974; Gillespie et al., 1985; Bodden, 1982; Porter and Rossi, 1985; GACR, 2007 etc.). Although except for some literature including Bozeman (1993), Kostoff's papers, and EERE (2004), most of the suggestions and principles pertain to the selection of research topic and the publication of scientific papers, these may also be very useful for improving the policy-level or programme-level expert review process.

496. For example, EPSRC of the UK suggests some good peer review principles for reviewing research proposals (See box 9). OMB of the US provides another example. *OMB Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information* **(**2001) set as general criteria for competent and credible peer review the following: (a) peer reviewers be selected primarily on the basis of necessary technical expertise, (b) peer reviewers be expected to disclose to agencies prior technical/policy positions they may have taken on the issues at hand, (c) peer reviewers be expected to disclose to agencies their sources of personal and institutional funding (private or public sector), and (d) peer reviews be conducted in an open and rigorous manner.

Box 9. EPSRC's Peer Review Principles

Transparency. Publish the criteria for assessing proposals and details of the peer review process before applicants submit proposals, defining how the assessment process will operate and be managed.

Appropriateness. Use a peer review process that is appropriate to the type of proposed research and in proportion with the investment and complexity of the work.

Managing interests. Ask all participants to declare interests when carrying out peer review activities so that any conflicts can be identified and managed.

Confidentiality. Treat proposals in confidence and ask those who advise us to do the same.

Expert assessment. Use expert peer reviewers, mainly from EPSRC's college of reviewers, to assess the individual merit of all proposals against the published criteria.

Prioritisation. Prioritise proposals for funding by assessing the merit of each proposal against that of others if its expert assessment has been sufficiently supportive.

Right to reply. Give principal investigators the right to reply to the expert reviewers' assessments when proposals are being prioritised.

Separation of duties. Separate peer review of proposals against the assessment criteria from making funding decisions. EPSRC staff will make funding decisions based on peer review advice, taking into account budgets available and the competing tensions between budgets. Those acting as peers will not also be responsible for authorising the funding decision.

No parallel assessment. Avoid carrying out multiple parallel assessments of a proposal's relative merit.

Source: http://www.esprc.ac.uk/ResearchFunding/ReviewingProposals/Principles.htm

497. What are key principles to high-quality programme/policy expert review? The section below suggests a list of principles and suggestions for good practice of programme/policy expert review based on a number of expert review literatures.

498. **Principle. The philosophy, the focus, the future uses of an evaluation must be understood and agreed upon by the relevant stakeholders in advance**. This is the foremost principle that applies to not only expert review but to every type of evaluation. Since expert review involves outsourcing to third-party experts by the evaluation manager, there is all the more need for the third-party experts to fully understand and agree upon the reasons for evaluation, the methods and principles guiding the evaluation as well as the utility of the evaluation.

- ***Suggestion. High-level policy makers or evaluation managers should clearly define the roles of each actor in the evaluation process and regularly monitor performance***. In all areas of public management, highest-level manager's encouragement and continuous interest in a task is a key factor for success. The evaluation task is more complicated than other tasks, not only because there may be a conflict of interest between the evaluators and the evaluated but also because it involves third-party experts. It is therefore indispensable for the high-level manager of an evaluation to clearly define the roles of each actor and to make sure that agreement is reached among the relevant actors concerning the objective as well as the philosophy of the evaluation well in advance.
- ***Suggestion. Provide pre-evaluation training program for the relevant actors.*** The training of relevant actors, that is, experts, evaluation staff (the secretariat), the evaluated (e.g. programme managers), processes and criteria could enhance the efficiency, the effectiveness as well as the receptiveness of evaluation.

- ***Suggestion. Before evaluation, select objective and useful evaluation criteria***. Evaluation criteria are important as they determine the focus as well as the scope of the evaluation. Consequently, it is imperative to provide clear evaluation criteria before embarking upon the evaluation.

499. **Principle. Qualified experts should be selected as evaluators.** The quality of the evaluators is dependent on the *professional competence* and *objectivity of the experts participating in the evaluation.* The panel chair and other experts should all have high professional competence in the areas in which they are required to make judgments in order to instil confidence in the stakeholders of the evaluation.

- ***Suggestion. In addition to technological experts, seek experts from diverse domains, including experts in social sciences and the economy.*** To make judgements on program's rationality as well as its socio-economic value, in addition to technological experts, it is desirable to have a group of experts from diverse domains, including economy, business management. This is a very important element in policy or programme evaluation
- ***Suggestion. Build a sufficiently large database of experts.*** For this, there needs to be regular monitoring of research personnel in various research institutions and universities. The data which should be collected on the personnel through monitoring are: past research experience, current research interests, field, affiliation, degree-granting institution, participating academic organisations and other detailed academic activities. These data allows one to infer what area of policy a particular researcher could evaluate based on their qualifications and what contributions they can provide if chosen as an evaluator.

500. **Principle. The risk of bias or conflict of interests should be reduced as much as possible.**

- ***Suggestion. Provide a bias statement for reviewers.*** That is, make that experts declare their interests to ensure that the panel's reputation for fairness is upheld. In principle, the evaluation manager should not appoint an evaluator who has a vested interest in the evaluated policy/programme or the evaluated institution. To ensure neutrality of the expert panel, make experts declare their interests relevant to the evaluation.
- ***Suggestion. Avoid "internal evaluators."*** Often, it is useful to include opinions of experts from another field/region or to have them in the panel. In particular, if there is no language-barrier or no additional cost, it is desirable to include foreign experts in the evaluation. In the case foreign experts are not well aware of the socio-economic conditions of the country in question, it is advised that they focus on the scientific or the technical aspects of the programme.
- ***Suggestion. Limit the number of evaluations or the duration of evaluating activities for the participating experts.*** If an expert participates in too many evaluations, the expert might develop a relationship with the evaluated bodies and might be susceptible to lobbying from them. Moreover, trapped in their judgments from previous evaluations, they might not be able to take a fresh look at similar programmes. However, to limit the duration of evaluating activities excessively may be counter-productive. For example, it might decrease their sense of responsibility. If the expert participates in the evaluation only once, their responsibility might be lower than if they were to participate in future evaluations. Therefore, one may consider an appropriate duration of evaluating activities for the experts.
- ***Suggestion. Prevent an expert who has expertise only in a particular domain from judging the quality or the value of what is being evaluated***. However, it is advisable that his/her opinions are transmitted to other experts in the panel to be used in the joint-decision making. For example, an expert on biotechnology evaluating a biotechnological programme might know more about the

technical aspects than other experts but may not be able to judge correctly the socio-economic value of the evaluated programme.

501. **Principle. The review should be conducted in a credible, fair, transparent manner with the highest degree of ethical standards**.

- ***Suggestion. Provide transparency in evaluation process and in evaluation results.*** Introduce transparency in evaluation principles, criteria, processes and make them accessible to all the actors and stakeholders in the evaluation so that they could prepare for the evaluation properly. After the evaluation, release evaluation results on-line, except those that may be confidential for national security reasons, and let those results be accessible to the evaluated bodies as well as the general public.
- ***Suggestion. Maintain high ethical standards***. To ensure that evaluators are free from personal bias, there is the option of requiring bias statements. Such declarations or statements may include clauses on overcoming personal biases as well as on the prohibition of any misuse of information obtained during the evaluation process, such as the use of such information for personal reasons or the release of such information without the permission of the relevant authority.

502. **Principle. The review should be based on objective evidence and information.**

- ***Suggestion. Provide, in advance, sufficient information on the evaluated policy/program to the evaluators.*** Judgement of the expert panel depends on their comprehension of given information. Therefore, providing sufficient information is as important as selecting qualified experts and one should not ask the experts to provide judgements which go beyond the scope of the provided information. For policy decision making, adjustment of programs, resources allocation, priority setting, it is useful to employ 3P analysis (positioning analysis, portfolio analysis, performance analysis).
- ***Suggestion***. ***If indicators or rating are used, test the validity and reliability of those indicators***. Indicators are important tools to ensure the objectivity of the evaluation. Therefore, before inferring evaluation results, it is necessary to have the evaluators go over the relevance and the reliability of indicators. In case modifications of one or more indicators are necessary, sufficient reason should be provided and such modification must be communicated to the evaluated bodies and be subject to their approval.
- ***Suggestion***. ***Encourage a maximum amount of dialogue and discussion.*** Ideas arising from discussions and dialogues as well as mutual-learning experience are the biggest advantages of expert review. Therefore, along the evaluation process, encourage as much discussion and dialogues among evaluators and the evaluated (programme managers) as possible. At this time, foregoing the introduction of experts from other domains and expanding the panel as much as possible would be an effective way to encourage productive discussions as well as the creation of new ideas.

503. **Principle. "One size does not fit all."**

- ***Suggestion. Complement expert review with quantitative methods to increase objectivity and scientific reliability of the evaluation.*** One may increase objectivity and accuracy of expert review evaluation by complementing it with quantitative methods such as bibliometrics or econometrics. Often, expert review is considered a particular way of research evaluation, evaluators often utilise case studies, benchmarking, surveys and other evaluation methods. Therefore, it is difficult to categorise expert-review as a separate evaluation method; it should be understood as a decision-making process that is complemented by various evaluation methods.

- ***Suggestion. Seek the type of expert review appropriate for the particular programme/policy.*** The review should be tailored to the aim of evaluation and the characteristics of the subject of evaluation. In other words, evaluation methods and results should be differentiated according to evaluation objectives. For example, if the primary objective is to set priorities, a scoring method could be used. If the improvement of a programme is the primary objective, opinions of expert review would be very important. Also, evaluation processes and the form of the final results should be tailored to the particularities of programmes.

504. **Principle. Evaluation efficiency may be increased through various measures.**

- ***Suggestion. Increase remote evaluation.*** Today, with the development of the Internet, there are many cases in which evaluating institutions distribute IDs and Passwords to the evaluators with which the evaluators may access data and submit reports online. To promote expert review efficiency, devise different technologies to enable evaluators to participate in the evaluation process from a distance.
- ***Suggestion. Build and operate evaluation management systems through internet-based technologies.*** Evaluation management systems (EMS) should cover fundamental information on the evaluation, including information on the pool of experts, evaluation data, evaluation principles, evaluation protocol, relevant analytic data, and evaluation results. To increase the utilization of EMS, evaluating organisms, evaluators and evaluation object should be able to use it freely. Admittedly, certain restrictions for security reasons may be introduced when necessary
- ***Suggestion. Minimize the part of the evaluation cost born by the subject of the evaluation.*** Especially, simplify administrative procedures and evaluation formats. In many cases, the evaluation subject is asked by the evaluators to provide administrative information irrelevant to the evaluation as well as unnecessary information. It is advisable to have the evaluation body to extract excessively complicated forms or unimportant information from the basic database and provide these to the evaluators themselves in order to reduce the administrative burden of the evaluates. In addition, having the evaluators provide clear reasons when they ask for additional information from the evaluation object may enhance cooperation on the part of the evaluation object.
- ***Suggestion. Design evaluation processes while paying attention to hidden indirect costs***. In expert review, the indirect costs such as the billable hours spent the evaluators are more important than direct costs such as venue and travel expenses.

Improve design of expert panels

- ***Suggestion.*** To ensure continuity, it is advisable to appoint someone who has participated in previous panels as the head of a panel.
- ***Suggestion.*** It is advisable that one-third to one-half of a review panel be carried over from one review to the next in order to enhance “new perspectives” as well as continuity.

Box 10. Summary of principles and suggestions

Principle. The philosophy, the focus, the future uses of an evaluation must be understood and agreed upon by the relevant stakeholders in advance.

- Suggestion. High-level policy maker or evaluation manager should clearly define the roles of each actor in the evaluation process and regularly monitor their performance.
- Suggestion. Provide pre-evaluation training program for the relevant actors.
- Suggestion. Before evaluation, select objective and useful evaluation criteria.

Principle. Qualified experts should be selected as evaluators.

- Suggestion. In addition to technological experts, seek experts from diverse domains, including experts **in social sciences and in economy.**
- **Suggestion**. Build a sufficiently large database of experts.

Principle. Reduce the risk of bias or conflict of interest as much as possible.

- Suggestion. Provide a bias statement for reviewers. That is, make that experts declare their interest to ensure that the panel's reputation for fairness is upheld.
- Suggestion. Guard against dysfunctional group dynamics.
- Suggestion. Avoid "internal evaluators."
- Suggestion. Limit the number evaluation or the duration of evaluating activities for the participating experts.
- Suggestion. Prevent an expert who has expertise only in a particular domain from judging the quality or the value of what is being evaluated.

Principle. The review should be conducted in a credible, fair, transparent manner with the highest ethical standards.

- Suggestion. Provide transparency in evaluation process and in evaluation results.
- Suggestion. Maintain high ethical standards.

Principle. The review should be based on objective evidence and information.

- Suggestion. Provide, in advance, sufficient information on the evaluated policy/programme to the evaluators.
- Suggestion. If indicators are used, test the validity and reliability of those indicators.
- Suggestion. Encourage maximum amount of dialogues and discussions.

Principle. "One size does not fit all."

- Suggestion. Complement expert review with quantitative methods to increase objectivity and scientific reliability of the evaluation.
- Suggestion. Seek the type of expert review appropriate for the particular evaluated programme.

Principle. Evaluation efficiency may be increased through various measures.

- Suggestion. Increase remote evaluation.
- Suggestion. Build and operate evaluation management system through internet-based technology.
- Suggestion. Minimize the part of the evaluation cost born by the evaluation object.
- Suggestion. Design evaluation processes while paying attention to hidden indirect costs.

Improve design of expert panels

- Suggestion. To ensure continuity, it is advisable to appoint someone who has participated in previous panels as the head of a panel.
- Suggestion. It is advisable that one-third to one-half of a review panel be carried over from one review to the next in order to enhance "fresh eyes" as well as continuity.

6. Concluding remarks

505. In the literature and in various workshops, there appears to be a common view emerging. First, despite problems in expert review including hollowing-out due to time constraints, rising financial costs, and the risk of conflicts of interest among expert reviewers, the expert review process remains as a fundamental mechanism for all stages of research planning and implementation as well as for both ex ante project selection and for ex post evaluation. Second, solutions are available to improve expert review process, including: making the process more transparent, providing clear objectives and guidelines to reviewers, using different tools (*e.g.* extended expert review processes involving non-scientific stakeholders) and using a variety of metrics and indicators. Thirdly, while indicators can strengthen and inform judgements, they do not form judgements by themselves. Making judgement still requires careful consideration to prevent perverse outcomes. Fourthly, there is a need to facilitate and improve the internationalisation of expert review because of increased international collaboration. However, caution is needed when policy is transferred across different cultures and contexts, particularly when understandings of policies are incomplete. Therefore, there is a need for a taxonomy of the internationalisation of expert review. Finally, one size does not fit all and hence a much better understanding of the design requirements for expert review is needed.

506. There is another important principle to be added to the emerging views above. A perfect evaluation system cannot exist; one must adapt the evaluation system to the environment. What was ideal in the past may be found to be no longer effective in the future. Also, when improving evaluation systems, it is important to take into account not only the opinions of the evaluation managers but also the opinions of the evaluated. This is because it is easier to find problems as well as solutions when one looks at the question from the perspective of a client. Admittedly, the opinion of the evaluated will differ according to the evaluation results. It is therefore important to strike a proper balance by considering the opinions of policy makers. For this, it is advisable that evaluation institutions carry out regular opinion surveys targeting evaluation participants and evaluates on expert review.

507. Because expert review involves more individual judgments than any other methods of evaluation, much like the functioning of an orchestra, co-operation among the conductor (review manager), the players (experts) and the audience (stakeholders) is necessary for a successful performance (evaluation). This may be said to be the foremost principle that should be respected for a successful expert review.

REFERENCES AND SUGGESTED FURTHER READINGS

Averch, H. (1990), "Policy Uses of 'Evaluation of Research' literature," *OTA Contractor Report*, July 1990.

Barker, K. (1994) "R&D Evaluation – Strengthening the impact of R&D evaluation on policy making: methodological and organisation considerations", Research Evaluation, Vol. 21, No. 6, pages 405-413.

Becher, Gerhard and Stefan Kuhlmann eds. (1995), *Evaluation of Technology Policy Programmes in Germany*. Kluwer Academic Publishers.

Bobe, B. & Viala, H. (1997) "Panels of experts of the European Commission" in Callon, M., Laredo, P. & Mustar, P. (eds.) The Strategic Management of Research and Technology – Evaluation of Programmes", Economica: Paris.

Bozeman, B. & Melkers, J. (eds.) (1993) Evaluating R&D Impacts: Methods and Practice, Kluwer: Boston.

Bozeman, Barry (1993), "Peer Review and Evaluation of R&D Impacts," in Bozeman, B. and J. Melkers (eds.), *Evaluating R&D Impacts: Methods and Practice*. Massachusetts: Kluwer Academic Publishers.

Branscomb, L., K. Morse, and M. Roberts (2000), *Managing Technical Risk: Understanding Private Sector Decision Making on Early Stage Technology-Based Projects*, NIST GCR 00-787, Gaithersburg, MD, April 2000.

Butler, Linda and Ian McAllister (2007), "Metrics or Peer Review? Evaluating the 2001 UK Research Assessment Exercise in Political Science," *http://repp.anu.edu.au/papers/2007_ukresearchassess.pdf*

Campbell, P. (2001) "Opinion: Introducing a new policy for authors of research papers in Nature and Nature journals," *Nature* 412, 751 (2001).

Campbell, Philip (2006), "Pressures on peer review at *Nature* and Nature Journals," Presentation in Prague Conference on Peer Review: Its Present and Future State. Oct. 2006.

Chráska, Pavel and Josef Syka (2006), "Peer Review Process in the Czech Science Foundation – Experiences and Challenges," Presentation in Prague Conference on Peer Review: Its Present and Future State. Oct. 2006.

Chubin, D. and S. Jasanoff (1985), "Peer Review and Public Policy," *Science, Technology and Human Values*, Vol. 10, No. 3, pp. 3-5.

Chubin, D. E. (1994), "Grants Peer Review in Theory and Practice," *Evaluation Review*, Vol. 18, No. 1, pp. 20-30.

Chubin, Daryl and Edward Hackett (1990), *Peerless Science: Peer Review and US Science Policy*. Albany: SUNY Press.

COSEPUP (1999), *Evaluating Federal Research Programs: Research and the Government Performance and Results Act.* Washington DC: National Academy Press. Available at *http://www.nap.edu/catalog/6416.html.*

Cozzen, S. E. (1987), "Expert Review in Evaluating Programs," *Science and Public Policy 14*, April.

Dixon, B. (1987), "Follow the Finnish Lead in Peer Review," *The Scientist*, May 4.

Donovan, Claire (2007a), "Introduction: Future pathways for science policy and research assessment: metrics vs. peer review, quality vs. impact," *Science and Public Policy 34(8)*, October 2007.

Donovan, Claire (2007b), "The qualitative future of research evaluation," *Science and Public Policy 34(8)*, October 2007.

EERE (2004), *EERE Peer Review Guide: Based on a Survey of Best Practices for In-Progress Peer Review*, August 2004.

Environmental Protection Agency (1997), "Peer Review and Peer Involvement at the U.S. Environmental Protection Agency." Available at *http://www.epa.gov/ordntrnt/ord/spc/perevmem.htm*.

Environmental Protection Agency (EPA). 1998. *Science Policy Council Peer Review Handbook*. Office of Science Policy. EPA 100-B-98-00. Washington, DC: EPA.

EPA (2000), *Peer Review Handbook*, Science Policy Council.

EPSRC (2008), "Peer Review Principles," *http://www.epsrc.ac.uk*

ESF MO Forum workshop on Peer Review Website (2008): *http://www.esf.org/activities/mo-fora/peer-review.html*

Forsdyke, D. R. (1991), "Bicameral grant review: an alternative to conventional peer review," *FASEB J. 5*, 2312-2314.

Forsdyke, D. R. (1991), "Bicameral grant review: an alternative to conventional peer review," *FASEB J.* 5, 2312-2314.

Forsdyke, D. R. (1993), "Bicameral grant review: how a systems analyst with AIDS would reform research funding," *Accountability in Research 2*, 237-241.

Forsdyke, D. R. (1993), "Bicameral grant review: how a systems analyst with AIDS would reform research funding," *Accountability in Research*, 2, 237-241.

Forsdyke, D. R. (1993). On giraffes and peer review. *FASEB.J.* 7, 619-621.

GACR (2007), *Peer Review: Its Present and Future State*, Czech Science Foundation.

Geisler, E. (2000), "The Metric of Peer Review," in E. Geisler, ed., *The Metrics of Science and Technology*. Westport, CT: Quorum Books.

General Accounting Office (1994). *Peer Review: Reforms Needed to Ensure Fairness in Federal Agency Grant Selection*, GAO/PEMD-94-1.

General Accounting Office (1999). *Federal Research: Peer Review Practices at Federal Science Agencies Vary*, GAO/RCED-99-99.

Gibbons, M. and L. Georghiou (1987), *Evaluation of Research: A Selection of Current Practices*. Paris: OECD.

Gieryn, Thomas F. (1995). "Boundaries of Science." pp. 393-443 in Sheila Jasanoff, Gerald E. Markle, James C. Petersen, and Trevor Pinch, eds., *Handbook of Science and Technology Studies*. Thousand Oaks, CA: Sage Publications.

Gillespie, G. W. Jr., D. E. Chubin, and G. M. Kurzon (1985), "Experience with NIH Peer Review: Researchers' Cynicism and Desire for Change," *Science, Technology and Human Values*, Vol. 10, No. 3, pp. 44-53.

Godlee, F. & Jefferson, T. (1999) *Peer Review in the Health Sciences*, BMJ Books: London.

Goodstein, T. (1995), "Ethics and Peer Review," *Biotechnology*, Vol. 13, No.6, p. 618.

Grigson, D. & Stokes, T. (1993) "Use of peer review to evaluate research outcomes," *Research Evaluation*, Vol. 3, No. 3, pp. 173-177.

Grol, R. and M Lawrence, eds. (1995), *Quality Improvement by Peer Review*. New York: Oxford University Press.

Gustafson, T. (1975), "The Controversy over Peer Review: Recent Studies of the Peer Review System Show that its Critics have yet to Make their Case," *Science*, Vol. 190, pp. 1060-1066.

Guston, David H. (2000a), "Expanding the Role of Peer Review Processes in the United States," paper presented to the US-European Policy Workshop on Learning from Science and Technology Policy Evaluation, Bad Herrenalb, Germany.

Guston, David H. (2000b). "Regulatory peer review." Paper presented at the annual meeting of the American Political Science Association, Washington, DC, 2 September.

Guston, David H. (2000c). *What is Peer Review?* Report submitted to the American Chemistry Council, May.

Hackett, E. J. (1997), "Peer Review in Science and Science Policy," in *Evaluating Science and Scientists* edited by M. S. Frankel and J. Cave. Budapest/Oxford, Central European University Press/OUP.

Hartmann, I. and F. Neidhardt (1990), "Peer Review at the Deutsche Forschungsgemeinschaft," *Scientometrics*, Vol. 19, No. 5-6, pp. 419-425.

Klahr, D., "Insiders, Outsiders, and Efficiency in a National Science Foundation Panel," *American Psychologist*, Vol. 40, No. 2, pp.148-154. Kostoff, Ronald N. (1993), "Evaluating Federal R&D in the United States," in Bozeman, B. and J. Melkers (eds.), *Evaluating R&D Impacts: Methods and Practice*. Massachusetts: Kluwer Academic Publishers.

Kostoff, R. (1998), "GPRA Science and Technology Peer Review"

Kostoff, Ronald N. (1993), "Evaluating Federal R&D in the United States," in Bozeman, B. and J. Melkers (eds.), *Evaluating R&D Impacts: Methods and Practice*. Massachusetts: Kluwer Academic Publishers.

Kostoff, Ronald N. (1997a), Research Program Peer Review: Principles, Practices, Protocols, *www.dtic.mil/dtic/kostoff/index.html*.

Kostoff, Ronald N. (1997b), "The Principles of Practices of Peer Review," *Science and Engineering Ethics*, Vol. 3, No. 1, pp. 19-34.

Kostoff, Ronald N. (2003), *Science and Technology Peer Review: GPRA*, Office of Naval Research.

Kostoff, Ronald N. (2004), *Research Program Peer Review: Purposes, Principles, Practices, Protocols*, Office of Naval Research.

Kruytbosch, Carlos (1989), "The Role and Effectiveness of Peer Review," Paper presented at the Ciba Foundation Conference on Research Evaluation.

Lee, M., K. Om and J. Ko (2000), *Theory and Practice of Peer Review*. Seoul: Pidi.

Luukonen-Gronow, T., (1987), "Scientific Research Evaluation: A Review of Methods and Various Contexts of Their Application," *R&D Management 17*.

Massimo, L. (1997) "Lessons from the experience of the European Commission" in Callon, M., Laredo, P. & Mustar (eds.) The Strategic Management of Research and Technology – Evaluation of Programmes", Economica: Paris.

McCullough, J. (1989), "First Comprehensive Survey of NSF Applicants Focuses on their Concerns about Proposal Review," *Science, Technology and Human Values*, Vol. 14, No. 1989, pp. 78-88.

Mehlhorn, Kurt (2006), "Peer Review in the Digital Age: Opportunities and Examples." Presentation in Prague Conference on Peer Review: Its Present and Future State. Oct. 2006.

Merton, R. K. (1973). *The Sociology of Science*. University of Chicago Press.

Moed, Henk 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)*, October 2007.

National Science and Technology Council (1996). *Assessing Fundamental Research*. Committee on Fundamental Science, Subcommittee on Research. Accessed online at *http://www.nsf.gov/sbe/srs/ostp/assess/start.htm*.

National Science Foundation (1999). *What Is the State Government Role in the R&D Enterprise?* Division of Science Resources Studies. 99-348. Arlington, VA: NSF.

Nature Peer Review Debate: *http://www.nature.com/nature/peerreview/debate*

Nightingale, Paul and Alister Scott (2007), "Peer Review and the Relevanc Gap: 10 Suggestions for Policy-makers," *Science and Public Policy34(8)*, Oct. 2007.

Noble, J. H. (1974), "Peer Review: Quality Control of Applied Social Research," *Science*, Vol. 185, pp.916-921.

NRC (1999), *Peer Review in Environmental Technology Development Programs,* Washington DC: National Academy Press. Available at *http://www.nap.edu/catalog/6408.html*

OECD (1998), *Policy Evaluation in Innovation and Technology: Towards Best practices*, Paris: Organisation for Economic Co-operation and Development.

OECD (2006), "Evaluation of Publicly Funded Research: Recent Trends and Perspectives," DSTI/STP/TIP(2006)3.

OECD (2007a), "Evaluating Government R&D Programmes: Lessons from Korean Experience," DSTI/STP/TIP(2007)7.

OECD (2007b), "Revised Proposal for the TIP Work on Evaluation of innovation policies in OECD Countries," DSTI/STP/TIP(2006)17REV1.

OECD (2007c), "Peer Review: Its Uses, Demands and Issues," DSTI/STP/TIP(2007)13.

OECD (2007d), "Summary of the workshop on Rethinking Evaluation in Science and Technology," DSTI/STP/TIP(2007)14.

OECD (2007e), "TIP activity on the Evaluation of research and innovation policies: current Status and next Steps," DSTI/STP/TIP(2007)15.

OECD-BMBF Conference (2005): *http://www.internationales-buero.de/de/2193.php*

Oh, Donghoon and YoungJun Kim (2006), "Overall Coordination of Government-Funded Research and Development Programmes in Korea," *Journal of MultiDisciplinary Evaluation*, 5 (September 2006), 153-170.

Ormala, E. (1989), "Nordic Experiences of the Evaluation of Technical Research and Development," *Research Policy*, Vol. 18, pp. 333-342.

Porter, A. & Rossini, F. (1985) Peer Review of Interdisciplinary Proposals" Science, Technology and Human Values, Vol. 10, No. 3, pages 34-42.

Porter, A. and F. Rossini (1985), "Peer Review of Interdisciplinary Proposals," *Science, Technology, and Human Values*, Vol. 10, No. 3, pp. 34-42.

Pouris, A. (1988), "Peer Review in Scientifically Small Countries," *R&D Management*, Vol. 18, No. 4, pp. 333-340.

Prague Conference (2006), "Peer Review: Its Present and Future State," Available at *http://www.pragueforscience.cz/Scientific-Programme.php*

Rigby, John (2002), Expert Panels and Peer Review," Fahrenkrog, Gustavo, Wolfgang Polt, Jaime Rojo, Alexander Tubke, and Klaus Zinocker eds., *RTD Evaluation Toolbox: Assessing the Socio-Economic Impact of RTD-Policies*. IPTS Technical Report Series, EUR 20382 EN.

Rossi, Peter H., Mark W. Lipsey, and Howard E. Freeman (2004), *Evaluation: A Systematic Approach*, Thousand Oaks: SAGE Publications.

Roy, R. (1985), "Funding Science: The *Real* Defects Of Peer-Review And An Alternative To It," *Science Technology & Human Values 52*, pp. 73-81.

Roy, Rustum. (1984), "Alternatives To Review By Peers - A Contribution To The Theory Of Scientific Choice," *Minerva 22* (3-4): 316-328.

Ruegg, Rosalie and Gretchen Jordan (2007), *Overview of Evaluation Methods for R&D Programs,* Office of Energy Efficiency and Renewable Energy, DOE.

Ruegg, Rosalie and Irwin Feller (2003), *A Toolkit for Evaluating Public RTD Investment,* NIST GCR 03-875, National Institute of Standards and Technology.

Scarpa, Toni (2006), "Challenges and Opportunities Facing NIH Peer Review: A Vision for Ensuring Its Strategic National Value," Presentation in Prague Conference on Peer Review: Its Present and Future State. Oct. 2006.

Scott, Alister (2006), "Peer Review and the Relevance of Science," *SPRU Electronic Working Paper Series*, No. 145.

Shapira, Philip and Stefan Kuhlmann eds. (2003), *Learning from Science and Technology Policy Evaluation: Experiences from the US and Europe*. Edward Elgar, Cheltenham.

The British Academy (2007), "Peer Review: the challenges for the humanities and social sciences," THE NATIONAL ACADEMY FOR THE HUMANITIES AND SOCIAL SCIENCES.

Wood, Fiona and Simon Wessley (2007), "Peer Review of Grant Applications: A Systematic Review," in Fiona Godlee, Tom Jefferson eds. (2007), *Peer Review in Health Sciences*. BMJ Books.

Wood. F. (1997), *The peer review process.* Canberra: Australian Research Council.

Worthen, Blaine R., James R. Sanders, and Jody L. Fitzpatrick (1997), *Program Evaluation: Alternative Approaches and Practical Guidelines*, Longman Publishers USA.

CHAPTER 4

IMPROVING EVALUATIONS THROUGH IMPACT ASSESSMENT OF PUBLIC R&D[41]

Impact assessment is part and parcel of evaluation. This chapter reviews recent and emerging impact assessment practices, including the main methodologies, and highlights their assumptions and limitations but also possibilities.

1. Introduction

508. Impact assessment is central to evaluation of public R&D. However, impact assessment is more than measuring success in meeting past objectives (ex post). It is also about determining where, who and how much to fund research and anticipate what society gets in return (ex ante). An impact analysis should help determine both the economic effects of public investment in R&D as well as the social impacts (e.g. better health outcomes).

509. Assessment of the impacts of public R&D is therefore closely intertwined with the evaluation of public R&D and should provide valuable feedback to the different phases of public policy formulation, including policy design. Public R&D impact assessment assists governments in their decisions to prioritise R&D resources, and can help them design research programmes. Moreover, assessment enhances public accountability, creates a better informed society and raises awareness of the contribution of public research to a country's economic and social development.

510. The chapter first defines the nature and scope of the potential impacts of public R&D, as well as the main challenges that practitioners face when identifying and assessing these impacts. It then distinguishes three main levels: *i)* overall public R&D investment in the research system; *ii)* public research organisations, including funding research councils, in relation to research carried out or funded by specific institutions; and *iii)* research programmes. It presents practices for assessing the impact of publicly funded and performed research and for assessing systemic impacts, *i.e.* those affecting the economy or society, as well as sector-specific impacts.

2. Defining the objects and impacts of R&D

511. There are many definitions of "impacts" that are used by evaluators and policy makers. In general, the definition used will depend on: 1) the nature of the impact: economic, scientific, technological, cultural, societal environmental, etc. 2) the scope of the impact: systemic, organisational, firm-based and 3) the timing of the impact: estimated, contemporary, ex-post. The academic literature provides various definitions of the types of impacts of science and technology (See Box 11).

41. This chapter draws and builds on the Chapter on Assessing the Socio-Economic Impacts of Public R&D included in the STI Outlook 2008 as well as the 2008 TIP workshop on Impact Assessment as well as the RIHR report on the evaluation of public research institutions.

Box .11. Eleven dimensions of the impacts of science

Science impacts: Research results have an effect on the subsequent progress of knowledge thanks to advance in theories, methodologies, models and facts. They affect the formation and development of disciplines and training and can also affect the development of research itself, generating interdisciplinary, cross-cutting and international research.

Technology impacts: Product, process and service innovations as well as technical know-how are types of impacts that partly result from research activities. There are few indicators for properly assessing this dimension, other than patents, at least until work based on innovation surveys results in analysis of outputs and impacts as well as innovation activity itself.

Economy impacts: These refer to the impact on an organisation's budgetary situation, operating costs, revenues, profits, the sale price of products; on the sources of finance, investments and production activities; and on the development of new markets. At the aggregate level, they can also refer to economic returns, either through economic growth or productivity growth, of a given geographical unit. It is probably the best-known dimension.

Culture impacts: These relate to what people often call public understanding of science, but above all to four types of knowledge: know-what, know-why, know-how and know-who. In other words, these are the impacts on an individual's knowledge and understanding of ideas and reality, as well as intellectual and practical skills, attitudes, interests, values and beliefs.

Society impacts: Research affects the welfare, behaviour, practices and activities of people and groups, including their well-being and quality of life. It also concerns customs and habits: consumption, work, sexuality, sports, food. Research can contribute to changing society's views and "modernise" ways of doing "business".

Policy impacts: Research influences how policy makers and policies act. It can provide evidence that influences policy decisions and can enhance citizens' participation in scientific and technological decisions.

Organisation impacts: These refer to the effects on the activities of institutions and organisations: planning, organisation of work, administration, human resources, etc.

Health impacts: These relate to impacts on public health, *e.g.* life expectancy, prevention of illnesses, and the health-care system.

Environment impacts: These concern management of the environment, notably natural resources and environmental pollution, as well as the impacts of research on climate and meteorology.

Symbolic impacts: These are the gains in areas such as credibility due to undertaking R&D or linked to universities or research institutions that offer gains in terms of potential clients, etc.

Training impacts: These are impacts of research on curricula, pedagogical tools, qualifications, entry into the workforce, etc.

All but the first three dimensions are somewhat new to statisticians, as they are less tangible and therefore difficult to measure or evaluate. This typology provides a checklist to remind evaluators that research affects areas other than those usually identified and measured in the economic literature.

Source: STI Outlook 2008 based on Godin and Doré (2006).

512. The different impacts can be diverse in scope as well as in nature. Impacts may accrue to society as a whole, to a particular group of people, to a research group or to enterprises or other institutions. Identifying the type of impact to be measured is crucial when deciding on the choice of methodology or methodologies for assessing the impact of public R&D.

3. Key challenges for assessing the socio-economic impacts of public R&D

513. It is difficult to determine and measure the various benefits of R&D investment for society. R&D spillovers and unintended effects are likely, many key scientific discoveries are made by accident or serendipity, and many applications of scientific research are found in areas very different from the original intention. Moreover, the time required for public R&D to generate its full benefits may be quite long, so that measurement of impacts may be premature and partial. Finally, the non-economic impacts of public research may be more difficult to identify and measure. For example, the measurement of health outcomes is not straightforward and complicates efforts to link health outcomes to public investment in R&D. Similar difficulties arise for linking investment in defence R&D to security outcomes or investment in energy R&D to energy security. Drawing on the STI Outlook, the most important challenges encountered

by science policy researchers and policy makers when analysing the impacts of public R&D can be summarised as follows:

- Causality problem: what is the relationship between research inputs, outputs, outcomes and impacts? No direct or unidirectional relationship
- Attribution problem: What portion of the benefits should be attributed to initial research and not to other inputs?
- Internationality problem: Role of spillovers
- Evaluation time scale problem: At which time should we measure the impacts?
- Definition of appropriate indicators

514. Owing to these challenges, analysis has traditionally focused on developing and collecting R&D input and output indicators and establishing a direct relationship between them. However, since many of the impacts of R&D only emerge over time, this type of analysis often ignores many of the long-term benefits of public R&D for a country's economy and society.

Box 12. The Traditional measures used for measuring "impacts" R&D

	R&D Inputs					R&D Outputs		
	Total Public R&D (GOVERD + HERD) 2005*	**GOVERD 2005* (% of GDP)**	**HERD 2005* (% GDP)**	**Basic Research 2005* (%GDP)**	**Researchers 2004* (per thousand of labour force)**	**Scientific articles per million population, 2003**	**Relative prominence of scientific literature, 2003**	**Share of PCT patents owned by Gov + HE (2002/04)**
Iceland	1.28	0.66	0.62	0.53	13	701.8		
Sweden	1	0.24	0.76		10.8	1142.8	0.86	0
Finland	0.99	0.33	0.66		15.7	997.9	0.83	0.4
Canada	0.9	0.18	0.72		7.3	783.2	0.85	10.3
France	0.79	0.37	0.42	0.52	7.3	516.2	0.76	10.8
Austria	0.77	0.12	0.65	0.39	6.6	604.4	0.8	1.1
Australia	0.76	0.28	0.48	0.42	7.9	791.2	0.71	10.3
Denmark	0.76	0.18	0.58	0.46	9.1	981.6	0.94	3.2
Germany	0.75	0.34	0.41		6.8	536.9	0.82	1.7
Netherlands	0.74	0.24	0.5		4.5	830.6	0.97	1.4
Japan	0.73	0.28	0.45	0.4	10.2	470.3	0.58	4.4
Norway	0.71	0.24	0.47	0.28	8.9	731.4	0.72	0.5
Switzerland	0.7	0.03	0.67	0.84	5.8	1153.5	1.15	2.2
United States	0.68	0.31	0.37	0.48	9.5	725.6	1.03	10

Source: OECD Main Science and Technology Indicators, 2008

515. Moreover, econometric analysis of the relationship between R&D and outcomes is typically based on a linear conception of innovation and the idea that innovation starts with basic research, followed by applied research and development and ends with the production and diffusion of new products and

processes in the economy. However, it is widely acknowledge that innovation is more complex, with multiple feedback loops between stages and actors, and innovation results from the interplay of public and private R&D investment, commercial interests and many other factors. As a result, a more comprehensive understanding of the effects of science and innovation requires a more encompassing approach to measuring and analysing innovation and the economic and social impacts that accrue to society.

4. Approaches to impact assessment of public research in OECD countries

516. Which methods can be used to assess the impacts of public R&D? Over the past decade, national governments and academics have carried out initiatives to develop new analytical techniques for assessing the impacts of public R&D investment, such as econometric analysis, data linkages approaches and case studies. The outcomes and robustness of such analyses are heavily influenced by the nature of these methods, the assumptions on which they rely and their inherent limitations Impact assessment methodologies are not universally applicable, but rather context specific depending on the objective of the impact assessment exercise, the timing of the exercise (ex ante and/or ex post); and the scope and nature of R&D.

517. The review done in TIP for the STI Outlook found top-down approaches, especially econometric and mathematical models, better suited to assess impacts affecting the whole research system and dealing with all types of research, both basic and applied. In particular, mathematical models, such as general equilibrium or similar models may be a good way to assess systemic impacts *ex ante*. On the other hand, when the subject of the assessment is a research programme and/or institution that aims at developing a specific type of technology with a clear industrial focus, bottom-up approaches may be favoured.

518. Some of the most promising and forward-looking practices include general equilibrium models, econometric analyses, data linkages and scientometrics methods, survey-based indicators combined with econometric analyses and case studies. These various methodologies are still evolving but until now, none of the available techniques has been able to capture the full range of impacts of public R&D on society, although they have opened new and encouraging lines of investigation. The following section reviews some of these approaches and their advantages and disadvantages.

4.1. Econometric-based impact assessments

519. Econometric studies have examined evidence on the contribution of R&D investment to economic growth both in microeconometric studies, which use data on firm and industry productivity to estimate the private and social returns to R&D investments, and in macroeconometric studies, which estimate the contribution of overall R&D investment to aggregate productivity.

520. The microeconometric studies have analysed productivity growth in private firms in a number of countries and for different periods of time. They have also assessed the presence of knowledge spillovers and the calculation of the social rate of return, *i.e.* the benefits that the private R&D investment generate for other firms located mainly, but not only, in their own industry. A seminal study by Lichtenberg and Siegel (1991) on the effects of private R&D investments on total factor productivity (TFP) found that the gross rate of return was significant and up to 35% for company-funded R&D. For publicly funded R&D, however, they found little significant impact on productivity. In 1994, Mamuneas and Nadiri also explored the social return of publicly funded R&D for US manufacturing firms, by estimating the reductions in costs associated with an extra dollar of public R&D investment. The results showed returns ranging from 8.7% to 5.8% and thus a positive social return to publicly funded R&D. Griliches (1986) also concluded that publicly funded R&D in industry had positive effects on productivity, although less than privately financed R&D.

521. In general, microeconometric studies have shown strong returns to private R&D investment and the presence of strong spillover effects that generate substantial economic benefits.[42] There is so far relatively little evidence on the impact of public R&D investments on private productivity growth, and the few existing studies provide inconclusive results. This may be because studies at the firm and industry level are unable to account for positive spillovers accruing from public R&D, which may only emerge at the national level. Moreover, as public research is often at the pre-competitive stage, the link to immediate commercial applications and productivity growth is likely to be less direct.

522. Macroeconometric studies analyse the effect of overall R&D on national productivity and can capture the full extent of knowledge spillovers to different firms and industries. These cross-country studies also make it possible to taken into account benefits that diffuse across firms and industries.

523. Many of these studies investigate both the social returns to national R&D investment and the spillover effects of foreign R&D. Coe and Helpman (1995) calculated the stocks of domestic R&D using the perpetual inventory method with an assumed depreciation rate ranging from 5 to 15%, and calculated the effects on total factor productivity for 22 OECD countries for the period 1971-90. They calculated a marginal rate of social return[43] of 123% for the seven large OECD economies and 85% for the others.

524. Because this study aggregated public and private R&D expenditure, the specific effect of public R&D expenditure on productivity growth was difficult to assess. A study by Guellec and van Pottelsberghe de la Potterie (2001) filled the gap and has become an extremely influential

525. However, the conclusions of this research have been challenged (Sveikauskas, 2007) owing to the lack of detailed microeconomic evidence on the specific mechanisms through which public science affects productivity growth, such as more rapid growth of high-technology industries. Moreover, Khan and Luintel (2006) introduced a number of other potential variables[44] that may explain productivity growth. They did not find that public R&D was a significant factor in productivity growth rates, suggesting that there was no direct link between the two. Finally, other macroecometric studies have also provided only limited evidence on the role of public R&D investment in productivity growth. OECD (2003) analysed different contributions to growth rates experienced in different OECD countries which might explain differences over time. Using cross-country regression analysis and a large set of variables that might explain observed differences in growth, the study concluded that private R&D has high social returns and contributes to economic growth, but that there is no evidence of this for government R&D. In general, the macroeconometric studies have reported high social rates of return, above 50% in many cases, showing the positive effect of overall R&D investment on productivity growth. However, they also suggest that public R&D does not contribute directly to economic growth but has an indirect effect via the impact on private R&D.

526. One limitation of these econometric studies is that they have ignored, at least until recently, the relationships among R&D actors which can provide insight on the innovation processes generated by R&D investment, as they take a relatively linear view of innovation. Besides, while they demonstrate associations between the variables, they seldom demonstrate a causal link. Moreover, they only focus on the relationship between R&D and increased output or productivity. Other objectives of research, such as national security, energy security, environmental protection or health and social cohesion, are excluded from the analysis as they are not captured in measures of economic growth. However, they need to be borne in mind when assessing the impacts of specific public R&D investment.

---

42. The benefits of this type of investment vary according to industry.

43. These estimates are calculated for the lower rate of capital depreciation of 5%.

44. These variables include determinants of productivity growth such education or public infrastructure.

4.2 Capitalisation of R&D

527. The econometric work is currently being complemented by growth accounting analysis which explicitly considers public and private investment in R&D as a source of productive investment. Inclusion of R&D in the national accounts stems from the need to move from a traditional view of R&D as current spending to a growing recognition that R&D should be seen as an investment in intangible capital, which expands a nation's knowledge stock, providing benefits over a number of years. Although R&D capital is commonly used to approximate knowledge stocks, its relationship to growth has not been a focus of national accounts.

528. To treat R&D as investment seems to be conceptually sound. In many cases, R&D closely resembles an investment in generating an asset, namely knowledge capital, that can be drawn on in the future to generate benefits in the form of new products or improved processes that reduce production costs. Therefore, treating R&D as investment expenditure may provide a consistent accounting link between the investment expenditure and the corresponding asset. However, R&D is not a straightforward investment. There is high risk associated with investment in R&D and its economic returns are not assured. Moreover, there are questions about which types of R&D should be subject to capitalisation. A number of issues therefore need to be addressed to ensure credible estimates of R&D capital formation.

529. Preliminary analysis for some OECD countries suggests that R&D investment may account for substantial shares of productivity growth. For the United Kingdom, Edworthy and Wallis (2006) give an estimated elasticity of 0.095% for R&D capital, which implies that a 10% increase in R&D capital is associated with a productivity increase of 0.95%. In the United States, a recent study carried out by the Bureau of Economic Analysis and the National Science Foundation (2007) estimated that R&D capitalisation resulted in an average increase in GDP of 2.9% between 1959 and 2004, and that current dollar private domestic investment in 2004 would be 10.6% higher than the currently published estimate. The results are more modest for the Netherlands. De Haan and Van Rooijen-Horsten (2005) conclude that the effect of capitalisation of R&D adjusts total gross domestic product (GDP) upwards by 1.1 to 1.2%. Equally, economic growth, measured by the volume increase in GDP, is hardly affected. Consequently, adjustments of net national income are also quite modest since upward adjustments of gross fixed capital formation (GFCF) are counterbalanced by negative adjustments from consumption of fixed capital. In principle, the capitalisation of R&D in the national accounts will also show the contribution of public investment in R&D to GDP growth, to the extent that public investment leads to goods and services that can be sold in the market.

5. Impact assessment of research councils and public research organisations

530. Detailed assessments of the impact of public R&D, at the level of individual institutions and programmes, have typically been more successful at identifying impacts (see Box 13) and sometimes evaluation is built into routine policy processes (see Box 14). Research councils and public research organisations can be differentiated according to their functions in the research system and the type of research they carry out. The national research councils (*e.g.* the Australian Research Council) mainly fund the research performed in the country, while public research organisations (*e.g.* Belgium's federal scientific institutes [EFS]) carry out research activities. Hybrids (*e.g.* the National Institutes of Health [NIH] in the United States) both fund and perform research. Some focus on basic research while others are industry-oriented. For example, the Australian Research Council focuses on basic research, the NIH on health and the EFS on space. The next section gives examples of funding and performing institutions engaged in general or sector-specific research and with or without a industry orientation.

Box13. Evaluating PRIs – Insights from the RIHR project

Public research institutions (PRIs), broadly defined as higher education or government research institutes, make a vital contribution to national innovation systems and their performance. They contribute to the formation of a skilled scientific and technological workforce, extend the boundaries of knowledge, and act as an important source of knowledge transfer for the innovation activities of firms. Understanding the exact nature and size of their impact on society and the economy has become an increasingly central question for policy makers in recent years. As governments seek to be more rigorous in their decisions about how much and where to invest in R&D, so they require more information about the contribution to growth and the social impacts of various programmes and institutions.

The RIHR work draws on evaluations of PRIs (excluding pure university institutes) to compare and contrast methodologies, highlight lessons learned regarding both PRI policy and evaluation processes, and examine how evaluation results are used in practice.

In general, the goals of evaluation of PRIs are to better understand the scale, nature and determinants of the return to investment in these institutions as well as to learn about any unintended effects. This information can be used to improve steering and funding decisions, as lessons are drawn from successes and failures. Some key questions that are (or should be) asked in such evaluations relate to the efficiency and effectiveness of institutions and their programmes, the rationale for government intervention and whether the original drivers are still valid, and the level of additionality that is achieved through government funding.

The dimensions of impact can be wide, and evaluation of PRIs has the potential to cover a broad range of less tangible issues, such as impacts on culture, societal views and organisations. These aspects are more difficult to measure, although advances have been made in some areas. Given the inevitable lack of precision in measurement, evaluation can only ever be a guidance tool for policy makers – it cannot provide definitive answers. Value judgements about the scale and direction of government funding for R&D will continue to play a role in steering and funding decisions.

The evaluations examined by RIHR addressed a variety of levels of PRI activity, ranging from the sector to individual projects. Impact or value added was the most common evaluation question to be explored, although the exact meaning of these terms was not always clear, followed by scientific outputs. The most common evaluation methodology was qualitative assessment, relying on interviews and questionnaires to stakeholders. Some evaluations used indicators to inform their assessment of performance. These were generally "backward looking", in terms of summing up past performance, and few evaluations attempted to capture the potential future impacts of research in their assessments. Most of the evaluations judged that the sector, institute or programme/project in question had been of value or had performed adequately; consistent with the choice of methodologies, this decision was often based on an overall judgement as to the costs, benefits and influences of the initiative, rather than a quantitative assessment.

There is scope for improving the use of evaluations for policy-making purposes. There was little information on how the findings and recommendations of the evaluations were practically used, which raises broader questions about why evaluations in general may be overlooked in steering and funding decisions. A useful step would be to improve the relevance of evaluations to decision-makers, by ensuring that evaluation methods and indicators keep pace with the changing environment within which PRIs operate, in particular capturing the increasing numbers of stakeholders and level of cross-sector activity. Raising stakeholder confidence in, and acceptance of, evaluation activities is also important, and could be improved by consciously involving stakeholders earlier in the process. A number of countries have some degree of "built-in" evaluation in their policy processes and these could perhaps be reviewed with these issues in mind.

Evaluations need to more explicitly address the issue of an underlying rationale for intervention. Circumstances change and evaluations need to consider whether a market failure still exists, how strong it is, and to what extent government intervention is still appropriate. In addition, since intervention can gradually change the behaviour of stakeholders in perverse as well as positive ways, it is important to review whether current approaches still work as desired. Clearly this is difficult, and the methodological challenges of creating a counterfactual are real. However, gauging the attitudes of firms and assessing their behaviour can provide some clues. This supports the use of qualitative information as well as quantitative methods of evaluation.

The evaluations highlighted some interesting issues and suggestions on how to improve the operation of PRIs and their programmes. While every country and innovation system is unique, and approaches cannot necessarily be replicated successfully in each one, it may be useful to note some of the issues and solutions that have been found. For instance, many evaluations pointed to the difficulties PRIs have in meeting the expectations of all their stakeholders. This is becoming more acute as the environment becomes more complex and numbers of stakeholders grow. Misalignment of stakeholder goals could be the reason why initiatives fail or produce suboptimal results, and a better understanding of stakeholders' motivations and likely behaviour would be a crucial tool in better policy design.

The setting of research agendas, and the related issue of levels of core funding, continues to be a difficult balancing act. The degree to which PRIs undertake fundamental research, and receive core/capability funding to do so, must be set against the needs of other contributing stakeholders, who may have different time horizons and different priorities. Some suggestions were to focus programming on the overall portfolio of research, rather than individual projects, so as to promote synergy and multi-disciplinarity and allow for a longer-term research focus. Allocating core funding at a higher level was also considered to help promote a strategic research approach and balance the influence of large stakeholders. Allowing flexibility in terms of changes to research agendas and stakeholder investments, while maintaining long-term funding stability, was a related challenge. A number of evaluations pointed to the dangers of "lock-in", but there was no clear solution.

An important issue highlighted by several evaluations was that research alone does not necessarily add value. Thinking about how results will be converted into further research advances or innovations must be an important part of the design of PRIs and their programmes. Dissemination strategies were observed to be inadequate in several cases and it appeared that specific activities were necessary to stimulate "valorisation". The exact approach is likely to differ according to the goals of the institute – for instance, creating "value" through commercialisation may not be appropriate for PRIs with the explicit goal of serving industry. Ensuring that industry has the absorptive capacity to utilise research results is also crucial. Some institutes directly engaged with firms to demonstrate research results and build capacity for future knowledge transfer. Further analysis would be useful in this area. Moreover, while the path to economic impact and effects on innovation is particularly difficult to measure this issue should not be ignored due to methodological difficulties. One way forward would be to take a longer term view and include more in-depth analyses in evaluations.

There continues to be scope for improvements to methodology and design of evaluation procedures, including the incorporation of potential future streams of costs and benefits arising from current and completed research. Advances have been made in measuring social and environmental outcomes and further work will be valuable. However, evaluations are not costless and quantification of outcomes can only be undertaken to the point that it remains cost-effective to do this. Regardless of the value and need for evaluation, the intensity of evaluations in terms of scope and frequency should be justified. Combining a variety of methodological approaches and including a range of stakeholders in the evaluation process can help to overcome individual shortcomings of various approaches and may be a useful way forward while new approaches are being developed. Not all evaluations used clear and time-consistent indicators of performance, and the integration of field specificities was not transparent in most cases. As a general observation, clear goal setting at the start of initiatives will aid in collecting data and indicators that will later help in evaluation exercises.

Finally, evaluation is now taking place in a more complex environment and new demands are being placed on evaluation exercises. Ensuring that evaluation approaches can take account of overlapping roles and responsibilities of stakeholders, multi-disciplinarity, globalisation and more complex funding arrangements will be essential if evaluation is to remain a useful tool for policy-makers.

Source: OECD 2009, Research Institutions and Human Resources (RIHR) project on Strengthening the Impact of Public Research Institutions.

Box 14. Building evaluation into policy processes

A certain degree of evaluation and/or performance monitoring of programmes and institutions is sometimes built into routine policy processes. Some examples from public research institutions (PRIs) are provided below:

Belgium: The Flemish Government concludes a multi-annual management agreement with all public research organisations (PROs), policy-related research centres and special institutes. Every 5 years the execution of the agreement and linked results are evaluated by an external partner supported by an international panel of experts. Following an in-depth evaluation, new agreements were concluded with all PROs in 2006-07. These contained result-oriented criteria such as patents, spin-off companies and publications, in return for which the PROs receive a yearly financial grant.

Finland: The performance management model designed by the Ministry of Finance seeks to improve accountability of public officials, including in publicly funded research organisations. As part of the model, basic criteria for performance are defined and included in legislation: policy effectiveness (or societal impact); operational efficiency; outputs and quality management; and management of human resources. In response, research institutions have developed methods for assessing and monitoring organisational impacts.

Italy: Reforms over the last decade have addressed the framework for evaluation of research. In 1999, reforms aimed to develop a governing structure for the research system that included national research policy evaluation and assessment. As part of this, a Committee for the Evaluation of Research (CIVR) was established. In 2003, an integrated system of research quality assessment was created and in 2007 the National Agency for the Evaluation of Universities and Research was assigned the task of assessing the quality of the research results produced by institutions as well as the efficacy and efficiency of their institutional activities.

Japan: Independent administrative institutions, which perform a diverse range of R&D activities to meet policy challenges, are designed to establish a medium-term goal of 3-5 years. At the end of this period, the institutions are required to conduct an overall review of the organisation and its operations. This is assessed by an evaluation committee within the supervising ministry and by the minister-in-charge. The institutions are also required to submit a report of their annual performance/results to the committee. Many national testing and research institutions, another category of PRI, also undergo organisational reviews every few years, in accordance with the General Guidelines on Evaluating R&D.

New Zealand: Under the Crown Research Institutes (CRIs) Act 1992, CRIs should promote and facilitate the application of the results of their research and technological developments. To shed light on this and the broader overall performance of CRIs, three groups of indicators are used: research application metrics, related to the transfer of results (applied to all CRIs); relationships/influencing role, related to how well CRIs are engaged in their sector (specific to each CRI); and measure of impact, related to the impact of selected research results or technologies that CRIs have applied or transferred over the previous five years (specific to each CRI). The Crown Company Monitoring and Advisory Unit (CCMAU) provides advice to shareholding Ministers on CRI performance, and monitors performance against targets.

Norway: A new core funding system for the research institute sector, to be administered by the Research Council of Norway, was introduced from 1 January 2009. This new scheme incorporates a tranche of performance-based basic funding (around 10%) that is based on institutes' production of scientific publications, co-operation with the higher education sector, income from the Research Council of Norway, income from abroad, and income from national research commissions.

Poland: Recent acts of law, particularly those published since 2001, have strengthened the importance of research/science/innovation policy evaluation and put more emphasis on the effectiveness of research, especially for socio-economic development. The Act of Law on the Financing of Science, which describes the main mechanisms for steering the activities of public research institutions, specifically includes the activities of the Evaluation Committee of Research Units.

United Kingdom: Research Council Institutes are reviewed by their parent Council every five years, both in terms of their research portfolio and their operational effectiveness.

Source: OECD, 2009. Information provided by RIHR country delegates for the Research Institutions and Human Resources (RIHR) project on Strengthening the Impact of Public Research Institutions.

6. Impact assessment of research programmes

531. Research programmes are one of the main instruments used by OECD countries to implement research and innovation policies. They may aim at funding basic or more applied research in a general or a specific sectoral context, with or without a commercial objective. Two of the most important research programmes in terms of resources are the European Union (EU) Framework Programme and the United States Advanced Technology Program (ATP). The nature and scope of the research carried out under these two programmes are very different.

6.1 The EU 7th RTD Framework Programme

532. The EU Research and Technological Development (RTD) Framework Programme (FP), the main multi-annual R&D funding programme at European level, aims at helping to meeting the EU's main goals. Since 1984, the Framework Programmes have played a leading role in multidisciplinary research and co-operative activities in Europe and beyond. The seventh Framework Programme (FP7) continues that task, and is both larger and more comprehensive than earlier Framework Programmes.

533. FP7 bundles all research-related EU initiatives together under a common umbrella which plays a crucial role in reaching the EU's goals of growth, competitiveness and employment. Running from 2007 to 2013, the programme has a budget of EUR 53.2 billion over its seven-year lifespan, the largest funding allocation yet. It funds both basic and applied research and aims at enhancing the research capacities and results of all stakeholders, *i.e.* private companies, individual researchers, universities, public research institutions and foreign actors.

534. The European Commission has attempted to assess the wider impacts of the Framework Programmes on the economy and society. The most significant studies have calculated the impacts on the economy through mathematical modelling. A study by the United Kingdom's Department of Trade and Industry[45] (DTI) analysed the impact on the United Kingdom's total factor productivity, using the model developed by Guellec and van Pottelsberghe de la Potterie (see Box 4.1). According to this study, the estimated annual contribution to UK industrial output would be GBP 3 billion, a very large economic return on UK Framework activity. Similarly, and using the same methodology, a study by the European Commission's Joint Research Centre at Ispra calculated the impacts of the Framework Programme on industry, measuring the increase in total factor productivity. The results seem to indicate that the effects are significant. For example, for Finland, the estimates suggest that 0.9% of annual industry value added is attributable to FP funding and many member states record even higher contributions. On average, it is estimated that EUR 1 of FP funding leads to a (long-term) increase in industry value added of between EUR 7 and EUR 14, depending on the assumptions and parameters used. This increase will be spread over a number of years, because there is always a time lag before R&D spending produces its economic effects.

535. In addition, the 7th Framework Programme has introduced an *ex ante* or prospective calculation of the impacts of expenditure. To do so, it uses a general equilibrium model called NEMESIS (see Box 4.9). This venture, while subject to further improvements, represents a qualitative jump in the *ex ante* impact assessment of research programmes and allows for estimating the benefits of an investment before they occur.

536. In order to assess the impacts of the new Framework Programme, the European Commission drafted three scenarios:

45. Its name has recently changed to the Department of Business, Enterprise and Regulatory Reform (BERR).

- The “do-nothing option” serves to analyse whether without EU intervention it is possible to reach the same objectives.
- The “business as usual option” continues the previous FP, with the same budget allocations, objectives, instruments, priorities and institutional actors.
- The “enhanced Framework Programme option” doubles the resources of the previous FP and is designed to better respond to the Lisbon Agenda objectives.

537. For these scenarios, the NEMESIS model can calculate the different sets of benefits that would accrue. As with all econometric forecasts, of course, the results must be interpreted with cautious, as it is hard to establish a linear causal relationship between specific policies and particular effects, and it is very difficult to quantify many predominately qualitative effects such as increased networking, improved absorptive capacity, strengthened research competencies of firms, or changes in behaviour. In addition to the economic gains, the FP could also have large social impacts, *e.g.* by increasing quality of life for society as a whole.

538. The US's Economic Assessment Office (EAO) tracks the progress of funded projects for several years after the ATP funding ends, and identifies the benefits, both direct and indirect, that ATP award recipients deliver. Direct benefits are achieved when technology development and commercialisation are accelerated, leading to private returns and market spillovers. Indirect benefits are delivered through publications, conference presentations, patents, and other means of dissemination of knowledge.

539. The EAO uses a variety of methods to “measure against mission” the results and impacts of the ATP’s investment. The methods range from early surveys used to generate immediate information to detailed case studies, statistical analysis, tracking of knowledge created and disseminated through patents and citation of patents, and informed judgements. However, as the evaluation of emerging technologies is a relatively new field, existing tools are modified, new ones are developed and/or existing methods are combined in new ways.

540. One of the EAO’s main methods, used on nearly 30 projects to date, is in-depth cost-benefit analysis. The case studies are based on interviews of funded companies, their customers and industry experts, and on other primary data collection activities, such as the Business Reporting System Survey (see Box 15). In the case studies, the benefits directly accruing from the ATP are estimated by the different stakeholders. The time at which the analysis is carried out is important. In general, *ex post* measurement of results already achieved (*e.g.* commercialised technology, sales of innovative products, reduction of costs due to process improvements) need to be combined with *ex ante* prospective analysis of the potential commercial benefits of the project.

Box 15. The Business Reporting System Survey

In early 1994, ATP implemented the Business Reporting System (BRS), a comprehensive data collection tool for tracking progress of its portfolio of projects and individual participants, from project baseline through closeout and into the post-ATP period, against business plans, projected economic goals and the ATP's economic criteria.

The survey is designed to capture economic and organisational changes that are expected in the award recipient population if progress is made towards the expected goals. The themes and topics defined by the goals are reflected in multiple lines of questions that vary in a logical progression over the survey period. Baseline information is collected from the initial survey, and follow-up questions in each area are included at the appropriate anniversary, closeout or post-project survey. Several variants of the surveys are used for different types of organisations . For example, participating non-profit organisations or universities are given a slightly different survey from that given to companies to reflect their specific roles in a project and their different organisational structures.

Intended for immediate use in project management and ATP evaluation, the data are also expected to support analysis of R&D behaviour and outcomes beyond ATP in the longer run.

541. The task is not simple. Prospective studies of project outcomes, particularly if performed before technical risks and uncertainties have been overcome and business risks significantly mitigated, may not generate credible or useful estimates of programme impacts even if they meet high standards of economic modelling and rigour. Probability distributions of long-term advanced technology project outcomes are extremely difficult to estimate. Given the uncertainties about these outcomes, at least some combination of retrospective and prospective analysis is appropriate as long as the analysis includes direct evidence of actual commercialised products or processes that incorporate the project-funded technology.

542. Sometimes, a project that achieves quantifiable economic benefits requires funding from multiple external sources, each of which is indispensable. A conservative approach to assessing the impacts of ATP funding is to allocate benefits in some equitable way among funding sources. Identification and attribution of benefits require matching the programme-funded projects to direct project outcomes, by tracing product outcomes back from company products to their origin in an R&D project and forward from the ATP-funded projects through the product development stages to identify the major contributions and an appropriate attribution of all or partial benefits to the ATP.

543. These studies are consistent with the Office of Management and Budget Circular A-94 recommendations for the use of cost-benefit analysis in general and of cash flow analysis methodology, of net present value (NPV), of cost-benefit ratios and of internal rate of returns. These are key metrics of programme outcomes. A few studies employ other quantitative methodologies, such as hedonic index models.

544. The results of individual cost-benefit studies can be aggregated to see the impact (usually prospective estimates) across ATP. The net social benefits from about 40 ATP projects, for which ATP provided USD 2.2 billion and industry provided USD 2.1 billion, are estimated at USD 18 billion.. However, as these projects were funded and studied at different times, the impacts computed in the different studies are not strictly comparable and their aggregation presents methodological problems.

7. Non-economic impacts

545. A substantial share of public R&D seeks to have an impact that goes beyond economic gains and increases the well-being of citizens. Environment, health, social development and cohesion are a few areas in which public R&D produces impacts that enhance quality of life. Cozzens (2007) classifies these benefits into two broad categories: the "what" and the "how" benefits. The "what" benefits deal with the overall status of individuals and cover elements such as health, education or environmental quality. The "how" benefits relate to the way we live our lives. Equity, democracy or community development are

examples of this dimension. Public research is conducted in a wide range of disciplines that have impacts that increase the well-being of citizens: health and environmental research, social science research, humanities, etc.

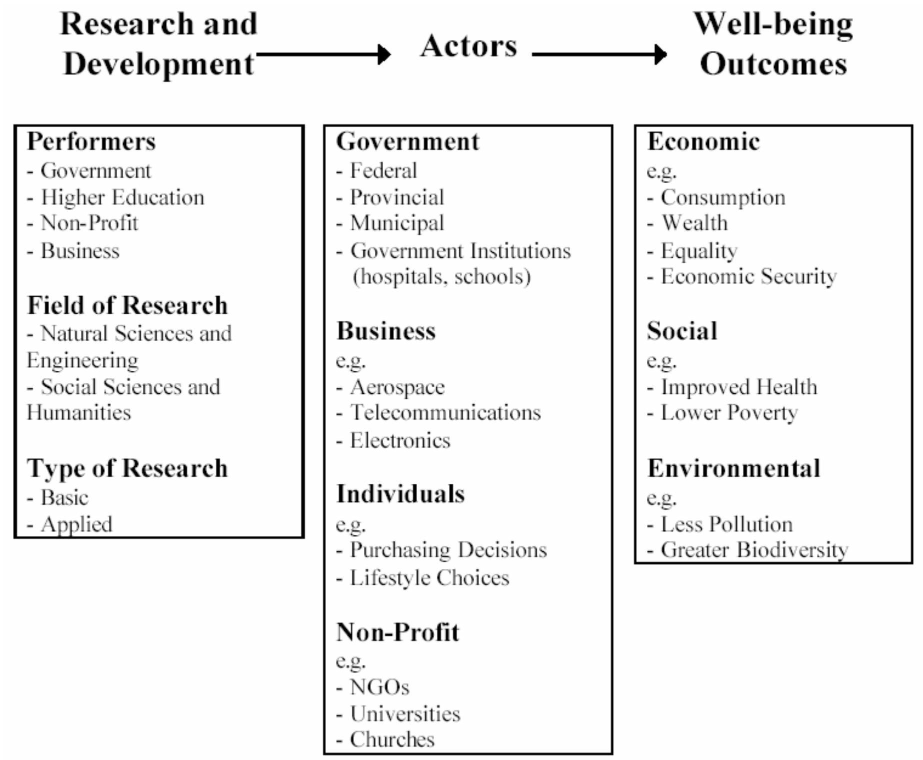
546. Unfortunately, the literature on the non-economic impacts of science is much less abundant and robust than the studies of economic impacts. Godin and Doré (2006) identify three main reasons for the scarce production of non-economic impacts studies. The first is that most measurement of science and research has been undertaken in an economic context. The second is that the economic dimension is often easier to measure. Finally, most of the outputs and impacts of science are intangible, diffuse and often occur with important lags. Although also difficult to measure, the economic dimension of science and technology[46] remains the least difficult.

547. Nevertheless, in recent years, researchers and governments have started to be interested in the non-economic impacts of public R&D, and progress has been made. There is a certain consensus among researchers that one of the first steps towards advancing understanding of the non-economic impacts of public R&D is to define a framework that links research investment and well-being (Sharpe and Smith, 2005). Cozzens (2007) argues that social outcome indicators for research are neither difficult nor rare and that there exist dozens of indicators relating to the public goals of research. In her view, what is lacking is not outcome indicators but the logic that connects research and innovation to outcome indicators.

548. Sharpe and Smith (2005) develop a basic general framework for assessing the impact of research on well-being. This basic framework (Figure 15) links research investment with well-being via the uses made by social actors of the increased knowledge generated by research. This general framework can in principle capture the impact of many different types of research investments used by different social actors to affect numerous dimensions of well-being.

46. Godin and Doré (2006) use the concept of science and technology, which is a broader concept than public R&D. However, they share the same problem in terms of impact assessments.

Figure 15. Framework for analysing the effects of research on well-being



Source: Sharpe and Smith (2005).

549. This model requires adopting a four-step approach in order to measure impacts on well-being and establish their connection to public research. These four steps are:

Define the broad domains of well-being (social, economic, environmental, etc.) that are of particular interest, as well as sub-domains within the broad domains (*e.g.* within the social domain, the sub-domains child well-being, education, etc.).

Choose concrete indicators that can capture the domains or sub-domains.

Identify research investments that influence or determine the chosen indicators and specify the paths through which these investments and the knowledge created affect the indicators.

Quantify the impact of particular research investments on the indicators of interest.

550. The model then should be able to use a mix of indicators to track changes in the desired outcome area and should make it possible to attribute the proportions of the changes to the research effort. Of course, the attribution of effects is not easy, especially given the diverse factors affecting the final outcome and the time that may elapse between the public investment and the perception of the impact. However, such attributions should be made possible thanks to the use of expert judgements, the timing of change or direct causal connections (Cozzens, 2007).

551. In health and environmental sciences, the development of metrics of social impacts is probably more advanced than in other fields, mainly because the causal relationship between investment and impact tends to be clearer and so is the attribution of benefits**.** However, in other cases, as the Allen Consulting

Group (2005) recognises, it is very difficult to express the primary social benefits[47] generated by using a common expression of value such as the social rate of return. In general, it must be realised that the most that can be done is to highlight where these impacts occur and articulate qualitatively the "value" of these impacts on society. To do this comprehensively, it would be necessary to "tell the story" of the impacts, and that is why the case study approach has mainly been adopted.

552. As a result, there is still a need to improve the models that link public R&D with well-being in order to overcome some of the difficulties inherent in this type of analysis. In particular, these models should emphasise the need to specify what specific research investments and what dimensions of well-being are of interest before undertaking any empirical work to estimate the impacts. Moreover, these models should deal with the problems of attributing the credit for impacts on well-being to public R&D. Several methods, such as the use of expert judgements, the timing of changes or direct causal connections can help, although the attribution can often only be made on the basis of disputable assumptions. Further work is needed to overcome these difficulties and obtain better estimates.

8. Conclusions

553. This chapter stresses the importance of understanding and measuring the impacts of public R&D investments in order to evaluate the efficiency of public spending, assess its contribution to achieving social and economic objectives and legitimise public intervention by enhancing public accountability. It has presented some of the most promising and forward-looking practices adopted in this respect: general equilibrium models, econometric analyses, data linkages and scientometrics methods, survey-based indicators combined with econometric analyses and case studies. These are a few of the analytical techniques that governments can use to assess the impacts of their spending on R&D. Other techniques, such as the use of experts (*e.g.* peer reviews), Delphi methods, technological foresight, systematic approach, sociological and socio-economic, longitudinal and historical methods are also options in the toolkit available for impact assessment.

554. The choice of methodology (methodologies) must be made in the context of an evaluation of specific research. An impact assessment exercise requires a deliberate selection of the dimensions that will shape the exercise. These are the timing (*e.g. ex ante,* monitoring, *ex post*), the object to be assessed (*e.g.* a research programme, public research organisation or a research system), and the specific nature of the research, *i.e.* whether it is basic science or technology development, and whether or not it is primarily industry-oriented.

555. When deciding which methodology to apply, it is also important to consider the scope of the impacts to be measured. Public R&D may have impacts at different levels of the economy or society and public R&D impact assessment exercises may focus on assessing the impacts of that investment on a specific sector, such as space or health, or on the overall economy or society. As a result, no single analytical method can be used in all contexts. In fact, methodologies tend to be quite context-specific and specific factors determine their appropriateness in a given situation.

556. This review found top-down approaches, especially econometric and mathematical models, better suited to assess impacts affecting the whole research system and that deal with all types of research, both basic and applied. In particular, mathematical models, such as general equilibrium or similar models, may be a good way to assess systemic impacts *ex ante*.

---

47. The Allen Consulting Group's classification uses "the human, environmental and social dimension of benefits" as the equivalent of what is here called the social or non-economic benefits of public R&D.

557. On the other hand, when the subject of the assessment is a research programme and/or institution that aims at developing a specific type of technology with a clear industrial focus, bottom-up approaches are favoured. Both identifying and measuring benefits by surveying potential users of the specific technology and the calculation of the impacts are easier. For large research programmes or institutions carrying out a wide range of research activities that are not particularly focused on specific technologies or industries, case study analyses that identify and quantify benefits and track them back to the original sources seem to be an option. Finally, case studies describing the main benefits, together with a narrative about these benefits, seem to be the only option for assessing the non-economic impacts of public R&D at present. In general, these methods seem to work better for *ex post* assessment. In the case of *ex ante* impact assessments, uncertainty about the type and nature of the benefits that may accrue and the time required for them to appear make these methods less accurate. As yet there are few *ex ante* impact assessments dealing with the specific impacts of research programmes or institutions. Most *ex ante* studies have focused on assessing systemic macroeconomic effects deriving from the research investment. Accurate *ex ante* identification of specific benefits and potential users is still limited.

558. When assessing the impacts of public R&D, it is also important to distinguish between publicly funded and publicly performed R&D. The objectives and scope of the activities differ, which may explain differences in returns to public resources. Publicly funded but privately performed R&D may have a more targeted objective and achieve more immediate results. On the other hand, publicly performed R&D may focus on basic research that might otherwise not be carried out and may take a long time to produce visible impacts, which may be more difficult to attribute to the original research. Therefore, distinguishing between publicly funded and publicly performed R&D when evaluating the impact of investments may provide a better picture of the returns.

559. This chapter has also shown that the various methodologies are still evolving and based on a series of working assumptions that must be borne in mind when drawing conclusions. Because of the many types of public R&D undertaken and the many different dimensions of well-being affected by these activities, it is very difficult to develop a framework that captures all the possible impacts of public R&D. As a result, until now, none of the available techniques has been able to capture the full range of impacts of public R&D on society, although they have opened new and encouraging lines of investigation.

560. In practice, since socio-economic impacts are complex and very different in nature, it is recommended to use a variety of methods to assess them. Where systematic and continuous assessments have been carried out using a range of methods, the coverage of impacts is better and the overall effectiveness and efficiency of the public investment can be better analysed.

561. Further work is needed on integrating different approaches and methodologies to create coherent impact assessment practices. More integrated frameworks using a combination of complementary methods should be explored. For now, no common framework for developing and using these analytical techniques has been agreed and international collaboration in this field is still scarce. The scope, nature and objectives of public R&D are diverse across OECD countries, as are national socio-economic demands for public research. Therefore, it may be difficult, due notably data and methodological limitations, to achieve full international comparability and benchmarks. However, this should not imply that countries should or can give up.

562. Finally, although methodologies for impact assessment remain a challenge, it is also crucial to recognise that some important values of scientific research will remain hard to quantify. Investment in some areas of basic science is primarily made to satisfy human curiosity and deepen our understanding of the universe. In some cases, such research may prove to have benefits beyond pure knowledge and the satisfaction of curiosity; in others, it may not. A related and perhaps larger challenge remains that in many cases the results do not feed back into the policy debate.

REFERENCES

Access Economics Ltd (2003), "Exceptional Returns: The Value of Investing in Health R&D in Australia", The Australian Society for Medical Research, Canberra.

Allen Consulting Group (2003), "A Wealth of Knowledge: The Return on Investment from ARC-funded Research", *Report from the Australian Research Council*, Canberra.

Allen Consulting Group (2005), "Measuring the Impacts of Publicly Funded Research", commissioned by the Australian Government, Department of Education, Science and Training.

Bureau of Economic Analysis and National Science Foundation (2007), "2007 Research and Development Satellite Account", mimeo.

Coe, D.T. and E. Helpman (1995), "International R&D Spillovers", *European Economic Review*, 39.5, pp. 859-887.

Cozzens, S.E. (2007), "Maximizing Social Impacts through Science and Technology: Carson's Quadrant", Presentation at the "Impacts of Science", 10-11 May, Vienna.

de Haan, M and M. van Rooijen-Horsten (2005), "Measuring R&D Output and Capital Formation in Open Economies" workshop on "The Measurement and Analysis of Innovation and Productivity Growth", Groningen, November.

Edworthy, E. and G. Wallis (2006), "Research and Development as a Value Creating Asset", UK Office for National Statistics (ONS), mimeo.

Godin, B. and C. Doré (2006), "Measuring the Impacts of Science: beyond the Economic Dimension", Working Paper, mimeo.

Griliches, Z. (1979), "Issues in Assessing the Contribution of R&D to Productivity", *Bell Journal of Economics*, 10.1, pp. 92-116.

Guellec, D. and B. van Pottelsberghe de la Potterie (2001), "R&D and Productivity Growth: Panel Data Analysis for 16 OECD Countries", *Oxford Bulletin of Economics and Statistics*, *OECD Economic Studies*, pp. 103-125.

Igami, M. and A. Saka (2007) "Capturing the Evolving Nature of Science, the Development of New Scientific Indicators and the Mapping of Science", OECD, STI Working Paper 2007/1, Paris.

Jacobsson, S. (2002). "Universities and Industrial Transformation: An Interpretative and Selective Literature Study with Special Emphasis on Sweden", *Science and Public Policy* 29.5, pp. 345-365.

Jaffe, A.B. (1986), "Technological Opportunity and Spillovers of R&D: Evidence from Firms' Patents, Profits and Market Share" *American Economic Review*, 76, pp. 984-1001.

Jankowski, J.E. (1991), "Do We Need a Price Index for Industrial R&D?", mimeo.

Kanninen, S. and T. Lemola (2006), "Methods for Evaluating the Impact of Basic Research Funding", Publications of the Academy of Finland, 9/06.

Khan, M. and K.B. Luintel (2006), "Sources of Knowledge and Productivity: How Robust is the Relationship?", DSTI Working Papers 2006/6, OECD Paris.

Kirschner, M. *et al*. (1994), "The Role of Biomedical Research in Health Care Reform", *Science* Vol. 266, No. 5182, pp. 49-51.

Kline, S. and N. Rosenberg (1986), "An Overview of Innovation", in R. Landau and N. Rosenberg (eds.), *The Positive Sum Strategy: Harnessing Technology for Economic Growth*, National Academy Press, Washington, DC.

Lichtenberg, F. and D. Siegel (1991), "The Impact of R&D Investment on Productivity: New Evidence using Linked R&D-LRD Data" *Economic Inquiry*, 29, pp. 203-228.

Lichtenberg, F. (1996), "The Effect of Pharmaceutical Utilization and Innovation on Hospitalisation and Mortality", NBER Working Paper 5418, Cambridge, MA.

Mamuneas, T.P. and M.I. Nadiri (1996), "Public R&D Policies and Cost Behaviour of the US Manufacturing Industries", *Journal of Public Economics* 63, pp. 57-81.

Meyer-Krahmer, F. and U. Schmoch (1998), "Science-based Technologies: University-industry Interactions in Four Fields", *Research Policy* 27(8), pp. 835-851.

Murphy, K.M. and R.H. Topel (2006), "The Value of Health and Longevity", *Journal of Political Economy* 114.5.

OECD (2002), *Frascati Manual: Proposed Standard Practice for Surveys on Research and Experimental Development*, OECD, Paris.

OECD (2003), *The Sources of Economic Growth in OECD Countries*, OECD, Paris.

OECD (2007a), *Main Science and Technological Indicators*, OECD, Paris.

OECD (2007b), *Science, Technology and Industry Scoreboard, OECD, Paris.*

Pavitt, K. (1998), "The Social Shaping of the National Science Base", *Research Policy* 27, pp. 793-805.

Productivity Commission (2007), *Public Support for Science and Innovation*, Productivity Commission, Camberra.

Salter, A.J. and B.R. Martin (2001), "The Economic Benefits of Publicly Funded Basic Research. A Critical Review", *Electronic Working Papers Series*, 34. Brighton, UK.

Salter, A.J. and B.R. Martin (2007), "The Economic Benefits of Publicly Funded Basic Research", presented at the conference The Impacts of Science, Vienna.

Sharpe, A. and J. Smith (2005), "Measuring the Impacts of Research on Well-being: A Survey of Indicators of Well-being" Centre for the Study of Living Standards- Centre d'étude des niveaux de vie", Ottawa, Canada.

Solow, R.M. (1956), "A Contribution to the Theory of Economic Growth", *Quarterly Journal of Economics*, 70.1, pp. 65-94.

Sveikauskas, L. (2007), "R&D and Productivity Growth: A Review of the Literature", BLS Working Papers 408.

Therrien, P. (2006), "Government Role in R&D and Result Indicators from Selected Federal Departments", *Blue Sky Conference, Ottawa.*

United States Senate (2000), "The Benefits of Medical Research and the Role of the NIH", Office of the Chairman, *http://jec.senate.gov*.

Veugeler, R. (2006), "Literature Review on M&A and R&D"in B. Cassiman and M.G. Colombo (eds.), *Mergers and Acquisitions: The Innovation Impact*, Edward Elgar, pp. 79-118.

Wisconsin Association for Biomedical Research and Education (1995), "Bioscience Research, Development and Industry: Impact on Health and Economic Growth in Wisconsin", *www.wabre.org.*

CHAPTER 5
PRIORITY SETTING IN SCIENCE AND TECHNOLOGY AND THE ROLE OF EVALUATION IN DECISION MAKING

This chapter presents some of the main issues policy issues surrounding priority setting and the role of evaluation in decision making. It discusses the role of the different actors in the system and the different tools used to help S&T decision makers set priorities. It then seeks to identify best practices for improving both the quality of *ex ante* evaluations and their usefulness in the policy making process, in particular priority setting. It also assesses the process of priority setting in S&T itself and identifies structural weaknesses as well as best practice solutions.

1. Introduction

563. As research and innovation take on a more central role in economic development but also as science and technology (S&T) are increasingly solicited to help governments meet societal challenges on a global scale, the setting of priorities for public R&D and innovation has become a more complex and urgent challenge. However priority setting is no easy task and requires not only political vision and clear societal goals but tools and mechanisms that can help government setting and implementing priorities, in order to allocate scarce public resources across areas, actors and types of research. Evaluation is one of these tools, in particular ex ante evaluations.

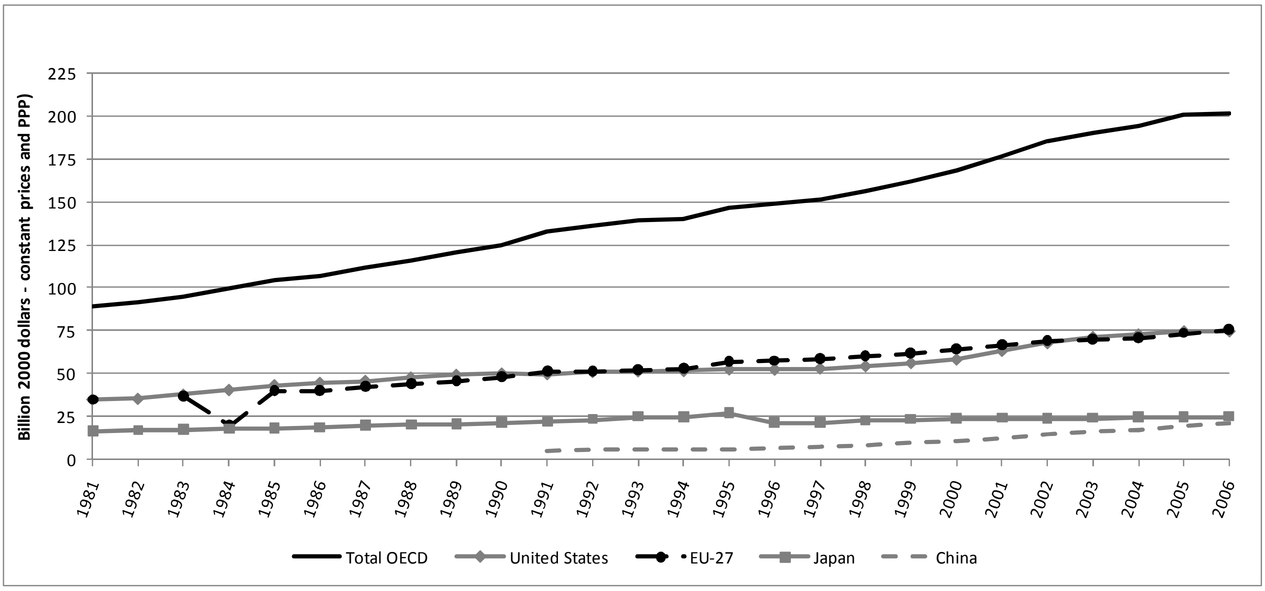
2. The rationale for priority setting

564. The rationale for setting priorities for public investment in science and technology continues to evolve. While scientific excellence continues to be the dominant rationale for funding, in particular for funding of basic and fundamental research, the focus on accountability and the social and economic benefits of research has become more pronounced over the past decades. Indeed over much of the post WWII period, it was largely taken as given that science should be performed without thought about its practical use, in order to expand knowledge and understanding of the natural world. Science would, through the mechanism of technology transfer, drive technological innovation and contribute to economic and social welfare (Bush, 1945). Implicit in this understanding was that the priorities for (basic) research should be set by researchers themselves.

565. In practice, this situation involved an implicit social contract between scientists and society. Over the past decades, the terms of this contract have changed as technological progress and a range of reforms to government in democratic societies have resulted in greater calls for accountability of publicly funded research. These demands not only focused on achieving pay-offs from public investment in research, e.g. in terms of “national competitiveness”, better health outcomes, etc, but also on accountability of science regarding the potential negative implications of certain scientific activities.

566. The pressure for greater accountability and economic benefits has increased alongside a rapid and absolute rise in public support to R&D, even if the share of government R&D has tended to fall relative to business R&D and higher education R&D. Still, the combined amount of public R&D expenditures (*i.e.* the sum of higher education R&D and government R&D) in the OECD area has risen from USD 89 billion in 1981 to USD 202 billion (in constant 2000 PPP USD) in 2006. Given the large amounts of public expenditures there is both a demand for evidence as to the outcomes and impacts of such investment and demands for more robust mechanisms through which governments set – or delegate to others— the priorities for research. Today, many countries are struggling with the challenge of allocating resources in a manner that improves economic performance without restricting the freedom of individual scientists to set their own research directions.

Figure 16. Trends in Total Public R&D Expenditures in the Triad and China, 1981-2006



Source: MSTI, 2008

3. Defining priority setting

567. Priority setting for science and technology can be defined as "the selection of certain activities at the expense of others with an impact on the allocation of public resources" (Polt, 2007). In short, it is the decision on who and what and how much research to fund and for how long. Priority setting, however, is not a mechanical process; instead it is a democratic process involving compromise among the different actors in the system (e.g. researchers, institutions, funding agencies, ministries). While priority setting remains a political bargaining process, there have been attempts since the 1980s and 1990s to improve the conceptual underpinnings of priority setting by greater use of technology foresight and technology assessment tools as well as through greater stakeholder involvement.

568. Priority setting is highly context specific and largely influenced, if not determined, by the institutional settings that govern science and technology as well as by the technological specialisation of different countries. Barré (2008) suggests that the basis for setting "national priorities" is to establish a rationale and a discourse about the national S&T policy as well as to give it political visibility. A secondary aim is to identify and highlight policy measures to improve the functioning of the national innovation system (e.g. universities and framework conditions for innovation policy) and finally, to give signals about the longer term shifts in relative funding among broad sectors. In contrast, priority setting at

sectoral level allows governments to give special attention to a few politically significant issues/sectors (challenges) and to design an integrated set of actions to address them.

569. Polt (2007) categorises three main dimensions of the priority setting processes, namely:

- *Types* of priorities: thematic priorities (scientific, technological, societal missions) or functional/generic priorities.
- *Levels* of priority setting for example national priority-setting exercises, institutional priority setting, etc.
- *Nature* of the priority setting process (e.g. top-down/expert-based vs. bottom-up/participatory, degree of formalization, mechanisms for implementation, evaluation).

4. The process of priority setting

570. The process of priority setting for research themes or areas can be divided into top-down approaches and bottom up approaches. Top-down mechanisms include governmental priorities expressed by government ministries which reflect strategic priorities such as economic development or public missions such as health. Bottom-up approaches essentially reflect the priorities of the producers of research, the researchers themselves, research institutions and funding agents.

571. With the exception of mission-oriented research institutions, bottom-up approaches have dominated thematic priorities for longer-term and fundamental research that make up the core funding of public research in many countries. Peer review of research publications has been the main basis on which research has been assessed to determine priorities. In practice, *ex post* and *ex ante* evaluations of research policies and instruments have had less an impact on the priority setting process. Nevertheless, although research performing institutions, especially universities, enjoy a high degree of autonomy, and set their own priorities according to their own criteria, the priorities of the public research funding agencies are inevitably reflected in the priorities of the performing institutions. For example, programme or project funding is often is tied to priorities of the funding agencies. These national priorities for societal objectives such as health and environment, as well as, broader objectives such as "knowledge-based" or "information" society can influence thematic research priorities at the operational level.

572. During the 1990s, many OECD countries began addressing gaps in their innovation system by making "functional" or structural issues a priority for policy action and (additional) funding. Functional priorities range from increases in research funding, strengthening university research, promoting basic research, to increasing women's participation, sustainable development, or specific technology areas (ICT and biotechnology). By identifying and setting these functional priorities policy makers aim to enhance the functioning of the national innovation system. Functional priorities have been expressed at national level in the context of various Green Papers and White Books with policy statements.

573. National priority setting can have many very different forms. At the macro level it can be expressed in government white papers, national innovation strategies or national S&T plans. At the operational level, priorities can be expressed via the missions of institutions, or through more flexible structures such as centres of excellence. More recently, governments have increasingly used instruments like research and technology programmes and performance-based contracting and public-private partnerships as more flexible ways of influencing the research agenda of research institutions (which in turn partly have a high degree of autonomy in setting their research agenda) Moreover, funding instruments also serve to adjust or set national priorities. Industry financing of public research or in public/private partnerships can also shift or align public priorities for research with business strategies, including both longer as well as shorter time horizons.

Table 6. Forms of Priority Setting

Strategic policy instruments	Institutional instruments	Funding instruments
Government White Papers Policies addressing specific industrial sectors, clusters (e.g. in the form of the 'Technology Platforms' established by the EU) Government procurement for specific sectors/technologies Strategic Research Agendas of research teams	Targeted Research and Technology Programmes Research Institutions with specific profiles and technological orientation Centres of Excellence	Budget plans & allocations Performance-Based Contracting for public research institutions Public-Private Partnerships Industry funding of public research

574. These (and yet other) means for priority setting have been used in varying degrees over time and in different innovation systems: *e.g.* in countries with a strong military focus of R&D, government procurement of R&D has typically played an important role. Countries with a strong focus on other large technological systems (like nuclear energy) have also had a strong focus on such mechanisms. Also, the setting up of dedicated public research institutions (sometimes with a quite narrow technological focus) was an early trend in the definition and implementation of S&T policy priorities. Again, these means are not confined to governments. Agencies and research institutions have themselves built up capacities to engage in priority setting and the implementation of priorities.

5. Actors in priority setting

575. Prioritising different fields of research is subject to a complex decision making process involving not only the scientific community but also stakeholders outside science including civil society (patient groups, industry, agriculture etc). Who are the stakeholders that take part in the priority setting process? All actors in research funding and performance take part in setting priorities. This includes the sponsors of research, the intermediary agencies, the research performing institution and the researcher him or herself. Depending on the institutional context, industry and the social partners as well as civil society also play an important role in the process, both directly and indirectly. Industry and civil society often participate in formal consultative mechanisms such as advisory councils or university boards etc. However, sometimes civil society actors such as private non-profit foundations (e.g. the Bill and Melinda Gates Foundation) help influence government priorities by making their own research priorities highly visible so as to influence public choices.[48]

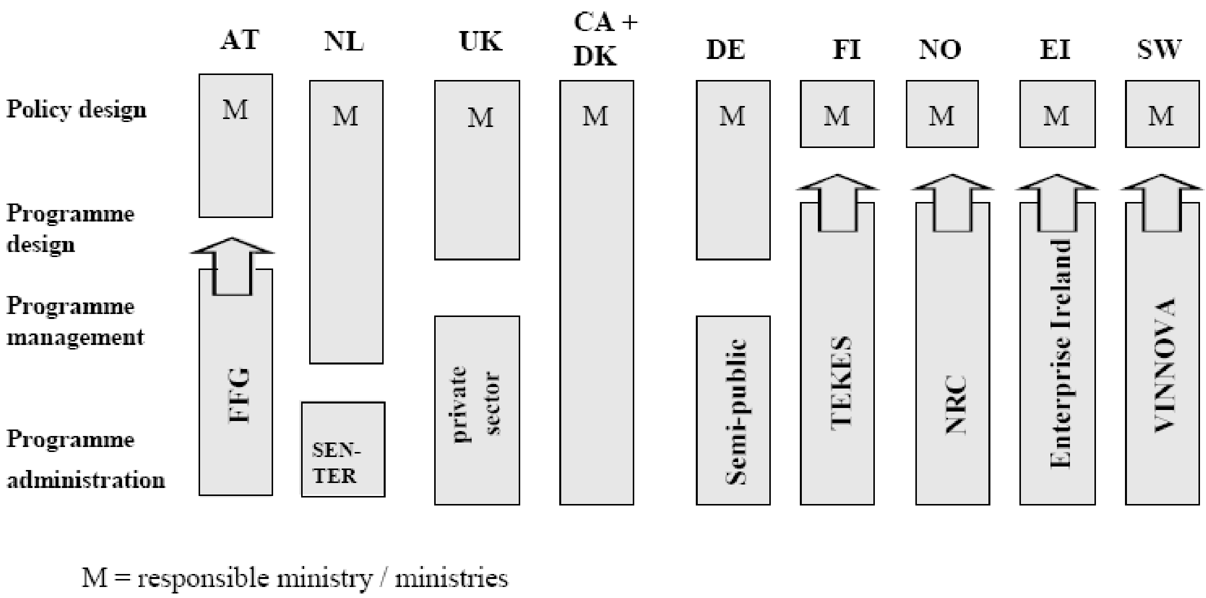
576. In most countries, priority setting is neither centralised nor does it occur primarily at the level of federal/central government. Rather, in the course of a growing division of labour in the policy systems, there are a number of actors involved in STI policy making, ranging from agencies (research councils, funding agencies) to regional governments. A number of approaches to set priorities can be observed, but nearly all include stakeholder involvement of some kind, including consultations with enterprises or business associations, as well as NGOs.

48 The role of such charities and foundations is growing, but remains poorly documented. MSTI data for the OECD as a whole show that the share of gross expenditure on R&D funded by the private non-profit sector rose from 2.3% in 1981 to 2.6% in 2006. However, there is considerable cross-country variation in this indicator, with the US share being the highest, at 4.3% in 2006.

577. A typical division of labour, that can be found in a number of countries, involves government (quite often with the help of science and technology policy councils) formulating broad policy orientations indicating budget frameworks and general goals in science, technology and innovation policy strategy documents, while more concrete priorities are decided at the level of individual ministries and funding agencies.

578. The trend towards devolution of competences to specialised agencies results in a build-up of competences both in evaluation and in priority setting at these institutions. Examples include TEKES in Finland, IWT in Belgium and FFG in Austria—all main funding institutions in their respective countries. Priority setting in such complex policy systems involves close interaction and bargaining between the 'principals' (governments, ministries) and the 'agents' (funding agencies, public research institutes).

Figure 17. Governance structure of technology policy in selected countries



Note: This Figure does not necessarily represent the current state of division of labour between actors. It is shown here as an illustration of the different possible configurations between actors.

Source: Arnold 2004

6. Institutional features and mechanisms for priority setting: some examples

579. Although priority setting is typically not centralised, some attempts have been made to centralise and especially co-ordinate priority setting. This may be done through research funding decision making mechanisms. Broadly speaking, there are countries in which the top-down approach dominates on the one hand, and on the other, those in which the bottom-up approach is more important. In all countries, both top-down and bottom-up forces exist, and some countries attempt to integrate the two approaches. The recent trend is that in many countries there seem to be increasing tensions and shifts in this balance, making priority setting a major policy issue.

580. In countries where the top-down approach dominates, the central government adopts explicit strategies, policies or plans that specify priority areas of research (e.g. Austria, Japan, and Norway). Most of these countries, as well as some others (e.g. Netherlands, Denmark, Germany and Korea) have some kind of central advisory body that makes recommendations about priorities.

581. In Japan, the Council on S&T Policy (CSTP) is chaired by the Prime Minister and meets every month. It is responsible for preparing the Plural-annual Basic Plan. The CSTP assesses the compatibility of each ministerial research programme or initiative with the Basic Plan. The plan has mostly transversal objectives, but also includes large sectoral priorities. Japan also produces a "White Paper" more oriented towards societal issues. Ministries are given a role similar to agencies, with which they work closely. Policy guidance and co-ordination is assured by the CSTP.

582. In France, the situation is evolving as the country moves from a system of dominant public research organisations (PROs) relative to universities, towards a more functionally driven design. Traditionally the steering and execution of research in France has been divided as follows:

- Steering of research (government);
- Programming of research (intermediate organisations, funding agencies); and
- Research performance (institutes, universities and their departments / units).

583. But in practice, the 'national S&T policy' has largely reflected the *ex-post* sum of the sectorally defined policies and strategies of the PROs. A major implication of the current reform embodied in the 2006 Law on Research is the need for an explicit *ex-ante* national S&T policy. This has arisen because of the need for performance indicators in the national budget, the demand for political transparency and rationale, and new players enter the steering and programming process (i.e. the ANR- *Agence Nationale pour la Recherche*).

584. In July 2008, France launched its "National Strategy for Research and Innovation" (NSRI) which is to be presented to the Council of Ministers in April 2009 and to be updated every four years. The National Strategy for Research and Innovation (NSRI) serves as the "steering" document. It sets the stage for an overall vision and multi-annual perspective that is coherent with the European Union's research and innovation priorities. The demand for 'priority setting' is really a demand for formulating a national policy which makes political sense by highlighting societal challenges the nation addresses in view of making priorities in the budgetary process.

585. At the other end, several countries use bottom-up, decentralised approaches. In the United States and Canada, the government advisory bodies for research are decentralised and serve different government agencies in priority setting. In other countries where a central advisory body does not exist, for example in Sweden, priority setting is left to individual government ministries and agencies.

586. In the United States, federal priority setting for research occurs at three levels: 1) in setting federal goals for research; 2) in the budget allocation processes for research within the White House and the Congress that in the aggregate produces the federal research portfolio and 3) in Federal agencies and departments focused on achieving their missions in alignment with the President's priorities for research. The agencies' advice on priorities comes from Federal Advisory Committees which are set up by different agencies that fund research. These committees make recommendations based on reports from the President's Committee on S&T, the National Academy of Sciences, and the President's science advisor, workshops organised by the agencies, and advice from professional societies. Membership of Federal Advisory Committees is supposed to include all stakeholders including industry. The US federal government-wide budget cycle allows agencies to coordinate their proposals and receive funding for identified priority areas.

587. In some countries, there is an explicit integration of top-down and bottom-up approaches. Germany, for example, has a decentralised research system with strong autonomy of public research institutions as well as universities. Priorities are set at the level of individual institutions as well as a result of dialogue between the government and the scientific community represented in research funding bodies and the publicly funded research performing institutions. However, despite its decentralised structure, Germany has a Science Council, which is an independent advisory body, consisting of representatives from the scientific community, government, business and civil society. It plays an important part in making recommendations on priority areas and conducts evaluations of research institutions and programmes. National action plans and priorities are implemented mostly through the programmes managed by the federal ministries. There is a significant effort made for inter-ministerial and inter-instrumental coordination. There is also an important role for the Research – Industry alliance (FWW), a committee of high level industrialists and heads of public research organisation.

588. In the United Kingdom, national S&T priorities are articulated through the Spending Review process of the Science budget and the Technology Strategy Board (TSB). Co-ordination is carried out by the Research Councils. In the UK, both research and innovation priorities are addressed in the total science budget. There are clear-cut roles for evaluation and management. The process allows for interactions and input from bottom-up. A key trend is the emphasis on stakeholder input and expert advice. Bibliometric tools are also used to monitor evaluation outcomes. An effort to monitor socio-economic effects before research is carried out is equally important. The challenge for the UK is to support decision making in strategic areas. Evaluations can play a role by proving input on the context and procedures for S&T policy making

589. In many countries the research priorities are directly linked to annual funding decisions which can create tensions between longer-term and shorter term objectives. In some cases, funding decisions are focused on increasing investment in new "initiatives" or "priority areas". Some of these are inter-agency initiatives (*e.g.*, the US National Nanotechnology Initiative) or involve the creation of new structures such as the Canadian Innovation Foundation, focusing on infrastructure and Genomics Canada.

7. Strategic policy intelligence and priority setting

590. Priority setting has always and foremost been a reflection of political priorities and a reflection of political bargaining processes. Yet from the onset, policy makers have also sought the support of different futures studies and tools of what nowadays is called 'strategic policy intelligence'. In the early phases, these tools included primarily technology planning and forecasting (e.g. in the 1960s, 1970s). Many governments use foresight processes as a part of their priority setting procedure or to stimulate dialogue. Canada uses different types of foresight adapted for various priority setting needs. The United Kingdom has had a government level foresight programme since the 1990s and the government departments are obliged to take account of foresight when developing their science and innovation strategies. Japan has been conducting periodic technology forecasting exercise using the Delphi method since 1970s. Korea also conducts foresight, and the results are implicitly integrated into national priorities by experts who are involved in evaluation and pre-budget review.

Technology road mapping

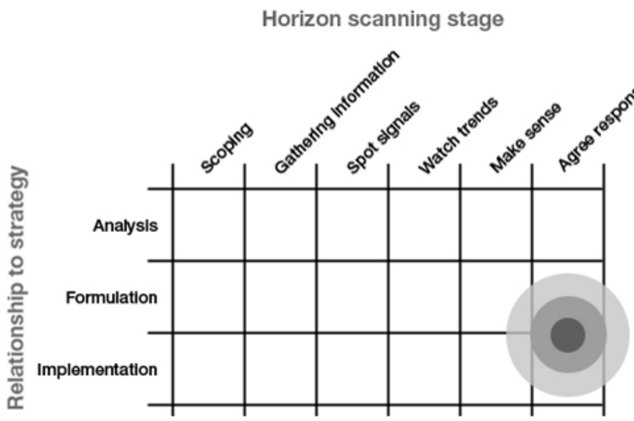
591. More recently, with the inclusion of strategic economic and societal goals in the S&T policy agenda, technology assessment, and technology road mapping were added to this toolbox in the 1980s and 1990s. Technology roadmaps are processes in which technologists try to determine the trajectory of a technology and the developments needed to keep it on its trajectory. Most roadmaps have been organized around existing industries or technologies, such as semiconductors, optoelectronics, aluminium, pulp and paper, electronic packaging. Some have included an assessment of future market needs, as well as the

technology developments. Many roadmaps have been done by and for industry, but the industries have often received government funding to support the road mapping process. Technology road mapping is especially important in complex technologies where there is a need to coordinate the development of many components and subsystems. Roadmaps identify the minimum performance needed for future technologies in order to be part of the system. Roadmaps appear to be most useful in established technologies that are evolving incrementally. They usually do not consider the social effects of the technology (Cheny, 2003). At OECD, an important effort to develop a technology roadmap concerns energy (IEA, 2008).

Horizon scanning

592. In recent years, horizon scanning exercises have become a mainstream activity, as a follow-up to futures and foresight exercises, notably in the United Kingdom and then emulated in other countries, including Australia, Denmark and the Netherlands. Horizon scanning is a distinct futures methodology that researches and draws out key trends which are on the margins of current thinking, but which will impact on people's lives in the future. Most horizon scanning exercises aim to provide advance notice of significant new and emerging risks and opportunities, to exchange information and to evaluate the potential impacts. This involves the review of a broad spectrum of information, beyond usual timescales and sources and the involvement of participants from various sectors of society. Smaller economies have perhaps been the most active with regard to using foresight and other futures oriented studies to inform priority setting because of the need to focus and get returns from relatively small investments.

Figure 18. Horizon scanning for policy making



Source: UK Foresight Centre 2008

Specialisation indices

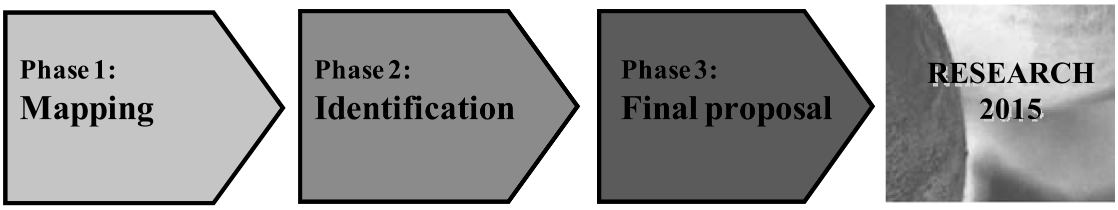
593. A more recent strategic intelligence tool has been the development of specialisation indices to help policy makers take stock of a country's absolute and relative strengths and weakness in a given research or technological field. Specialisation indicators are useful in assessing characteristic patterns of distribution in national and regional research and innovation systems (differentiation strategies). Different national specialisation patterns reflect R&D intensity, position in global R&D networks, historical hysteresis etc. Ideally they can help identify priorities that can increase critical mass and create areas of comparative advantage but their interpretation must take into account other factors including political decisions (*e.g*. on nuclear energy) or public choices.

Box 16. Prioritising Strategic Research in Denmark – RESEARCH 2015 Initiative

Background: The Danish parliament (Folketing) decided that the foundation for setting priorities for strategic research – research within specially prioritized areas of society – should be improved. Strategic research can be both fundamental and application oriented but essentially is problem-oriented and interdisciplinary.

Focus/goal: Mandated by the parliament, the Danish Agency for Science, Technology and Innovation carried out the exercise and involved broad range of stakeholders.

Structure/approach: A mapping was carried out between March and October 2007 and involved an international literature scan by the OECD, broad consultations with societal stakeholders as well as input from government ministries on strategic themes. Benchmarking tools where used to assess the capacity for Danish research to pursue research priorities. .



Result: 21 Themes were identified as being directed at important societal challenges; where research-based knowledge is an important means; broad enough to ensure competition amongst research institutions and concrete enough to form the basis for coherent research programmes. The final catalogue of options will be used in political negotiations to set priorities for strategic research.

Implementation: The proposals will be directly implemented by the Danish Research Council for Strategic Research. The RESEARCH2015-catalogue proposals do not aim to set priorities for allocating core funding (i.e. general university funds) but should nevertheless serve to inspire the direction of research in universities. It is expected that the exercise will be repeated in four years.

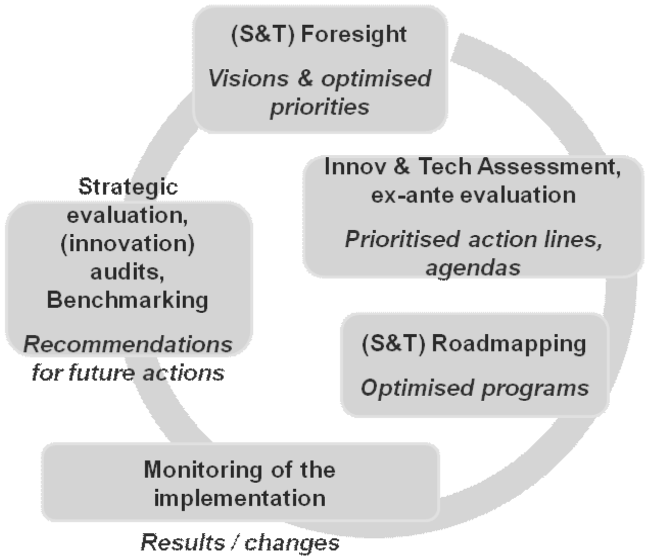
Lessons: It is important to balance short and long term perspectives. One must also consider political drivers versus challenge/opportunity drivers. In addition, there is a risk of reproducing existing priorities versus identifying new themes. There is also a need for improved documentation and data to assess current and potential strengths. Finally, the process requires clear political commitment.

Source: OECD based on Hoff (2008)

594. The combination of these various tools to help set priorities is known as "strategic policy intelligence" which can be defined as "the set of activities to search, process, diffuse & protect information in order to make it available to the right persons at the right time, so that they can take the right decisions". Related to Research & Innovation (RTDI) policies this includes such policy support instruments as foresight & technology assessment, monitoring, benchmarking, regional innovation auditing, and strategic evaluation (Acheson, 2008).

Box 17. Strategic Policy Intelligence in Ireland

Like many countries, Ireland has turned to technology foresight exercises to help set priorities. Prior to the mid-1980s, spending for science, technology and innovation was 'derived' from the wider policy objectives of each Department – there was no clear system discernible in relating STI expenditure to these policy objectives. The results of the first foresight exercise highlighted the significant investment needed for basic research in niche areas of ICT and biotechnology, two generic technologies underpinning strategically important sectors in the Irish economy. Since the first foresight study in the 1990s, the conceptual and empirical basis helping policy makers set priorities has evolved to include other tools such as technology assessment and technology road-mapping.



Source: Acheson (2008)

595. Another element of strategic policy intelligence is evaluation, both ex post and ex ante. Until recently, evaluations were not systematically used by policy makers in the process of priority setting. This can partly be explained by the fact that evaluations in the past were more successful in the assessment of the rationale, the implementation and the assessment of goal attainment and less so in assessing outcomes and impacts. Also, most evaluations have been of an ex-post nature, while ex-ante evaluations would lend themselves more easily for the purpose of priority setting.

596. Overall, the circle of people involved in the various forward-looking exercises has been considerably expanded from experts in the field of the respective technology to the broader public in more recent foresight exercises. Nevertheless, there has been a clear trend towards more "expertise-supported consultation mechanisms" for priority setting.

8. Data requirements for priority setting

597. Improving the empirical basis for priority setting, especially as regards budgetary choices, depends on highly quality and timely data, both quantitative and qualitative. The data requirements are multiple. On the input side, this implies understanding the amount of R&D funded through public budgets. While aggregate data are reported by countries, including to the OECD, there remain gaps in our understanding of public R&D by scientific and technological field and socio-economic objective. The multidisciplinary nature of research complicates this further: research in sub fields of mathematics can advance research and innovation in nanotechnologies or life sciences but may not be accounted for as such.

598. On the output side, evidence on the outcomes and impacts of public support to R&D in various fields are important to determine whether to invest more in research area "A" versus research in area "B". Over the years, countries have made attempts at improving the measure of scientific outputs such as scientific publications, graduates, and even patents in order to identify areas of national strengths and weakness. But using such benchmarking to decide on funding decisions has its inherent limits: it is extremely difficult to determine with any accuracy where the high impact and high return investments may lie (see also Chapter 4 of OECD (2008)). Improving data and analysis on both the input and output side will necessarily require more work developing up to date field definitions and taxonomies and as well as greater co-operation in the collection of data across different agents (funders and performers). The report of the OECD Blue Sky II conference discusses this issue in more detail (OECD, 2007).

9. International dimensions of priority setting

599. There are also important international dimensions to national priority setting. First, foreign priority setting has a direct and indirect impact on national priority setting exercises. The most direct impact is via the competition among scientists themselves in the creation and discovery of new knowledge. Strength in knowledge production in one country – specialisation – can influence the direction of specialisation in another country with different financial and intellectual resource endowments. Foreign funding for research via multinational firms or public research organisations may also have an indirect impact on the direction of research in the receiving country by signalling user demand in a given area. EU Structural Funds and Regional Funds have arguably shaped the direction of research priorities in the new member states.

600. There is however a risk that such external priorities substitute for internal and national priority setting processes with a resulting shift in resources towards areas of knowledge production that are of less concern to "national" stakeholders but that are nevertheless important in a globally linked research community. Regulations in one country that limit certain areas of research (e.g. stem cells) can also create incentives and opportunities for boosting research in that same area in another country with different regulations. While science has always been international, globalisation and the Internet have accelerated the movement and exchange of ideas and people. Priority setting exercises, like the evaluation of research itself, therefore increasingly involve taking into account global trends as well as direct input from foreign scientists and experts.

601. Another international dimension in the priority setting process concerns the setting of priorities for regional or global challenges or for large research infrastructures where international co-operation is necessary to shoulder the high development costs. In the context of the creation of the European Research Area, there has been an increasing focus on ways to better coordinate the national research programmes of the Member States. This ranges from the definition of agreed common priorities (common vision) as set out in the Lisbon Strategy and European Framework programmes to the implementation and the common monitoring or evaluation of national and common programmes. The challenge in setting priorities for international collaboration is balancing a global and co-ordinated approach with a differentiated approach in relation to the different types of research. Furthermore, international priority setting must take into account the different technological specialisation of partnering countries, the need to foster both competition and co-operation among research teams and between and between bottom-up research initiatives and top-down strategic guidance.

Box 18. Priority Setting through the European Commission Framework Programmes

The 7th Framework Programme for Research and Technological Development (7th FP) is the EU's main instrument for funding and steering of research policy. The 7th Framework Programme differs from the previous Framework Programme as it was explicitly designed to support the Lisbon strategy and as such is focused “on innovation and knowledge for growth" in the context of the European Research Area (ERA) - the internal market for knowledge in Europe. The FP7 is organized into four specific programmes, corresponding to four major objectives of European research policy, namely a) co-operation in research and discovery; b) ideas, which refers to the establishment of the European Research Council that will support individual research grants c) people, which refers to Marie Curie actions to improve human resources in S&T, and d) capacities, which refers to research infrastructure, regional and international co-operation.

6th EC Framework Programme (2002-2006)	7th EC Framework Programme (2007-2013)
Budget : EUR 17 billion	**Budget: EUR 53.2 billion**
1.Life science 2.Information society and technologies 3.Nano, Mat. Process 4. Aeronautics, Space 5.Food quality and safety 6.Sustanable Development, global changes, ecosystems 7.Citizens and society	1. Health (€6 bn) 2. Food, agriculture and biotechnology (€1.9 bn) 3. Information and Communication technologies (€9.1 bn) 4.Nanoproduction (€3.5 bn) 5. Energy (€2.3 bn) 6. Environnent (including climate change) (€1.8 bn) 7. Transport (including Aeronautics) (€4.1 bn) 8 Socio-economic sciences and humanities (€0.6 bn). 9. Security (NEW! compared to FP6) (€1.4 bn) 10. Space (€1.3 bn)

The thematic research programmes are subsumed in the “co-operation” pillar (EUR32.3 billion) which refers to gaining leadership in key scientific and technology areas by supporting co-operation between universities, industry, research centres and public authorities across the EU and with the rest of the world. Trans-national cooperation will remain the main instrument for carrying out research activities. This programme consists of ten different thematic research areas as listed above. Despite the greater number of priorities in FP7, there is a great deal of continuity between the thematic focus of the two programmes. However, the FP7 represents a 63% increase in public spending compared to FP6.

Source: European Commission (2008)

10. Evaluation and feed-back into policy design

602. Against the background of increasing support for research and innovation as well as increasing demands for accountability and socio-economic impacts, the question of how ex ante and ex post evaluations relate to the selection of priorities in research and innovation has become central to policy makers. The issues of priority setting and evaluation are two distinct issues with their own dimensions. Priority setting is the conscious selection of activities at the expense of others with an impact on resource allocation in contrast with the type of priority setting that takes place in a self-organising system. Priority setting is concerned with questions such as shall we invest more in basic research or innovation? What are the technologies that have greater private and social returns? Or, in the case of an institution, shall we invest in earth observation platform or a particle collider? The priority setting process has several dimensions which vary with emphasis over time. Historically, thematic priorities like technological

priorities have dominated. Then came along mission-oriented priorities to respond to societal demands; more recently, other types of priorities have come to attention of policy makers such as "functional" priorities or those that affect the functioning of the system (e.g. shall we focus on policy strands instead of technologies?).

10.1 Decentralisation of priority setting

603. Today there are more actors involved in priority setting than in previous decades with different approaches and different methodologies in their respective fields. Priority setting is increasingly decentralised following the devolution of competences and the creation of funding agencies. The process of priority setting also differs. Sometimes it is top down with expert-based procedures or bottom up input. Sometimes the process in enshrined in formal law or mechanisms, sometimes it is more informal. This raises the question of coherence between priority setting and policy making. A key question is how policy structures can be designed to ensure coherence in the process of priority setting? Policy councils and research funding agencies as well as large public research organisations all support priority setting at the national and sub-national levels and these councils have to interact. Another development is that the balance of power in priority setting seems to be shifting. Funding agencies have more control in priority setting but this may give rise to conflicts with political setting at ministerial level. In quite a number of countries these funding agencies carry out foresight studies that also have a bearing on priority setting.

604. Another issue is that the conceptual underpinning of priority setting remains rather weak. Most lists of priorities or critical technologies remain at levels which provide only a poor guide for policy. In the 1960s priority setting was seen as a planning exercise. This shifted in the 1990s in the face of a complex environment characterized by greater use of technology assessment and foresight exercises as well as of impact assessments and greater stakeholder involvement. Today priority setting is an expert supported consultative mechanism that relies on some of the same tools used in the past. The one constant is that priority setting remains a process of political bargaining.

605. To what extent can evaluations improve priority setting? Evaluation is one of the potential instruments for priority setting and can help improve the quality of priority setting. Evaluations can be *ex ante* and look at the potential impacts up-front or *via* in interim stage (during the lifespan) or *ex post;* looking at the impact and goal attainment. This would possibly have a feed back in some way or another.

606. Despite this potential, evaluation has not been used for priority setting. The main challenge is to be able to compare between alternatives. Most evaluations have focused on a single policy measure but not in relation to other measures. Can we assess future impacts of technological developments and the policy interventions in order to allow for priority setting? What is the scope of evaluations today? They address appropriateness, quality, and efficiency of implementation and assessment of additionality. There has been some progress especially for assessing quality and efficiency of intervention measures. In terms of appropriateness there has been some progress using *ex ante* assessments (e.g. programme logic, rationale). Also there has been progress on impact and *behavioural additionality* but less with regard to *output additionality*.

607. Of course if one wants to prioritise, one would need to know the potential impacts of different measures (e.g. the net present value of alternative investments). There have been both meso and micro level estimates of the impacts on different measures, say on productivity. These studies look at the relation between R&D and impact at various levels. Progress has been slower on the meso level of the programme and technologies due to the effects of time lag effects. In short, current methodologies for measuring the social rate of return of individual projects or subsidies to specific technologies are not satisfactory. For example, studies of the ATP programme in the US find that on average the return is positive but the range is so large that such conclusions cannot be used to select specific investments. There are limits to *ex ante*

evaluation in priority setting. Although this is not to say that evaluation cannot be useful and can inform policy more than it has in the past. For example, *ex ante* evaluation can give a good idea about the rationale of programmes. As we have only a few examples of looking at policy measures in context (e.g. compared to other measures) such evaluation is more useful but mainly for legitimising policy interventions, not for priority setting.

10.2 Gap between evaluation methodology and practice

608. Another problem is that most evaluations do not use quantitative tools which make comparisons difficult. Policy makers should have realistic expectations as to what evaluations can and cannot do. First, the information requirements of evaluations with respect to behavioural additionality far exceed what is available and evaluations can be more costly than the measure being examined. As a way forward, policy makers should:

- Push the envelope on evaluation. For example some researchers have tried to use instruments that are wide-spread in economics (e.g. option value approaches or microeconomic modelling using CIS data). The most promising avenue would be carrying out evaluation in context, in a systemic perspective. One example is an evaluation of subsidies in relation to R&D tax credits to see if these instruments are complements or substitutes.
- Another option is to embark on an evaluation of the priority setting process itself. It would be worthwhile expanding qualitative analysis to see which of the priority setting process can be good practice for other countries. The benchmark would be a priority setting process that actually has an impact on the direction of technology and on specialisation.

609. As highlighted previously, Ireland also has many actors involved in priority setting. There are some nine ministries dealing with S&T with different budgets and these ministries also have their own agencies.

610. Although Ireland used a top-down and bottom-up technology foresight exercise in the process that led to establishment of Science Foundation Ireland, some 90 % of the technology foresight was not implemented. The foresight exercise served mainly as a technology intelligence tool. Another example was the expert review of Science Foundation Ireland in 2005. The Foundation is seen as a key pillar of Ireland's national innovation system and funding has gone to EUR 1.4 billion in the latest seven year national plan. Policy makers are, however, asking questions on the return on investment, especially in the case of funding for basic research.

611. Previous evaluations of business R&D support measures were not satisfactory because the instruments were often changed by the agencies and because of time delays. In reality, the strategy for science and innovation (*e.g.* shared vision, roadmaps and capital investment mechanisms) was mainly input focused but it did not address clear targets. In response to this, Forfas is attempting to establish a framework for understanding and evaluating policy goals and performance (*e.g.* inputs, outcomes, outputs, etc).

612. Forfas has also put together a schedule for the Ministry and the agencies for a wider consultation. Technology Ireland has agreed to the schedule and its implementation in the agencies. A key lesson is that "buy-in" is necessary to have an evaluation implemented to make sure people are collecting the right indicators so that it can inform policy in the future. There is also a need to educate actors on what is an evaluation. And then there is the territorial issue: 9 out of 15 Ministries and 40 to 50 agencies are guarding the activities that they have in the S&T area. On the positive side, there has been at least more verbal support for evaluation at the level of the administration, led by the department of finance which is more

focused on the "bang for the buck". In concluding, Jacqueline Allan noted that a new system of evaluation and indicators can influence the system but that it requires "buy-in" at various levels.

613. One of the challenges in thinking about evaluation in a "systemic" perspective is that it requires better definitions of innovation systems. But it also requires the establishment of relations between different policy mixes and the policy outcomes, which for the moment is out of the reach of many researchers and policy makers. However, with more modest targets, it may be possible to make modest steps towards evaluation that considers its context.

614. One issue of contention is that priority setting is often seen from an economic perspective and assumes the goal is to generate greater social returns. But often, priorities can be set in a different way or with a different goal such as to strengthen existing capabilities. Some countries like Sweden and Finland have focused priority research on topics that reflect resource-based or industrial comparative advantage (e.g. wood and fisheries). One question that remains outstanding is whether priorities should reflect socio-economic needs such as ageing or global warming, or whether countries should first explicitly formulate their priorities?

615. Another issue is the relation between changing priorities and spending. Priority in one field often has consequences for spending in other fields. For example, an increased focus on nanotechnology may result in greater attention or priority in a given area of mathematics. In practice, priority setting is set according to what is important. How a certain field becomes to be considered as important can results from relevance, political choices or a perceived gap etc. Evaluations can provide some lessons in this context because it is possible to analyse a country's NIS and determine gaps, such a poor industry-science relations or low rates of new firm creation. Policy analysis can contribute, but not in helping make portfolio investment decisions. It will not be possible with any form of *ex ante* evaluation to reach an investment decision of the type reached by investment bankers. The problem is that this remains a prominent wish among policy makers. Science councils in the UK are increasingly asking about the economic return to basic research. Caution is important here; how to approach priority setting in this context? Accordingly some experts suggest that priority setting should be set in a different light, not just as tool to determine the economic returns on investment.

10.3 Evaluation and feed-back into policy design

616. The interrelation between evaluation and the implementation of measures is perhaps the most difficult challenge that faces policy makers. In many countries, and despite the emergence of an evaluation culture, expert opinion continues to predominate in the types of evaluation used by policy makers. In some countries there is a tension between policy makers who propose policy alternatives and central agencies asking about the rationale for policy intervention. Policy development has its own well developed process. Policy objectives are not clear and are often all encompassing. For example, high level indicators like HERD as a share of GDP do show a direct correlation with policy interventions but others do not. Often, quantitative econometric models are not used and what remains is expert opinion. In fact, S&T policies have not been developed like other policies where ex ante policy evaluation is the norm. Yet S&T also requires a systems approach and governments will need a suite of policy initiatives to influence innovation.

617. Ex ante evaluation of policy options are forward looking but the majority of tools used are mainly qualitative measures like stakeholder consultations, focus groups, to determine which programmes have provided the greatest benefit to stakeholders. Expert opinion and review are other common tools. Foresight is useful for developing policy alternatives. Similarly, an analysis of past experience in terms of successful or unsuccessful policy is also used.

10.4 Use of quantitative evaluations

618. As regards quantitative models, "logic models" are increasing used for evaluations by academics and experts, but these have not yet been transferred to the policy groups. Other quantitative methodologies include cost-benefit, totally quantitative or qualitative tools or a hybrid of both. Often these are attempted, but not completely developed due to time constraints and the inability of these tools to project what the benefits will be in the long term. By the time analysts determine the impacts, policy makers have moved on to a new policy measure. The cost analysis of implementation is also a consideration in implementing a policy decision that may require a new administrative body such as an agency. In Canada two new agencies have been created to address specific policy concerns (*e.g.* the Canadian Innovation Foundation and Genomics Canada). As regards spill-over effects, the methodology is not well established but spill-over effects are significant. Some attempts have been done to map this in the private sector where spill-overs have a greater impact on business decisions.

619. In practice, and despite the emergence of new tools, policy analysis lacks robust measures. While qualitative analysis is important and needs to be undertaken, there is a need to have more robust measures in order to be able to evaluate a "portfolio" of measures. Some work done at Simon Fraser on Policy Options Analysis and the Meta-choice framework look at mutually exclusive policy alternatives. When selecting the goals against which to measure the policy alternatives one needs to use different quantitative and qualitative tools. One also needs to value the predicted impact against a set of goals. For multi-goal analysis it is important to determine the policy criteria against which a programme or policy will be measured.

620. Another tool that emerged is the "Weighted Criteria Framework" which consists of defining a set of criteria to serve as measurement tool. For this, it is necessary to define objectives in order to measure policy outputs. There is also a need to articulate how the success or failure to define the criteria will affect the measures. How can one determine that the right criteria are selected? Whether or not one set of criteria or another is used usually depends on policy goals. In this model if a weighted criteria is to be used then the weights must be determined by the extent to which they will influence a policy outcome. Again this process requires the combination of different matrices to determine the policy options. Another lesson is that evaluation must be aligned with policy goals. Some elements will have more risks (*e.g.* cost, or public perception or ability to implement) than others. Increasing this makes it easier to define policy options vis-à-vis central agencies.

621. What is the value of *ex ante* besides determining differences in policy options? First, it formalises the policy rationale and provides a line of evidence for ex post evaluations of the programme. When ex ante and ex post evaluations are combined, this enables policy evaluation at the macro level which can be combined with programme evaluations. What is often missing is the rationale for choosing that particular programme in relation to a policy goal.

Box 19. Feed-back of evaluation in setting national priorities in Canada

The development of Canada's 2007 Science and Technology Strategy, which drew on expert opinions and foresight exercises, provides an interesting case study to illustrate some of the challenges in linking evaluations to policy making. The Council for S&T identified key areas in which Canada had strength and this was reflected in the priorities of the final strategy. A number of surveys were carried out to determine the efficiency of current programmes and of programme delivery mechanisms in response to fragmentation within the access to research dollars (*e.g.* the case of applicants having to apply for research funding at one Council and then to the Canadian Innovation Foundation for infrastructure funding). In addition, an analysis of best practices in programme elements was carried out. Macro level indicators such as OECD S&T indicators were used to compare Canada to other OECD countries. There was also extensive use of key stakeholder roundtables, involving industry or the scientific community as well as calls for written submissions. Programme evaluation reports did influence some of the programme elements attached to the 2007 strategy. However there was no direct intervention from evaluation to policy making. Attempts to evaluate previous strategies (*i.e.* the 1996 and 2002 strategies) failed due to lack of "buy-in" to collect data, although there is some hope that it can be done for the 2007 strategy. In conjunction with the focus on evaluation, there will be a greater focus on ex post programme evaluation but there will be a feed-in to improve policy decisions. *Ex ante* evaluation remains weak however. New rules to require cost- benefit analysis of all decisions will be enforced on a stricter basis in light of performance measure frameworks of regulatory decisions.

In conclusion, although there has been progress on the development of conceptual and theoretical frameworks for providing analytical input to policy actions, in practice the take up has been slow. The interest of the international community and OECD in particular, is essential. If this bears out, it is hoped that the next national S&T strategy will have a greater focus on ex ante evaluations of policy options.

Source: Newell (2007)

622. The issue at stake is really dual – evaluation and the engineering of policy making. The coherence between policy tools, values and objectives and the involvement of actors are important. If we look at ex ante evaluation of policy tools, of the relevance and efficiency of policy instruments, we see that the logic is based on resource intensiveness. There is also a question on the coherence of the policy mix of instruments used to address the same target group. There are intrinsic weaknesses to national strategies given the complexity of the NIS, as well as new modes of knowledge production and new modes of governance. So the national strategy is trying to control the system and understand how change came about. But it is difficult to control. In using ex ante evaluations, it is important to understand which policy goals one seeks to achieve.

623. There is also a need for balance between new public management tools and national innovation systems. The main challenge for ex ante evaluation is to be developed upfront with the policy. What we want from ex ante evaluation are indicators and data that can reduce information asymmetries and legitimize policy interventions.

624. The example of the Italian National Research Plan for 2005 and 2007 is telling. It was developed to set up strategies and priorities by a panel of experts nominated by the ministry of research. They stated clear objectives mainly related to the EU Lisbon objectives which were based on a combination of different policy measures. The approach combined inputs from research ministries, stakeholders etc. mainly through consultations and reporting. This exercise was also characterised by use of international data and indicators. But due to a low effort in the implementation phase, ex post evaluation was not developed as part of the policy process. In conclusion, there is a need to integrate evaluation into the process of policy making. For now, traditional peer review remains dominant in Italy and there is little use of *ex post* evaluations in practice.

625. The use of evaluation in policy making does not follow formal procedures. Approaches are used depending on the nature of policy. In South Africa, the evaluation tools most used are used are expert panels and groups. In this context, the setting of the terms of reference (ToR) for the group is important.

Credibility is equally important and this requires the involvement of local groups as well as international experts to ensure credibility. In South Africa the democratisation process helped develop a culture of consultation. Representativeness and inclusiveness are important. There are a number of ways to do this, such as focused discussion with stakeholder groups or consultative conferences in order to receive input from stakeholders and this includes written input. Foresight exercises and technology audits are also used and help formulate national S&T strategies. One problem however is that such a process involves large parts of society but thus far the feed-back to participants has been weak. Another form of evaluation input comes from advisory bodies and boards of agencies, such as the National Research Foundation which provides grants. These boards are very involved in evaluation. Ultimately, stakeholder input is valuable for getting political decisions that stick.

626. Good practices in the ex-ante evaluation of policy measures are those which (a) identify the policy context in which the new policy is to be created, (b) analyse the new balance of the policy mix which would emerge for each policy alternative and (c) make explicit the assumptions under which the new balance of the policy mix would be justified.

627. With regard to the issue of how does priority setting operate at institutional level, it is important to note that PROs operate in so many different contexts that it is difficult to express general points about priority setting. In principle it is performed within the frame of the strategy of the institution, which provides the guidelines for operational decisions. Each PRO has its own internal process, which also depends on the personality of the decision-makers involved.

628. The issue of the positioning of the PROs within the system is a central and difficult one. The question at stake is the role performed by PROs as compared to Contract Research Organisations and as compared to Universities; the situation is even more complicated in France, since PROs have also, to a certain extent, the function of a funding agency. Therefore, having a proper understanding of the function each PRO plays within the system is very important; otherwise there can be misinterpretations of the relevant relations with the other parts of the system, resulting in inadequate policies.

629. Priority setting is part and parcel of the policy making process but it is not synonymous. Each funding source in the UK has priority setting capabilities. A key trend is the emphasis on stakeholder input and expert advice. Bibliometric tools are also used to monitor evaluation outcomes. An effort to monitor socio-economic effects before research is carried out is equally important. The challenge for the UK is to support decision making in strategic areas. Evaluations can play a role by proving input on the context and procedures for S&T policy making.

630. In summary, the key findings of TIP work on Evaluation and Priority Setting, can be summarised below:

- Although priority setting and evaluation interact in policy making, they remain two distinct dimensions of policy making with their own characteristics and internal processes.
- In line with the increase in actors in S&T policy making (e.g. regional governments, separate funding agencies, etc.) priority setting has become more complex and involves more actors using different approaches and methodologies. This increases the need for mechanisms to ensure greater coherence between priority setting and policy making.
- Despite the emergence of new quantitative tools for evaluation, the conceptual underpinning of priority setting remains rather weak and expert opinion continues to predominate in the types of evaluation used by policy makers to make policy decisions.

- Improving the process of priority setting through the use of ex ante evaluations will require political buy-in from the different stakeholders and commitment to invest in resources and skills including the creation of indicators to monitor policy effectiveness.
- The process of priority setting in S&T could itself be the subject of evaluation in order to identify structural weaknesses as well as best practices.
- The interest of the international community and of the OECD in particular, is essential to improve the ability of countries to develop and use ex ante evaluation in priority setting and policy making in general.

REFERENCES

Acheson, H. (2008), presentation at the Joint TIP/RIHR Workshop on Enhancing Research Performance through Evaluation and Priority Setting

Arnold, E. (2004), Innovation Governance: Typologies and Principles, European Trend Chart Policy Workshop, European Commission. Brussels

Barré, R. (2008) presentation at the Joint TIP/RIHR Workshop on Enhancing Research Performance through Evaluation and Priority Setting

Bush V. (1945) Science, the Endless Frontier: A Report to the President. Washington, D. C.: Public Affairs Press. www.nsf.gov/od/lpa/nsf50/vbush1945.htm.

Cheney, D. W. The Emergence of S&T Priorities in the United States, SRI International United States. Presentation at the Second International Conference on Technology Foresight – Tokyo, 27-28 Feb. 2003

European Commission (2007), *Europe in the Global Research Landscape*

Hoff, A. (2008), Presentation at the Joint TIP/RIHR Workshop on Enhancing Research Performance through Evaluation and Priority Setting

IEA, Energy Technology Perspectives (2008), Paris.

National Institutes of Health (1997), Setting Research Priorities at the National Institutes of Health, (www.nih.gov/news/ResPriority/priority.htm).

National Science Board (2001), Federal Research Resources: A Process for Setting Priorities. Washington, DC: U.S. Government Printing Office, 2001. (NSB 01-156).

National Science Board (2001), The Scientific Allocation of Scientific Resources

OECD (1991), *Choosing Priorities in Science and Technology*, Paris.

OECD (2003*), Governance of Public Research*, Paris.

OECD (2005) *Governance of Innovation Systems*. Volume 1: Synthesis Report, Paris.

OECD (2007), *Science, Technology and Innovation Indicators in a Changing World: Responding to Policy Needs*, Paris.

OECD (2008), *Science, Technology and Industry Outlook 2008*, forthcoming, Paris.

Polt W. (2007), Issues in Evaluation and Priority Setting, Room Document for the TIP/RIHR Workshop on Evaluation and Priority Setting.

Teich, A.H. (1994) "Priority-Setting and Economic Payoffs in Basic Research: An American Perspective", Journal of Higher Education, Volume 28, Number 1, pages 95-107, Springer.