

Proceedings from the 2000 U.S.-European Workshop on

# Learning from Science and Technology Policy Evaluation

Bad Herrenalb, Germany

Edited by Philip Shapira and Stefan Kuhlmann

 **Georgia Institute**  
of **Tech**nology  
School of Public Policy

 **Fraunhofer**  
Institute  
Systems and  
Innovation Research

# **Learning from Science and Technology Policy Evaluation**

Philip Shapira and Stefan Kuhlmann (Editors)

**Proceedings from the 2000 US-EU Workshop on Learning from Science and Technology  
Policy Evaluation, Bad Herrenalb, Germany.**

School of Public Policy, Georgia Institute of Technology, Atlanta, USA  
and the  
Fraunhofer Institute for Systems and Innovations Research, Karlsruhe, Germany

2001

An electronic version of this volume of proceedings is available through the Worldwide Web at:  
**<http://www.cherry.gatech.edu/e-value> or <http://www.isi.fhg.de/ti/departm.htm>**

© 2001 Georgia Tech Research Corporation and the Fraunhofer Institute for Systems and Innovation Research.

## Contents

---

	<i>Page</i>
Table of Contents	iii
Preface	vi
Acknowledgements	viii
<b>1 Introduction</b>	<b>1-1</b>
Learning from Science and Technology Policy Evaluation	1-1
U.S.-European Comparisons and Opportunities <i>Philip Shapira, Georgia Institute of Technology, USA and Stefan Kuhlmann, Fraunhofer Institute for Systems and Innovation Research, Germany</i>	
<b>2 Emerging Paradigms for Evaluating Research, Innovation, and Technology Policies in the United States and Europe</b>	<b>2-1</b>
The Academic Policy Analyst as Reporter: The Who, What,...and How of Evaluating Technology Programs <i>Irwin Feller, Pennsylvania State University, USA</i>	2-1
Societal Challenges for R&D Evaluation <i>Arie Rip, Twente University, The Netherlands</i>	2-18
Evaluation of Research and Innovation Policy in Europe – New Policies, New Frameworks? <i>Luke Georghiou, PREST, University of Manchester, UK</i>	2-42
Frameworks for Evaluating Science and Technology Policies in the United States <i>Susan Cozzens, Georgia Institute of Technology, USA</i>	2-54
Discussion	2-57
<b>3 Evaluation Of Public Research, Innovation and Technology Policies in The European Union: Roundtable on Trends and Issues</b>	<b>3-1</b>
Socioeconomic Evaluation of the Fifth Framework Programme <i>Liam O’Sullivan, European Commission</i>	3-1
Evaluation of National Policies – Recent German Experiences <i>Christian Uhlhorn, Federal Ministry for Education and Research, Germany</i>	3-2
Understanding the Cultural Logic of EU Science and Technology Programmes <i>Joseph Cullen, Tavistock Institute, UK</i>	3-5
Discussion on Roundtable Presentations	3-7
Commentary <i>Nicholas Vonortas, George Washington University, USA</i>	3-10

	<i>Page</i>
4	<b>Public Research, Innovation and Technology Policies in the US: Federal Policy Developments and Evaluative Approaches</b> 4-1
	Developments in Federal Science and Technology Policies in the United States 4-1
	<i>Christopher Y. Hill, George Mason University, USA</i>
	Assessment of Federal Lab-Industry Interactions 4-12
	<i>James Dietz, Georgia Institute of Technology, USA</i>
	The Expanding Role of Peer Review Processes in the United States 4-31
	<i>David Guston, Rutgers University, USA</i>
	Commentary on the Session and Additional Discussion 4-47
	<i>Arie Rip, Twente University, The Netherlands</i>
5	<b>Evaluation of Large-Scale Programs in Europe</b> 5-1
	Evaluation of the BRITE/EURAM Programme 5-1
	<i>Laurent Bach, Marc-Jacques Ledoux and Mireille Matt, Université L. Pasteur, France</i>
	Evaluation of the EU's 4 <sup>th</sup> Framework Programme in Finland 5-25
	<i>Terttu Luukkonen, VTT Technology Group, Finland</i>
	Evaluation of Nanotechnology Competence Centers 5-40
	<i>Susanne Bühner, Fraunhofer Institute for Systems and Innovations Research, Germany</i>
	Commentary and Additional Discussion 5-43
	<i>Louis Tornatsky, Southern Technology Council, USA</i>
6	<b>Evaluation of Regionally-Based Science and Technology Programs</b> 6-1
	Evaluation of Industrial Extension and Technology Deployment Programs in the United States 6-1
	<i>Philip Shapira, Georgia Institute of Technology, USA</i>
	Evaluation of Regional RIT Strategies in Europe 6-32
	<i>Patries Boekholt, Technopolis, The Netherlands</i>
	Science and Technology Infrastructure Investments for Rebuilding Distressed Communities: A Case Study in the United States 6-43
	<i>Peter Blair, Sigma Xi, The Scientific Research Society, USA</i>

	<i>Page</i>
7	<b>Evaluation in Multi-Goal Partnerships</b> 7-1
	Recognizing and Assessing Competing Values in S&T Organizations and Implications for Evaluation 7-1
	<i>Gretchen Jordan, Sandia National Laboratory, USA and L.D. Streit, McNeil Technologies, USA</i>
	System Evaluation of Publicly Supported Industrial Collaborative Research 7-25
	<i>Heike Belitz, German Institute for Economic Research, Germany and Hans-Peter Lorenzen, Federal Ministry of Economics and Technology, Berlin, Germany</i>
	Discussion 7-36
8	<b>Policy, Institutional, and Portfolio Evaluations</b> 8-1
	Evaluating Scientific and Technical Human Capital: An Event History Approach 8-1
	<i>Barry Bozeman, Georgia Institute of Technology, USA and Monica Gaughan, Oglethorpe University, USA</i>
	Assessment of the Evaluation of Leibniz-Institutes - External Evaluation and Self-Evaluation 8-16
	<i>Martina Röbbecke and Dagmar Simon (WZB, Germany)</i>
	University-Industry Benchmarking 8-24
	<i>Louis G. Tornatsky, Southern Technology Council, USA</i>
	Evaluation of University Research in UK, Netherlands, Austria and Germany 8-43
	<i>David Campbell, Institute for Advanced Studies, Austria</i>
	Discussion 8-59
9	<b>Appropriate Methodological Matrices for S/T Evaluation</b> 9-1
	Real Options Approach for Evaluating Public Sector R&D Investments 9-1
	<i>Nicholas Vonortas and Matt Lakey, George Washington University, USA</i>
	Assessing Programme Portfolios via Multi-Module Approaches 9-14
	<i>Ken Guy, WiseGuys, UK</i>
	Making Evaluation Useful for Decision Making 9-24
	<i>John Barber, Department of Trade and Industry, UK</i>
10	<b>Lessons from Existing Networks– Roundtable</b> 10-1

	<i>Page</i>
11	<b>Evaluation as a Source of “Strategic Intelligence”</b> 11-1
	The Triple Helix of University-Industry-Government: Dynamics of Innovation Spaces and Implications for Policy and Evaluation 11-1
	<i>Henry Etzkowitz, State University of New York, USA</i>
	Evaluation as a Source of “Strategic Intelligence” 11-21
	<i>Stefan Kuhlmann, Fraunhofer Institute for Systems and Innovations Research, Germany</i>
	Commentary and Discussion 11-60
	<i>Irwin Feller, Pennsylvania State University, USA</i>
12	<b>New Opportunities for Transatlantic Collaboration – Roundtable</b> 12-1
	Discussion and Workshop Conclusions 12-1
	<b>Appendices</b>
	Workshop Agenda A-1
	Workshop Participants A-2

## **Preface**

---

In September 2000, the School of Public Policy at Georgia Institute of Technology (Georgia Tech) and the Fraunhofer Institute for Systems and Innovation Research (ISI) co-sponsored a US-European Workshop on Learning from Science and Technology Evaluation. The workshop was held at the Evangelische Akademie in Bad Herrenalb, Germany. Some 30 American and European participants were involved in the workshop, including senior and junior scholars and representatives from science and technology organizations interested in evaluation. In addition to discussing state-of-the-art evaluation research, the group considered possibilities for future US-European science and technology policy evaluation linkages.

The Bad Herrenalb workshop provided a unique opportunity for transatlantic dialogue and the benchmarking of evaluative approaches in the field of science and technology policy. We are pleased, through this volume of proceedings, to make available a complete set of the workshop papers and discussion summaries to the broader community interested in this field. Much was learned at the workshop through comparisons of concepts, perspectives, methods and results. But it was also realized that much needs to be done in the US and Europe (as well as elsewhere) to improve the context, practice, and utility of evaluation in the science and technology policy sphere. Workshop participants agreed that further transatlantic collaboration could be most helpful in making progress on these challenges, through such means as ongoing exchange, benchmarking, networking, and joint methodological and project activities. In further developing such collaborations, we invite dialogue and engagement with others to build on the foundations of open exchange, constructive criticism and shared purpose established at Bad Herrenalb.

*Philip Shapira*  
*School of Public Policy*  
*Georgia Institute of Technology*  
*Atlanta, USA*  
*Email: ps25@prism.gatech.edu*

*Stefan Kuhlmann*  
*Fraunhofer Institute for Systems and*  
*Innovation Research*  
*Karlsruhe, Germany*  
*Email: sk@isi.fhg.de*

## **Acknowledgements**

---

American participation in the US-European Workshop on Learning from Science and Technology Evaluation was made possible by support from the United States National Science Foundation (Award 9906139). The VolkswagenStiftung (Volkswagen Foundation) sponsored European participation in the workshop. Supplementary support from the Georgia Tech Foundation with the support of Georgia Tech's Ivan Allen College allowed additional young Georgia Tech scholars to participate in the workshop. We acknowledge our deep thanks to these sponsors.

The administration of the workshop was most ably supported by Edeltraud Geibel, Meike Urrista-Carillo and Markus Peter at the Fraunhofer Institute for Systems and Innovation Research and by Chris Fehrenbach and LaTissia Caldwell at Georgia Tech. James Dietz, Jakob Edler, Sybille Hintze and Dara O'Neil and served as note takers, allowing an accurate record to be made of the discussions. Again, we are most grateful for this assistance.

Finally, we sincerely thank the Bad Herrenalb workshop participants, not only for their papers but also for their time, energy and enthusiasm in contributing to the event's success.

# 1. Overview of Workshop and Aims

---

## Learning from Science and Technology Policy Evaluation U.S.-European Comparisons and Opportunities

### **Philip Shapira**

School of Public Policy, Georgia Institute of Technology  
Atlanta, Georgia, USA  
Email: ps25@prism.gatech.edu

### **Stefan Kuhlmann**

Technology Analysis and Innovation Strategies  
Fraunhofer Institute for Systems and Innovation Research  
Karlsruhe, Germany  
Email: sk@isi.fhg.de

---

## **Introduction**

In both the United States and European, increased attention has been focused in recent years to policies for research, innovation, and technology (RIT). With heightened global economic competition, policymakers have sought to focus RIT policies in ways that will enhance the performance of national and regional economies. In addition, new patterns of industry collaboration and commercialization, developments in information exchange and knowledge transfer, and the pace of progress in science and technology itself have stimulated the sponsors and performers of research to review their priorities, approaches, and institutional structures. Budgetary pressures and changes in public management approaches, including a renewed emphasis on “performance,” have also strengthened policy-driven demands for accountability and better value from RIT investments.

The increased attention focused on RIT by policymakers, in turn, has spurred greater interest in the field of research, innovation, and technology policy evaluation. It is frequently recognized that many aspects of RIT policy are intrinsically hard to evaluate. Nonetheless, evaluators are continually pushed to provide concrete assessments of the effectiveness of RIT policies and particular programs and to offer timely feedback and guidance that can be used in policy decision-making as well as program oversight and management. In the United States and Europe, a growing number of researchers and practitioners have specialized in the RIT evaluation field. Despite this growth, the RIT evaluation field is still at an early stage of development, with many debates about the appropriateness and value of particular evaluative methods and about how to connect evaluation with improved

## *1 Overview of Workshop and Aims*

policy making. Moreover, the individuals and organizations involved in RIT evaluations tend to be dispersed and fragmented. In particular, there are relatively few *organized* opportunities for intensive comparative discussion about issues and opportunities in the field between American and European RIT evaluators and policymakers.

The authors of this paper propose a new collaborative initiative to bring together American and European researchers, policymakers, and other stakeholders involved in the evaluation of research, innovation, and technology policies. We believe this initiative is timely given parallel developments in the expansion of commercially focused research, innovation and technology policies in the US and across Europe, including ongoing reviews of US frameworks for RIT and the forthcoming 5th EU framework program. The first step in this initiative is an initial workshop to examine what can be learned from recent science and technology policy evaluation approaches in the US and Europe and to consider implications for future evaluation strategies and framework policies.

At the workshop, we aim (through papers around key themes) to benchmark the current state of the art; but we would also aim to discuss potential ideas, opportunities, projects, and needs, for future work. Thus, it is hoped that the workshop will lead to the subsequent development of a multi-year US-European network that would aim to share best practices, review methods, examine policy implications of evaluation, and stimulate new collaborative research projects. The particular form that this transatlantic network might take will be one of the issues discussed at the workshop.

### **Background**

A number of analysts have argued that policies for research, innovation, and technology in advanced economies are shifting into a new phase of development. For example, Caracostas and Muldar (1998) suggest that since the end of World War II, the research and innovation policy framework has gone through two major stages: first, from an emphasis on basic science and defense needs to, second, a more recent focus on key technologies and industrial objectives. Now, a transition to a third phase is underway – with a thrust on innovation and societal goals. Others have advanced similar ideas, with general agreement that traditional models of science and technology policy have been superseded. In their place, new paradigms have emerged that typically involve more complex and iterative perspectives on the RIT process, emphasize innovation, competitiveness, and societal development, and stress performance in meeting broadened policy goals (Kodama 1991, Freeman 1991, Tassej 1992, Crow 1994, Galli and Teubal 1997).

A series of factors have driven these changes in intellectual and policy perspectives. International economic competition has intensified, particularly in technology-

## *1 Overview of Workshop and Aims*

intensive sectors such as electronics, computing, and new generation vehicles. Simultaneously, new geographical clusters of technological capability have developed and more intense transnational economic and technological linkages have been forged. It is no longer possible for individual countries to dominate major industrial and technological sectors (in the way the United States did in the 1950s and 1960s). Nor is it realistic to expect that strength in science alone will “automatically” result in strong industrial performance. One consequence of this has been a shift in science and technology policy in recent years in the US and Europe, with relatively less pure science, more attention to applied research and, and a greater emphasis on diffusion (Branscomb and Keller 1998, Caracostas and Muldar 1998). New patterns of industry collaboration and commercialization have emerged, with the growth of industry consortia, university-industry linkages, public-private partnerships, and multi-national research programs such as those sponsored by the European Union or the Intelligent Manufacturing Systems project. Rapid developments in information exchange and knowledge transfer, and the pace of progress in science and technology itself have stimulated the sponsors and performers of research to review their priorities, approaches, and institutional structures. Included in this has been a rethinking of governmental roles, not only in terms of partnerships to speed diffusion, but also in establishing improved framework policies for research and innovation and in prioritizing research areas and targeted technologies.

Examples of the new policy approaches at the federal level in the US include the development of the Advanced Technology Program, the Manufacturing Extension Partnership, and the United States Innovation Partnership (Shapira, Kingsley and Youtie 1997). Comparable programs have been developed in Europe and in other advanced industrial economies (Chang 1998). U.S. states have also increased their investments in technology development programs in recent years, including university/non-profit centers, joint industry-university research partnerships, direct financing grants, incubators, and other programs using science and technology for economic development (Coburn and Berglund, 1995; Berglund, 1998). In Europe, member states of the European Union have increased generally budget shares allocated to research policy over the last decade, while the Europe Union has itself mounted several waves of applied research and technology initiatives within the context of successive framework programs (Caracostas and Muldar 1998), meanwhile covering a small but increasing share of national RIT activities (Reger and Kuhlmann 1995, Larédo 1995, Georghiou et al. 1993).

These developments in policy paradigms and governmental roles have been accompanied by changes in the context for evaluation. In Europe, since the 1980s, various national governments forced the evaluation of effects and socio-economic impacts of RIT programmes (Georghiou 1995, Meyer-Krahmer and Montigny 1989). German Federal Ministries (for research and technology, for economic affairs) supported independent impact evaluation studies of many major RIT programmes – seeds for an emerging "evaluation culture" (Becher and Kuhlmann 1995, Kuhlmann 1995). New efforts to reform the public sector and to control its

## *1 Overview of Workshop and Aims*

1995). New efforts to reform the public sector and to control its cost have led not only to the promotion of partnerships to leverage resources, but also to a greatly increased concern with efficiency and performance measurement (Osborne and Gaebler 1993, Shapira, Kingsley and Youtie 1997). The federal government and a majority of US states are now pursuing performance-based management and budgeting systems – and such systems are increasingly being applied to RIT policies and programs. For example, one recent survey of US state science and technology programs found that most of states have some type of method for collecting performance data or conducting a program evaluation. Yet, despite greater “data” collection, few states have well-conceived science and technology policy evaluation plans. Activity reporting, client survey data, and informal client contact are the most commonly used evaluation methods; more systematic evaluation approaches are less common (Cozzens and Melkers 1997). Only in part is this due to lack of funding or interest; there are also complex issues about how best to apply evaluation methodologies to assess the often diffuse and indirect effects of technology promotion policies.

The desire for enhanced measurement of RIT policies and programs is also evident at national and, to some extent, international levels. In the United States, federal pressure on RIT sponsors and performers to demonstrate the relevance and value of what they do has grown. This is due not only to new legislative requirements, for instance through the Government Performance and Results Act, or even tighter budget constraints, but also to heightened political debate about to focus public investments in scientific, technological, and economic competitiveness in the new era. A case in point is the US Advanced Technology Program (ATP) – a controversial federal partnership program with industry to develop and commercialize high-risk enabling technologies to support US competitiveness. The ATP aims at long-term projects that are not only risky but are likely to have diffuse spillovers. Nonetheless, Ruegg (1998) observes that the ATP “has met nearly continuous demand for measures of impact of the program since the day it was established.” Yet, the ATP is not an isolated case. In other federal programs, as well as in comparable RIT programs in Europe, at the European Union and national government level, the demand for evaluation has increased.

However, significant “supply-side” issues remain in the development and application of appropriate methodologies to meet the increased demand for RIT measurement and evaluation. RIT policies pose many measurement challenges for evaluators. It is usually difficult to fully assess all benefits and costs of RIT programs, related spillovers, counter-factual explanations, and “soft” institutional and learning effects. Such evaluation challenges increase as RIT programs become more complex in efforts to address multiple objectives, involve partnership consortia, or to promote the capabilities of research networks and clusters. Again, the ATP case is illustrative. Extensive efforts to improve evaluation methodologies have been required, only some of which have so far come to fruition. Moreover, even where

## 1 Overview of Workshop and Aims

comprehensive evaluative strategies are developed, issues remain in the communication and use of results. A major issue is how to better link evaluation strategies with program learning and continuous improvement, as well as evaluation's more conventional application to program justification (see Shapira and Youtie 1998).

The RIT policies of the European Commission provide another illustrative case: evaluation has been a legislative requirement for European RIT programmes since the early 1980s. The European Commission services have acquired a solid experience in the field of research evaluation since then. For the first three framework programs (FPs), more than 70 program evaluations and more than 40 supporting studies have been carried out, all in all involving more than 500 European experts. For the 4th FP, the changing S&T environment and the increasing pressure for timely, independent evaluation has led to the implementation of the current evaluation scheme which is based on two activities: for each program, a *continuous monitoring* with the assistance of *experts external* to the Commission Services, and, with multi-annual intervals, a *five-year assessment* conducted by external experts. While this scheme fulfilled its duties, the continuously increasing need to better demonstrate the usefulness of a European public RIT policy call for an enhanced measurement of results and impacts. The next 5-year assessment and the forthcoming 5th Framework Programme offer the necessary opportunities. A set of "*criteria*" has been developed ("European value added" and the subsidiarity principle in relation to national efforts; social objectives; economic development and scientific and technological prospects) that calls for a further development of the evaluation practice, concerning in particular the socio-economic aspects of the programs. Changes in the objectives and, in particular, a shift towards broader socio-economic targets require a *redesign* of the assessment mechanisms and related ways of data provision currently available: it is likely that a new, comprehensive program assessment scheme will be implemented.

Furthermore, the new center-left governments in the major European countries (UK, France, Germany) are putting a stronger emphasis on public intervention in technology and innovation, but – different to previous social democratic governments – clearly based on a strong performance management; there are signs that advanced program evaluation practices will gain a strong thrust. This development parallels the "reinvented government" and performance-focused management approaches that, as already noted, have been a growing feature of U.S. science and technology policies over the past decade (Shapira, Kingsley, and Youtie 1997).

As communities of evaluators are everywhere being encouraged (if not pushed) to enhance their methodologies, windows of opportunity are opened up for EU-US comparisons, contrasts, and mutual learning. There are considerable needs and demands for the benchmarking of RIT evaluation strategies, for exchange and learning about best and innovative practices in the field, and for the development of collaborative projects to improve practice. Such comparisons have to be qualified by con-

## *1 Overview of Workshop and Aims*

trasts in innovation systems, including variations in scientific, technological, and industrial conditions, as well as by differences in specific policies and implementation methods. Nonetheless, there is a convergence of broad RIT objectives in the US and Europe. Each “bloc” is intensively pursuing new science and technology frameworks, policies, and partnerships to enhance technological standing and industrial competitiveness in the global economy, raise payoffs from R&D investments, and promote innovation and venture creation. Equally, on both sides of the Atlantic, there are increased demands for RIT performance measurement and evaluation. This situation offers rich opportunities for comparison and mutual learning and the improvement of practice. In the past, US science and technology policy evaluators and the policymaking community in general tended to focus on their own US ‘national’ system of innovation, with comparative work mainly organized through individual transatlantic contacts. We now sense greater interest, need, and opportunity for comparative transatlantic initiatives and projects - and we would like to encourage and support this in an institutional form through the workshop and subsequent follow-on activities. In turn, European feedback confirms that the workshop will be timely: after a period of intensive internal building of European-level science and technology policy projects and evaluation networks, we sense a keen readiness to deepen collaboration with US colleagues.

### **The Workshop: Learning from Science and Technology Policy Evaluation**

The US-European Workshop on Learning from Science and Technology Policy Evaluation aims to advance the processes of exploring the opportunities for mutual learning and improved practice described above. This workshop is a new initiative to bring together American and European researchers involved in the evaluation of research, innovation, and technology policies. In addition to researchers in academic institutions, the workshop also involves researchers and analysts from other organizations involved in evaluation, programs, and policy in the RIT field. Those involved represent an interdisciplinary group, including specialists in science and technology studies, economics, program evaluation, and public policy.

The workshop has multiple, interrelated aims, as follows:

- To analyze, better understand, and contrast the characteristics of the need and demand for RIT performance measurement and evaluation within the US and European innovation and policy making systems. This involves consideration of how RIT evaluation strategies are affected by changing paradigms about the role and operation of science and technology policy and broader trends in public management, economic integration, and knowledge flows.
- To assess and contrast current US and European RIT evaluation practices in selected comparable fields, including approaches, agencies of implementation, principal metrics, methodologies, and communication and use of results.

## *1 Overview of Workshop and Aims*

- To identify and review innovative, “leading-edge” evaluative methodologies in critical RIT fields, especially to understand the potential of such new approaches to enhance feedback for program learning and improvement as well as to better determine economic and social impacts.
- To consider strategies, research approaches, and other changes that could lead to improvements in RIT evaluation design, methodologies, and policy utility.
- To identify and discuss needs and opportunities for collaborative US-European projects in the field of RIT evaluation, including the development, value, and sustainability of ongoing US-European RIT evaluation learning networks.

The workshop has been organized as a collaborative initiative between two principal organizations. On the US side, the lead organization is the School of Public Policy (SPP) at Georgia Institute of Technology (Georgia Tech). SPP is recognized as one of the prominent centers for science, technology, and information policy in the US and has taken the lead in engaging other US research and policy institutions and individuals in the workshop. On the European side, the lead organization is the Fraunhofer Institute for Systems and Innovation Research (ISI) in Karlsruhe, Germany. ISI is a public non-profit institution that is widely recognized as a major European center for research and analysis in the field of science and technology policy. In the developing and implementing the workshop, Georgia Tech and ISI are able to build on existing collaborative linkages. Researchers from these two institutions have collaborated in a series of projects and publications on science and technology themes in recent years. Each organization is also extensively involved in national and international S&T networks.

For the American side, the United States National Science Foundation has provided support for the participation of US evaluation researchers in the workshop. The Georgia Tech Foundation has also provided support for Georgia Tech researchers and students. The Volkswagen Foundation has provided support for European participation.

In inviting workshop participants, we have sought to incorporate a range of perspectives from within the US and European RIT evaluation and policy communities able to contribute to the workshop themes. Participant selection will be based on demonstrated and ongoing interest in collaboration on key US-EU RIT evaluation themes. We have included both senior and junior scholars. Participant invitations are extended based on agreement to prepare and present a written paper contribution on an agreed topic and to submit this paper for inclusion in the workshop proceedings and follow-up publications, with appropriate revision as necessary. Participants are also expected to serve as commentators and reviewers of other workshop papers.

We recognize that it is important not only to distinguish the UE-EU workshop from other exchange mechanisms, but also to take into account and build upon other

## *1 Overview of Workshop and Aims*

mechanisms to enhance the value of the workshop and ensure that follow-up activities will be successful. The workshop is organized primarily as an opportunity for researchers and analysts in RIT evaluation to (1) engage in high-level, in-depth, and critical discussion on the current state of the field; (2) present leading-edge research and cases; (3) stimulate comparative learning about research and practice between the US and Europe; and (4) explore possibilities for follow-on collaborative US-European projects and networks. The US-EU workshop is not replicated or duplicated by other structures. Some existing networks within Europe bring together European researchers engaged in science and technology assessment and policy evaluation, such as the European Research, technology and Development (RTD) network or the European Advanced Science and Technology Policy Network. However, these networks do not involve US participants and they tend to focus around the specific requirements of the European Union. In the US, there are also nascent networks of RIT evaluators, for example through the topical interest group on research and technology policy of the American Evaluation Association (AEA). Conferences on either continent, such as the AEA Annual Meeting or its European counterpart, allow some Transatlantic exchange, although generally such exchanges have limited participants and are loosely structured.

Significant government-to-government exchange mechanisms for science and technology policy exist, and they provide a context for the workshop. For instance, under the Joint US-European Union Action Plan to expand relationships, a conference on Transatlantic Science and Technology Cooperation was held in June 1998 at the National Academy of Sciences in Washington, DC. This conference was large (more than 200 US and European invitees), and focused broadly on emerging developments in information technology, transportation, climate change, and life sciences. Evaluation methodologies for policies within these themes was not the predominate concern. Moreover, it is important to note that agendas and discussion at the workshop will not be exclusively oriented or constrained by the politics of intergovernmental governmental. However, the likely growth of Transatlantic research collaboration stimulated by such intergovernmental exchanges raises challenges in evaluating multinational policies and projects that the workshop will need to consider.

An active plan of dissemination, publication, and follow-up collaboration forms an integral part of the workshop project.

All participants invited to the workshop will be expected to prepare a written paper on an agreed topic as part of a workshop panel theme or keynote presentation. Participants will be asked to use this paper as an occasion to crystallize and advance new (i.e. unpublished or not yet widely presented) concepts, ideas, methods, and results related to issues, challenges, and opportunities in the RIT evaluation field. All papers will be shared with other participants, and presented and discussed at the

## *1 Overview of Workshop and Aims*

workshop. Participants will be required to submit their papers prior to the workshop in hard copy and electronic format.

At the workshop itself, a rapporteur system will be used to track workshop discussions and conclusions. At each session, an individual will be designated to serve as a rapporteur and to take notes of the proceedings. In the final section of the workshop, discussions will be held not only on workshop findings and conclusions, but also on needs and possibilities for new collaborative US-EU initiatives and projects on specific RIT evaluation challenges and opportunities.

Following the workshop, several specific dissemination strategies will be pursued. After the workshop, the papers will be posted to a Web site, along with a report of the conference discussions, conclusions, the final communication, and listing of participants. The widespread availability of the Internet and Worldwide Web reduces the need to pre-produce a large number of paper copies of the proceedings. Nonetheless, we will produce and make available paper copies to workshop participants, sponsors (the National Science Foundation and the European workshop co-sponsor), the libraries at Georgia Institute of Technology and the Fraunhofer Institute for Systems and Innovation Research, and to other requesters who are not easily able to access the papers through the Worldwide Web.

Following the workshop, further publication opportunities will be pursued for the workshop papers, in addition to the proceedings. There are likely to be opportunities for either a journal issue or an edited book. In pursuing these opportunities, we will offer editorial comments, seek peer review and encourage authors to revise their papers to ensure high quality.

Although the workshop stands as a valuable activity in its own right, it is an explicit aim to further enhance the value of the workshop by using it as platform for follow-up collaborative activities. In preparatory interactions, the actual workshop, and subsequent follow-up activities, opportunities for follow-on collaboration will be actively pursued. In particular, the workshop seeks to motivate a new round of interaction between evaluators in the RIT field in the US and Europe. There will be a full discussion at the workshop of plans and opportunities for future, ongoing collaboration.

As part of this discussion, one of the immediate areas will be to focus on targets of opportunity for focusing future research. Here, workshop participants will be encouraged to identify unresolved issues, promising new evaluative methods, or comparative approaches where joint US-EU science and technology policy evaluative projects would be worthwhile. It is intended that this process will stimulate new collaborative project proposals.

## *1 Overview of Workshop and Aims*

In addition, workshop participants will be asked to consider and guide the development of an ongoing US-European RIT evaluation network or learning consortium. This consortium would aim to stimulate the sharing of experience, the benchmarking of results, the promotion of collaboration, and the advancement of evaluative practice in the science and technology policy field. The US “anchor” for this consortium could be Georgia Tech, while ISI would function in a similar role for Europe. Workshop participants would be encouraged to continue to be engaged, although involvement in the consortium would also be open to interested others in the US and European science and technology policy evaluation communities. At the Fall 2000 workshop, participants are asked to initiate a discussion on the potential activities, benefits, costs, and organizational forms of further collaborative activities. As well as physical meetings, we will explore possibilities for virtual networking using readily available electronic communications technologies. As an outcome of this discussion, it is hoped that a framework for further cooperation can be agreed among participants, leading to a collaborative proposal (involving others) for a multi-year US-European science and technology policy evaluation and learning network. We believe that there are sponsors and RIT policy organizations in both the US and Europe that would be receptive to such a proposal.

## **References**

- Becher, G., and Kuhlmann, S., (eds.) 1995, *Evaluation of Technology Policy Programmes in Germany*, Kluwer Academic Publishers, Boston, Dordrecht and London.
- Berglund, D., 1998. *State Funding for Cooperative Technology Programs*. Columbus, OH: Batelle Memorial Institute.
- Branscomb, L., and Keller, H., 1998. *Investing in Innovation: Creating a Research and Innovation Policy that Works*, MIT Press, Cambridge, MA.
- Caracostas, P. and Muldur, U., 1998, *Society, The Endless Frontier*, European Commission, Directorate-General XII, Science, Research, and Development, Brussels.
- Chang, C., 1998. “A New Lexicon and Framework for Analyzing the Internal Structures of the U.S. Advanced Technology Program and its Analogues Around the World,” *The Journal of Technology Transfer*, 23,2, summer.
- Coburn, C., and Berglund, D., 1995. *Partnerships: A Compendium of State and Federal Cooperative Technology Programs*. Columbus, OH: Batelle Memorial Institute.
- Cozzens, S., and Melkers, J., (1997). “Use and Usefulness of Performance Measurement in State Science and Technology Programs,” *Policy Studies Journal*, 25, 2.
- Crow, M., 1994. “Science and Technology Policy in the US: Trading in the 1950 Model,” *Science and Public Policy*, August.
- Freeman, C., 1991. “Technology, Progress, and the Quality of Life,” *Science and Public Policy*, 18(6).

## 1 Overview of Workshop and Aims

- Galli, R. and Teubal, M., 1997. "Paradigmatic Shifts in National Innovation Systems," in Edquist, C. (editor). *Systems of Innovation: Technologies, Institutions, and Organizations*, Pinter, London.
- Georghiou, L., 1995, "Research Evaluation in European National Science and Technology Systems," *Research Evaluation Vol. 5, No. 1*, 3-10
- Georghiou, L. et al., 1993, *The Impact of European Community Policies for Research and Technological Development upon Science and Technology in the United Kingdom*, London (HMSO)
- Kodama, F., 1991, *Emerging Patterns of Innovation: Sources of Japan's Technological Edge*, Harvard Business School Press.
- Kuhlmann, S., 1995, "Patterns of Science and Technology Policy Evaluation in Germany," *Research Evaluation*, Vol. 5, No. 1, 23-33
- Larédo, P., 1995, *The Impact of Community Research Programmes in France*, Paris (Les Presses de l'Ecole des Mines)
- Meyer-Krahmer, F., and Montigny, P., 1989, "Evaluations of innovation programmes in selected European countries," *Research Policy*, 18, vol. 6, 313-331
- Osborne, D. and T. Gaebler (1993), *Reinventing Government: How the Entrepreneurial Spirit is Transforming the Public Sector*, Plume, New York.
- Reger, G., and Kuhlmann, S., 1995, *European Technology Policy in Germany*, Physica-Springer, Heidelberg (series "Technology, Innovation, and Policy" vol. 2).
- Ruegg, R., 1998, "Symposium Overview," Symposium on Evaluating a Public Private Partnership: The Advanced Technology Program, *The Journal of Technology Transfer*, 23, 2, summer.
- Shapira, P. and Youtie, J., 1998, "Evaluating Industrial Extension: Results, and Insights from the Georgia Manufacturing Extension Alliance," *The Journal of Technology Transfer*, 23, 1, spring.
- Shapira, P., and Youtie, J., 1998, *Manufacturing Modernization: Implications of Evaluation Results for Program Improvement and Policy Development*, Proceedings of Fourth Workshop on the Evaluation of Industrial Modernization Programs, Aberdeen Woods, Atlanta, 1997. School of Public Policy and Economic Development Institute, and the Georgia Institute of Technology, Atlanta, GA. Available at <<http://www.cherry.gatech.edu/workshop>>.
- Shapira, P., Kingsley, G., and Youtie, J., 1997. "Manufacturing Partnerships: Evaluation in the Context of Government Reform," *Evaluation and Program Planning*, 2, 1, 103-112.
- Tassef, G., 1992, *Technology Infrastructure and Competitive Position*, Kluwer, Norwell, MA.

## **2. Introduction: Emerging Paradigms for Evaluating Research, Innovation, and Technology Policies in the U.S. and Europe**

**Session Chair: Philip Shapira (Georgia Institute of Technology)**

---

### **The Academic Policy Analyst as Reporter: The Who, What, . . . and How of Evaluating Science and Technology Programs**

**Irwin Feller**

Pennsylvania State University, University Park, PA, USA

Email: IQF@PSU.EDU

---

#### **I. Introduction**

Opening, agenda-setting papers are challenging undertakings. To be informative, they require a framework that encompasses the multiple perspectives and topics to be covered in succeeding sessions. They also require the proper tone: not too solipsistic as to suggest that the author already possesses the answer(s) to the issues or conundrums that will be subsequently addressed by others; not too bland or deferential as to suggest the absence of possible substantive differences between his/or her assessments and theirs. Added to this is the requirement to say something insightful or fresh, a truly difficult task given the expertise of the participants in the workshop and the availability of several excellent treatments and survey articles on many of the issues it will address (e.g., Georghiou and Roessner, 2000).

Influenced by my son's experiences as a newspaper reporter in compressing considerable amounts of information into both space and time while striving always to be accurate and objective, I have adopted the reporter's algorithm of what, where, when, who, why, and how to structure these comments. As a good reporter knows, these elements are interwoven: any of these questions can serve as a starting point; one question leads logically to the others; the complete story is known only when each question is answered.

I must quickly add two disclaimers. Lest I be challenged about adherence to a reporter's code of professional conduct, first, I freely admit that the boundaries I draw between facts and interpretations are as much symbolic as substantive; I leave the role of desk editor in separating the two to other workshop participants. Second, I draw primarily on a set of U.S. and state S&T programs with which I have been

involved. These experiences encompass a goodly range of NSF, NIST, USDA, NASA, and DOE programs, but obviously do not encompass all such undertakings—limited treatment of NIH programs and lack of treatment of national laboratories being among the more evident omissions.

## **II. Where, Who, When, Why, What, and How**

### **Where?**

To lead off this account, I start with an answer to what would appear to be a simple question: Where is evaluation of S&T programs occurring? The answer is everywhere, at least as suggested by the countries and governments participating in this workshop. Relatedly, given the federal structure of the U.S. political system and the decentralized, organizationally wide-ranging character of U.S. funding of science and technology, it applies, too, across levels of government, a swathe of agencies, and a large and diverse set of S&T program strategies (Melkers and Roessner, 1997).

### **Who?**

To an economist, the question of “who?” immediately decomposes into “who demands” and “who supplies” evaluation. The answer to who demands evaluation is implicitly subsumed under where—namely, many levels of government and many agencies.

The answer to who supplies evaluation also is straightforward, albeit laden with implications for the later discussion of how. Formal executive or legislative entities—such as U.S. General Accounting Office, Office of Management of Budget, intra-agency evaluation offices, the National Academy of Sciences-National Research Council complex, think-tanks and consulting firms within and without the Washington, DC, Beltway, management consulting firms, organized academic units, and individual researchers and consultants—are but a few of the large number of performers of evaluation studies.

The result is a large and heterogeneous number of performers. One consequence of large numbers is quality variation, both within and across methodologies. Another consequence is competition among suppliers. Put differently, nothing approaching a dominant institution or evaluation methodology exists.

### **When?**

“When?” would seem to be another seemingly simple question to answer. The answer is now. As many observers have noted, we live in an age of assessment. But to an inquisitive reporter, to answer now immediately raises the question, why not before?

## *2 Paradigms for Evaluating Research, Innovation, and Technology Policies*

Indeed, it is obvious that evaluation of S&T programs is not a “new” phenomenon. Historians of federal support of S&T have identified many “early” proto-evaluative studies. For example, the 1884 Allison Commission’s review of the Signal Service, Geological Survey, Coast and Geodetic Survey, and the Hydrographic Office of the Navy Department was precipitated by a search for “greater efficiency and economy of administration of the public service” (Dupree, 1957). Without too much stretching of the term evaluation, a more detailed reading of the historical record would likely produce even earlier precedents.

To focus on the more contemporary period, several authors have detailed the presence of evaluative studies of U.S. S&T programs in the 1970s and 1980s. Guston, for example, in describing increasing demands by the U.S. Congress for accountability of academic science and attempts “to increase information about scientific productivity for policy analysis,” notes that the Public Health Service Act was amended in 1970 to set aside up to one percent of NIH appropriations for program evaluation (Guston, 2000, p. 79). Roessner (1989) and Cozzens (1997) review S&T program evaluations of the 1980s with varying attention to the pre-1980s. Braun, combining when and why, points to the 1980s as a singular period. He notes that:

... the evaluation of funding programs was hardly developed until the end of the 1980s and even with standardized evaluation reports it is hard to tell if the distribution of research funds could have been optimized by different allocations or alternative use of funding instruments. During the period when the science-push image ruled governmental considerations in research policy-making, the reputation of mission-agencies among the scientific community became, therefore, the accepted assessment criteria for policy-makers (1993, p. 156).

Older observers, such as myself, are tempted to find the cultural and institutional origins of program evaluation in Rivlin’s 1971 call for systematic experimentation, in which an innovation would “be tried in enough places to establish its capacity to make a difference and the conditions under which it works best,” and with “controls to make the new method comparable with the old method or with no action at all” (1971, p. 87). Indeed, terms like “experimental” in the title of various S&T-related programs, as in the National Science Foundation’s Experimental Program to Stimulate Competitive Research, or the widespread use of the metaphor “laboratories of democracy” to describe American state government university-industry-government programs would seem to suggest an experimental basis to recent U.S. S&T initiatives, thus clearly calling for evaluation (or hypothesis testing) as an integral part of an overall strategy. (However, few of these programs, in fact, are true experiments in the case that they can be terminated if found not to work.)

### **Why?**

Dating of when is far more than an academic parlor game, for dates provide a key to the question of why. Many answers have been proffered. A brief summary would include the following.

- 1.) Generic questioning, if not outright skepticism, about the capacity of the public sector to improve on the private sector's allocation of resources in any domain, not excluding science and technology programs—with government failures, rent-seeking, pork barrel behaviors being as pervasive as the (Arrow-Nelson) market failure settings conventionally used to justify public support of basic research and selected technologies (Cohen and Noll, 1991; Wolf, 1988).
- 2.) The absolute scale of public expenditures on S&T, which even in periods of expanding government revenues, and even more so in periods of stagnant budgets, calls attention to the opportunity costs of such outlays.
- 3.) Movement, in the U.S. at least, of S&T programs outside the pale of the more widely accepted economic justifications for public support of fundamental research into promotion of civilian technologies, NIST's Advanced Technology Program being the ritualistic battleground over which these boundary wars are fought.
- 4.) High on any list of Why must be the eroding political acceptance of claims by the scientific and technological communities, or at least the academic portions of these communities, to be self-policing. Both Guston and Braun have recently used principal-agent frameworks to argue that as one or more branches of government, the principal, comes to lose trust in the ability of the research community, the agent, to be self-policing, it imposes new procedural and administrative requirements to ensure accountability (Braun, 1993; Guston, 2000). Evaluation is such an instrument.
- 5.) Last as a source of Why, one cannot avoid mention of the Government Performance and Results Act (GPRA), with its requirements upon federal agencies to present systematic statements of goals and objectives, to link budget requests to objectives, and to document results from prior expenditures.

I place GPRA last for two reasons. First, assessment of whether GPRA will produce significant improvements in the management of science and technology programs, or indeed in any field remains an open question, one on which experts on the legislation disagree (Behn, 1994; Cozzens, 1997; Radin, 2000).<sup>1</sup> Second, and more di-

---

<sup>1</sup> E.g., Radin [2000]: "Viewed as whole, GPRA has failed to significantly influence substantive policy and budgetary processes. Instead, its use of administrative rhetoric has caused it to collide with institutional, functional, and policy/political constraints that are a part of the American decision-making system [p. 133]."

## 2 Paradigms for Evaluating Research, Innovation, and Technology Policies

rectly for this workshop, I do so because my several recent experiences with program evaluations and GPRA within NSF and NIST suggest that although GPRA is rife with implications, both positive and negative, for future evaluative undertakings, it has not been a major precipitant of large-scale evaluations to date. Many of the recent evaluations of NSF's major programmatic initiatives, such as of its Engineering Research Centers, Science and Technology Centers, the Experimental Program to Stimulate Competitive Research (EPSCoR), and the Industry-University-Cooperative Research Centers program, predate GPRA and are only loosely connected to the GPRA process. Similar statements hold for NIST's ATP and MEP programs.

Again, based on my observations and participation in NIST-MEP and NSF program review activities, it also appears that GPRA has yet to be the basis of any significant advance in evaluation methodology. NSF's GPRA report for FY1999, for example, relies heavily on Committee of Visitors and Advisory Committee reports, which in turn are based primarily on peer review assessments and a modicum of data collection procedures that, as noted in the NSF GPRA FY1999 report, are subject to expert biases and unevenness in data quality (National Science Foundation, 2000).

GPRA, however, is serving to highlight the shortcomings of existing data collection procedures and the strengths and weaknesses of existing methodologies for evaluating programs. It also seems to be encouraging increased experimentation in agency procedures that would permit more "controlled" tests of the "contribution" or "value-added" of the agency's programs. Thus, its impacts in terms of the frequency and standards for evaluation may yet to be felt.

### **What?**

"What?" is presented here in terms of "what use" is made of evaluation studies. This approach to the question has both *ex ante* and *ex post* dimensions.

A useful answer to What as an *ex ante* question is found in Goldenberg's characterization of the three faces of evaluation: 1) to learn about a program's operations (does it work?; how can it be made to work better?); 2) to control the behavior of those responsible for program implementation (modify objectives; alter activities, reallocate resources, reassign responsibilities, etc.); and 3) to influence the responses of outsiders in the program's political evaluation (i.e. create the appearance of a well-managed organization; preemptively set metrics and methodologies; preempt external evaluations) (Goldenberg, 1983; see also Sapolsky, 1972). Each of these faces is visible in formal program reviews of many S&T programs, site visits, and conditional release of tranches of funds.

Given the above answer to "Why?" – namely, that evaluation is increasingly a means by which skeptical if not outright critics of public support of science and technology challenge either the existence or level of funding of these programs, it

## 2 Paradigms for Evaluating Research, Innovation, and Technology Policies

appears that the face that must be shown frontally to the audience is increasing that of program justification. Thus, NIST's Manufacturing Technology Centers program started out under the Omnibus Trade and Competitiveness Act with congressional stipulations that centers be subject to an external reviews before they could receive a second three-year tranche of funding. ATP, of course, has struggled for continued existence since its inception. Lacking the building base of grassroots support in the small business community and state government that MEP has accumulated, it has been forced to turn to increasingly sophisticated evaluation techniques to demonstrate its value-added.

The *ex post* question is this: what use is made of evaluation studies after they are completed? And indeed, an extensive literature on knowledge utilization points to the limited and highly selective use of social science research in general, and evaluation studies in particular (Webber, 1991–92; Weiss 1979, 1980). More to the immediate point is Tolstoy's observation that doing good may not make you happy, but doing evil will surely make you unhappy. A counterpart albeit at times inverted proposition exists in the evaluation of S&T programs: a "positive" evaluation pointing to program effectiveness and/or efficiency is not a guarantee that a program will pass political scrutiny. A "negative" evaluation pointing to the lack of program effectiveness/and/or efficiency/or whatever is no predictor of program termination or contraction.

Many factors enter into this inversion but two warrant particular notice. First, evaluation studies seldom yield unmixed findings concerning outcomes, especially of programs with multiple objectives. Second, as the political agenda and character of stakeholders change, so too can the objectives. Thus, programs can always be attacked for missing the new target; alternatively, new targets can always be identified to directly correspond to a program's activities.

*Two quick examples.* Few U.S. S&T programs have faced as much political opposition as has ATP. Few federal S&T programs have invested as heavily in evaluation and have been as successful in recruiting the leading researchers on the returns to public investments in R&D-based programs—Darby and Zucker, Feldman, Griliches, Jaffe, to identify only a few who have studied ATP—to examine its impacts. Econometric evaluations of the ATP program, particularly of its spillover benefits, along with some striking case histories—the proverbial "golden nuggets"—suggest that there may indeed be value-added to the program. Against this backdrop of often exacting methodological scrutiny, consider the following statement from the House Appropriations Committee report for FY2001:

The advocates for the ATP program have always had to answer fundamental questions, such as whether the program achieved results that could not be achieved through the private marketplace; whether it funded technology development and commercialization that would not be undertaken but for the existence of the pro-

## *2 Paradigms for Evaluating Research, Innovation, and Technology Policies*

gram; and whether the Federal government should play a role in picking technologies to be developed and then funding that development at substantial expense.

After many years in existence, the program has not produced a body of evidence to overcome these fundamental questions about whether the program should exist in the first place (U.S. Congress, July 19, 2000).

Similar discrepancies between the conclusions of evaluation studies and policy outcomes are reported for state government S&T programs. North Carolina's MCNC (Microelectronics Center of North Carolina) was, "(b)y many measures" highly successful, but lost its state funding in the mid-1990s in part because of "political issues" associated with changes in legislative leadership, but also because the measures did not connect with the broader agenda of policy makers in economic performance, particularly job creation (Carlise, 1997).

Contrast this attitude with the relationship between evaluative studies of the Small Business Innovation Research Program and EPSCoR and congressional actions. Wallstein's research, for example, raises questions about the net economic impacts of SBIR, suggesting that federal funding crowds out industry spending, and that "Simply because SBIR funded a commercially viable project does not mean that the program stimulated innovation or commercialization" (1998, p. 206; also, Lerner, 2000). This (and related) findings have done little to lessen congressional adoration of the program.

The Yin-Feller evaluation of the NSF EPSCoR program at times is cited as demonstrating that the program is accomplishing its original and still primary objective of increasing the share of NSF (and federal government) academic R&D funds received by states that historically have received low percentages of these awards (Yin and Feller, 1999). A closer reading of their findings indicates that only 10 of the 18 states and Puerto Rico that participated in the program had percentage increases above the national average increase in federal funding of academic R&D, whether this national average is computed for all 50 states or for all non-EPSCoR states. The nine states whose percentage increases fell below the national average had declines in shares (although it could be argued that the program's impact for these states was to keep shares from falling even further) (Feller, forthcoming). Findings like these, however, are essentially irrelevant to continued pressures to increase the number of states eligible for EPSCoR support, the number of agencies required by Congress to adopt EPSCoR-like programs, and total funding for the aggregate of EPSCoR-like programs.

These vignettes, of course, serve mainly to confirm a well-known proposition, but one that cannot be overemphasized too frequently: Evaluations or related research, however well done, are only one element in political decision-making.

## *2 Paradigms for Evaluating Research, Innovation, and Technology Policies*

There is another aspect to “what use” is made of evaluation studies that warrants attention, namely, the differential use made by the sponsors of summative studies that document aggregate impacts and formative studies that point to alternatives to dominant agency strategies. Studies of the first type are important—indeed possibly essential, as just suggested—to demonstrate a program’s effectiveness to executive and legislative bodies, thereby justifying continuing and possibly increased funding levels. Studies of the second type, which may be nested within the first type, highlight issues related to program design and implementation. Findings related to such issues may not be visible to those who hold the purse strings, and therefore may be ignored by program sponsors.

*Again, two quick examples.* Recently, as part of its deletion of a sunset provision on the maximum number of years that a manufacturing modernization center could receive NIST-MEP funding, the U.S. Congress required NIST to submit an annual report on its evaluation procedures and also to have this report vetted by an external review panel. Included in the NIST report are accolades from nationally renowned external reviewers: MEP’s evaluation design is described as “an extremely impressive example—an almost unique example—of well designed very large scale evaluation of very large scale programs” by Michael Scriven, President, American Evaluation Society. MEP’s contribution to increased productivity, value-added, sales, and other economic outcomes also are documented in a series of well-crafted survey, benefit-cost and econometric studies (e.g., Jarmin, 1999; Oldsman, 1996). MEP has thus been served well politically by evaluations of its program impacts.

Less apparent are the impacts of evaluations upon MEP’s strategies for targeting clients and providing services. Luria, for example, has argued that, nationally, the clients of these centers comprise two distinct sets, which he labels “high-productivity” and “low-road (low-wage) shops among MEP’s clients. Too many of the extension centers’ services, according to Luria, are directed at “quick-hit” projects that appeal to the latter set of firms. Such projects “result in clients’ achieving greater sales and employment growth than non-clients, but not to relative increases in productivity, wages, or profits” (Luria, 1997, p. 99). The prescription is a “substantially reoriented mix of services, with better matching of projects to more thoughtfully attracted clients.” Acceptance of this prescription at either the national or center level appears slow.

Relatedly, in its concluding section, the Yin-Feller evaluation of NSF’s EPSCoR program calls for reexamination of the criteria by which a state or university is deemed eligible to participate in the program. It also observes that “Since the evaluation has demonstrated that the program is successful in improving the R&D competitiveness of participating states, consideration of graduation criteria or university transition from EPSCoR support seems relevant” (1999, p. 35). These recommendations have gone nowhere; they fly in the face of political realities.

**How?**

“How?” is both the easiest and most difficult of the questions to answer. How has been articulated in several compendia authored and edited by participants in this workshop (e.g., Bozeman and Melkers, 1993) and others. Indeed, considering only the U.S. participants, the techniques represented at this workshop are quite wide ranging: bibliometrics, value-mapping, benchmarking, benefit-cost analysis, social capital, and social savings, to name only a few that have distinct headings.

An ecumenical spirit is certainly appropriate when viewing the distinctive contribution of each of these techniques to evaluating the complex, multi-faceted, and multi-objective character of public sector S&T programs. For example, the prospective impact of Narin, Hamilton, and Olivastro’s (1997) use of citation measures from patent statistics to demonstrate the societal returns to public investments in fundamental research may equal that of such classic economic studies on the social returns to basic research, represented by Griliches’ studies on hybrid corn and Mansfield’s work on the contribution of academic research to industrial innovation.

Moreover, ecumenicism, along with eclecticism and syncretism, may be the only effective evaluation research strategy in coping with the nettlesome methodological and data issues found in complex and multi-objective S&T programs. As my colleagues Amy Glasmeier, Mel Mark, and I have written, echoing Cronbach, “The standard for future action is not a single flawless study that satisfies all strictures, but rather a succession of studies that critique and improve upon each other as they collectively advance toward norms of formal evaluation methodology” (1996. p. 318).

But appreciation of the need for and potential contribution of different methodologies should not lull one into analytical complacency or polite methodological correctness. Differences exist about the construct validity, operational feasibility, and decision-making relevance of the different methodologies represented at this workshop.

Within-group differences about the validity, truthfulness, or importance of particular techniques or findings, of course, are the stuff of normal science. Cole’s dissection of competing theories of the sociology of science (1992), David, Hall, and Toole’s (2000) critique of econometric studies of the question of whether public R&D spending is a complement of or substitute for private R&D spending, and Meyer’s critique (2000) of Narin, Hamilton, and Olivastro’s use of citations from patent filings, represent debates largely entered into and addressed to other specialists within the same research traditions.

This workshop is not structured along such lines; instead, it is a group assembled to discuss the diversity and heterogeneity of evaluative approaches. Thus, it invites

## *2 Paradigms for Evaluating Research, Innovation, and Technology Policies*

either a serial presentation of different techniques, essentially a menu of options that may tantalize one or more to change tastes, or explicit cross-group assessments.

Anyone who has experience with both the research and policy literature on the diffusion of technological innovations, understands that cross-disciplinary exegesis is a formidable and at times fractious task (Downs and Mohr, 1976; Mowery and Rosenberg, 1979; Tornatzky and Klein, 1982). Specifically, in the context of evaluations of S&T programs, cross-technique assessments pose questions about whether (or the degree to which) the metrics proposed in alternative methodologies accurately capture the objectives being sought by decision-makers, the presence or consistency of theoretical linkages between and among the input and output variables, and the relationships between measured outputs and desired outcomes.

Although clearly the more demanding alternative, this latter approach, I believe, makes for a more informative story. To provide detail to the story, I draw upon my current research to assess two of the techniques represented at this workshop: benchmarking and social capital.

### **III. Benchmarking**

If one starts, as I do, from Hirschman's perspective that organizations operate on the basis of "X-efficiency" rather than "allocative efficiency," benchmarking can be a useful diagnostic to identify shortfalls in potential performance and as an organizational prod to search and learning behavior about best-practice techniques elsewhere. Learning, in turn, may lead to changes in objectives, policies, strategies, organizational form, and resources, any or all of which can move an organization in new directions and/or towards production efficiency frontiers.

But benchmarking as a decision-making or evaluation methodology related to S&T programs has many pitfalls. Benchmarking can be, and indeed has been, measurement without theory. Causal connections between the things being measured and efficient attainment of stated or nominal objectives at times have been loosely or incorrectly stated. The result is attention to irrelevant or wrong variables and relationships; to the extent that benchmarking is adopted as an organizational tactic without explicitly identified and validated causal linkages, its effects can be dysfunctional. It can induce or compel subunit optimization based on specific benchmarks that in fact impede an organization's attainment of higher order objectives. (Or as Demming has phrased this point, measures of productivity do not lead to improvements in productivity.)

Thus, in the case of university contributions to technological innovation and technology transfer, a frequently used benchmark—patents per R&D dollar—fails to account for significant variations in industrial and university propensities to either patent or to license patents across patent utility classes. Relatedly, the use of univer-

sity license income, either as an absolute measure or as a ratio divided by some measure of R&D activity, as a benchmark to gauge institutional performance fails to take into account that some major research universities knowingly trade off license income for increased sponsor research support (Feller, Feldman, Bercovitz and Burton, 2000; Southern Technology Council, 1997). In such cases the benchmark is the outcome of a complex set of decisions, and is as likely to be the residual outcome of higher order objectives than the decision variable that the organization has sought to optimize.

#### **IV. Social Capital**

A different set of difficulties is associated with the use of the concept of social capital as an evaluation approach (Dietz, 2000; Fountain, 1998). As with benchmarking, my comments reflect a mixed assessment, pointing to what I perceive to be the insights offered by the approach but suggesting also the pitfalls and dead-ends contained in its use. At its simplest, according to Dietz, “social capital can be thought of as the stock of good will/mutual trust accruing from cooperative relationships among two or more parties” (2000, p. 139). Debate exists about the analytical or empirical groundings of the social capital framework, as articulated by Coleman, Fukuyama, and especially Putnam (Cf. Bridger and Luloff, forthcoming), but the immediate focus here is on the use of the approach in program evaluation. In general, whereas benchmarking suffers from the defect of being overly specific, social capital suffers from conceptual and operational diffuseness.

The most direct application of the social capital model appears in recent treatments of the EPSCoR program, where it is presented as both a complement to and as an independent objective to the program’s stated objective of increasing the share of NSF (and federal government) academic R&D awards received by participating states. Increased cooperation and trust among universities within a state flowing from programmatic requirements of the EPSCoR program are clearly evident in a number of states. The program’s requirements for a single statewide application rather than applications from each eligible university and for the formation of a single steering committee have brought together representatives from universities that had not only historically competed for state funding but also had developed patterns of aloofness and acrimony. Cross-campus collaboration flowing from the EPSCoR program has improved the general standing of the state’s universities with state government (albeit not necessarily to significant increases in state appropriations). In Alabama and other states, it has made possible consortium-based proposals across institutions involving a pooling of expertise and research facilities that allowed the state’s universities to compete for major national awards from federal agencies for which no single university would have been competitive and which would not formerly have not possible absent the trust and cooperation engendered by the EPSCoR program.

## *2 Paradigms for Evaluating Research, Innovation, and Technology Policies*

But even abstracting from continuing questions about its core concepts, the social capital model has severe limits as either a programmatic guide or as an evaluation methodology. As a programmatic guide, it is beset with problems related to eligibility or graduation criteria (how does one determine among a pool of potential participants which ones require boosts in social capital?; how does one know when funded sites have achieved threshold levels of social capital so as not to require further support?; do differences exist in the quantity of social capital in “have” and “have-not” states?). As both a programmatic guide and as an evaluation approach, it runs the risk of confounding necessary conditions, process variables, and possible spillover benefits with sufficient conditions, other “necessary” conditions, and intended or realized outputs and outcomes.

In the EPSCoR case, for example, considerable enhancement of trust and cooperation among institutions may flow from the program but still not produce sufficient support from state governments to enable the participating universities to offer competitive faculty salaries or develop competitive research facilities that are the critical determinants of the ability of these universities, singly and collectively, to compete for federal academic research awards. The social capital ethos is also redolent with goal displacement, or what Wildavsky has termed strategic retreat from objectives, in which programs redefine their objectives over time to be those that their activities appear to be producing rather than those with which they started.

### **V. Coda: International Comparisons**

Absent from the paper to this point has been any explicit comparative, U.S.-EU cast. It could have easily been presented to an all-U.S. audience addressing U.S. S&T policies. This seeming provincialism, however, relates more to the facts cited than to its analytical structure. Details of science and technology programs (and personal experiences) aside, its framing structure of who, what, where, when, why, and how, I believe also can be applied to the EU setting. More importantly, I believe that much of its contents applies to evaluation of S&T programs in the EU, although this is a proposition that requires testing.

Three specific comparative aspects of the paper warrant highlighting.

First, given the preponderance of academics among U.S. participants and my singling out of differences as well as complementarities in the evaluation methodologies they represent, it is worth noting again the decentralized character of the U.S. evaluation system. Multiple sponsors fund multiple researchers located in multiple institutions; the result is a diverse, at times competitive evaluation marketplace, in keeping with the characteristics of the U.S. political and academic systems. Methodological orthodoxy, thankfully, is impossible to establish. I would welcome hearing the EU workshop participants comment on the range of techniques found in the European setting.

Second, the objectives of seemingly similar U.S. and EU S&T programs can differ. For a time, for example, NSF and the EU expressed interest in comparing what seemed to be similar initiatives to stimulate the development of scientific and technical capabilities in “have-not” or “peripheral regions.” As noted, NSF’s EPSCoR program was originally designed to enhance the research competitiveness of those states whose universities historically had received low percentages of federal academic research funds. Over time, the program took on responsibilities, or at least a rhetoric, of contributing to state regional development. The EU’s STRIDE program (Science and Technology for Research Industrial Development in Europe) was more avowedly an economic development strategy, with universities occupying a lesser role. But central to the differences behind the two programs was STRIDE emphasis on collaboration of researchers and institutions across national borders, a characteristic of many EU S&T programs. This emphasis followed, as one EU representative phrased it, from one of the program’s objectives of building Europe.

Third, important differences exist in institution frameworks between the U.S. and EU, such as in systems of property rights, that can affect interpretations of evaluations of S&T programs conducted with identical methodologies. For example, a sizeable portion of evaluative assessments of current U.S. science and technology policy programs revolves about the use of patent data and attendant patterns of citations to other patents and various forms of published documents to assess direct and indirect impacts. Differences, however, exist between the U.S. patent system, on the one hand, and those of the EU and nation states in Europe, on the other. The EU patent system provides for opposition challenges to the granting of a patent, which is not provided for in the U.S. Relatedly, the number of citations to the “prior art” in EU patterns is below that in the U.S. (Harhoff, Scherer, and Vopel, 1999). Thus, any attempt to compare impacts of specific types of programs using comparable techniques may be confounded by different institutional contexts.

## References

- Behn, R. (1994). “Here Comes Performance Assessment-and it Might Even Be Good for You.” Pp. 257–264 in A. Teich, S. Nelson and C. McEnaney (eds.), *Science and Technology Policy Yearbook-1994* (Washington, DC: American Association for the Advancement of Science).
- Bozeman, B., and J. Melkers (1993). *Evaluating R&D Impacts: Methods and Practices* (Boston: Kluwer Publishers).
- Braun, D. (1993). “Who Governs Intermediary Agencies: Principal-Agent Relations in Research Policy Making.” *Journal of Public Policy*.
- Bridger, J., and A. Luloff (forthcoming). “Building the Sustainable Community: Is Social Capital the Answer.” *Sociological Inquiry*.
- Carlisle, R. (1997). “The Influence of Evaluation on Technology Policy-Making and Program Justification: A State Level Perspective.” Pp. 119–126 in P.

## 2 Paradigms for Evaluating Research, Innovation, and Technology Policies

- Shapira and J. Youtie (eds.), *Learning from Evaluation Practices and Results* (Atlanta, GA: Georgia Institute of Technology).
- Cohen, L., and R. Noll (1991). *The Technology Pork Barrel* (Washington, DC: Brookings Institution).
- Cole, S. (1992). *Making Science* (Cambridge, MA: Harvard University Press).
- Cozzens, S. (1997). "The Knowledge Pool: Measurement Challenges in Evaluating Fundamental Research Programs." *Evaluation and Program Planning* 20:77–89.
- Cozzens, S., and J. Melkers (1997). "Use and Usefulness of Performance Measurement in State Science and Technology Programs." *Policy Studies Journal* 25:425–435.
- David, P., B. Hall, and A. Toole (2000). "Is Public R&D a Complement or Substitute for Private R&D: A Review of the Econometric Evidence." *Research Policy* 29:497–530.
- Dietz, J. (2000). "Building a Social Capital Model of Research Development: The Case of the Experimental Program to Stimulate Competitive Research." *Science and Public Policy* 27:137–145.
- Downs, G., and L. Mohr (1976). "Conceptual Issues in the Study of Innovation." *Administrative Science Quarterly* 21:700–714.
- Dupree, A. Hunter (1957). *Science in the Federal Government* (New York: Harper & Row).
- Feller, I. "Elite and/or Distributed Science." In M. Feldman and A. Link (eds.), *Innovation for a Knowledge-based Economy* (Boston: Kluwer Academic Publishers), forthcoming.
- Feller, I., A. Glasmeier, and M. Mark (1996). "Issues and Perspectives on Evaluating Manufacturing Modernization Programs." *Research Policy* 25: 309–319.
- Feller, I., and J. Nelson (1999). "The Microeconomics of Manufacturing Modernization Programs." *Research Policy* 28: 807–818.
- Feller, I., M. Feldman, J. Bercovitz, and R. Burton (2000). "A Disaggregated Examination of Patent and Licensing Behavior at Three Research Universities." Paper presented at the Western Economic Association Meeting.
- Fountain, J. (1998). "Social Capital: An Enabler of Innovation." Pp. 85–111 in L. Branscomb and J. Keller (eds.), *Investing in Innovation* (Cambridge, MA: MIT Press).
- Georghiou, L., and D. Roessner (2000). "Evaluating Technology Programs: Tools and Methods." *Research Policy* 29:657–678.
- Goldenberg, E. (1983). "The Three Faces of Evaluation." *Journal of Policy Analysis and Management* 2: 515–525.
- Griliches, Z. (1958). "Research Costs and Social Returns: Hybrid Corn and Related Innovations." *Journal of Political Economy* 66: 419–431.
- Guston, D. (2000). *Between Politics and Science* (Cambridge, UK: Cambridge University Press).
- Harhoff, D., F. Scherer, and K. Vopel (1999). "Citation Frequency and the Value of Patented Innovation." *Review of Economics and Statistics* 81:511–515.

## 2 Paradigms for Evaluating Research, Innovation, and Technology Policies

- Jarmin, R (1999). "Evaluating the Impact of Manufacturing Extension on Productivity Growth." *Journal of Policy Analysis and Management* 18: 99–119.
- Kingsley, G., B. Bozeman, and K. Coker (1996). "Technology Transfer and Absorption: An 'R&D Value-Mapping' Approach to Evaluation." *Research Policy* 25: 967–995.
- Lerner, J. (2000). "The Problematic Venture Capitalist." *Science* 11 Feb., 2000, p. 977.
- Luria, D. (1997). "Toward Lean or Rich? What Performance Benchmarking Tells Us about SME Performance, and Some Implications for Extension Center Services and Mission." Pp. 99–114 in P. Shapira and J. Youtie (eds.), *Manufacturing Modernization: Learning from Evaluation Practices and Results* (Atlanta, GA: Georgia Institute of Technology).
- Luria, D., and E. Wiarda (1996). "Performance Benchmarking and Measuring Program Impacts on Customers: Lessons from the Midwest Manufacturing Technology Center." *Research Policy* 25:233–246.
- Mansfield, E. (1991). "Academic Research and Industrial Innovation." *Research Policy* 20:1–12.
- Melkers, J., and D. Roessner (1997). "Politics and the Political Setting as an Influence on Evaluation Activities: National Research and Technology Policy Programs in the United States and Canada." *Evaluation and Program Planning* 20:57–75.
- Meyer, M. (2000). "Does Science Push Technology? Patents Citing Scientific Literature." *Research Policy* 29:409–434.
- Mowery, D., and N. Rosenberg (1979). "The Influence of Market Demand Upon Innovation: A Critical Review of Some Empirical Studies." *Research Policy* 8:103–153.
- Narin, F. K. Hamilton, and D. Olivastro (1997). "The Increasing Linkage between U.S. Technology Policy and Public Science." *Research Policy* 16:317–330.
- Nathan, R. (1988). *Social Science in Government* (New York: Basic Books).
- National Science Foundation (2000). *GPRA Performance Report FY1999* (Washington, DC: National Science Foundation).
- Oldsman, E. (1996). "Does Manufacturing Extension Matter? An Evaluation of the Industrial Technology Extension Service in New York." *Journal of Policy Analysis and Management* 25: 215–232.
- Radin, B. (2000). "The Government Performance and Results Act and the Tradition of Federal Management Reform: Square Pegs in Round Holes." *Journal of Public Administration and Theory* 10: 111–135.
- Rivlin, A. (1971). *Systematic Thinking for Social Action* (Washington, DC: Brookings Institution).
- Roessner, D. (1989). "Evaluating Government Innovation Programmes: Lessons from the US Experience." *Research Policy* 18:
- Sapolsky, H. (1972). *The Polaris System Development* (Cambridge, MA: Harvard University Press).

- Shapira, P., J. Youtie, and J. Roessner (1996). "Current Practices in the Evaluation of US Industrial Modernization Programs." *Research Policy* 25: 185–214.
- Southern Technology Council (1997). *Benchmarking University-Industry Technology Transfer in the South, 1995-1996*
- Tornatzky, L., and K. Klein (1982). "Innovation Characteristics and Adoption-Implementation: A Meta-Analysis of Findings." *IEEE Transactions on Engineering Management* EM-29: 28–45.
- U.S. Congress, House of Representatives, Departments of Commerce, Justice and State, The Judiciary, and Related Agencies Appropriations Bill, Fiscal Year 2001, 106<sup>th</sup> Congress, 2d Session.
- U.S. Department of Commerce, National Institute of Standards and Technology (1999). *The NIST Manufacturing Extension Partnership: A Network for Success, A Report to Congress*.
- Wallsten, S. (1998). "Rethinking the Small Business Innovation Research Program." Pp. 194–222 in L. Branscomb and J. Keller (eds.), *Investing in Innovation* (Cambridge, MA: MIT Press).
- Webber, D. (1991–92). "The Distribution and Use of Policy Knowledge in the Policy Process." *Knowledge and Policy* 4:6–35.
- Weiss, C. (1979). "The Many Meanings of Research Utilization." *Public Administration Review* 39: 426–431.
- Weiss, C. (1980). "Knowledge Creep and Decision Accretion." *Knowledge: Creation, Diffusion, and Utilization* 1: 381–404.
- Wildavsky, A. (1979). "Strategic Retreat on Objectives: Learning from Failure in American Public Policy." Pp. 41–61 in *Speaking Truth to Power* (Boston, MA: Little, Brown and Company).
- Wolf, C. (1990). *Markets or Governments* (Cambridge, MA: MIT Press).
- Yin, R., and I. Feller (1999). *A Report on the Evaluation of the National Science Foundation's Experimental Program to Stimulate Competitive Research* (Bethesda, MD: COSMOS Corporation).
- 

## Discussion of Irwin Feller's paper

**John Barber** (Department of Trade and Industry, UK): I agree that there is a problem if evaluation documents are not published. If we rely on economic theory to provide a cast iron rule or a rationale for policy actions, we would do nothing. But what happens when market theory implies that something happening in a market should not be happening? When it comes to evaluation, you have incomplete information from a policy point of view. And we have to take action from incompleteness. The real, practical question to ask is: Is there a reasonable expectation this program will be successful?

## *2 Paradigms for Evaluating Research, Innovation, and Technology Policies*

**Irwin Feller:** We need to consider evaluation as a learning modality. And for that, we must have credible theory. For example, a study of small business programs by David Birch has shown that they have created jobs, but the study was shoddy in quality. But this study has been referenced many times in other subsequent papers and articles. We must be willing to change theories when they prove unworkable and seek broader theories. Evaluation is theory building. The agency may want to know if the program works, but for R&D evaluators such studies provide an opportunity to understand complex social and economic environments.

**David Guston** (Rutgers University, USA): The literature on the use of S&T knowledge in policy environments can be characterized as both over- and under-critical. Where does improved evaluation fit?

**Irwin Feller:** I would place it in the competitive nature of the policy process. Some critics will oppose this viewpoint. Evaluators must focus on reducing the number of opportunities to criticize their findings (which is a tactic used by opponents to neutralize the findings). Evaluators should force themselves to meet the standards of peer review.

## Societal Challenges for R&D Evaluation

**Arie Rip**

University of Twente, The Netherlands<sup>1</sup>

Email: a.rip@wmw.utwente.nl

---

### State of the Art

Let me start by recapitulating the history which has shaped RTD evaluation as we know it now. After the Second World War, funding agencies were established providing grants for open-ended research, and a tradition of ex-ante assessment of research proposals was developed. Specifically, on the basis of peer review and procedures to ensure fairness – in dividing the spoils.

The focus was on inputs: feeding the geese of science in the hope they would produce golden eggs (Figure 1). The view of scientific research in those decades is captured by the phrase “Science, The Endless Frontier,” after the title of Vannevar Bush’s 1945 Report to the USA President. Science had to be fed, and ex-post evaluation was not necessary.



**Figure 1: Science as the goose with golden eggs**

---

<sup>1</sup>This version of the paper, as presented at the US-European Workshop on Learning from Science and Technology Policy Evaluation, is not final and bibliographic references are not provided.

## *2 Paradigms for Evaluating Research, Innovation, and Technology Policies*

These times have gone. The American political scientist Don Price, in an article in 1978, emphasized that it was not either ‘Endless Frontier or Bureaucratic Morass’ as scientists themselves tended to think, but politicians and publics holding scientists to their promises (of golden eggs). Promises of contributions to economic growth (competitiveness) and to sustainability now have to be delivered – and that was a reason for an interest in RTD evaluation to emerge (on top of the general notion of accountability when spending public money). How to do such evaluations was not immediately clear. Funding agencies, public universities, and public research institutes had no experience, and actually resisted the first attempts to include societal merit or relevance as a criterion for funding and eventual evaluation.

There was another tradition of research funding and evaluation, however. Contract research and programs with a concrete mission specified the products that were expected, and there would be customers to assess ex-post what was delivered. In the 1960s and 1970s, mission-oriented programs like the Apollo Program to put a Man on the Moon and the so-called War on Cancer in the USA were very visible, although there were also support for distributed networks like that of photovoltaic research in the USA. Within the framework of such programs, a variety of research would be carried out, some of it supported because of its overall promise, some other parts addressing questions derived from the goals to be achieved. Ex-post evaluation was straightforward: check whether goals were achieved (had a man gone to the moon? was cancer mortality reduced?).

From the early 1970s onward, one sees also government R&D stimulation programs aiming at generalized relevance rather than a concrete mission. For example, information and communication science and technology are to be stimulated, not just to make industry become or remain competitive -- the political legitimation of the program --, but also to create capacity and commitment and a collaborative culture. By now, such R&D stimulation programs are seen as attempts to shift the research system in a strategic direction, not just to procure a set of desirable research results. The commitments to new and relevant research directions should remain and be effective after the programme has run its course. In other words, such programs are not just instruments serving particular goals, they have become part of the institutionalized landscape of science.

Ex-post evaluation of stimulation programs was done almost from the beginning, and is now standard. It is reasonably straightforward, even if a judgmental component is necessary. The European Union and some European countries (UK, Nordic countries) have been important in developing methodologies of program evaluation.

Thus, methods and practices have evolved, and have been perfected, in two different traditions. The landscape of science has been changing, however, and the existing methods are not sufficient anymore. The first, and most concrete change is the emergence grants and programs for strategic research, which had to be evaluated.

## 2 Paradigms for Evaluating Research, Innovation, and Technology Policies

What happened was that the robust methodologies developed earlier (peer review of proposals and of ex post evaluation of programs and contracts by expert panels and professional evaluators) were taken up, expanded and modified, to address the new challenges for the evaluation of strategic research. The Table below gives an overview of the present state of the art.

	<b>Grants for open-ended research</b>	<b>Grants and programs for strategic research</b> <i>(new challenge)</i>	<b>Contracts and R&amp;D programs</b> <b>(stimulation, mission)</b>
<b>Ex ante assessment: appraisal</b>	Peer review of proposals <i>(strongly developed)</i>	Peer and (some) user review of scientific and societal quality; spokespersons asked to comment. Priority themes are referred to (cf. foresight)	Prerogative of customer/sponsor (ad-hoc, often in-house appraisal)
<b>Ex post assessment: evaluation</b>	Only through track record criterion used with later proposals	Expert panels estimate strategic value achieved. Exploration of methods to trace uptake and impact.	Expert panels and professional evaluators measure and evaluate output, goal attainment <i>(strongly developed)</i>

There are other changes, in particular the interest in the functioning of national research systems and the emergence of system-level evaluations (of university research, as in UK and the Netherlands, of public research institutes, as in Germany). One can also see signs of stronger links between evaluation and policy: evaluation is not just a check how money was spent (accountability) and whether the goals were achieved, but addresses questions how appropriate the policy or program was and what might be a follow-up. This might evolve further and allow the integration of R&D evaluation and other policy instruments dealing with strategic aspects of S&T, like foresight and technology assessment.

The interest in strategic research, that is, the combination of long-term investment and expected relevance, is one indicator that a new social contract between science and society is being concluded, and a new regime (Strategic Science, rather The Endless Frontier) is emerging. In the definition of Irvine and Martin, strategic research is basic research carried out with the expectation that it will produce a broad base of knowledge likely to form the background to the solution of recognized current or future practical problems. Under the new regime of Strategic Science scientists are allowed to limit themselves to delivering new options, rather than actual innovations. Also, there is less of an exclusive focus on economic values (“wealth crea-

## *2 Paradigms for Evaluating Research, Innovation, and Technology Policies*

tion” as it is phrased in UK science policy), and “quality of life” and “policy relevance” are important reasons to be interested in scientific research.

There are other changes in the social contract of science:

- society is less fatalistic about impacts and risks (cf. molecular biology and genetic modification) and wants some TA done (including ethical aspects);
- and society wants expertise (up to “sound science”) even in the face of large uncertainties.

This has implications for public scrutiny of science, for extended peer review and the involvement of old and new stakeholders more generally. It has also to do with public understanding of science, with new interactions in the risk society, and with continuing trust in science even if specific developments may be criticized.

All these changes have implications for RTD evaluation. Ex-post evaluation had, and has, a strong focus on goal achievement (effectiveness) and management (efficiency). For research performance institutions (university departments, public research institutes), this focus has been adapted by taking the mission of the institution as the reference for evaluation. By now, the community of professional evaluators has developed competencies in policy-oriented and system-level evaluation. Evaluating social impacts is tried (up to attempts to find indicators), but is fraught with difficulties. The recent changes relating to new stakeholders, public debate about scientific developments, and expertise under uncertainty, have not yet been taken up in RTD evaluation, but are seen as a challenge.

The state of the art of RTD evaluation is not just a matter of evaluation methods and approaches. It has specific features related to the character of its object (for example, the open-ended character of RTD, the predominance of skewed distributions of quality and impact, the importance of peer review), and its present state derives from responses to questions and challenges raised by the sponsors of, and audiences for, science, and is located in longer-term changes in the social contract of science.

To discuss the evolving state of the art further, I identify four challenges for RTD evaluation. The first two have been taken up already, and relevant competencies are in place, at least with professional evaluators. For the third, impacts of open-ended R&D, a lot of work has been done, and we understand the problems even while we cannot resolve them. The fourth challenge is how to take new stakeholders and new dimensions into account in RTD evaluation; this is new territory.

### **Challenge 1: R&D evaluation has to address strategic issues**

Ex-post evaluation in general has its roots in accountability (cf. role of General Accounting Offices in evaluation) and in so-called punitive evaluation (an evaluation is called for to justify a decision to close down something). In some policy domains,

## *2 Paradigms for Evaluating Research, Innovation, and Technology Policies*

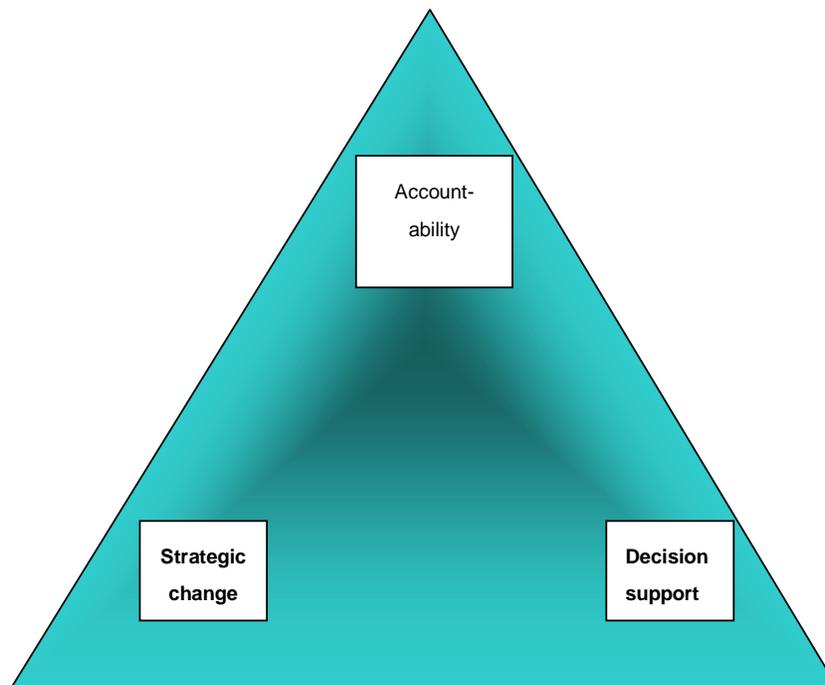
in particular welfare and education, evaluation of programs has become important and is a “business” in its own right. Most handbooks on evaluation derive from the latter tradition.

In the evolution of R&D evaluation there is a similar link with accountability, but also an interest in strategic issues (in the small and in the large). The particular shape R&D evaluations took was also determined by the particularities of the national and international research and innovation systems in which they were embedded. (The theme issue of *Research Evaluation*, April 1995, has descriptions and analysis of different R&D evaluation cultures in European countries.)

Recently, there are two trends: an interest in strategy and learning, and a return of accountability pressure. The former is highlighted in this quote from a report by experienced professional R&D evaluators:

“The major rationale for evaluations has shifted and evolved from a desire to legitimate past actions and demonstrate accountability, to the need to improve understanding and inform future actions. Correspondingly, the issue focus of evaluations has broadened away from a narrow focus on quality, economy, efficiency and effectiveness, and towards a more all encompassing concern with additional issues such as the appropriateness of past actions and a concern with performance improvement and strategy development.” (Final Report of the EU-TSERP Network on Advanced Science and Technology Policy and Planning, June 1999, p. 37) (The last part is argued for by professional evaluators and some policy makers, but not yet widely practiced.)

There is a further distinction: the learning that is made possible through R&D evaluation can be oriented towards the maintenance and improvement of existing systems, or can serve an interest in strategic change. Thus, the functions of RTD evaluations can be located according to a triangle spanning up the three main roles of evaluation. Any concrete evaluation exercise can be characterized by positioning it in the triangle.



**Figure 2: The triangle of types of R&D evaluation**

The triangle is not just a metric. It also presents the tensions and pressures at work around R&D evaluations, and can thus also be used as a diagnostic tool, and stimulate reflection. The corners of the triangle are “ideal” types (as Max Weber used this concept) but examples of the pure cases can be found in practice. They can be characterized as follows:

(1) *accountability* (“what did you do with the money?”), which leads to *audit-type methods*. For RTD, e.g. public institutes, regular audit methods can be used, but the bookkeeping is often extended to include research inputs and research outputs. In the 1990s, accountability itself has taken on a new complexion, with the establishment of Government Performance and Results Acts in a number of countries (the terminology used here is that of the 1993 USA Act). This has created tensions, especially for the basic research institutions; in the USA, the National Science Foundation was allowed to use adapted performance indicators. The existence of such Acts implies that criteria and sometimes measures will be specified explicitly, and become binding (to a certain extent).

In the present political and policy climate, relevance to society will be an important criterion, and there will be pressure to operationalize it. Such pressures occur also without there being an Act, for example in the university research assessment exercise in the Netherlands (not in the UK!): it is a combination of diffuse credibility pressures and strategic action of scientific entrepreneurs and science officials. In

## 2 Paradigms for Evaluating Research, Innovation, and Technology Policies

other words, the possibility of scrutiny drives actors to prepare performance indicators, and to make sure that they actually perform.

(2) *decision support* (ad hoc or on a regular basis), in the small (supporting management of a research institution) or in the large (offering system-wide data). R&D evaluation then uses combinations of quantitative and qualitative methods, and attempts to assess effectivity. Judgement is involved, for example by peers, but this is often transformed into scores.

Simplest example is the decision whether to continue a project, a program or a special center. Without such an explicit stake, there is decision-support type evaluation on a regular basis, say of all publicly funded research institutes, or of all universities. Such evaluations produce information (evidence, “intelligence”) which is then available in the system, and can, but need not be taken up at various levels. In the UK, the Research Assessment Exercise is directly linked with funding decisions (which explains some of its special dynamics, up to the new trade in visible and productive researchers and research groups across universities). In France, the *Conseil National d’Evaluation* produced university evaluation information since 1984; such information is now used in the *contrats d’établissement* negotiated between each university and the Ministry.

(3) *strategic change, or better, input into strategic change and policies*. Actors (policy actors or others) can try to effect strategic changes in the research system, or in the direction RTD is developing, or in the organization of research performing institutions, which leads to an interest in evaluating appropriateness of policy goals and of promises of actual and possible directions of RTD and the RTD system. In addition, there will be an interest in assessing how well embedded the changes are, i.e. whether they will continue without continual policy attention. In the Netherlands, the metaphor of “being anchored” in the research system is used, and evaluators are asked to check the extent of anchoring.

Strategic policies may be predicated on the visions and ideologies of policy makers (as when New Zealand overhauled its research system to make it fully business-like, with universities having to capitalize their assets), on *mimesis* (imitating what is happening elsewhere, because this is seen as a model (USA for Europe) and/or to avoid decision regret), and on a diagnosis of transformations in science and society which one wants to follow, or even lead (that is how the Mode 2 thesis of Gibbons et al. (1994) has been taken up in some countries). The questions, often more like background concerns behind the explicit terms of reference for the evaluation, will be different depending on how the strategic policy is positioned (focusing on effectiveness, on comparison, and on achieving the new state, respectively).

Reflecting on the three types of R&D evaluation, and the idea of a triangular metric to characterize concrete R&D evaluation exercises, one sees two tensions, one ver-

## *2 Paradigms for Evaluating Research, Innovation, and Technology Policies*

tical, the other horizontal. The vertical tension is between control, served by audit-type evaluations for a sponsor or other “principal”, and emergent learning, served by data and judgments made available, and combining ex post and ex ante evaluations. The bottom line of the triangle in Figure 2 has types of R&D evaluation with still some control elements, reflecting the history of R&D evaluation. In general evaluation, there is the proposal of Guba and Lincoln to have fourth generation evaluation, serving a variety of customers and stakeholders with varying perspectives.

The horizontal tension relates to the stance taken toward the changing RTD system in its contexts. When transformation is foregrounded, evaluation will focus on strategic changes (lefthand corner), when productive functioning of the evolving RTD system is the main concern, decision support is the task of evaluation (righthand corner). In most countries, both kinds of R&D evaluation will be visible at the same time depending on the level and the segment. For example, in the 1980s and early 1990s many European and North-American countries had policies to turn around public research institutes to a more entrepreneurial style, while decision support R&D evaluation continued (although with additional performance indicators like the percentage of the total budget earned as external money).

R&D evaluators have to work in concrete situations, and will be exposed to the tensions, and the actual and potential conflicts related to the stakes of the evaluation. Their actions and the output of their evaluation will change the situation. In other words, there are politics involved. The micro-politics of evaluation, relating to the interests of the actors involved and their strategic positioning, are well known, and can be located in the vertical tension. There are also macro-politics of evaluation: how is the evaluation located with respect to overall changes and their perceived value, and how much of the diagnosis of the situation by the evaluator herself can be taken up. This refers to the horizontal tension, and to the role of the evaluator as a modest change agent. It also creates openings for new challenges to RTD evaluation to enter and be taken up, for example the increasing importance to take concerns of new stakeholders into account (interactively and/or through new criteria used in the evaluation). For example, the interest, in Europe, in issues of science and governance, spurred by the recent Communication of the European Commission “Towards a European research area,” implies a reconsideration of the baseline of the triangle.

### **Challenge 2: Improving (national) research systems**

National governments, as providers of funds for R&D, but also from a sense of responsibility for science and its role in society, are interested in the health of the science system, in productive institutional structures and in actual productivity and relevance of science. These concerns lead to various science policies and technol-

## *2 Paradigms for Evaluating Research, Innovation, and Technology Policies*

ogy and innovation policies, and to activities to maintain and sometimes transform the institutional landscape of national RTD systems, with its academic institutions, big public laboratories, RTD stimulation programs, and special research centers. The idea of a national research system works out differently in unitary or in federal systems. In Germany, for example, the universities are the responsibility of the Länder, not of the Federal Government. Note that the recent moves toward internationalization and towards rationalization (up to the concept of a regional innovation system) will introduce shifts in the division of responsibilities.

R&D evaluation can be, and is occasionally being, used to evaluate science and technology policies. There are also examples of system-level evaluations, for example of the whole range of publicly supported research institutes (in Germany) or of a portfolio of R&D stimulation programs (in Finland). In addition, professional evaluators, when evaluating a specific program or institute, will often locate their conclusions and recommendations in a diagnosis of the overall system, to make them more appropriate.

Thus, R&D evaluation can contribute to improving national research systems (or national research and innovation systems), by showing what works and what does not, as well as by directly addressing the challenge of maintaining the health and relevance of the system and possibly improve it.

The need to maintain the system and keep it healthy has a conservative element to it, and there is then a tension with the other need (acutely felt by critics and some policy makers) to modify and improve, and sometimes overhaul, the system. This works out differently in different national systems. In the UK, the 1993 White Paper announced major changes, pushing research institutes and funding agencies to become more strategic and entrepreneurial – in this actually happened. In contrast, as Van der Meulen and Rip have shown, in the Netherlands the characteristics of the system make it difficult for policy makers to introduce strategic change in a top-down manner.

As I noted already when discussing the R&D evaluation triangle, R&D evaluation is located in this tension, and the same approaches will be viewed and applied differently depending on whether they are positioned at the side of maintaining what exists or at the side of trying to change it. Evaluations of the system of university research, and evaluations of public research institutes are examples. The response to increasing pressures for evaluation can be defensive. An example of spokespersons for established science can react is drawn from the Proceedings of a conference in October 1994, where "the world science leaders" met in Jerusalem to discuss strategies for the national support of basic research: "[A]s research resources tighten, science will have to define and fight for its priorities." Assessment of science, however distasteful, had to be taken up in earnest now: "[I]f we do not measure ourselves, somebody else will - "upper management," the government, funding agencies, who-

## *2 Paradigms for Evaluating Research, Innovation, and Technology Policies*

ever - and they will probably do an even worse job of it.” R&D evaluators will then be seen as those who will do an even worse job ...

What concerns me here is not the micro- and macro-politics of R&D evaluation, but the question whether it is possible to evaluate the research system as a whole, or at least introduce partial evaluations of the system into regular R&D evaluation. This requires systematic diagnosis, a combination of judgment and evidence (which is characteristic of all R&D evaluation, I would add).

There is a tradition, established already in the 1960s by OECD and UNESCO (each in their own way) of country reviews of science policy and the functioning of the research system. Panels of experts then look at data and indicators, do interviews, and use their experience and insight to give a diagnosis and recommendations. Implicit in this approach is that there is something like an “ideal” national research system, or at least a limited variety of such systems depending on the “resource base and assets” of the country (to use a term from management studies) and on its (evolving) contexts. In practice, the research system of a country deemed to be successful is often taken as a model, as the embodiment of an ideal system. In Europe, reference to the United States has often played such a role.

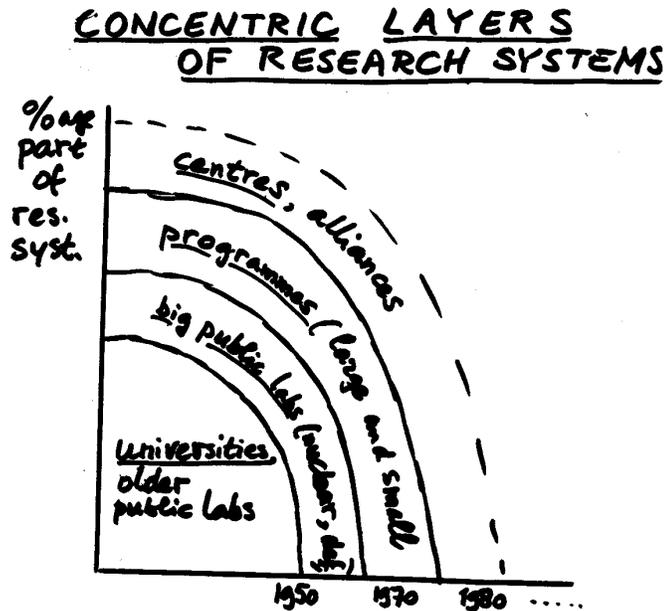
R&D evaluations have contributed insights, especially when there were opportunities for cross-national comparisons. This can take the form of asking experienced R&D evaluators to present a picture of their national research system and critically evaluate it, as was recently done by the French Commissariat du Plan to support the development of their own diagnosis of the French research system. Interestingly, Larédo and Mustar, the editors of the book based on these reports, conclude that the success of the United States had (and has) more to do with its resources and its investment in military research, than with having a research system which is adequate for the 21<sup>st</sup> century.

Another contributory strand is the (fashionable) interest in, and studies of, national (and regional) innovation systems, now also including studies of knowledge flows within the system, and of the learning economy. There can be overlap with R&D evaluation studies, both in data and analysis (for example, effects on research productivity and wealth creation of a particular institutional set-up), and in background diagnosis of the nature of the system and its evolution.

Among the various attempts to meet this challenge to R&D evaluation, I will here single out a diagnostic tool for the institutional landscape of a national research system, and related to that, a tentative assessment of modern and post-modern research systems.

The diagnostic tool was inspired by a figure (in a study by the Fraunhofer Institute for Systems and Innovation Studies in Karlsruhe) which visualised the key compo-

nents of the German research system (Figure 3). Over time, further layers were added to a research system emerging in the late 19<sup>th</sup> century and then consisting of research universities (a new phenomenon) and classical public research institutes (state physical laboratories, national bureaus of standards, geological surveys). The picture of concentric layers may not be completely correct, but is a useful reduction of historical complexity, and turns out to be applicable to other Western countries than Germany.



**Figure 3: Concentric layers in national research systems**

After the Second World War, a range of new public research institutes were established to address the potential of scientific advances, atomic energy being the most striking one. These institutes, and the increased support for research in universities (often guided through government funding agencies), were the two ways in which the regime of Science, The Endless Frontier, was solidified in the institutional landscape of national research systems. As I noted earlier in this paper, smaller and larger programs with a strategic thrust appeared in the 1970s, and became an accepted feature of the institutional landscape. During the 1980s, research centers with special funding (and consortia and alliances functioning as distributed centers) emerged as the “major science policy innovation of the 1980s”, as the OECD called it. These programs and special centers can be seen as institutional differentiations of the sci-

## *2 Paradigms for Evaluating Research, Innovation, and Technology Policies*

ence system to take up the challenge of increasing interaction with society. (No further concentric layer is indicated in Figure 3, even if further differentiations may be expected.)

Why is the diagram in Figure 3 a diagnostic tool for R&D evaluation? It transforms a historical evolution of national research systems into a specification of a “well-founded” research system (just as one can speak of a well-founded laboratory or institute): all these layers must be present (and functional) to have a productive national research system. If one or more is absent, the research system can still function, but cannot adequately take up challenges. In the Netherlands, for example, the move toward specialized research centers has been absent until recently (compared with countries like the UK, Australia and the USA, and also with Germany, where Max-Planck-Institutes and other institutes could play a functionally equivalent role, even in the conservative structure of the German system), and this explains why the widespread interest of scientists in more strategic research has not led to major shifts: the institutional forms to do so were not easily available. Another type of application of the diagnostic tool is in the South-African research system, where for historical reasons the big public laboratories remained dominant, and neither programs nor centers were part of the institutional landscape. Now that the new South-African government introduces science policies to mobilize the research system, the brunt of these policies falls upon the universities and the public research institutes. The latter try to accommodate, some reasonably successfully. The former, the universities, have a difficult time, because there are no protective concentric layers to handle the translation between ongoing research and societal relevance.

The traditional role (and skills) of an R&D evaluator will become broader when using this or other diagnostic tools: she will be a competent commentator and constructive critic, rather than just evaluating past performance.

My use of the diagnostic tool implies that I assume there is an ideal research system, at least the modern research system emerging in the late 19<sup>th</sup> century and being consciously shaped after the Second World War. For countries outside the “charmed circle” of North-West Europe and North America, the transition to this modern research system is an important challenge in itself. The shift away from patronage and the introduction of peer review in the Mediterranean European countries are one example. In less-developed countries, building up a modern research system is an important task. Especially when ambitions to compete globally are present, there is a problem of moving goal-posts. The modern research system that is aspired to is the system in place in North-America and North-West Europe in the 1990s. By the time less-developed countries have emulated this model, the leading countries will have moved on ...

What are the present changes within the “charmed circle”? My earlier discussion of the new regime of Strategic Science can encompass these changes, but did not actu-

## 2 Paradigms for Evaluating Research, Innovation, and Technology Policies

ally specify them sufficiently. One trend is toward interaction between government, industry and academic research, highlighted in the metaphor of a triple helix, as introduced by Leydesdorff and Etzkowitz. Outsourcing of industrial research nearby or inside universities is one example, joint centers between French CNRS and industry (and sometimes *grands organismes*) is another example. The intensive collaboration in South-Africa between a public research institute (CSIR) and a university (University of Pretoria) is a further example.

Another change is the general trend towards heterogeneity in knowledge production, as highlighted in the idea of a Mode 2 of knowledge production introduced by Gibbons et al. (and criticized by other science policy analysts). Rip and Van der Meulen have argued that central control, typical for modern research systems, is not necessary any more for productive functioning, and that Japan and the Netherlands are examples of post-modern research systems, where aggregation processes take over from top-down policy making and implementation. Their claim is interesting, but should be reconsidered in the light of their own admission that national research systems with strong aggregation (as in the Netherlands) can also be conservative (cf. above).

- A third change is the opening up of the boundary between science and society. The role of industrialists in defining agendas and priorities for science is accepted already, but there are new stakeholders starting to play similar roles. Patient associations in medical research, and environmental groups (up to Greenpeace) in a variety of research areas are taking a lead. In parallel (and sometimes overlapping) is the reconsideration of the role of (traditional) scientific expertise. All of this is visible (even if programmatic) in the EU activities to create a European Research Area (starting with the “Communication” prepared by Commissioner Busquin (January 2000), accepted at the Lisbon meeting of April 2000, and elaborated in guidelines “to make a reality of the European Research Area, October 2000). Interestingly, there are concrete proposals to stimulate expertise & competence centres, and European Reference Networks – with the proposed European Food Authority as the first example. This could lead to a further concentric layer in national and European research systems.
- R&D evaluation is in ambiguous position. Its tools and competencies are not geared to the evaluation of research systems and their evolution (partly because the commissioners of R&D evaluation studies did not want input at this level). But professional R&D evaluators (individually and through interaction with colleagues) can build up experience and insight into these issues, and use it occasionally. And such insights (individually and/or available in the professional community) are drawn upon in offering conclusions and recommendations in specialized R&D evaluation studies.
- Just as engineers may have (and further develop) privileged insights in construction (of bridges, of buildings, of machines), which they use to do bet-

## *2 Paradigms for Evaluating Research, Innovation, and Technology Policies*

ter in concrete design jobs, so R&D evaluators will have background understanding of the nature and functioning of (national) research systems. This will colour their conclusions and recommendations (and improve the quality). At the same time, their role is also to help actors in making sensible decisions-in-context, and to do this, they deliver a report with recommendations which might make a difference. The general understanding that has been achieved need not be made visible. For the quality of the work of professional R&D evaluators, however, there must be ways and means to build, enhance, and transfer such general understandings. R&D evaluation training courses, if not too “technical”, are one route to do so.

### **Challenge 3: Expected and unexpected impacts of open-ended R&D**

While the open-ended nature of R&D may well be recognized, the implications should be taken up systematically, both in what is asked of R&D evaluators and in how they address their task. If R&D is not a routine activity with specifiable outputs, one should be reluctant to specify performance indicators, one should emphasize learning rather achieving the goals originally set, and one should be careful (in all senses of the word) in tracing impacts of R&D.

As to performance indicators, Susan Cozzens has set out the problems when discussing the GPRA in the USA:

The major conceptual problem is that with any research activity, results are unpredictable and long-term, placing limitations on the usefulness of the roadmap/milestone approach. As one NSF official puts it, ‘We cannot predict what discoveries are going to be made, let alone say when we are halfway through one’. Annual monitoring indicators are thus quite likely to focus on some less important aspect of research than discoveries and advances. [i.e. content]

For academic research, performance indicators are then often replaced by activity indicators, like the number of articles. (And this then leads to efforts to meet the performance indicator, for example increasing the number of articles by putting fewer results in each article – until the “Least Publishable Unit” is reached.) For applied research, intermediate performance indicators are sometimes used, for instance the very dubious measures based on patents. As Cozzens notes, there are studies showing the economic and strategic importance of R&D in the aggregate, using some such indicators. But the link at the level of projects and programs is uncertain.

[..] [T]o-date, not a single U.S. funding programme has adopted any of them to measure its achievement of programme goals. The reason is that the links

## *2 Paradigms for Evaluating Research, Innovation, and Technology Policies*

between programme activities and economic outcomes are too complex and unpredictable, especially for fundamental research.

A second issue deriving from the open-ended nature of R&D is the occurrence of goal shifts. In the standard view, evaluations should take the original goals as the standard against which to measure goal achievement. While learning will occur, also about the feasibility and desirability of these original goals, and some adaptation is acceptable, one cannot allow any adaptation -- actors could then redefine whatever they did as goal achievement. In the case of R&D, the argument for accepting shifts in goals is stronger, although there must still be some limits. R&D program managers can actively pursue learning, including goal shifts, as a way to do better. Or accept goal shifts after the fact, as happened for example in the UK Alvey Program, where collaborative culture was taken as an important goal when it became clear that the original goal of making British ICT industry competitive could not be achieved. For R&D evaluation it is then important to not only compare outputs and outcomes with a set of goals agreed upon in the beginning of the evaluation study, but to also check the overall strategy of program management. If the program aims to accommodate to what is being learned in the course of doing the research, goal shifts are actually part of the strategy (and one can evaluate how well the learning process was done). An extreme case of accommodation occurs in the open competitive funding programs of funding agencies, where proposal pressure from below (plus selection on quality) determines what are the priority areas of research. In contrast, program management can aim at orchestration, that is, structuring their interaction with researchers and research institutions in such a way that the goals of the program will be achieved, even if in a round-about way. If the original goal is not achieved, this will then be a failure. The learning that occurred must be directed to understanding better how to “manipulate” the relevant actors.

A third issue has to do with impacts. While the question of impacts has become increasingly important, there are no short-circuiting approaches which can reduce complexity so as to allow the measurement of impacts to become a routine business. Even in the case where actors have already reduced the complexity, for example for publications in physical and life sciences (in English-language journals), where there is refereeing of manuscripts, rules for citing previous work, and a Science Citation Index, the evaluation of scientific impact has its difficulties. (An additional problem is the skewed distribution of number of citations to authors.)

Already in the 1980s, Georghiou discussed the ‘impact gap’ in R&D evaluation: desirable impacts are the goal of a program or intervention, but they are difficult to characterize and to capture. The connection between research and effects is diffuse and indirect. This remains the case, even if there is now more effort, from the part of researchers, research institutions and program managers, to realize the desirable effects through dissemination and interaction with potential users and intermediaries.

## *2 Paradigms for Evaluating Research, Innovation, and Technology Policies*

A key point is that eventual effects, or impacts, require other inputs than just the conclusion of an R&D project and a first round of dissemination. Even within science, the uptake of results is an active process, and the user of such results always combines them with results from other projects to create a new combination which is relevant for the user's own work. The general point is that impacts are co-produced, with the research project or program or policy being only one of the contributing factors. This creates a problem for R&D evaluation when certain impacts have to be attributed to earlier research. (There have been attempts to link innovations to earlier, possibly basic research, with little more result than reinforcement of convictions about the relevance or irrelevance of basic research, depending on the boundaries set for the exercise and the causal linkages that were accepted. The difficulty of attribution, and the battle for legitimation involved, is well brought out in Jerry Ravetz's aphorism: "Science takes credit for penicillin, but Society takes the blame for the Bomb.")

The co-production of impacts of a cluster of research can be followed, but this requires a case-study approach to capture the complexity. As well as time: it is not just a matter of transmission and uptake according to a quasi-linear model, but also, and in particular, of relevance to a problem or situation elsewhere. In retrospective studies of impact of R&D projects funded by a Dutch funding agency for technical sciences, where users had been involved with the projects from the beginning, it sometimes took ten years or more for impacts (inclusion in a product or process or service) to be realized.

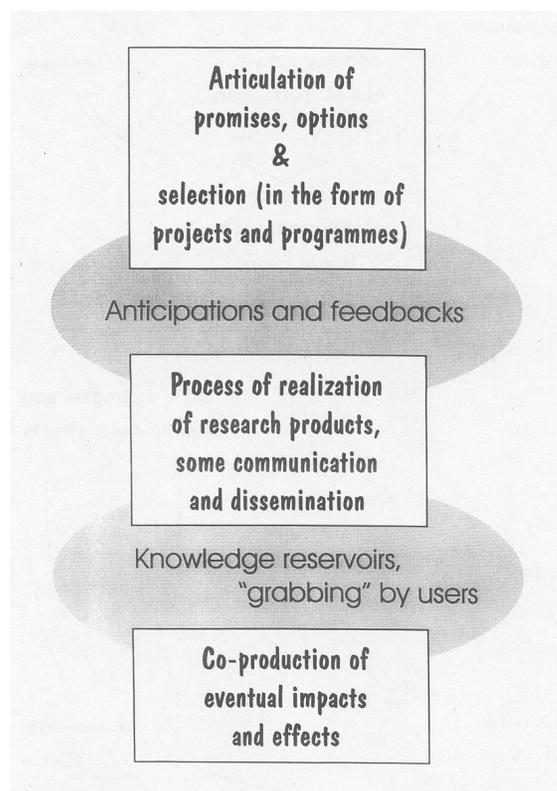
- For R&D evaluators, there is thus a problem: not only must they rely on limited data about impacts (case studies are too expensive), the assessment of impacts is required long before the whole range of possible impacts can materialize. Policy makers and other decision makers want evaluation results to fit their timing and policy cycles. Thus, there is a trade-off: do an early impact assessment to have an input in decision making, but the extent of impact identified will then be less than the project or program will "deserve". For scientific impacts, there are good arguments to use a three-year time window to capture sufficient impacts (because there have been studies showing citation impact to peak two or three years after publication – but this is different in different disciplines!). The situation becomes progressively more difficult for non-academic impact in articulated domains and sectors (medical/health; industrial sectors like chemistry, pharmaceuticals), and in diffuse domains (like social science and policy)

One could draw the conclusion that evaluation of impacts is important but impossible. But our understanding of the linkages between research and impacts has increased, and the policy pressures to work towards impacts have created situations where impact can be traced more easily, and/or more resources are available to trace

## 2 Paradigms for Evaluating Research, Innovation, and Technology Policies

diffuse impacts. In practice, we can handle some the problems, or at least be clear about the limitations of the approaches that are tried out.

Instead of an overview of the literature on impact assessment,<sup>1</sup> I will give an impression of our understanding of links between research and impacts, that is the patterns in the co-production, with the help of a figure showing research adding to reservoirs of knowledge, users fishing into these reservoirs, and impacts being co-produced (Figure 4).



**Figure 4: Chain-linked co-production of impacts**

In concrete domains, the various elements shown in the Figure in general terms can be specified, and data can sometimes be collected for them. In this way, some indications of the nature and extent of co-production and impacts can be given. In the UK, the Economic and Social Research Council has commissioned pilot studies doing just that. The first results imply that one can indeed learn something, even for the diffuse domain of non-academic impact of social sciences, but that it takes quite an effort, also because the connections between research and impact are not always

---

<sup>1</sup> Recently, some overviews have been published, for example in a report by an EU ETAN working party and in an article by Georghiou and Roessner in *Research Policy*, April 2000. See also the special issue of *Research Evaluation*, April 2000.

## *2 Paradigms for Evaluating Research, Innovation, and Technology Policies*

clear to the actors. Regular evaluation of non-academic impact, if possible at all, will be too expensive. But for selected programs or clusters of projects, the effort may be worthwhile.

To round off this brief discussion of impact assessment, I note that R&D evaluation plays a dual role here. The tracing of impacts and identification of links with research allows feedback into policy and decision-making, and provides some accountability in a world where research is expected to produce desirable results. But doing such impact assessments then also has a symbolic function, to show that “we” want to do well in the Age of Strategic Science, and hopefully, have some success stories to tell. The UK Economic and Social Research Council has taken a lead in non-academic impact assessment because it feels pressures to legitimize its funding and be a credible partner in discussion.

R&D evaluators attempting to do impact assessments will not only address a complex issue, but also be exposed to similar pressures, and face choices for example in how to report the patchwork of evidence about impacts. The micro-politics and macro-politics of R&D evaluation become intertwined.

### **Challenge 4: New stakeholders, new dimensions**

There is a fourth cluster of challenges, less well defined, but important already even if the technical evaluation questions have not yet been taken up systematically. This cluster has to do with the opening up of the traditional boundaries between science and society, both in terms of actors who can get involved in science, and the knowledge that is produced and its effects. There is increasing involvement and interference of old and new stakeholders, from industrialists sitting in advisory committees as well as participating in projects, to patient associations not only collecting funds and helping to set priorities in medical research, but also participating in review as “experience-based experts”. While opening up science is seen as important for reasons of democracy, it will be clear that difficult issues about knowledge enter as well as soon “experience-based expertise” is accepted: which experiences and types of argument get a hearing, and which are still out of bounds?.

For R&D evaluation, questions of extent and quality of interactions can probably be handled in terms of the explicit or implicit goals involved. The more difficult challenge is the extent and quality of the new inputs (broader notions of expertise) and the emergence of further dimensions against which R&D is judged by society. The interest in Ethical, Legal and Social Implications of R&D has already led to a new acronym, ELSI. Issues of uncertainty and risk are very much on the policy agenda, and new institutions and networks are set up to provide evidence and judgment, hopefully with sufficient authority. The pace with which this is happening in Europe is particularly striking.

## *2 Paradigms for Evaluating Research, Innovation, and Technology Policies*

In time, evaluators will be asked to do evaluations, of various scope, and they may not be prepared for the new tasks. It is possible to speculate about approaches and competencies required. I suggest that evaluators should focus on what one might call procedural competencies (but with a substantial component). My model is the way professional R&D evaluators now handle peer review in determining scientific quality. They are not in position to judge scientific quality themselves, and cannot reduce the quality assessment to analysis with the help of (often bibliometric) indicators. So they turn to the “peers” – but they themselves have a responsibility for selecting the peers, for structuring and managing the review process, and for estimating the value of the result. Such “orchestration of peer review” is now part of the tool kit of professional R&D evaluators, but it is not a simple task, and requires judgment on their part (up to judgments that the peer review did not go well and that the results should not play a role in the overall evaluation).

For the new issues, one can imagine that there are procedural aspects as well, but based on an understanding of what is happening. For example, to understand the role of new stakeholders one can think of so-called “hybrid forums” and their dynamics. The R&D evaluator can then inquire how far these dynamics have been recognized and made productive (proceduralization of the object of evaluation), and use similar dynamics when setting up a forum to judge societal quality (proceduralization of the evaluation approach).

In this brief discussion of the fourth cluster of challenges, I cannot cover all the complexities. Instead, I will just highlight a few points, concentrating on epistemic and ethical issues.

The quality of the input from old and new stakeholders is an issue in practice, and R&D evaluators will be confronted with it as well, both in what they have to evaluate (their object) and in how they can evaluate (their approach). This runs from difficulties how to accommodate extended peer review, and how to check for quality if traditional scientific indicators are not applicable, to issues of quality of expertise in public arenas, and the tensions between democracy and privileged expertise. R&D evaluators cannot avoid the question of what is “sound science”, and this will lead them into difficult epistemological questions (for which they may not be prepared). An illustration of the difficulties is given in the box.

In 1996, US Representative George E. Brown, Jr. (Democrats), chaired a working party analyzing Congressional hearings and debates on environmental advice. In its report, *Environmental Science Under Siege*, it showed how, in the hearings convened by the Republican-dominated Energy and Environment Subcommittee, “again and again, like a mantra, we heard calls for ‘sound science’ from Members who had little or no experience of what science does and how it progresses.” *Who can be against ‘sound science,’ one might ask. But see what is meant:* Brown and his working party show that ‘sound science’ is equated with direct observations, rather than models and statistical analysis – let alone theory, which is a term of abuse. The epistemic politics involved are clear when subcommittee members are seen to argue that government should intervene on environmental problems only after incontrovertible direct observations confirm the problem’s existence, and Brown comments that such a standard would make it impossible to prevent environmental harms in advance.

The quality of expertise has become contested, and peer review is called for before the evidence and understanding is taken up in regulatory and other decisions. This shifts the contestation to the issue who is admissible as “peer”. Part of the problem, as suggested already by George E. Brown, Jr., is a lack of understanding of how science works. In particular, its open-ended character and the possibility of revising insights on the basis of new data and/or new theory. Within scientific specialties, at the research front as it were, practitioners accept the uncertainty inherent in their work, and can handle the insecurity. The idea then is that these characteristics of science at-the-research-front should be understood and accepted more widely, so that the claims on science would be more realistic (no pressure to have “one-handed” scientists, giving unambiguous advice).

Interestingly, in our modern risk society lay persons, the wider public, appear to accept the uncertainties involved in regulation, and handle them fatalistically. The cartoon (Figure 5) highlights this attitude. It is the world of professionals outside a specialty, and the policy makers needing support for their decisions, which have a low tolerance for uncertainty in what science offers. They need to act, in their own world, and want their inputs to be robust. In the sociological literature, this difference in tolerance of uncertainty has been visualized as “the trough of (tolerance for) uncertainty”, being high at the research front, low for professionals and policy makers (direct users), and moderate (and varying) for indirect users. For R&D evaluation, and my suggestion of partial proceduralization, the recurrent pattern shown by the “trough” allows checks of behavior and quality not in absolute terms, but relative to the world in which the expertise functions.

**EXPERT AND LAY PERCEPTIONS OF TECHNOLOGICAL RISK**



*Give me an overview. Would these do me some good for a while before they're declared dangerous?*

Figure 5.

A similar approach is possible for quality of expertise, when it is recognized that expertise relevant for action has to build on many different inputs, and critically synthesize them. Evolving networks are the carriers (and then also the outcomes) of joint – possibly contested – learning. The notion of ‘learning’ is attracting a lot of interest (cf. the ‘learning’ economy), and the possible bias of thinking only in terms of harmonious working toward shared goals is increasingly recognized. Key points are that joint learning is open-ended (nobody knows the “right” answers yet), that its progress can be traced as increasing articulation and alignment, and that new forms of (distributed) quality control have to be found. The joint-learning perspective does not, by itself, resolve such issues, but it indicates a direction to go. In addition, we should recognize how joint learning (and its quality control) takes place in an existing division of labor (between experts, policy-makers and publics), in force-fields of interests and values, and is refracted by them.

The term ‘distributed intelligence’ refers to the circumstance that relevant information and knowledge is not available as such, but is distributed, across places, across actors, over time. There are local efforts at producing and improving knowledge,

## 2 Paradigms for Evaluating Research, Innovation, and Technology Policies

and additional efforts are necessary to bring these together. Information and communication technologies have increased the possibilities to do this. The phenomenon of distributed intelligence requires synthesis. And synthesis relevant to the purpose at hand. For policy making, Jasanoff's evocative phrase "serviceable truth" comes to mind, but the point is general, and also refers to the reviews and synthetic pieces produced within a discipline. Synthesis of distributed intelligence requires competencies.

If one takes the distribution of intelligence seriously, quality control and assurance become distributed as well. There is no obvious center or epistemic authority anymore. The notion of experience-based expertise is now recognized in medical and health care research, in response to the increasing *political and epistemic power* of patient associations.

R&D evaluation can then focus on the processes through which synthesis (and packaging) is arrived at, and trace efforts at distributed quality control. For example, whether good-faith efforts have been done.

My third and final point is about ELSI, in particular the recent interest in ethical aspects. This has to do with real concerns, like impact of R&D on moral values and norms (think for example of controversies about evolutionary theory and sociobiological approaches), and the ethical discussion that accompanies the development of novel scientific insights and technologies, in particular with respect to possible impacts. It also has to do with the fact that stakeholders raise ethical arguments from different perspectives, and make debates hard to resolve. Decision makers, authorities, other stakeholders all hope that something can be done to make the debates more tractable, and the inclusion of ethical aspects in R&D evaluation is one of the straws to grasp. In the background, there is also the cultural issue that ethics is presented as a counterforce to so-called autonomous science and technology.

- The issue of "co-production" and "co-evolution" returns: we are part of a kaleidoscopic and conflictual social experiment, and notions like 'knowledge society' and 'risk society' indicate increasing self-reflection. Ethics can play a productive role in this self-reflection. The question then is whether (and if yes, how) such issues can be taken up in R&D evaluation. Guba and Lincoln's suggestion of fourth-generation evaluation could, in principle, address the issues of conflicting and contested perspectives, it has remained a programmatic call and no practice has evolved. Similarly for hermeneutic policy analysis and evaluation.

My suggestion would be to shift the focus from the values and perspectives, to emerging divisions of responsibilities for doing things well. Impacts are co-produced, and responsibilities for impacts differ. I quoted already Jerry Ravetz's aphorism: "Science takes credit for penicillin, but Society takes the blame for the Bomb." In general, one can speak of a division of responsibility, in the same way

## 2 Paradigms for Evaluating Research, Innovation, and Technology Policies

we speak of a division of labor in production, making up the social order. The traditional division of responsibility was that scientists and technologists produced new options, industrialists took them up, and society had to take care of impacts, somehow. The 1933 World Fair in Chicago celebrated a century of progress, and took as its motto: “Science Finds – Industry Applies – Man Conforms.”

This division of responsibilities is not accepted anymore, on two counts. First, anticipation of impacts is now expected (witness the rise of technology assessment), regulatory agencies are more pro-active, and actors, including scientists, can be criticized for not being concerned with possible impacts. The action of molecular biologists in the early 1970s to call attention to possible dangers of recombinant-DNA research, and to consider a moratorium on such research, was an important event, and set a precedent. Second, there are the new, and newly recognized, risks of long-term and often uncertain, but possibly large consequences of human activities, for example of low-level exposure to chemicals and radiation. Such new risks had not been adequately addressed, because no one is responsible for them. But they are now recognized, by the public, policy makers, scientists and not to forget, insurance companies, and there are attempts to define new responsibilities.

The precautionary principle is one possible guideline in addressing new risks. One could see it as a macro-ethical stance. Its political, legal and economic aspects have been debated, and there is now some acceptance, at least in Europe. The responsibilities involved are larger than the drafting of regulation, however. It must be linked with the general idea that scientists are responsible for early warning, even if this is based on necessarily speculative theories and models. It is also linked with the recognition of the importance of interaction with old and new stakeholders and various publics.

What we see here is an emerging “constitution” for our technological risk society. This is a *de facto* constitution: it is not laid down by law although elements may find their way into laws and regulations. But it can be forceful in guiding action, just as cultural norms and values in all walks of life are forceful. Appraisal of such an emerging constitution is important, and social and political philosophy can contribute importantly.

In this broader picture, R&D evaluation can be framed as the question how actual and emerging divisions of responsibilities – as part of the social order – work out. If a division of responsibilities has become more or less articulated, or if it is specified as a goal, evaluators can check how well things have actually turned out. If not, they can still identify and assess what is happening in terms of evolving divisions of responsibilities. In both cases, their work helps to increase reflexivity in our late-modern societies.

## *2 Paradigms for Evaluating Research, Innovation, and Technology Policies*

This is a process approach, and does not enter into debates about the importance of one or another value. For some issues, there is general acceptance, for example the requirement of informed consent and the protection of privacy. R&D evaluators can then check whether the rules were followed. Other issues are still open, and sometimes contested. Such questions can also be taken up in an R&D evaluation, but at arms-length. As it were in “boxes”, where actors (directly or as reported by the evaluator) can have their say without their values being endorsed by the evaluator (other than her identifying them as sufficiently important to be mentioned).

The substantial task that the evaluator can address is an appraisal of the evolving division of responsibilities. The present interest in science and governance supports the inclusion of such a task, and identifies important elements like interaction with stakeholders and citizens. Evaluators will then not only be professionals who do a good job, but also intellectuals who are motivated by the possibility to contribute to the good life.

---

### **Discussion of Arie Rip's paper**

**Maryellen Kelley** (National Institute of Standards and Technology, USA): At NIST's Advanced Technology Program, we are looking at programs in an institutional context. What is the program doing in that context? When in theory mode we can forget this. Everything related to a program has a legacy with which the evaluation must contend. If one talks about change in altering these institutions, how does policy do that?

**Arie Rip:** Usually policies don't make changes, except by small increments over time. In Europe, evaluators are more willing to be a participant in the institutional context.

**Louis Tornatsky** (Southern Technology Council, USA): The Belgian centers of excellence are all going to follow a trajectory. Not all of those will meet the optimal level. How do you account for this under the proposed framework?

**Arie Rip:** There is much less variability in quality than you would expect.

## Evaluation of Research and Innovation Policy in Europe – New Policies, New Frameworks?

**Luke Georghiou**

PREST, University of Manchester, UK

Email: Luke.Georghiou@man.ac.uk

---

### Introduction

In a transatlantic forum, the concept of European frameworks for evaluation is an appealing one, but as with so many things European the situation is more complicated, involving collaborative European schemes, national and regional initiatives. There is an evolving framework for the evaluation of research and technological development sponsored by the European Union and frameworks for other European co-operative schemes at the supra-national level. Within these there are some common trends, but essentially they remain distinct. Several more detailed accounts indicate the historical development, current status and future options for these systems of evaluation<sup>1</sup>. Some of the emergent themes are discussed below but probably the most recurrent concern arises from ongoing efforts to improve the level of information and analysis on the socio-economic impacts of collaborative research programs. In practical terms the main effect of this has been a move towards the more systematic collection of information on effects achieved at project level after completion. The supra-national element is generally addressed by the involvement of an independent panel with membership drawn from several countries. The panel has responsibility for interpreting the information collected. This modified peer review structure copes adequately with issues of scientific quality and administration but is generally operating at or beyond the limits of its competence when it comes to the socio-economic dimension.

---

<sup>1</sup> For historical accounts of the evolution of evaluation of the Framework programme see Olds, Beverley J. (1992) *Technological Eur-phoria? An analysis of European Community science and technology programme evaluation reports*, Rotterdam: Beliedsstudies Technologie Economie and Georghiou, Luke (1995) *Assessing the Framework Programmes - a meta-evaluation*, *Evaluation* Vol. 1, No.2, October 1995 pp.171-188, for descriptions and analyses of the current status see Fayl G., Dumont Y., Durieux L., Karatzas I. And O'Sullivan L. (1998) *Evaluation of research and technological development programmes: a tool for policy design*, *Research Evaluation* Vol.7 No.2, August 1998, Georghiou L. (1999) *Socio-economic effects of collaborative R&D – European experiences*, *Journal of Technology Transfer* 24: 69-79 and [Technopolis STOA report], and for a perspective on the future see Airaghi A., Busch N.E., Georghiou L., Kuhlmann S, Ledoux M.J., van Raan A.F.J. and Viana Baptista J. (1999) *Options and Limits for Assessing the Socio-Economic Impact of European RTD Programmes*, ETAN, Commission of the European Communities, January 1999

## 2 Paradigms for Evaluating Research, Innovation, and Technology Policies

The concept of a European framework for evaluation is severely tested when addressing the national level (which accounts for well over 90% of public funding in Europe). A comparative project in the mid-1990s came to the conclusion that evaluation approaches remain intricately bound up with national administrative cultures, with interchange taking place through the medium of EU programs and networks but mainly at a methodological level<sup>2</sup>. Three main groupings emerged, countries with centralized frameworks (notably the UK and France, countries where evaluation was well-established but uncoordinated across ministries and agencies (Germany and the Netherlands) and countries where rigid legislative frameworks for science policy left little room for an evaluative culture to emerge (most of Southern Europe). These broad generalizations disguised significant variation in local practice. For example centralization for the UK meant the promotion of standards and best-practice methods by the Cabinet Office (Prime Minister's office) and Treasury and a focus on programs with testable objectives. By contrast, in France evaluation was under the aegis of two high level committees appointed by the National Assembly and addressed whole institutions. The comparative project did not cover Nordic countries, where a distinctive style involving heavy use of panelists from overseas has long been present<sup>3</sup>. Since that time, Spain has tended to pull away from the other southern countries, with evaluation becoming far more embedded as standard practice.

In view of the lack of a common framework, this paper will instead examine the implications of some trends in science, technology and innovation policy for the practice of evaluation. At a European Union level the changes are still at level of planning and articulation. However, a document published by the Commissioner for Research at the beginning of 2000, *Towards a European Research Area*<sup>4</sup>, has launched a debate which is unlikely to leave the Commission's research activities on the trajectory they have occupied for the past fifteen years; that is to say confined to the two instruments of support of collaborative research and work in the Commission's own laboratories. As well as exploring some of the ideas in this document, some related trends will be discussed:

- Globalisation of science and industry;
- Increasing interdisciplinarity;
- Increased emphasis on economic exploitation of research;
- Increased emphasis on achievement of social goals;
- Emergence of new rationales for innovation policy and new instruments such as foresight;

---

<sup>2</sup> See the papers in the special edition of *Research Evaluation*, April 1995.

<sup>3</sup> Ormala Erkki (1989) Nordic experiences of the evaluation of technological research and development, *Research Policy* 18 (6) pp.343-59, and Luukkonen Terttu.

<sup>4</sup> Commission of the European Communities, Communication from the Commission to the Council, the European Parliament, the Economic and Social Committee and the Committee of the Regions, *Towards a European Research Area*, Brussels, 18 January 2000, COM (2000) 6.

## *2 Paradigms for Evaluating Research, Innovation, and Technology Policies*

- Institutional reform and new public management in science, or the auditors' challenge.

These are discussed in turn.

### **Globalization of science and industry**

There is ample evidence to confirm the growing internationalization, or globalization, of scientific and industrial activity. On the scientific side this is manifested in two ways:

- Growth of international collaboration, as evidenced by co-publication, by the emergence of international collaborative programs and by increased mobility of scientists; and
- Growth of international comparison of scientific performance, as evidenced in the frequent publication of bibliometric comparisons and studies linking science to competitive performance.

Globalization of industry is a large theme which cannot be dealt with adequately here save to note two aspects:

- the linkages between science and industry are no longer necessarily confined within a Nation State, even if that State remains the main source of financial support for science; and
- a growing rationale for support for science is its ability to attract and retain industrial investment, particularly but not exclusively in R&D facilities and high technology.

All of these themes are echoed in the European Research Area document which after making unfavorable comparisons between the collective research performance of Europe and the USA and stressing its economic significance in the context of a knowledge-based society then puts forward related proposals. These include seeking to extend and upgrade the level of collaboration to encompass national programs, promoting mobility of researchers, attracting researchers from the rest of the world.

What then are the implications for the practice of evaluation? At a first level the implications have already been felt in that a substantial number of evaluations now concern international programs, including those with a global scope such as the International Human Frontiers Science Program and the Intelligent Machine Systems program. Such studies have had to bring new items onto the agenda for investigation, notably the costs and benefits of team working in widely different research cultures. The trend to comparative studies has been caught up with the wider fashion for benchmarking studies, which are migrating from the industrial domain to

## 2 Paradigms for Evaluating Research, Innovation, and Technology Policies

public policy<sup>5</sup>. The language of benchmarking, “relevant performance indicators”, “qualitative understanding of best practice”, “monitoring mechanisms”, is close if not the same as the language of evaluation. The broader relation of evaluation to performance indicators is discussed in a later section of this paper. For benchmarking as currently envisaged, the main difference lies in the scope of what is examined, the items for comparison being features of national systems of research and innovation rather than programs or institutions.

The industrial aspect of globalization also has implications for the scope of evaluation. If science has a rationale of attracting links with overseas firms, the focus of measurement turns either to the *returns* available from individual cases (for example flows of license income or investment in the national science base), or else to evaluation of the *capability* of the science base of a country. Evaluation will be more concerned with stocks than flows.

### **Interdisciplinarity**

The growth of interdisciplinarity or multidisciplinarity as the organizational unit of scientific research has been one of the principal phenomena of the past decade. The scope of what is envisaged under these terms has also broadened – at one time interdisciplinarity was usually understood to involve co-working with someone from a neighboring sub-discipline. With the arrival of problem-centered research (as identified by Gibbons et al<sup>6</sup>) thematic organization of research in areas such as ageing has brought together teams ranging from cellular biology through to social science concerns with care and behavior. What might be termed “massive interdisciplinarity” is also present in realms such as e-commerce where technology and content span engineering and the humanities.

Again, what implications for evaluation? Here the main challenge is to the assumptions which underpin peer review. Long-established criticisms of peer review suggest that peers tend to be drawn from the centres of disciplines and mark down activities which fall between them. The existence of this bias is widely believed to be the case by academic researchers in the UK when they are being rated in the national Research Assessment Exercise (RAE)<sup>7</sup>, despite frequent attempts by the agen-

---

<sup>5</sup> See for example the document on CORDIS from the Commission entitled Comparing performance: a proposal to launch a benchmarking exercise on national R&D policies in Europe.

<sup>6</sup> Gibbons et al, *The New Production of Knowledge*, Sage, 1994

<sup>7</sup> The RAE is a quasi-peer review exercise. It adopts the format of peer review in that subject panels read a sample chosen from four publications for each academic entered as their principal means of awarding quality ratings (which determine the selective allocation of block-funding). However, the current state of knowledge means that it is inconceivable that six or seven people, with little or no support from referees, would be competent to assess the entire spectrum of work in disciplines such as Chemistry, Economics or English, to name but three. The exercise is intended to award its highest

## 2 Paradigms for Evaluating Research, Innovation, and Technology Policies

agencies which run the exercise to deny that there is a problem. A recent study by PREST<sup>8</sup> indicates that the problem is most acute where there is in effect a new discipline emerging, for example Development Studies which currently falls between economics and sociology.

The real problem comes when the alternatives are considered. The number of interdisciplinary combinations tends to infinity and even the current operational subset would be well beyond the capabilities of a bureaucratic exercise to accommodate. Should problem-centered work be regarded as the equivalent of a new discipline or should the component parts be disaggregated for separate review? Whatever the answer, peer review exercises are likely to become ever more complex to set up.

Peer review faces further challenges both from the increasingly competitive nature of science (with even international peers now facing implications from decisions giving resources to potential rivals or collaborators) and from a possible threat to the “gold standard” of peer review, the refereed journal article. Changes in publication behaviour were the subject of a survey of editors of the journals most frequently cited in the RAE<sup>9</sup>. While all saw growing importance for electronic publication, the consensus was that peer review was and will be the main function of journals (with scientific communication falling behind). The exception came from editors of leading medical journals who believed that peer review could become divorced from publication, perhaps being the province of learned societies. They saw rapid announcements in specialist journals, with (their own) broader and higher status publications reporting to a wider audience on what was interesting in these. On balance, the “gold standard” looks set to survive but if it did not peer review both in its direct form and in its indirect manifestation of bibliometrics would be severely impaired.

### **Increased emphasis on economic exploitation of research**

The growing use of economic rationales to support increased public funding for research has its natural corollary in the desire for evaluations to ascertain whether the promised benefits are actually being delivered. Successive studies and reviews<sup>10</sup> have drawn two main conclusions:

---

ratings for work of “international quality” but thusfar has neglected to consult any peers outside the United Kingdom (a very limited consultation will take place in the next exercise in 2001).

<sup>8</sup> PREST, Changes in the Research Landscape and the Impacts of the Research Assessment Exercise, 2000, downloadable from the website of the Higher Education Funding Council for England <http://www.hefce.ac.uk/>

<sup>9</sup> PREST *ibid.*

<sup>10</sup> See for example U.S. Congress, Office of Technology Assessment. 1986. *Research Funding as an Investment: Can We Measure the Returns?* Washington, DC: Office of Technology Assessment.; Cozzens, S., Popper, S., Bonomo, J., Koizumi, K., and Flanagan, A. 1994. *Methods for Evaluating*

## 2 Paradigms for Evaluating Research, Innovation, and Technology Policies

- The general case for the importance to the economy of research is strong; but
- Methods for demonstrating the specific contribution of a research program are partial and in general insufficient to overcome fully the barriers of timing, attribution and appropriability, or put another way, to incorporate an interactive rather than linear model of innovation.

A finding which seems to sum up a part of the European experience concerns the issue of additionality, or the difference made by public funding<sup>11</sup>. Traditional approaches (still recurring in economic studies) see additionality as purely an input issue, that is to say how much extra a firm spends on research in relation to the subsidy it receives. Proponents of this approach presume an intimate knowledge of the decision-making processes of the organization. A more fruitful question is that of output additionality, or what has been produced incrementally as a result of the funding. This still tends to focus excessively upon the production of artifacts and encounters a problem discussed in the next paragraph. Empirical work on support for collaborative R&D in Europe has led us to propose a third category, that of behavioral additionality. This term is intended to encompass the situation where the firm reports that it would have done the work without public funding but it would have done it differently, for example less quickly, without the input of collaborators or with a reduced scientific content. In other words, the firm's behavior has been modified. In many cases these effects, notably new linkages (but for small or new firms also the monitoring procedures) persist well beyond the project and arguably can lead to the greatest effects.

A further element in the economic evaluation of research funding arises from the "project fallacy", the fallacy being on the part of policymakers who expect that the contract they have made with a company (or even an academic) will result in a uniquely attributable set of outputs and effects which may then be compared with the inputs. In practice, it is often the case that from the point of view of the sponsored organization, the contracted work is a contribution towards a larger and longer-running project (or raft of projects) from which the eventual innovation will be produced. Hence, the sponsored body may be tempted either to over-attribute effects to the public input, or else to artificially identify a set of deliverables and outcomes to satisfy the sponsor. The implication for evaluators is that it is insufficient to take a project focus, the being instead for a broader picture of the role of the public contribution in the sponsored organization's strategy. On the other hand this raises issues of legitimacy of access for the evaluators.

---

*Fundamental Science*, Washington, DC: RAND/CTI DRU-875/2- CTI.; Georghiou L and Roessner D, Evaluating technology programs: tools and methods, *Research Policy* 29 (2000) 657-678

<sup>11</sup> The issues raised in this paragraph are discussed much more fully in Buisseret, Timothy, Cameron, Hugh and Georghiou, Luke (1995) What difference does it make? Additionality in the public support of R&D in large firms, *International Journal of Technology Management*, 10(4-6) pp.587-600.

A final word on economic evaluation of research concerns the means by which the benefits of scientific research are realized. Increasingly these involve institutional innovations such as the establishment of intermediary companies, incubators for new forms, novel forms of venture capital etc.. Evaluating the economic effects of research involves understanding the effectiveness of these pathways, all of which are far from the traditional linear model perspective.

### **Increased emphasis on achievement of social goals through research**

A trend visible in the Fifth Framework program and also at a national level in Europe is the increased emphasis given to research as a means to achieve social goals. The book sometimes referred to as the ideology of the Fifth Framework Program paraphrases Vannevar Bush in its title: *Society, the Endless Frontier*<sup>12</sup>. By social goals is meant improvements to environment, health and safety, as well as more explicitly societal aims such as reduction of unemployment and social exclusion. An example of the type of issues addressed is shown in Figure 1, an extract from PREST's methodology for the evaluation of finished projects in the Telematics Applications Program in the Fourth Framework Program<sup>13</sup>.

The study in question dealt with a program that mainly addressed the applications of tailored telecommunications services for a public sector audience, with sub-programs in fields such as health, libraries, transport and public administration. It sought to identify economic, organizational and social effects. The latter, in common with economic effects, can take some time to be manifested. To address this, a two-part methodology was developed. First a questionnaire was used to screen all 100 or so completed projects (a very high response rate was achieved in this exercise). On this basis the forty projects with the greatest reported effects were subject to detailed study<sup>14</sup>. It was at this point that an innovative methodology was applied. The assumption was made that the attitude of various levels of public authority would be the critical factor affecting the diffusion and uptake of initially successful project results. To explore this, project participants were taken through four basic scenarios spanning the range of attitudes and then asked to explore with the interviewer the implications of these over a ten to fifteen year period for each dimension of impact (including social). While highly labor intensive, with interviews lasting

---

<sup>12</sup> Caracostas P and Muldur U, (1997), *Society the Endless Frontier – a European Vision or Research and Innovation Policies for the 21<sup>st</sup> Century*, Commission of the European Communities EUR 17655

<sup>13</sup> A series of reports detailing the full results of this study (TAP-ASSESS) are available from a link on PREST's website given at the front of this paper.

<sup>14</sup> In evaluation of research purposive sampling of this kind is more effective than random sampling given the typical concentration of a large proportion of effects in relatively few projects – see Georgiou L, Journal of Technology Transfer op.cit.

**Figure 1 Example of Social Effects Framework for Telematics Program**

Please provide figures on the societal achievements relevant to your project. Provide figures for the expected impact to be achieved within three years. You may use selected measures in the parenthesis.

	Current	Expected within 3 years
Improved service/higher consumer satisfaction (results from service/consumer satisfaction survey, reduction on operating cycles)		
Improved access to services (increased number of people using the service)		
Improved safety (reduction in number of casualties)		
Better informed consumers (number of people with access to information, reduction in errors by misinformation)		
More active citizens participation (increase of participants in specific issues)		
Greater trust in security and reliability of electronic interactions (number of people exchanging information/messages)		
Reduced social exclusion (number of vulnerable people benefiting from application)		
Reduced crime (number of on-line alerts or calls)		
Greater equality between European regions (reduction in regional differences in access)		

Source: Miles I and Barker K, Socio-economic and Industrial Assessment of FPIV Telematics Applications Projects completed between 1996 and 1998 – Guide for face to face interviews, 14 June 1999

several hours, the approach did allow a more dynamic perspective on effects to be achieved on a timescale which delivered results of use to policymakers.

A wider consequence of the trend towards including social issues within the scope of program evaluations is the extension of the relevant group of stakeholders. To the normal research-industry-government triangle is added a wide range of voluntary organizations, including consumer and pressure groups, charities, and regulatory agencies. Many of these will have little familiarity with research and it may be difficult for them to provide useful feedback. It is also the case that research often is a very minor factor by comparison with other policies – this is particularly the case

with employment, where direct effects tend both to be small and to involve creation of new jobs or safeguarding of existing ones. It is rare for an innovation to attributable result in jobs for the unemployed.

### **Emergence of new rationales for innovation policy and new instruments such as foresight**

A conclusion of the 1995 review of evaluation in Europe was that evaluation would need to keep pace with the emergence of new instruments for innovation policy. Most of these instruments implicitly or explicitly are founded upon a more sophisticated rationale for government intervention than the traditional market failure arguments. In particular, they draw upon the notion of system failure, whereby the analysis is that a lack of bridging institutions to link actors within the innovation system creates barriers to innovation. The role of the state is to provide bridging institutions of this type. One of the more prominent policies of this type has been the emergence of foresight programs, particularly in their manifestation in the UK where promotion of networks has over time become the prime goal<sup>15</sup>. The driver for foresight has been the provision of a shared space in which firms who have to innovate in concert with their customers, suppliers, academic collaborators and regulators can develop strategies which reduce uncertainty<sup>16</sup>. Put another way, foresight is wiring-up the national innovation system<sup>17</sup>.

How then should foresight be evaluated? In a recent project in collaboration with Wiseguys Limited, PREST developed a framework for evaluation of the UK program. One of the key design considerations was to make allowance for the fact that the Program relies to a great extent upon volunteers and upon the formation of networks. Both of these require a light touch so as not to disturb the effect being evaluated. One means adopted was to engage the panels which drive the program in the development of both the framework for evaluation and the eventual measures which would be applied. This would build their commitment as stakeholders in the evaluation and help them to understand how it would benefit them. A further consideration was the importance of process issues – a great deal of foresight effort is involved with building a foresight culture and fostering particular ways of thinking. With a narrow base of expertise available and a typically two-year turnover among participants it was essential that the evaluation should provide some means of capturing, codifying and disseminating the knowledge base of the program. For this and other reasons a real-time evaluation was recommended.

---

<sup>15</sup> In the first round of the UK Foresight Programme promotion of networking between science and industry was given equal prominence with setting priorities for science funding but the second round of the Programme is optimised for networking rather than for priority-setting.

<sup>16</sup> Georghiou L, The UK Technology Foresight Programme, Futures, Vol 28, No.4, pp359-377

<sup>17</sup> Martin B and Johnston R....

The hardest task was to reconcile the specific needs of foresight with broader government practice designed for more direct forms of business assistance and still founded upon a market failure rationale. The eventual compromise was to describe the Foresight Program in terms of an indicator framework which covered the main dimensions of performance and intended impacts. The evaluation design team stressed the need to underpin all of these indicators with case studies that would facilitate interpretation of the data.

### **Institutional reform and new public management in science, or the auditors' challenge**

In this concluding section, this paper will discuss some challenges to evaluation arising from broader changes in public sector management. While the most prominent case is the UK, elements of these effects have been evident elsewhere in Europe. The impetus has come from the withdrawal of central government from the direct management of many of its functions, including research. The model has been one of contractualization (with operational parts of government turned into executive agencies or in some case privatized)<sup>18</sup>. There are direct implications for evaluation. At one time a public laboratory in full government ownership was considered to be a unit for institutional evaluation at various levels. The legitimacy for government to do this has been removed as a result of the changes described above. Evaluation must now strictly follow the contracts which government provides for the provision of scientific services. This means that institutional evaluation is replaced by program evaluation. The labs have only to account for what they have done with the resources provided under the contract. However, this takes us back to the project fallacy problem raised above. There may of course be privately commissioned institutional evaluations serving the management interests of the labs themselves but these have yet to emerge.

Declining public legitimacy for evaluation is also becoming an issue for universities. While universities are technically independent organizations in the UK, their past dependence upon block government funding has led them to allow almost unlimited audit of their activities in teaching and research, whether government funded or not. However, some leading institutions have been developing alternative sources of income to the extent that government has become a minority funder. If this trend continues, it is unlikely that these institutions will continue to tolerate accountability requirements beyond those directly associated with funding. Holistic exercises such as the RAE and its teaching analogues may not survive.

A final word on the issue of performance indicators and evaluation. Are these complements to evaluation or inferior substitutes? Georghiou and Roessner argued:

---

<sup>18</sup> Boden R., Gummert P.J., Cox D. and Barker K., Men in Grey Suits....Men in White Coats, *Accounting, Auditing and Accountability Journal*, 1998

“The short-term yet aggregate perspective of GPRA’s performance reviews conflicts with the longer-term, program level, yet context-dependent perspective of technology program evaluations. The fundamental requirement for the design of a performance indicator regime is a clear understanding of context, goals and the relationships which link goals to effects. Whether this important distinction will be recognized and dealt with by government officials and the evaluation community remains to be seen.”

Research and innovation policy is more an exception than a rule in being evaluated by specialists with their own values and methods. The auditors community has advanced from its traditional territory of financial accountability and efficiency and is increasingly interested in outcomes. Reduction of research and innovation activities to a set of easily understood performance indicators has appeal for politicians and other senior administrators. The danger is twofold, firstly that the wrong indicators will be suggested because of lack of understanding of the underlying models, and secondly that the indicators will keep breaking away from their context and will be used to the detriment of the research community. The most significant challenge for the research evaluation community is to demonstrate that we can offer a more accurate and credible alternative which uses performance indicators where they help but is not driven by them.

---

## Discussion of Luke Georghiou's paper

**Terttu Luukkonen** (VTT Technology Group, Finland): The EU approach is embracing diversification of goals and therefore the subjects of evaluation are diversifying and countries need to learn from each other’s approach to evaluation and methods. There is greater professionalization and diversification of methods. Also, the Nordic model is of the past.

**Luke Georghiou:** I accept there has been diversification but there are still a lot of expert judgment-based approaches.

**Irwin Feller** (Pennsylvania State University, USA): There seems to be some tension between productive outputs as a subject for evaluation and the interdisciplinary movement of indicators. Is there a tension between them?

**Liam O’Sullivan** (European Commission): There is more emphasis on the social aspects in the EU than in the US, where economics still predominates. To draw a conclusion, the nature of the research impact is more socioeconomic than economic. When it comes to the methods problem, you can’t apply methods in a general sense,

## *2 Paradigms for Evaluating Research, Innovation, and Technology Policies*

it must be in some context. Is that a fair reflection? Is socio-economics an issue in the US?

**Maryellen Kelley** (National Institute of Standards and Technology, USA): Yes, for the Advanced Technology Program, the central question is what is society getting for its investment in research.

**John Barber** (Department of Trade and Industry, UK): There are a few things a program can be sold on. Despite this, there is a movement to get all evaluation information on one side of a sheet of paper so politician-policymakers can consume it. The question asked by politicians is, what has the program done for my constituents?

## Frameworks for Evaluating Science and Technology Policies in the United States

**Susan Cozzens**

Georgia Institute of Technology, Atlanta, USA

Email: susan.cozzens@pubpolicy.gatech.edu

---

*Editors' Note: The following is a summary of the presentation made by Susan Cozzens.*

Three key components in evaluations are the 1) frames (the scope of the phenomena that is to be analyzed); 2) frameworks (ground rules for making decisions); and 3) framework assumptions (inherent values).

Research and innovation policy should include the regulation of technology areas. Regulation is often not included in the analysis, but if evaluation includes all areas that affect S&T, then regulations also need to be included. Additionally, government procurement of technologies is important to assess in policy evaluation as government spending can help to shape markets. For example, government purchases of recycled paper create a higher demand for recycled goods. In the U.S., the framework is the Constitution and its values, which should be used as the framework in S&T policy evaluation. The goal of S&T policies is often to produce capacity for development, which is often achieved through incentives such as R&D tax credits.

It is important to recognize the difference between policy and program evaluation. Policy evaluation uses more peer review than program evaluation. There has been a shift in the ground rules of the framework for S&T policy evaluation without a parallel shift in the assumptions of the evaluation. The old framework comes from Vannevar Bush's concept that investments in R&D will have benefits in society. The new framework focuses on performance review and outcomes for the public so there is a mismatch between the old and new frameworks' goals.

Do we need to redraw the big picture? Are there any frameworks to talk about? Frameworks need to be categorized in terms of other concepts such as frames (which address the scope), frameworks (that include decision rules), and framework assumptions (which are values based).

## *2 Paradigms for Evaluating Research, Innovation, and Technology Policies*

Are the terms S&T policy and research, innovation, and technology policy synonymous? In addition to research, innovation, and technology, our framework should include these government areas:

- Regulation—e.g., drugs, telecommunications, transportation, food
- Procurement—the purchase of technology for government uses such as defense and health.

Both of these areas could fall under innovation policy, but we don't normally include them when we use the term.

- Government can shape a lot of the technology policies by creating markets and reshaping them.
- Widening frame—the US Constitution dictates a societal role that evolves with a changing society.
- Frameworks—research and innovation policy make investments and provide incentives. We invest in human resources and human capital.
- Major shift—investment in research in creating capacity and not just producing knowledge.
- Incentives—R&D tax credits.

How can these be evaluated?

Policy evaluation and program evaluation—policies that don't take programmatic form:

1. Peer review—this is a sacred cow
2. US Office of Management and Budget (OMB) Circular A-21
3. Bayh-Dole Act (1980)—reworked structure of intellectual property rights from federally-funded research
4. R&D tax credits
5. Patent policy.

Programs [policies in programmatic form]:

This is the heart of research and innovation policy and evaluation. We are really tinkering with the allocation of funds. Very little attention is devoted to defense research as an S&T policy matter despite its large share of federal funds.

Essentially what has happened on the US S&T policy front is a shift in the framework ground rules without a shift in the framework assumptions.

The original flavor of the frameworks and framework assumptions include:

## *2 Paradigms for Evaluating Research, Innovation, and Technology Policies*

- A pluralist federal system
- General Congressional oversight
- Project grant/peer review for individual investigators
- Existence of government labs.

What has happened is a macro shaping of policy with micro autonomy. The old Vannevar Bush framework can be summarized as “autonomy for prosperity.” The new framework includes performance indicators, one-year goals, and annual performance reports. The focus is on outcomes, social goals, and frameworks. But there is an obvious mismatch between the new and old frameworks, including:

- Heavy management at the input side but emphasis on results.
- A strong tradition of autonomy versus a one-year time framework.
- An increase in the legitimacy of stakeholder involvement.
- The making of researchers into strategic inputs with goals for their own research.
- The new frame is being beaten back however. If GPRA loses, we lose a focus on outcomes.

Despite the new and old frameworks, we still really don't know the impacts of S&T policy on society (in a broad sense). The challenge is to develop the tools for examining societal impacts.

---

### **Discussion of Susan Cozzen's presentation**

**David Guston** (Rutgers University, USA): There are reasons why economic measures get greater attention. We created easy indicators that have political economy and a constituency for them. Are there ways to structure a political economy for social indicators?

**Susan Cozzens:** It starts with the political system that may or may not decide to link S&T indicators to social indicators. What no one has figured out is how to link them through political interest.

**John Barber** (Department of Trade and Industry, UK): It is not the initial innovation that has societal impacts but its contribution over say 40-50 years. Benefits are very long-term. Even after 100 years, the impact of electricity was still being felt. Innovation surveys foreshadow that industry itself is what shapes innovation and you can't leave out the technology link because it is too important to the story of research. Science, technology, and innovation—you need all three.

**Susan Cozzens:** However, this “long-term only” benefits view lets us too far off the hook and limits what we look at by the assumptions that are made.

---

## **Open Discussion: Emerging Paradigms for Evaluating Research**

**John Barber** (Department of Trade and Industry, UK): S&T policy evaluation and S&T policy are directly linked. One refers to the other. What do we teach the students? What we teach the students has moved from science policy (1970s-1980s) to technology policy (1980s-1990s) to innovation policy (mid 1990s). Research is important, but innovation can occur without research. If evaluation is reflective of policy, what you are looking for in the program changes over time. Economists have taken up this role because we have moved from basic science policy to a focus on innovation and each has its own ideas of quality.

**Stefan Kuhlmann** (Fraunhofer Institute for Systems and Innovations Research, Germany): Science, research, innovation, the growing interdisciplinarity, strategic research — these all mirror changing perceptions of how innovation comes about. And these are mirroring real changes in the research organization itself. A key concept for today is innovation systems. Innovation systems cover all institutions involved in the production and utilization of new processes and products in one way or another. More and more policy processes are and need to be related in order to better understand innovation between and across organizations. We have to take the innovation systems approach into account

**Terttu Luukkonen** (VTT Technology Group, Finland): This discussion of social benefits brings us to the issue of the distribution of the benefits of science. For example, medical science may be advanced, but it may not benefit a large segment of the population.

**Louis Tornatsky** (Southern Technology Council, USA): I am bothered by the concept of S&T as an abstract thing. We need to keep in mind that science and technology are enacted by people. How do people do science and technology? They carry knowledge from one organization to another.

**Gretchen Jordan** (Sandia National Laboratory, USA): Diffusion is a central issue in the societal benefits of S&T and a lot of that is related to the role of government. So if you are trying to connect social outcomes and S&T, you need to look at the diffusion of innovation.

## *2 Paradigms for Evaluating Research, Innovation, and Technology Policies*

**Philip Shapira** (Georgia Institute of Technology, USA): In Irwin Feller's model, what is it that drives change? Is it only macro-political forces? Or does the evaluation community also learn and evolve and proactively make changes?

**Irwin Feller** (Pennsylvania State University, USA): The accumulation of 30 years of research has affected the way policymakers think and the decisions they make. For example, the average staffer working for Congress has assimilated works like Rogers' diffusion thesis and that is shaping policy. We have convinced policymakers that if you increase R&D expenditures economic growth will increase. Now we have to look at it for real. The ATP program has been proven to work through the evaluations that have been conducted. Yet Congress is convinced that it doesn't have the evidence to support that conclusion. At the same time, the Small Business Innovation Research SBIR program – which is more politically and ideologically popular – is not held to that evidential standard because it is more in accordance with our ideological notions about the economy and the role of government. A lot of attention has been paid to socioeconomics and to persons in the activity of research and to distributive policies. The criteria are quality and contributions to society. Yet, there is little understanding of these two criteria. The scientific community has not bought into this notion. If we in social science cannot figure out the societal implications, then how can natural scientists?

**Luke Georghiou** (University of Manchester, UK): In the social impact evaluation approach, the consumer is the sponsor. The effects of programs can be thought of as coming out of government to the public and back through government for evaluation. Government must serve as the proxy for the consumer.

**Arie Rip** (Twente University, The Netherlands): Indicators are too simplistic. There is a saying that suggests there is no indicator that cannot be undermined by physicists. In the collection of indicators, the selecting of parts of the system has implications for what we can and cannot see. How can we keep an eye on the complexity?

**John Barber** (Department of Trade and Industry, UK): All indicator systems that are to do good must do it quickly, because the indicators will be short-lived and change with the changing landscape otherwise.

**Louis Tornatsky** (Southern Technology Council, USA): Who pays and what is the relationship with what is being evaluated? Better to be a third party.

**Susan Cozzens** (Georgia Institute of Technology, USA): Outcomes indicators must be stimulated by an audience in the public sector in Europe and by the private sector in the US. The recipients of output indicators are different in the EU and the US. Terttu Luukkonen referenced social benefits and Irwin Feller suggested that in cost-benefit analysis costs are easy to capture but benefits are hard to measure. But

## *2 Paradigms for Evaluating Research, Innovation, and Technology Policies*

in a socially-derived view of science, maybe it is the costs that have been harder to measure, whereas the benefits are relatively easy to measure.

### **3. Evaluation of Public Research, Innovation and Technology Policies in the European Union: Roundtable on Trends and Issues**

**Session Chair: Susan Cozzens (Georgia Institute of Technology)**

---

#### **Roundtable Presenters**

##### **Socio Economic Evaluation of the European Framework Program**

Liam O'Sullivan, European Commission.

##### **Evaluation of National Policies – Recent German Experience**

Christian D. Uhlhorn, Federal Ministry of Education and Research, Bonn, Germany.

##### **Understanding the Cultural Logic of EU Science and Technology Programs**

Joseph Cullen, Tavistock Institute, London, UK.

##### **Commentary on the Session**

Nicholas Vornatas, Georgia Washington University, Washington, DC, USA.

---

#### **Socio Economic Evaluation of the European Framework Program**

**Liam O'Sullivan** (European Commission). Until recently, the traditional mode of evaluation of EU RTD programmes has mainly been by peer review which befits the largely scientific objectives of the Framework Programme up to now. The strong focus on scientific objectives meant that a system whereby best-qualified experts pronounce on research quality was adequate to the purpose. There are however, increasingly explicit links between scientific and economic objectives in the formulation of Research and Technological Development (RTD) policy. For its part, the Fifth Framework Programme includes competitiveness in its objectives and has the new socio-economic criteria built into its design while for their part, firms recognise that competitiveness is increasingly dependent on knowledge-generation capacity and less on traditional, cost-based determinants. Such a change in policy emphasis implies a corresponding change in evaluation methodology. In the sense that the Framework Programme has evolved into a multipurpose, multi-actor policy tool, the more horizontal focus of its objectives has meant that less reliance can be placed on peer review alone as an evaluation methodology. The extension of policy aims requires a broader approach involving different disciplines including not only

the ‘hard’ sciences but also the social sciences in the search for more objective tools to deliver better quality evaluation.

---

## **Evaluation of National Policies – Recent German Experience**

**Christian Uhlhorn** (Federal Ministry of Education and Research, Bonn, Germany).

### **I. Recent German experience**

1. Germany developed –over decades - a quite complex, multifaceted scientific system, which reflects the federal political system
2. The German federal structure and the German legal framework (German constitution/”GG”) made the system quite stiff
3. Local interest competes with international challenge
4. On the occasion of (re)unification (1990) it was decided, to let the structure of the German science and research system unchanged (the research institutes of the former GDR have been evaluated and fitted into the system in a very short time)
5. International competition / globalisation exerts more and more pressure on national systems
6. Germany discussed and collected experience with a full set of instruments of policy evaluation, but not all desirable policy actions have been taken
7. It is still an unsettled question which instruments shall be used regularly

### **II. Preliminary note: The subjects are different, so are the ways to tackle the problem:**

- Institutional structures/”landscape”
  1. German science council (Wissenschaftsrat) evaluates systemic questions, e.g. co-operation between universities and “national labs” (Helmholtz-Zentren)
  2. Peer review committees perform “system evaluations” (MPG, DFG, FhG), but also Wissenschaftsrat (HGF)
  3. Fed.Gov. via Hochschulrahmengesetz (legal framework act for universities) and States via Landeshochschulgesetze and administrative procedures induce structural changes at universities, guided by expert advice and endless discussions between various groups of direct interest.
- R&D Institutes, regularly or on specific request
  1. Wissenschaftsrat (WR) evaluates all WGL-Institutes (former “Blue List”), recommends closure, restructuring, shrinking or continued financial support
  2. MPG reformulated own evaluation rules in the nineties, still most important principal of MPG: ex-ante evaluation of the scientific potential of people (leading scientists)

### *3 Public Research, Innovation, and Technology Policies in the EU*

3. FhG uses mainly indicators, e.g. earned money, co-operations.
4. HGF-centres have individual rules of evaluation, originating in the seventies, but BMBF also requested several peer reviews during the last decades. Special scientific policy advice was given for large investment decisions.

- Objectives / research goals

Full set (probably most complete set of instruments/methods) of ex ante evaluations, either in the full responsibility of research organisations or on request of/funded by BMBF. Numerous studies of the type “ex post-evaluation” and also “meta-evaluations.”

- Support / funding instruments

1. SME-support
2. Start-up-support
3. Co-operative research projects

BMBF (now also BMWi) call for tender / kind of system evaluation for programmes

Research audits as a new tool for (research) goal and (funding) instruments evaluation: not yet approved.

### **III. As a starting point for improved policy evaluation, the role government wants to take has to be defined:**

Three Governmental roles in a competitive market system

- “Generous Patron (Maecenas)” / sponsor of science  
i.e. government finances a global budget that can be spent self organized in the full responsibility of the science community. Government asks only for the output product of international competitive excellence. Money could be earmarked for certain disciplines, but then apparently useful results for applications are hoped for. Examples: DFG, MPG, research budget of universities
- “Buyer of – ordered – results” / contract awarding agency  
Government wants results to strengthen competitiveness of important economic sectors (e.g. biotechnology) or for urgent solutions to problems (e.g. HIV)
- Architect of structures or of restructuring  
Government wants to establish best practice (e.g. start-up support, virtual networks of competence), new modes of co-operation between various players (regional innovation networks, inducement of industrial districts), scientific infrastructure (e.g. high speed data links, supercomputing centres, technology transfer agencies)

### *3 Public Research, Innovation, and Technology Policies in the EU*

Which roles governments take depends apparently on the subject of support policy, mixing of roles has a high probability.

#### **IV. Most important questions:**

- Do we support the most important (what are the criteria?) R&D-subjects appropriate and in an earliest possible stage?  
Germany tried very hard in the nineties, to get profound answers to this simple but difficult question
- Can we attract the best researchers world-wide
- Do we support the brightest young people at the earliest?
- Do we have the best infrastructure?
- Do we have the best working conditions for creative R&D-personnel?

#### **V. Learning from evaluation experience of the past decade**

Considerable effort is needed in the design phase of evaluations to answer the following questions:

1. What are the right (detailed!) questions to ask?
2. Are there lessons already learned before?
3. What are the political risks of the possible outcome of the exercise?
4. Who can be trusted, which results are non-debatable, viable, durable?

#### **VI. Suggestion: A report system can be used as a policy tool**

1. Regular reporting creates regular rethinking of policies
2. Raises obligations to formulate policy directions
3. Leads to discussions with parliament and the public/media
4. Manages public awareness

#### **VII. Another road: High ranking policy committees**

Science, research, technology and innovation policy needs regular rethinking of

- What we want
- How we do it

What is the best procedure to get all relevant players involved?

- To foster the motivation
- To start in new directions of research
- To change things and thinking

Germany tried several models

- Committee on “pure basic science” - Members: scientists, Nobel prize winners
- Strategy council - Members: Business (CEO's), scientists

### *3 Public Research, Innovation, and Technology Policies in the EU*

- Technology council- Members: Business (CEO's), unions, scientists (also as representatives of their organisations)
- Alliance for work - Members: business associations, unions

#### **VIII. Résumé**

There are many – partly competing – “national policies” and so are the requirements and difficulties of evaluating them, and more important, to manage a consensus on a better future policy.

Always Einstein's statement should be kept in mind: “Do everything as simple as possible, but not simpler!”

---

## **Understanding the Cultural Logic of EU Technology Programs**

**Joseph Cullen** (Tavistock Institute, London, UK). This presentation, as the title suggests, focuses on two aspects of science and technology policy evaluation: firstly, the distinctive policy and evaluation context of European R&D Programmes, and secondly, the social and cultural ‘construction’ of that environment. The presentation argues the case for an evaluation approach that links outcome-oriented assessment methods to those that aim to unpack the political discourse around ‘science’ and ‘technology’.

The presentation begins with a look at the contrasting meanings of ‘programme’ within the US and the EU. In the American context, Programmes are typically pragmatic. They are situated in a bounded conceptual, methodological and operational space. They are driven by a relatively coherent underlying theoretical and conceptual base. They are hypothesis-focused, and they are outcome-oriented – especially in terms of providing evidence to support or reject their underlying hypotheses. In contrast, EU R&D Programmes typically occupy an evolving and contested space. They encapsulate more explicitly an ideological and political discourse, rather than ‘value-free’ aims and objectives. They are, in practice at any rate, more about ‘learning’ than about directly measurable outcomes – although they frequently emphasise outcomes and impacts.

The current policy focus for EU science and technology innovation can be characterised in terms of a ‘Holy Trinity’ of over-arching objectives. These are: Promoting economic competitiveness; Reducing barriers and promoting mobility for people, goods and services; Combating social exclusion.

### *3 Public Research, Innovation, and Technology Policies in the EU*

These over-arching objectives, and the ‘ideology-focused’ policy environment in which they operate create, it is argued, particular tensions and conflicts that cannot help but shape in part the evaluation agenda of programmes. These tensions are driven by:

- ‘Learning Patrimony’ – the cultural traditions of the different EU member states that have evolved historically to shape understandings about evaluation, and methodological preferences
- Evolving and contested innovation ‘objects’ – S&T goals are hardly ever agreed in common. There are many different variations of technological ‘metaphors’ that shape the intended outputs of Programmes. Programmes (and hence evaluation objects) are therefore constantly ‘moving targets’
- ‘Knowledge Creep’. This transient and evolutionary state in turn engenders ‘knowledge creep’ – the slippage of objectives and goals (and hence evaluation objectives) as the development of the Programme creates new knowledge and ideas.
- Multiple realities – since there is no common position and common set of perceptions about ‘technology’, there can be no one single technological reality. hence, there are ‘multiple realities’ that need to be evaluated.
- Multiple stakeholders – it also follows that there are different constituencies involved in Programmes who will each have a different requirement for evaluation approaches and results.
- Incorporated in ‘Programme Architecture’ – the architecture of the Programme (its project selection procedures; rules’ administration; governance) will reflect these tensions.

As a result, S&T Programmes within the EU environment reflect what might be termed ‘schizophrenic evaluation’. They typically have to be all things to all stakeholders. They call for the application of approaches and methods that can sometimes be seen as diametrically opposed: Formative v Summative evaluation; Normative v output-based evaluation; Industrial v ‘social’ goals; detached v embedded and user-involvement centred; linear v iterative.

Given this complexity, I argue that a ‘Cultural Logic’ analysis can help contribute to the evaluation of such Programmes – even with regard to outcome and impact measurement. Cultural logic analysis derives from the hermeneutic tradition in evaluation, and is influenced by the work of writers such as Habermas and Strydom. It reflects both ‘synchronic’ and ‘diachronic complexity’ – the dynamics of both the processes of science and technology development, and the different cultural ‘life worlds’ in which they operate. The cultural logic of a Programme is expressed in the ‘innovation image’ of technologies that are developed.

As an example of Innovation Image, the presentation refers to the DELTA programme. This was a European Union programme implemented in the mid 1990’s. It

involved 22 pilot projects, and its main objective was to promote economic competitiveness for Europe through applying advanced telecoms and telematics systems to learning technologies. The kinds of 'innovation images' that reflect the cultural logic of the DELTA programme include: the Electronic European Open University; SME's and Learning Technology Centres; The Office Classroom; The Lonely Learner. These 'images' underline a fundamental premise of 'hermeneutic' approaches to the evaluation of S&T programmes: that innovation is always 'Social'. ICTs are socio-technical systems. They frequently integrate stable technologies adapted to new institutional arrangements or new economic enterprise, rather than develop entirely 'new' technical artefacts. They rarely involve 'new forms of learning' – but generally make conventional learning more efficient, accessible.

The social basis of S&T policy, it is argued, strongly suggest the need for a more sophisticated evaluation methodology. This methodology can learn from the hermeneutic and ethnomethodological traditions. It would need to accommodate multiple realities. Its evaluation object is an 'evolving object'. Its target groups are 'moving targets'. Such an evaluation approach would have to be: multi-purpose, including operational, summative, and learning evaluation. Its indicators would combine elements like effectiveness; coherence; social capital and capacity building. The kinds of evaluation methods adopted would include: Discourse & content analysis; Observation; Surveys; focus groups; expert panels; modeling and simulations; participatory evaluation (involving real people). This kind of evaluation stance is very difficult to operationalise in practice. As an example, the presentation briefly summarises the 'SEAHORSE' project. This is another EU project (in the Telematics Applications Programme). It is intended to develop support and information systems for PWHA (people with HIV and AIDS). It adopts a self-help approach to R&D and to evaluation, with a reliance on 'collaborative knowledge production' between experts and users. The evaluation focus is on building evidence base through peer review. The key problems experienced in this type of participative evaluation focus on: access (many user/collaborators do not have access to on-line peer-review systems); language – the language of science and evaluation is typically discordant; motivation and reward- what's in it for them?. The presentation concludes with a call for a paradigm shift in the way S&T policy evaluations are carried out.

---

## **Discussion of Roundtable Presentations**

**Irwin Feller** (Pennsylvania State University, USA). In the Fifth Framework, what is the attention paid to the linkages among the partners over the twenty years? To what extent is there any evidence of the impacts on the national innovation assessment on inter-organizational frameworks?

### *3 Public Research, Innovation, and Technology Policies in the EU*

**Liam O'Sullivan.** The Framework program was designed to be integrated with national policy goals. Additionally, the European Research Area concept was developed to not have contradictions between national and EU goals and to enhance and exploit complementarities between national and European activities.

**Philip Shapira** (Georgia Institute of Technology, USA). Is there a question about whether Germany gets good value for the money it contributes to the EU R&D programs? Can you even address this question?

**Christian Uhlhorn.** One must compare the amount of money flowing back to Germany to the money the German government spends in the same scientific and technological area. In the past, they only looked at the IT sector but for most other areas, the money coming from Brussels is only marginal for the scientific community.

**Liam O'Sullivan.** The finance ministries are more involved than in the past. One of the problems with subsidiaries is that at the national level, a judgment is made about how much money is expected from Brussels and the budget is then adjusted accordingly at the national level. However, this “subtractionality” is not what Brussels had in mind.

**Louis Tornatzky** (Southern Technology Council, USA). What actually is the difference between national and European programmes. Does the EU mandate collaborative projects?

**Liam O'Sullivan.** The most important difference indeed is that there is a notion of international collaboration, which requires participation by at least two nations.

**Terttu Luukkonen** (VTT Technology Group, Finland). We should note that the Framework Programme is not the only transnational collaboration project in Europe.

**Henry Etzkowitz** (State University of New York, USA). In response to the mention of in Liam's presentation of “downward flexibility” of wages as a way of curing unemployment – doesn't that mean lower pay?

**Liam O'Sullivan.** This is not an EU policy program, but more of a theoretical notion coming from neoclassical economic literature. In order for technological change to go through smoothly, there must be increasing demand and decreasing wages over time for an industry. The effect is to create temporary unemployment, which can become permanent if it is not addressed. The telecom industry is a good example. There is no way to set out in a clear way what happens to policy in a theoretical framework. It fits into multiple theoretical models, but not as easily in

### *3 Public Research, Innovation, and Technology Policies in the EU*

Europe as within the US system. There is a theory, but then there is the actual practice.

**Philip Shapira** (Georgia Institute of Technology, USA). I would like to probe what is meant by the concept of “program” in Europe and are programs really different in the US and in Europe, as Joseph Cullen suggested? It may not be accurate to describe US programs simply as “theory driven.” Theory is usually driven by prior practice. Similarly, theory is clearly present in the prior European examples. Also, although evaluation should involve learning, I wonder if it is culturally convenient for European programs to emphasize the terminology of learning rather than evaluation, because formal evaluation can be a difficult and sensitive process.

**Joseph Cullen.** In the EU context, there may be something driving the innovation focus. It may be driven by cultural contexts that may allow more space for evolution than in the US context.

**Arie Rip** (Twente University, The Netherlands). Especially when you talk about methodology, it is realistic that the people who pay for evaluations want a particular report oriented to their sponsor needs. How do you deal with user participation?

**Joseph Cullen.** Use evaluation peers for discussion online to find possible methods. The outcomes are not really understood by the users anyway.

**Liam O'Sullivan.** Theories are not consistent with each other. Neo-classical, microeconomics, and others have some very different approaches. Evaluators need to explicitly state what the theories are to help avoid confusion.

**Peter Blair** (Sigma Xi, USA). We may be using a way too narrow view of the word “program” as the evolved program may have absolutely nothing to do with original concept of the program’s goals. For example, look at the manned space program, which had multiple goals. In the US, we cannot too strictly apply theoretical definitions.

**Luke Georghiou** (University of Manchester, UK). Whatever the theories of the program, the system is still very hard to direct. We could say that the Fifth Programme only has one direction but you can’t really have a range of theories unless you have a range of instruments. In the EU you have only one basic instrument, i.e. transnational co-operation.

**Barry Bozeman** (Georgia Institute of Technology, USA). One of O'Sullivan's slides had economic growth and quality of life as different concepts. If they are not the same things, should we realize that quality of life means different things to different people? What is a useful approach to looking at providing balance?

**Liam O'Sullivan.** Consistency of quality of life and economic growth are both objectives, but they are not the same. What are the limits to economic growth? When does it justify deterioration of quality of life (e.g. environmental, sustainable growth)? The deeper problem of sustainable growth is something that needs to be explored.

---

## **Commentary on the Session**

**Nicholas Vonortas** (George Washington University, USA). I know both the US and European systems from theory and practice. The US system seems to be much more economic-focused which is seen in the quantitative approach taken in evaluations. Proposals reflect this difference. In Europe the methodology is much softer but looks further ahead than in US: in the US, the methodology is more detailed and complex because of the shorter term of project. Ex-ante project appraisal is not usually thought of when people think of an evaluation, but is a useful approach to help people decide where to invest. There has been a convergence of S&T policy and economic policies, which were, previously not together, even though people realize the two are interlinked. Innovation is much bigger than simply technology as it has a social component. However, its evaluation is driven by indicators that may not exactly be correct. Because we are measuring R&D, we are led to believe that technology is the most important part of innovation. Quantitative indicators as for R&D are extremely important and available, but we must be careful to the extent that they can show true innovation.

---

## **4. Public Research, Innovation and Technology Policies in the United States: Federal Policy Developments and Evaluative Approaches**

**Session Chair: Susan Cozzens (Georgia Institute of Technology)**

---

### **Developments in Federal Science and Technology Policies in the United States**

**Christopher T. Hill**

Vice Provost for Research and Professor of Public Policy and Technology  
George Mason University, Fairfax, Virginia, USA  
Email: chrishll@erols.com

---

#### **The Framework**

At any time, many different initiatives are under consideration as candidates as new elements of science and technology policy in the United States. To ascertain which of these are significant, however, requires an examination of these initiatives in light of a model or framework that addresses how public policy for science and technology is made and how such policy is affected by the larger social context and by the factors that tend to drive policy change. I begin with consideration of one such framework for analysis. This framework depends on a simple set of assertions: that science and technology responds largely to public anxieties arising outside the realm of science and technology per se. New scientific understanding and new technological applications, in turn, create new anxieties that call for further policy responses.

Direct examination of the history of science and technology policy in the United States in the twentieth century illustrates that public support for scientific research and technology development has been driven by and contingent on perceptions that the results of R&D are directly useful in addressing important national problems. This perspective has been particularly evident since the successful mobilization of the nation's scientific and technical resources during the Second World War. Funding to support R&D and the education of scientists and engineers in diverse fields has tended to grow in response to a series of national challenges, such as the Cold War threat from the Soviet Union, the launch of Sputnik, the environmental and energy crises,

#### *4 Public Research, Innovation and Technology Policies in the US*

and the challenges of cancer and AIDS.<sup>1</sup> When the nation is worried, science tends to thrive--and usually with little or no regard to systems of accountability and external scrutiny of achievement.

This perspective on public and political support for science and technology tends to discomfit the scientists and engineers themselves, who much prefer to see themselves as offering to society a cornucopia of new opportunities to improve the human condition--science truly offers, in this view, an "endless frontier" of challenge and opportunity. From this perspective, research and development are benign and benevolent activities that represent some of the best of human endeavors. Viewed in this way, science and technology deserve unreserved societal support, unconstrained by rules that seek to preserve the status quo and unfettered even by venal considerations of costs and benefits. The grand project of science and technology development should be held accountable only in the final analysis of whether society is better off, not whether specific research programs and projects pay their way.

The same history shows that American society has not viewed scientific research and technology development only as benign and benevolent activities. Society has recognized that the power to remake the world as we would want it also embodies the power to make a world quite different from that most of us would prefer. The mad scientist bent on controlling or destroying the world has been a staple of popular culture for a century, sharing the shelves with literatures, both utopian and dystopian, projecting worlds based on science far different from anything we have ever known or would want to live in. Technological advances like nuclear fission, genetic engineering, and antibacterial drugs have also threatened nuclear holocaust, new ways to discriminate against individual and groups based on their susceptibility to genetic diseases, and the risk of widespread bacterial resistance and the recurrence of endemic infections beyond our control. The progress of science--and increasingly its conduct--and the applications of new technology are deemed too important not to be brought under societal control and to be made accountable to our political processes and, through them, to our values and preferences.

In a nutshell, the views of the American public toward science and technology might best be characterized by quoting former President Ronald Reagan in a different but equally significant context--when it comes to science, "trust, but verify."

---

<sup>1</sup> For an elaboration of this concept of the dynamics of support for science and technology, see (Press report on allocating federal funds). Also see (Hill background paper for the Press report)

### **Tensions Between the Public and the Experts**

The tensions inherent in these divergent perspectives on the role of science and technology in society are apparent throughout the federal policy making process, the federal scientific and technological institutions, and the patterns of funding for scientific and technological endeavors.

For example, policymakers tend to prefer to support R&D through mission agencies established to address identifiable problems and hope to see results flowing from those R&D programs that directly address and solve the problems driving the missions. Substantial funds flow to Department of Defense R&D to provide for a strong national defense and biomedical research in the National Institutes of Health is increasingly well-funded because advances in that arena promise to diagnose and cure a wide range of human diseases. The scientific and technical communities, on the other hand, prefer to have R&D supported through agencies, such as the National Science Foundation and the Department of Energy's fundamental research programs, which select fields and projects to support based heavily on their promises to add to the storehouse of human knowledge rather than on their promises to solve particular problems.

Likewise, working scientists and engineers chaff under societally-imposed limits on the resources available to them and under rules intended to constrain inquiry and application in the interest of preserving cultural norms and directing technology away from domains deemed undesirable by influential elements of society. In the 1960s and 1970s, the public's growing demands to rein in the environmental, health and safety consequences of technological applications led to a host of new federal regulations in these areas and to a predictable resistance to their adoption by many in the scientific and technical worlds who feared limitations on their charter to study, invent and apply new ideas. More recently, the public's concern has turned to such matters as opposition to certain uses of the Internet and to the uncontrolled application of genetic engineering to treat disease and to modify foodstuffs. Once again, the possibility of socially-imposed controls on such activity has highlighted strong differences of perspective between the general public and the scientists and engineers most involved.

### **The Policy Framework and Recent Accountability Expectations**

Within this framework of analysis of science and technology policymaking in the United States, formal demands such as those under the Government Performance and Results Act (GPRA) can be viewed as adding to a pre-existing accountability system, rather than as bringing a formerly unaccountable system under control for the first time.

#### *4 Public Research, Innovation and Technology Policies in the US*

For mission-oriented R&D programs, for example, expectations have usually been high, oversight intense, and results vigorously presented and critiqued. The essential role of independent scientific and technical judgment in conducting such assessments has been realized and adopted through such institutions as peer review, open publication, and the creation of elaborate social mechanisms for structuring such judgment via the National Research Council, the scientific and engineering societies, and federal science advisory boards. These mechanisms are not perfect by any means, yet they reflect serious commitments to trying to ensure that mission R&D programs fulfill their promises. These mechanisms serve the collateral purpose of ensuring that narrow conceptions of how to plan and carry out complex technical endeavors do not so encumber them as to guarantee mediocre results and modest accomplishments. In addition, congressional oversight of progress toward objectives and of the processes under which programs are implemented provides some assurance that society's expectations are being accomplished and costs are reasonable.

For more fundamental inquiry, the institutions described above are not as effective as they are for mission oriented science. Basic research investments by society are still premised heavily on the open-ended promise that "science will yield good things for us." More focused social objectives of fundamental scientific inquiry are often not clearly stated, or state-able, in advance, and, of course, without clear objectives no accountability system can work well. Independent scientific and technical judgment is also of limited value, for, when the pursuit of the "unknown" is the activity, even those "in the know" have only limited insight into whether a program might, or has, worked. Congressional oversight of fundamental research can also be of only limited effectiveness. Members and staff have little capacity for determining whether the promises made to them about discoveries yet unmade are reasonable or accurate. As a result, such oversight as occurs tends to focus on marginal matters such as personal venality in program management, while most "oversight and investigation" over the years has amounted instead to celebrations of the prior accomplishments of those under "investigation."

The use of scientific and technical knowledge in the making and executing of more general public policy poses a different set of challenges to new simplified accountability frameworks such as GPRA. When scientific and technical knowledge are used in, for example, setting occupational health standards or deciding whether to invest in a new weapons system, the main challenge to society is to determine the state of scientific and technical understanding, and the nature of its attendant uncertainties. In such cases, scientific and technical truth is rarely dispositive; that is, whatever “facts” are agreed upon by all sides to a policy debate are rarely sufficient as a guide to action. Instead, various parties to the policy debate are able to marshal a reasonable set of “facts” from among the available scientific and technical evidence to support the actions and outcomes they prefer for other reasons. In this light, for example, the task of the lobbyist is not to distort scientific “truth” but to exploit the scientific uncertainties on behalf of the client’s position. When the issue under consideration is a policy matter, it is ultimately up to policymakers to make judgments about where the scientific evidence is the stronger and to incorporate that view of the evidence in making a decision.<sup>2</sup> If the policy decision has been delegated to a regulatory authority, as so often occurs, the general framework of the Administrative Procedure Act provides for a stylized set of mechanisms that enables all sides to present evidence as to the state of the science, that requires the policy maker to take that evidence into account, and that opens the process to challenge in the courts to help ensure that decisions are based on the evidence that is presented. If the issue under consideration is an enforcement or implementation decision pursuant to a policy decision, direct recourse to the courts is available as a mechanism to ensure that scientific understanding has been reasonably used.

Thus, it should be apparent that diverse accountability structures have been in place for some time to try to ensure that policymaking in regard to the support of science and technology and to the use of scientific and technical information is accountable to the public through the legislature, the administrative agencies, the courts, and independent experts in science and technology itself. From this perspective, it seems to me, the challenge to GPRA advocates is to demonstrate how the formalized, accounting-based systems they advocate can make a contribution to improved decision making that is commensurate with the effort required and the resources consumed in the process.

### **Some Notable Developments in U.S. Science and Technology Policy**

This part of this paper addresses recent developments in U.S. federal science and technology policies that support of science and technology and that are intended to control how science is done and technology is used. It also addresses practices af-

---

<sup>2</sup> One of the key contributions of the former Office of Technology Assessment was to assist congressional decision makers in balancing the competing claims of interest groups as to the nature of current knowledge and expected future developments.

#### *4 Public Research, Innovation and Technology Policies in the US*

fecting how scientific and technical information are used in policymaking. Finally, it addresses the apparent role of science and technology in the 2000 presidential election campaigns.

##### ***Policies to Support Science and Technology***

The dominant mechanism used to support the scientific and technical enterprise is funding for research and development. In this domain, the usual policy question is, “who gets how much to do what?” Here is where our national anxieties are most apparent--R&D funding priorities are heavily influenced by perceptions of the nation’s most pressing problems and of the potential of R&D to help solve them. During the past two decades, attention in the United States has shifted from the energy crisis, to the “Evil Empire,” to the competitiveness challenge, to concerns for improved health and higher quality of life. R&D funding has shifted accordingly. At present, funding for defense R&D is in slow decline while funding for biomedical R&D is exploding. Almost no attention is being paid to funding to address energy supply and demand, while the strength of the U.S. economy during the past nine years has essentially driven concern for competitiveness into the “dustbin of history.” These trends show no sign of abating, regardless of the outcome of the presidential campaign.

The growth in biomedical research funding has been phenomenal during the past several years. A series of high-profile instances of various diseases and conditions, such as Alzheimers, spinal cord injury, prostate cancer, and AIDS seems to drive ever higher spending for health research. The two Houses of Congress, and the Administration, seem to vie with each other for “leadership” in addressing health issues by proposing ever-higher budgets for such work.

While the source of America’s economic strength during the past decade remains elusive, it is certainly manifest as a sustained growth in productivity. And, while explaining productivity growth is nearly as challenging as explaining economic growth, there seems to be widespread agreement that some combination of industrial restructuring, new approaches to management of business and industry, and the adoption and use of information technology in its several guises has been most responsible. It is also widely understood among policy makers that the federal government played an essential role in the development of the Internet and related information technologies, and that this development owed little to organized efforts to find new technologies to enhance competitiveness. This understanding, coupled with the robust performance of the economy, has largely driven policy initiatives to support industrial technology off the policy agenda.

An important development whose full significance is not fully realized, in my view, is a change in the long-established consensus about the role of the National Science Foundation in supporting R&D. Since the late 1950s, the “central function” of NSF

has been to support research in the core disciplines underlying twentieth century modernization, including such fields as mathematics, physics, chemistry, earth sciences, economics, and--reluctantly for many years--engineering. Increasingly, however, and to an extent that seems much more pervasive than in early experiments, NSF has taken upon itself the role of encouraging multidisciplinary and systems approaches to complex scientific and technical problems. NSF announces various high-profile initiatives of short (two or three year) duration, such as the KDI program, that receive intense response from the R&D community in the form of ideas and proposals. Sometimes these initiatives seem only vaguely defined, with the portfolios of funded projects seeming less revolutionary than the NSF rhetoric that called them forth. By the time the performing community has had an opportunity to ascertain what NSF “really had in mind,” the initiative has come to an end and a new one is on the way. This practice seems to pay less direct attention to the core disciplines than in the past and may be encouraging the pursuit of the trendy program of the day at the expense of sustaining the nation’s academic research infrastructure. This trend is particularly disturbing in view of the fact that support for several of the core disciplines by various mission agencies is also in decline.<sup>3</sup> In what may be an important harbinger of change in this area, leaders of the mission agencies such as NIH have recently called for more investment in a wider array of fields of science and engineering than those funded by the mission agencies themselves.

The national enthusiasm for collaborative forms of research and development shows no sign of abating. If anything, corporate alliances, public-private partnerships, and university-industry collaborations are now the norm, not the exception. While some complain that industry partnerships have compromised the traditional mission of the university, or that national security may be compromised by federal laboratory agreements with companies, or that industry is profiting overly much from exploiting biomedical inventions made at universities with federal support, nevertheless the collaborative approach to research and especially to technology development is the approach of choice.

(As a personal observation, the growth of collaborative research has greatly complicated the life of research administrators and “chief technical officers” of modern universities. We now spend much of our time on collaborative matters including recruiting partners, drafting and negotiating partnership agreements, working through conflicts about intellectual property ownership and benefit, and attending to the conflicts of interest that so often arise. These activities have undoubtedly increased the administrative part of the overhead, or “facilities and administration,”

---

<sup>3</sup> See the analyses by the Academy’s Committee on Science, Engineering and Public Policy and by Merrill and McGeary.

#### *4 Public Research, Innovation and Technology Policies in the US*

costs that universities have been allowed to claim on federal contracts during a time when such costs have been artificially capped at a rate of 26 percent. In addition, many industrial partners refuse to accept as legitimate costs of doing business the federally negotiated overhead rates that universities seek to charge.)

Another important development in federal support of research is the increasing use of congressionally directed spending, also known as “earmarks” or “pork,” to support specific projects at specific institutions. On the one hand, this process is seen as subverting the orderly administration of agency programs and as bypassing the peer and merit review processes they use to ensure quality. On the other hand, it can also be seen as an acknowledgment by political leaders that such projects are worthy of the expenditure of the political “capital” that is necessary to incorporate them into appropriations acts of Congress. Earmarking may also serve as a signal to Congress that new agency programs should be established to respond to the needs evidenced in the earmarking process. For example, some argue that the Technology Reinvestment Project adopted in the early Clinton administration was one approach to trying to channel the demand for earmarked defense reversion projects through a more orderly process of consideration. And, earmarked science projects at universities might be reasonably viewed as a substitute for the practice adopted in earlier decades of establishing a plethora of permanent mission-oriented federal laboratories in particular congressional districts and states.

Despite the manifest importance of information and computational sciences and technology to the current economic boom in the United States, only limited efforts have been made to expand federal research and development programs in related areas such as electrical and computer engineering, computational sciences, and information and computer security. This relative inaction reflects in part the view, mistaken in my opinion, that the information revolution has proceeded in spite of, not because of, government support for new technology. It also reflects a widely held view that further research and development efforts are not needed to spur the information economy; in fact, many noteworthy “IT” firms do little or no research at all. Instead, their business is based on design and implementation of new business models or on installing new systems incorporating standard technological elements in new ways. In addition, major new R&D programs, like most new spending programs, have not been popular in Congress throughout the Clinton administration.

On the whole, the outlook for federal financial support for R&D and related education is for continued slow growth in the total, a continuing shift toward biomedical and health related research, and modest increases in support for interdisciplinary and problem-oriented collaborative research activities.

***Policies to Address the Conduct of Science and the Application of Technology***

Policymakers are showing a modestly renewed interest in policies to address the directions in which new technology is going, as well as in guiding the conduct of inquiry itself.

For example, there is growing concern over how the promise of financial rewards to institutions and to individual academic inventors may be creating undue incentives for researchers to abuse the rights of persons who are subjects of their research. These concerns are focused on, but not limited to, biomedical research and clinical trials of new drugs and medical procedures. They also are addressed to concerns about unethical treatment of research subjects by social scientists and by fundamental scientists and engineers doing research on genomics using person identifiable data. Recently, the oversight of such research by sponsoring federal agencies was elevated in status in the Department of Health and Human Services, and the new director of the cognizant office has made public statements suggesting a new rigor in oversight of such research. As a consequence of several widely-reported temporary closures of research programs involving human subjects at universities by the Office for the Protection from Research Risks, new scrutiny is being given to such research by institutions as well as by funding agencies, and new requirements are being imposed on the processes used by local institutional review boards and for the training of such boards as well as individual investigators in the principles of human subjects protection. Furthermore, there is some reason to expect that financial conflicts of interest in conducting and exploiting federally funded research may get additional scrutiny in the near future.

The question of assuring the accuracy and objectivity of scientific and technical information continues to be a matter of considerable debate and concern. Two years ago, for example, Congress passed a new law applying the provisions of the Freedom of Information Act to research data obtained using federal funds. The act would make such data that is in the hands of the performing institution subject to a FOIA request through the sponsoring agency. The Office of Management and Budget ultimately adopted procedural rules pursuant to the act that would appear to limit the reach of FOIA-based access to research data to cases in which the data are subsequently used to make federal regulatory policy. It is not unreasonable to expect a court challenge to the OMB rules by the same sorts of interests that sought the new law originally, and OMB's position could well be found to be overly limiting of the access rights extended by the law. While the particular motivations for the adoption of this new legislation are the subject of some debate, it can be viewed in the larger context of efforts by society to ensure that alleged scientific facts used in public policymaking are of high quality; i.e., that so-called "junk science" not be used to make public decisions.

#### *4 Public Research, Innovation and Technology Policies in the US*

The significance and quality of scientific information and its utility for decision making has been called into question in a very different way by the case of the former Los Alamos National Laboratory scientist, Wen Ho Lee. While the facts of the allegations made against him have not yet been tested in a court of law, it is claimed by the government that Dr. Lee downloaded classified information to his personal computer that could have been used to design and construct a nuclear weapon. On the other hand, competent scientists experienced in weapons design have maintained the information he downloaded would not give another country the information to achieve this task. It will be interesting to see how the courts address the implied question of how much does one need to know to build a nuclear weapon and how much of that knowledge has been or can be captured in a computer program.

#### **Science and Technology Policy in the 2000 Presidential Campaign**

The platforms of the two major candidates, Al Gore and George W. Bush, include statements concerning their views on selected science and technology issues. Their positions are published on the Web.<sup>4</sup>

Setting aside the substance of their positions, it is interesting that neither candidate has a statement of his position on “science and technology (policy)” as such. Instead, Gore’s is covered under the rubric, “Internet and Technology,” and Bush’s is covered under the rubric, “Old Truths for the New Economy.” (Actually, the Bush statement is found on a Republican National Committee site.) However, it is also significant that both candidates do include some aspects of support for science and technology in their issue statements. In prior years’ elections, advocates for science and technology have often been in the position of pressing such concerns on candidates who have been indifferent-to-reluctant to incorporate these topics in their issues. As a general observation, the Gore position is considerably more detailed than the statement of the Bush/Republican National Committee position.<sup>5</sup>

Gore’s statement covers a wide range of issues, from funding fundamental research at universities to urging the World Trade Organization to make “cyberspace” a “duty-free zone.” It addresses several detailed policy issues under each of the following topics:

- Creating the High-Tech, High-Wage Jobs of the 21<sup>st</sup> Century
- Closing the Digital Divide and Creating Digital Opportunity

---

<sup>4</sup> For Al Gore’s position, see: [http://www.algore.com/internet\\_and\\_technology/](http://www.algore.com/internet_and_technology/). For George W. Bush’s, see: <http://www.rnc.org/2000/2000platform2>. Accessed Sept. 1, 2000.

<sup>5</sup> In the final paper, it may prove illustrative to include a detailed “side-by-side” comparison of the two candidates’ positions. On the other hand, such a comparison will be moot by the time the final paper appears in print.

#### *4 Public Research, Innovation and Technology Policies in the US*

- Protecting our Values in a Technological Era
- Development of Broadband Access and Next Generation Internet
- Creating an E-Government for the 21<sup>st</sup> Century
- Promoting an “E-Society”
- Investing in Science and Technology for our Future.

Bush’s statement is more constrained and is incorporated in a broader statement about economic policy. Portions somewhat comparable to the Gore agenda appear as detailed issues under the following topics:

- Trade: The Force of Economic Freedom
- Technology and the New Economy: The Force for Change
- Privacy and Secure Technologies.

#### **Concluding Remarks**

In this brief paper, I have attempted to identify some of the more visible and important directions in U.S. federal science and technology policy at the opening of the new millennium. I argue that the policy agenda is established largely in response to public anxieties about important threats to national well-being. Because the current times are perceived as relatively benign, the forces driving major new federal investments in science and technology are attenuated. As a result, except in the field of biomedical research, agency and program budgets are not growing and in some cases are even shrinking substantially. Apparently, federally supported R&D does not thrive in good times.

However, the major candidates for President of the United States have disseminated detailed positions concerning support for science, the encouragement of new technology, and the need to ensure that citizens are protected from some of the undesirable consequences of the use of new technologies, especially of the Internet. That they are doing so reflects, in part, the growing acceptance that science and technology policy are now central aspects of public policy making, in a way that they have not previously been. It may also reflect the vigorous pursuit of political and financial support from the “high-tech” entrepreneurial corporate communities across the nation who would expect to see such positions.

A final point of this paper is the observation, made at several places throughout, that accountability systems have long been in place to try to ensure that mission oriented R&D programs accomplish their objectives and that fundamental research programs are well managed and directed. It is not apparent that more formal and mechanistic systems based, for example, on GPRA tools, can provide for improved accountability of federal R&D programs or that they can avoid creating undesirable barriers to effectiveness of those programs.

## **Factors Affecting Technology Transfer in Industry—Federal Laboratory Partnerships**

**James S. Dietz**

School of Public Policy, Georgia Institute of Technology, USA

Email: gt7631a@prism.gatech.edu

---

### **Introduction**

Companies work with federal laboratories in cooperative arrangements for any number of reasons—some intend to develop new products, while others do not. Some seek technical assistance, unique laboratory expertise, or the use of specialized research instruments (Bozeman and Papadakis, 1995; Roessner, 1993; Roessner and Bean, 1991). Others enter into these arrangements to perform joint research and development (R&D) with the explicit intention that it will lead to new or improved products or processes.

It seems reasonable to believe that there may be some factors that characterize those industry-federal laboratory interactions that do or do not lead to product development. And, likewise, it would be reasonable to expect that certain factors may affect the firm's original purpose or objective in partnering with the federal laboratory. A better understanding of these factors can lead to better policy choices and more effective industry-government technology transfer practices. This paper examines the role of the companies' characteristics, their research strategies, and the nature of their relationships with the labs in seeking answers to these questions.

### **Background**

Laws<sup>6</sup> passed throughout the 1980s and early 1990s were designed, first, to permit federal laboratories to work with the private sector and, later, to mandate it. Prior to this, restrictions were placed on these activities chiefly for public goods and national security reasons (Lee, 1994). But in the late-1970s, as perceptions mounted that the competitive edge of U.S. industry was eroding, calls became frequent and vocal for a more cooperative relationship between the public and private sectors in hopes that greater commercialization of government technologies would result in gains in U.S. economic competitiveness. By the early 1990s, as the Soviet Union

---

<sup>6</sup> For a summary of federal technology transfer legislation, see, for example: Coburn, 1995; Lee, 1994; Bozeman and Coker, 1992; Bagur and Guissinger, 1987.

fell while U.S. budget deficits did not, the “cooperative technology paradigm” (Bozeman, 1994; Crow and Bozeman 1998) took root. Cuts in defense spending paired with an evolving interest in civilian technology development and the need to maintain laboratory technical capabilities resulted in a host of new lab-missions.

Characterized as scientific “treasure chests” (Ham and Mowery, 1995; Bozeman and Wittmer, 1996) by some and “sleeping giants” (Sedaitis, 1996) by others, the labs were seen as an obvious source for new and exploitable commercial and technological riches. The not entirely accurate image became commonplace that technology transfer would lead somewhat effortlessly to new commercial products for the benefit of U.S. industry and the economy in general.

As used most often, the term technology transfer implies the active giving on the part of the lab and a rather passive taking on the part of the company (Zhao and Riesman, 1992; Rogers et. al., 1998; Bozeman, 1994b). This is often referred to as “off-the-shelf” technology transfer (Bozeman and Coker, 1992). However, for the majority of federal laboratory–company interactions examined as part of this and previous studies, this term is not descriptively accurate. Many partnerships are highly cooperative, research intensive exercises that take quite some time to surrender their riches, if ever (Bozeman and Papadakis, 1995).

Then, in 1994 the policy landscape shifted. Republicans took control of the U.S. House of Representatives, and one of their first targets (with mixed outcomes) was the cooperative technology paradigm. At present, the overall technology policy direction remains unclear, but one thing can be said for certain: many fewer laboratory–company partnerships are in effect today (Rogers and colleagues, 1998) than was the case when these data were collected (Berman, 1994), making this study a central source for determining the effectiveness of policy strategies stemming from the cooperative technology paradigm. Furthermore, little empirical evidence outside this study is available upon which to judge the track record of federal laboratory–industry partnership outcomes. While several studies have been undertaken that examine the relationship of federal laboratories with industry (e.g., O’Keefe, 1982; Roessner and Bean, 1990, 1991, 1994), surprisingly little research provides any clue as to what factors are associated with commercialization of products and processes that result from these interactions.

### **Research Procedures and Methods**

With this dearth of empirical evidence in mind, in 1994, the Federal Laboratory Study<sup>7</sup> was conducted to explore the nature of federal laboratory partnerships with the private sector. Data were collected from industrial organizations that had inter-

---

<sup>7</sup> For a more extensive discussion of data collection procedures, see Bozeman and Papadakis, 1995.

#### *4 Public Research, Innovation and Technology Policies in the US*

acted with federal laboratories during the five years prior to the study. The data set includes 229 industry–federal laboratory projects between 219 companies and 27 federal government laboratories.

The focus of the study was not just on technology transfer activities per se, but on commercial interactions more generally, including cooperative R&D, technical assistance, personnel exchange, resource sharing, and the transfer of "know how" and processes. The study also included items about the characteristics of the companies themselves, their objectives in working with the labs, how well those objectives were met, and the extent of commercialization due to those interactions.

Even though the interaction (not the laboratory or the company) was the unit of analysis, decisions had to be made about which of the federal laboratories to include, which would ultimately yield a sample of companies to survey. One possible approach, random selection, was not used because it was known from prior studies (U.S. GAO, 1989; Bozeman, 1994) that most federal laboratories do not often perform cooperative research with the private sector.

Instead, a two-stage sample was drawn. First, a sample of federal laboratories was taken from a U.S. General Accounting Office study of 330 labs with R&D budgets greater than \$100,000 and subject to the Federal Technology Transfer Act of 1986<sup>8</sup>. Despite using not particularly stringent criteria, only 54 of the labs could be considered "commercially active"<sup>9</sup> and half (27) agreed to participate<sup>10</sup>. In the second stage, the researchers mailed 694 questionnaires to companies identified by the participating federal laboratories as having had technical interactions with them between the years 1989-93. The effective sampling population was 544<sup>11</sup>. A total of 229 usable surveys were returned, resulting in an effective response rate of 42 percent. For this paper, comparison of means tests and logistic regression were used in analyzing the results.

#### **Research Questions and Hypotheses**

---

<sup>8</sup> This Act amended the Stevenson–Wydler Technology Innovation Act to authorize cooperative research and development agreements (CRADAs) between federal laboratories and other entities.

<sup>9</sup> Labs that were included met any one of the following criteria: at least 12 personnel visiting from industry, 6 CRADAs with small business, 6 CRADAs with other U.S. business, 8 total patents issued by the lab prior to 1986, 8 total patents issued by the lab after 1986, or \$50,000 in royalty income.

<sup>10</sup> These include labs of the Department of Energy, Department of Defense, Bureau of Land Management (of the Department of the Interior), NASA, Centers for Disease Control, Food and Drug Administration, and the National Institutes of Health (of the Department of Health and Human Services).

<sup>11</sup> Cases were deleted from the sampling population due to postal returns (25); communications that the person to whom the questionnaire was addressed was no longer with the company (and no one else had knowledge of the project) (47); company out of business (20); respondents indicated either no project with lab or project did not fit researchers' requirements for inclusion in the sample (e.g., work was as vendor for lab) (58).

#### 4 Public Research, Innovation and Technology Policies in the US

The chief objective of this research is to pursue more closely the effect of company intention on partnership outcome. When companies work in partnership with federal laboratories, what role does their original intention play vis-a-vis organizational characteristics, the nature of their collaborative arrangements with the labs, or the research strategies they pursue? The answer can advance our understanding of the technology transfer process which may be crucial for effective policymaking. Table I classifies partnerships in terms of the companies' original intention versus the actual outcome of the partnership.

- "No Intention-No Product"—those where the company had *no intention* to develop or improve a product or process and *did not*,
- "Intention-No Product"—those where the company *intended* to develop or improve a product or process but *did not*,
- "No Intention-Product"—those where the company *did not* intend to develop or improve a product or process but *did*, and
- "Intention-Product"—those where the company *intended* to develop or improve a product or process and *did*.

Approximately 39 percent of the companies that set out to develop or improve a product or process in partnership with a federal laboratory, succeeded by the time of the study<sup>12</sup> (see Cell A of Table I). And, likewise, 61 percent of the companies with that same intention failed to develop products or processes (these are the companies in Cell B of Table I). Interestingly, more than one-third of the companies that developed or improved products or processes as a result of their partnership *had no original intention* to do so (Cell C).

From this simple analysis, the nature of the relationship between company intention and partnership outcome is not clear. Perhaps there is no direct relationship. Or, perhaps the relationship is more subtle and complex. Figure 1 presents a path model of the technology transfer process developed to guide this research. In brief, the model posits that certain organizational characteristics are related to the intention to develop or improve a product or process; that intention is related to the choice of the R&D strategy; and that organizational characteristics, intention, and R&D strategy affect outcome of the partnership (see Table II for a definition of variables).

Research on product development and on innovation more generally has focused on organizational characteristics such as size, age, and capability (e.g., Schumpeter, 1942; Cyert and March, 1963; Dougherty and Hardy, 1996; Chandy and Tellis, 1998). In regard to these characteristics, the model asserts that small companies (Link and Bozeman, 1991; Balachandra, 1996; Damanpour, 1996) will be more

---

<sup>12</sup> The period studied includes interactions that were active between the years 1989 and 1993

likely to work in partnership with federal laboratories with the intent to develop or improve products or processes (Roessner and Bean, 1991; 1994), mostly because they are less likely to have the necessary in-house resources than larger companies. Likewise, younger companies—being more agile and eager in responding to opportunities—will be more likely to intend to develop or improve products and processes (Bozeman and Coker, 1992; Bolton, 1993; Dougherty and Hardy, 1996; Crow and Bozeman, 1998). And, finally, R&D-intensive companies will possess greater research absorptive capacity and boundary spanning capabilities which will make them more likely to pursue product or process development (Kingsley, Bozeman, and Coker, 1996; Mowery, Oxley, and Silverman, 1996; Turpin, Garrett-Jones, and Rankin, 1996; Ingham and Mothe, 1998).

The second part of the model involves the effect of company intention on choice of R&D strategy. The model asserts that companies performing research toward the development side of the R&D continuum (including commercial-applied, development, and testing) will be more likely to develop products or processes than companies focusing on basic research (Bozeman and Wittmer, 1996; Piper and Naghshpour, 1996; Spivey et. al., 1997; Ingham and Mothe, 1998; Rogers et. al. 1998). And, because projects with more collaborative forms of research will have more intense interaction between company and lab personnel they will be more likely to result in a new or improved product or process (Teece, 1992; Roessner, 1993; Berman 1994; Hameri, 1996; Geisler, 1997; Aldrich et. al., 1998; Rogers et. al., 1998).

This leads to the following hypotheses that are highlighted in Figure 1:

**H<sub>1</sub>:** *Organizational characteristics* of the firms will affect their intention to develop or improve products or processes. Specifically, small businesses will be more likely to do so as will firms with higher R&D intensity. Older companies will be less likely to intend to develop products. (Hypothesis 1 is represented on Figure 1 as the dotted lines.)

**H<sub>2</sub>:** The *intention* to develop or improve a product or process will affect the nature of the company-federal laboratory partnership and the choice of R&D strategy. Those companies that intend to develop or improve products or processes (a) will be more likely to perform R&D toward the development side of the R&D spectrum and (b) will be more likely to engage in cooperative forms of R&D with their federal laboratory partners. (Hypothesis 2 is represented on Figure 1 as the non-bolded lines.)

**H<sub>3</sub>:** The *development* or improvement of a product or process (i.e., the actual outcome) will be (in part) determined by intention, organizational characteristics (noted in hypothesis 1), research strategy (development R&D), and by the nature of the collaboration between the company and the federal laboratory (cooperative R&D). (Hypothesis 3 is represented on Figure 1 as the bolded lines.)

## Findings

A qualitative choice model was used to test the hypotheses<sup>13</sup> (see Table III). In Model 1, the relationship between three organizational characteristics (small business, age, and R&D intensity<sup>14</sup>) and intention to develop a product or process is examined. The coefficients on small business and R&D intensity were statistically significant, but the age coefficient was not. Thus, companies with 500 or fewer employees had 2.7 times greater odds of having the intention to develop products or processes than companies with more than 500 employees, holding age and R&D intensity constant. The probability that a small company has the intent to develop or improve a product or process as a result of its lab partnership is estimated to be approximately .73. R&D intensity was negatively related to intention, meaning that more R&D-intensive companies were less likely to have product or process development as one of their objectives in working with federal labs<sup>15</sup>.

In the full model (Model 2) all exogenous and endogenous independent variables, including intention, were added and the partnership outcome (whether or not a new or improved product or process resulted) was entered as the dependent variable. Only small business and development R&D were statistically significant. For companies that performed development R&D, the odds they developed new products or processes were more than double the odds for companies that did not perform development R&D holding all other variables constant. The odds that a small business developed a product or process were more than double that of larger businesses holding all else constant. A 95 percent confidence interval was constructed around

---

<sup>13</sup> Qualitative choice models estimate the odds of an event occurring given a host of factors. That is, the dependent variable in the regression model is dichotomous, making OLS regression undesirable. Two qualitative choice models commonly used—probit and logit—are based on sigmoidal distributions of the normal cumulative distribution function and the logistic distribution function respectively. The difference in estimates is usually negligible, making the choice one of convenience. A logit model was used in this paper because it is more traditionally used with observational data (as opposed to experimental data which more often uses a probit model).

<sup>14</sup> R&D intensity was entered into the model in natural log form. Including the variable in the model with the incorrect functional form would bias the coefficient estimates, and even in large samples, they would be inconsistent. Moreover, estimates of the coefficient standard errors would be biased and inconsistent. The difficulty in testing functional form in logistic regression has been well-documented, and save significance tests, Menard (1995) suggests that theory is perhaps the strongest criterion for form specification.

A linear relationship between R&D intensity and the probability that a company would have the intention to and/or succeed in developing new products and processes from partnerships with government labs seems unlikely. Natural log form assumes that the effects of R&D intensity are most pronounced at lower levels and as a company becomes more and more R&D intensive the effects taper off.

<sup>15</sup> Because of the functional form of R&D intensity, the interpretation of the coefficient is difficult. The coefficient on LN R&D intensity tells us that as the natural log of R&D intensity increases by one unit, the log odds of intention fall by .2614.

the difference in estimated proportion of small versus large companies that developed products or process. As a result of their federal laboratory partnerships, small businesses developed products or processes 12 to 36 percent *more* often than did large companies.

Interestingly, the intention to develop a product or process was not significant<sup>16</sup> ( $p = .58$ ). When cooperative R&D and development R&D are regressed on intent the coefficients are significant. The odds, then, that companies that intended to develop products or processes engaged in cooperative R&D with the federal lab were about 70 percent greater than firms that did not intend to develop products or processes. Likewise, the odds that firms that intended to develop products or processes engaged in development R&D are 2.7 times that of firms that did not intend to develop products or processes. And, as we have already established, companies that performed development R&D are much more likely to develop products or processes from their lab partnerships. So, it is the case that intent has an effect on outcome, but it is only indirect in nature: companies that intend to develop products, perform development R&D, and companies that perform development R&D succeed in developing products. Then what role does intention play in developing products or processes? A logistic regression was run dropping intention (Model 3) from the full model, resulting in almost no loss in explanatory power of the model<sup>17</sup>. The answer, then, seems to be remarkably little direct effect but indirect effect through the intervening variable, development R&D.

To investigate this relationship between intent and partnership outcome, it is perhaps necessary to return to the typology presented in Table 1 and compare the characteristics of firms, the nature of their collaboration with the federal laboratory, and the R&D strategies across company groups.

First, it is useful to compare the two groups of companies that developed or improved products or processes—where Cell A companies held this as their original intent, while Cell C companies did not. Cell C companies tended to be: (1) similar to Cell A companies in organizational characteristics, but (2) perhaps less likely to have engaged in collaborative processes with the federal laboratory, and (3) less likely to have performed commercial-applied research, development, and (perhaps) testing (see Table IV).

---

<sup>16</sup> Why might this be the case? It is unlikely due to measurement error because (1) the questionnaire item is quite clear on this point, referring to the original intent at the start of the project explicitly, and, (2) if there was response error one would expect that it would favor the actual outcome or likely outcome of the project, in which case intent would be a powerful predictor of outcome. Multicollinearity is of course always a possibility, yet its highest correlation with any other variable in the model is just .23. It is possible that there is some omitted variable that is correlated both with intention and outcome or some other model misspecification. To check for spurious effects the model was run without the intervening variables (i.e., cooperative R&D and development R&D), the coefficient on intent was not statistically significant.

<sup>17</sup> The difference in the -2 log likelihood did not change (while Chi-square d.f. 1, alpha = .05, is equal to 3.84).

A second useful comparison is between these same companies (Cell C) and companies that experienced the complete opposite of circumstances (i.e., companies that had originally intended to develop products or processes but did not [Cell B]). Cell C companies tended to be (1) younger and perhaps more likely to be small in size than Cell B companies, (2) less likely to engage in collaborative processes with the federal laboratory, and (3) perhaps less likely to have performed commercial-applied research, development, and testing (see Table V).

In general then, companies that did not originally set out to develop products or processes but did (Cell B), can be characterized as young and often small in size, but low-intensity in terms of their collaborative efforts and in terms of their R&D strategies. These companies may be beneficiaries of “off-the-shelf” technology transfer. Perhaps they set out to partner with federal laboratories for other reasons, but soon found out that there was potential awaiting them in product and process development.

Third, is the comparison of companies that intended to develop a product or process and did (Cell A) with companies that intended but did not (Cell B). Cell A companies tended to be (1) younger and more often small in size than Cell B companies, (2) not especially different in terms of the collaborative nature of their partnerships with the federal labs (Cell B companies being perhaps slightly more active), but (3) much more likely to have performed research toward the development side of the R&D spectrum (see Table VI). In general, Cell A companies tend to be young and often small in size, and they tend to emphasize commercial-applied research, development, and testing more than any other group. Cell B companies were generally older and more often large in size than the other groups of companies, and they tended to focus not as much on R&D toward the development side of the spectrum as Cell A companies, but more so than Cell C companies. They may have also stressed collaborative partnering arrangements more so than did the other two groups.

## **Conclusions**

There is evidence here to suggest that some of the characteristics we expect to be associated with technology transfer were in fact confirmed: company size and age. But, one conclusion was not expected and is not intuitive, at least not at the surface. This conclusion concerns the role of intent on outcome in product or process development or improvement in partnerships between companies and federal laboratories. Specifically:

- Small businesses had greater odds of engaging in federal laboratory partnerships for the explicit intent to develop or improve products or processes than companies of the same age and R&D capacity.

#### *4 Public Research, Innovation and Technology Policies in the US*

- Companies that intended to develop products or processes had a greater likelihood of employing commercial-applied, development, and/or testing forms of R&D than firms that do not possess such intention.
- Firms that perform development R&D have greater odds of producing a product or process from their partnership than firms that do not perform these forms of R&D.

There is also evidence to suggest that complementary roles may be important especially in who performs what type of research (Bozeman and Wittmer, 1996; Piper and Naghshpour, 1996; Spivey et. al., 1997; Ingham and Mothe, 1998; Rogers et. al. 1998). Perhaps, in the case of the Cell B companies the roles that were played by each organization were not complementary or were “too” collaborative. Conversely, is it possible for a company–federal laboratory relationship to be “passive enough” to unexpectedly result in a new or improved product or process? That is, were these (Cell C) companies beneficiaries of the “off-the-shelf” mode of technology transfer? It seems from the evidence presented in this paper, that success occurs at both extremes with failure in the middle. That is, some companies that did not intend to develop products or processes, did not often pursue collaborative-style partnerships with federal laboratories, and did not often perform research toward the development side of the spectrum were nonetheless successful in developing products or processes. At the same time, companies that intended to develop products or processes and were at the other extreme in terms of collaborative partnering styles and R&D strategies, were also successful.

There are several implications of this work for policymakers, federal laboratories, and companies who wish to work with the labs. First, all parties need to reexamine what constitutes effective collaboration and cooperative R&D arrangements in company-federal laboratory partnerships. Second, cooperative R&D did not enter the model in explaining the development of products or processes in these partnering arrangements. This might have been because of any number of statistical reasons or it might mean that “more” here is not “better.” Third, is the lesson of intent itself. Companies and federal laboratories should not be overly closed-minded about all they expect to gain from the relationship. There is evidence from this study to suggest that intents and expectations evolve and are often abandoned in favor of new opportunities that arise in the partnering process.

## Acknowledgements

I would like to thank Barry Bozeman for the use of the data that form the basis of this paper and for his helpful comments. I also thank Gregory Lewis and Lynne Austin for their helpful comments.

## References

Aldrich, H.E., M.K. Bolton, T. Baker, and T. Sasaki (1998) "Information Exchange and Governance Structures in U.S. and Japanese R&D Consortia: Institutional and Organizational Influences." *IEEE Transactions on Engineering Management*, 45(3): 263-275.

Bagur, J., and Guissinger, A. (1987) "Technology transfer legislation: An overview." *Journal of Technology Transfer*, 12:51-63.

Balachandra, R. (1996) "International Technology Transfer in Small Business: A New Paradigm." *International Journal of Technology Management*, 12(5-6): 625-638.

Berman, E.M. (1994) "Technology-Transfer And The Federal Laboratories - A Midterm Assessment Of Cooperative Research." *Policy Studies Journal*, 22(2): 338-348.

Bolton, M.K. (1993) "Organizational Innovation and Substandard Performance: When is Necessity the Mother of Innovation?" *Organization Science*, 4(1): 57-75.

Bozeman, B. (1994) "Evaluating Government Technology Transfer: Early Impacts of 'Cooperative Technology' Paradigm." *Policy Studies Journal*, 22(2): 322-337.

Bozeman, B. and K. Coker (1992) "Assessing the Effectiveness of Technology Transfer from U.S. Government R&D Laboratories: The Impact of Market Orientation." *Technovation*, 12(4): 239.

Bozeman, B. and M. Papadakis (1995) "Company Interactions with Federal Laboratories: What They Do and Why They Do It." *Journal of Technology Transfer*, 20(4): 64-74.

Bozeman, B. and D. Wittmer (1996) "Technical Roles, Technical Range and Success of Federal Laboratory-Industry Partnerships." Unpublished Manuscript.

Chandy, R.K. and G.J. Tellis (1998) "Organizing for Radical Product Innovation: The Overlooked Role of Willingness to Cannibalize." *Journal of Marketing Research*, 35(Nov): 474-487.

Coburn, C. ed. (1995) *Partnerships: A Compendium of State and Federal Cooperative Technology Programs*. Columbus, OH: Battelle Press.

#### *4 Public Research, Innovation and Technology Policies in the US*

Crow, M. and B. Bozeman (1998) *Limited by design : R&D laboratories in the U.S. national innovation system*. (New York : Columbia University Press).

Cyert, R. and J. G. March (1963) *A Behavioral Theory of the Firm*. Englewood Cliffs, NJ: Prentice-Hall.

Damapour, F. (1996) "Organizational Complexity and Innovation: Developing and Testing Multiple Contingency Models." *Management Science*, 42(5): 693-716.

Davis, J.A. (1985) *The Logic of Causal Order*. Sage University Paper 55. Newbury Park: Sage.

Dougherty, D. and C. Hardy (1996) "Sustained Product Innovation in Large Mature Organizations: Overcoming Innovation-to-Organization Problems." *Academy of Management Journal*, 39(5): 1120-1153.

Geisler, E. (1997) "Intersector Technology Cooperation: Hard Myths, Soft Facts." *Technovation*, 17(6): 309-320.

Ham, R. M. and D. C. Mowery (1995) "Improving Industry-Government Cooperative R&D: You Won't Find A Treasure Chest of Market-Ready Technologies at the National Labs, But You Can Find Opportunities for Productive Collaboration." *Issues in Science and Technology*, 11(4): 67.

Hameri, A.P. (1996) "Technology Transfer Between Basic Research and Industry." *Technovation*, 16(2): 51-57.

Ingham, M. and C. Mothe (1998) "How to Learn in R&D Partnerships?" *R&D Management*, 28(4): 249-261.

Kingsley G., B. Bozeman and K. Coker (1996) "Technology transfer and absorption: An 'R&D value-mapping' approach to evaluation." *Research Policy*, 25(6): 967-995.

Lee, Y. S. (1994) "Technology Transfer And Public Policy in an Age of Global Economic Competition: Introduction to the Symposium." *Policy Studies Journal*, 22(2): 260-266.

Link, A. (1991) "Innovative Behavior of Small-Sized Firms." *Small Business Economics*, 3: 179-184.

Menard, S. (1995) *Applied Logistic Regression Analysis*. Sage University Paper 106. Thousand Oaks: Sage.

Mowery D.C., J. E. Oxley, and B.S. Silverman (1995) "Strategic Alliances and Interfirm Knowledge Transfer." *Strategic Management Journal*, 17(Winter): 77-91.

#### *4 Public Research, Innovation and Technology Policies in the US*

O'Keefe, T.G. (1982) "Evaluating Technology Transfer: Some Problems and Solutions." *Journal of Technology Transfer*, 6(2): 53.

Piper, W. S. and S. Naghshpour (1996) "Government technology transfer: The effective use of both push and pull marketing strategies." *International Journal of Technology Management*, 12(1): 85-94.

Roessner, J. D. (1993) "What Companies Want From The Federal Labs." *Issues in Science and Technology*, 10(1): 37-42.

Roessner, J. D. and A. Bean (1990) "Industry Interactions with Federal Laboratories." *Journal of Technology Transfer*, 15(4): 5.

Roessner, J. D. and A. Bean (1991) "How Industry Interacts with Federal Laboratories." *Research Technology Management*, 34(4):22.

Roessner, J. D. and A. S. Bean (1994) "Patterns of Industry Interaction with Federal Laboratories." *Journal of Technology Transfer*, 19(4): 59-77.

Rogers, E. M., E. G. Carayannis, K. Kurihara, et al. (1998) "Cooperative research and development agreements (CRADAs) as technology transfer mechanisms." *R&D Management*, 28(2): 79-88.

Sedaitis, J. B. (1996) "Waking the sleeping giants: Commercializing state R&D in the United States and Russia." *Technological Forecasting and Social Change*, 53(3): 293-308.

Schumpeter, J.A. (1942) *Capitalism, Socialism, and Democracy*. New York: Harper.

Spivey, W.A., J.M. Munson, M.A. Nelson, and G.B. Dietrich (1997) "Coordinating The Technology Transfer and Transition of Information Technology: A Phenomenological Perspective." *IEEE Transactions On Engineering Management*, 44(4): 359-366.

Teece, D.J. (1992) "Competition, Cooperation, and Innovation: Organizational Arrangements for Regimes of Rapid Technological Progress." *Journal of Economic Behavior and Organization*, 18: 1-25.

Turpin, T., S. Garrett-Jones, and N. Rankin (1996) "Bricoleurs and Boundary Riders: Managing Basic Research and Innovation Knowledge Networks." *R&D Management*, 26(3): 267-282.

U.S. General Accounting Office (1989) *Technology transfer: Implementation status of the federal technology transfer act of 1986*. Washington, DC: U.S.G.P.O.

**Table I. Original Intention of Company versus Project Result at the Time of the Study**

		Intended to develop or improve a product or process	
		Yes	No
Resulted in a new or improved product or process	Yes	<i>Cell A</i> “Intention-Product” n = 52	<i>Cell C</i> “No Intention-Product” n = 27
	No	<i>Cell B</i> “Intention-No Product” n = 80	<i>Cell D</i> “No Intention-No Product” n = 64

**Table II. Definitions of Variables**

Variable Name	Mean	Range	Definition
Intention to Develop [product or process]	.59	0-1	Original intention of the firm in partnering with the federal lab. Coded as intention to develop a new product/process or to improve an existing one (1) or other (0).
Small Business	.52	0-1	Firms with 500 or fewer employees (1) or more (0)
Age of Company	41.40	3-150	Total number of years the firm has been operating.
R&D Intensity	29.66	0-100	Total number of firm employees working in R&D as a percent of the total number of employees.
Development R&D	.67	0-1	Firms performing commercial-applied research, development, or testing (1); or basic or precommercial-applied (0).
Cooperative R&D	.63	0-1	Lab partnerships that included CRADAs and/or joint or cooperative research other than a CRADA (1), all others (0).
Developed Product	.35	0-1	Firms that developed a new product/process or improved an existing one (1) during the partnership or other (0)

**Table III. Logistic Regression Models**

Independent Variable	Beta	S.E.	Dependent Variable
<b>Model One</b>			Intend to develop
Small Business*	.9933	.4049	
Company Age	.0045	.0058	
LN R&D Intensity*	-.2614	.1248	
Constant**	-.7563	.4575	
<b>Model Two</b>			Developed Product
Small Business**	.8221	.4423	
Company Age	-.0086	.0067	
LN R&D Intensity	.0452	.1387	
Intended to Develop Product	.1966	.3560	
Cooperative R&D	.1569	.3561	
Development R&D*	.7845	.3845	
Constant**	-1.5203	.6199	
<b>Model Three</b>			Developed Product
Small Business*	.8712	.4325	
Company Age	-.0082	.0066	
LN R&D Intensity	.0357	.1375	
Cooperative R&D	.1761	.3534	
Development R&D*	.8185	.3796	
Constant*	-1.4919	.6162	
<b>Model Four</b>			Cooperative R&D
Intended to Develop Product*	.5702	.2832	
Constant	.2029	.2131	
<b>Model Five</b>			Development R&D
Intended to Develop Product*	.9827	.2952	
Constant	.1366	.2137	

**Notes:**

\* Statistically significant difference  $p < .05$

\*\*Statistically significant difference  $p < .10$

All models significant at .05 level.

**Table IV. Company characteristics, collaborative nature of partnership, and research strategy: No Intention-Product companies versus Intention-Product companies**

	CELL C (NO INTENTION- PRODUCT)	CELL A (INTENTION-PRODUCT)
<b>Company Characteristics</b>		
Age (mean, in years)	28.7	29.9
Small Business	64.0	72.3
R&D Intensity	32.7	32.1
<b>Collaborative Nature of Partnership</b>		
Cooperative R&D	59.3	63.5
Company Had Previous Informal Contact with Lab?	30.8	41.2
Lab Tech Transfer Staff Involved?*	7.7	32.0
<b>R&amp;D Strategy</b>		
Company Performed Commercial-Applied R&D*	22.2	62.0
Company Performed Development*	37.0	70.0
Company Performed Testing	33.3	54.0

**Notes:**

All entries are percent of company group exhibiting characteristic unless noted otherwise.

\* Statistically significant difference  $p < .05$  \*\*Statistically significant difference  $p < .10$

**Table V. Company characteristics, collaborative nature of partnership, and research strategy: No Intention-Product companies versus Intention-No Product companies**

	CELL C (NO INTENTION- PRODUCT)	CELL B (INTENTION-NO PRODUCT)
<b>Company Characteristics</b>		
Company Age* (mean, in years)	28.7	52.5
Small Business	64.0	42.7
Company R&D Intensity	32.7	25.3
<b>Collaborative Nature of Partnership</b>		
Cooperative R&D	59.3	72.5
Company Had Previous Informal Contact with Lab?***	30.8	56.3
Lab Tech Transfer Staff Involved*?	7.7	32.9
<b>R&amp;D Strategy</b>		
Company Performed Commercial-Applied R&D	22.2	39.2
Company Performed Development	37.0	48.1
Company Performed Testing	33.3	40.5

**Notes:**

All entries are percent of company group exhibiting characteristic unless noted otherwise.

\* Statistically significant difference  $p < .05$  \*\*Statistically significant difference  $p < .10$

**Table VI. Company characteristics, collaborative nature of partnership, and research strategy: Intention-No Product companies versus Intention-Product companies**

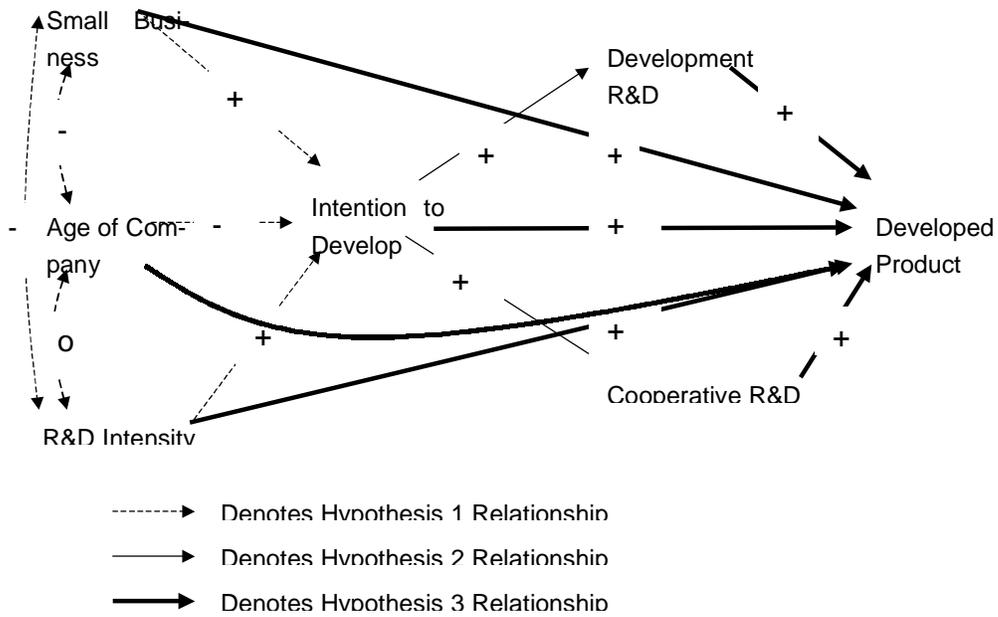
	CELL A (INTENTION- PRODUCT)	CELL B (INTENTION-NO PRODUCT)
<b>Company Characteristics</b>		
Age* (mean, in years)	29.9	52.5
Small Business*	72.3	42.7
R&D Intensity	32.1	25.3
<b>Collaborative Nature of Partnership</b>		
Cooperative R&D	63.5	72.5
Company Had Previous Informal Contact with Lab?	41.2	56.3
Lab Tech Transfer Staff Involved?	32.0	32.9
<b>R&amp;D Strategy</b>		
Company Performed Commercial-Applied R&D**	62.0	39.2
Company Performed Development*	70.0	48.1
Lab Performed Development	54.0	40.5

**Notes:**

All entries are percent of company group exhibiting characteristic unless noted otherwise.

\* Statistically significant difference  $p < .05$  \*\*Statistically significant difference  $p < .10$

**Figure 1: Hypothesized Model**



## **Discussion of Dietz's paper**

**Louis Tornatzky** (Southern Technology Council, USA). Perhaps the intention variable of the firm was influenced by the outcome. They may not remember/report correctly their intention to develop.

**David Guston** The intent of the partner is important because they may only want a CRADA to get scientific information from the federal lab because they may not have another way of acquiring this information.

**Irwin Feller** (Pennsylvania State University, USA). It is a question of how deep inside the firm this relationship goes. Penetration inside the firm may actually be quite modest. It is a low-cost way to scan the environment for firms.

**James Dietz** One of the survey questions asks how many layers within the company were involved in the process.

**John Barber** (Department of Trade and Industry, UK). The degree of commitment by the firm may be influenced by their motives for coming into the partnership. Firms need to know about certain technologies and one way to obtain this information is to enter into a collaborative venture. This may not be because they want to enter into the market now but because they may in the future and want to be an intelligent investor.

**Louis Tornatzky**. Did you look at the financial commitment of firms?

**James Dietz**. No, but it could be done.

## **The Expanding Role of Peer Review Processes in the United States**

**David H. Guston**

Bloustein School of Planning & Public Policy, Rutgers, The State University of New Jersey, USA

Email: [guston@rci.rutgers.edu](mailto:guston@rci.rutgers.edu)

---

### **Introduction**

The diverse processes of peer review are familiar to science policy. Peer review processes serve a critical role in the allocation of such scarce resources as research funds and journal space, as well as in helping to produce knowledge on which researchers rely (Chubin and Hackett 1990). In part because of the perceived success of peer review in these roles, reformers have sought to harness peer review to help produce knowledge on which policy makers can rely, for the ultimate purposes of improving decisions, reducing the occurrence of legal challenges and other procedural obstacles, and achieving other political goals (Jasanoff 1990).

This paper will explore what appears to be the increasing domain of peer review processes in science policy. It is an early portion of a two-pronged research agenda that seeks to elaborate both the logical structure and the detailed procedures of peer review and the use of scientific expertise in the policy process. In elaborating peer review's logical structure, the project builds on previous work on the logical structure of science policy more generally. Following the traditional between "policy for science" and "science in policy" (Brooks 1968), this earlier work applies a principal-agent perspective to focus exclusively on the delegation inherent in funding research (Braun 1993; Guston 1996b; van der Meulen 1998; Caswill 1998). Applying the same principal-agent perspective to "science in policy," the agenda seeks to articulate a more sophisticated analytical framework for peer review (and other uses of expertise) that both scholars in several disciplines and practitioners in advice-producing and -consuming arenas can appreciate.

In elaborating the detailed procedures of peer review, the agenda extends an approach to studying science and scientists by following them out of the laboratory (Latour 1987) and into arenas in which they are called upon to come to judgments in a more public and accountable way. It also extends an approach to studying science and scientists that emphasizes observing how, through their rhetoric as well as the establishment of policies and procedures, scientists attempt to demarcate their vision of scientific from non-scientific activities (Gieryn 1995; Jasanoff 1990).

Such detailed scrutiny, for example, can yield a more complete, scholarly understanding of consensus and consensus formation in science (Kim 1994), as well as create precise, policy-relevant knowledge about the decision making of expert advisory committees (Guston 1999). Ultimately, the project will synthesize both perspectives in a way that has been done for “policy for science” (Guston 2000a).

This paper provides preliminary descriptive work for the agenda. It begins by defining peer review and recounting a very brief history of its use by the federal government of the United States. It then describes the expansion of its domain in several areas: the allocation of federal funds, the evaluation of research programs, the evaluation of knowledge inputs to policy, the admission of expert testimony in federal courts, and in state science policy. The paper concludes with a brief evaluation of these trends.

### **Definition and History of Peer Review**

As the General Accounting Office (GAO 1999) has found, there is no single definition of peer review used across government agencies in the United States. Consequently, the practice of peer review varies within certain wide parameters (Guston 2000c). GAO (1999:3) found, however, that “all of the agencies’ definitions or descriptions of peer review contained the fundamental concept of a review of technical or scientific merit by individuals with sufficient technical competence and no unresolved conflict of interest.” This definition is sufficient for the purposes of this paper, but key to its further usefulness is the specification of exactly what constitutes technical or scientific merit, who the competent persons are and how are they selected, what conflicts of interest need to be resolved in which ways, and how the process of review itself relates to actual outcomes. I take up some of these questions elsewhere (e.g., Guston 2000b).

Over the long span of history, peer review as a method of evaluation has clearly expanded its domain. Among the earliest examples of peer review for proposed research projects in the United States include the Smithsonian Institution, which created an advisory committee for reviewing and recommending funding proposals in the 1840s; the Navy Consulting Board, which in 1915 began reviewing requests for funding from inventors (Savage 1999); and the Hygienic Laboratory, predecessor to the National Institutes of Health (NIH), which pioneered peer review in 1902 with a congressionally mandated advisory committee (Smith 1990).<sup>18</sup> The National Cancer Act of 1937 and the Public Health Service Act of 1944 legitimated the use of advisory councils to award grants to extramural researchers, and NIH fully established its peer review system with the creation of its Division of Research Grants (now the Center for Scientific Review) and the original review groups after World

---

<sup>18</sup> An early example of peer review for publication occurred when President Thomas Jefferson, who commissioned the Lewis and Clark expedition, requested the American Philosophical Society to review the expedition’s report prior to its publication (Savage 1999:5).

War II (Smith 1990). At the National Science Foundation (NSF), peer review was only implicit at the start, although NSF's governing body, the National Science Board (NSB), understood that after staff members evaluated proposals, they would send them to advisory bodies before presenting them to NSB for statutory approval (England 1982).

### **Expansion of Peer Review in the Allocation of Federal Funds**

One of the critical conflicts in science policy at the time of the founding of NSF was between different schemes for allocating federal research and development (R&D) dollars: Would the wartime practice of distributing funds only to the elite universities continue, or would there be some geographic formula (as had been the mode with agricultural research) that might redress some of the distributional inequities? Although the former model triumphed, and was buttressed by the peer review system, the political demand for the redress of geographic inequities did not disappear. Instead, it manifested itself, in part, in the earmarking of federal appropriations to specific projects. Thus, earmarking (or porkbarreling) is seen as competitive with peer review, and the level of earmarks are anxiously traced in the science policy community. Journalists from *The Chronicle of Higher Education* conduct an annual survey of earmarks to academic institutions. In recent years, the amount of those earmarks identified by *The Chronicle* have totaled in the upwards of \$500 million, showing a generally increasing trend (Savage 1999:3). The latest figure for fiscal year (FY) 2000 is slightly over \$1 billion (Brainard and Southwick 2000). This figure represents only about 3% of federal R&D spending at colleges and universities, but it is widely agreed to be an underestimate.<sup>19</sup>

Identifying the amount of peer reviewed research within total R&D is difficult (Smith 1990). A generation ago, Harvey Brooks (1982) estimated the fraction of peer-reviewed research in total national R&D expenditures at about 5%. This figure assumes that no peer review occurs in the large fraction (then about two-fifths; now, about two-thirds) of R&D funded and performed in-house by corporations – a notion disputed by Lewis Branscomb (1982). Whereas, Brooks lamented the small fraction; Rustum Roy (1985:73), a critic of peer review, estimated with some satisfaction that “at least 90% of the total money spent on research and development in the United States” is allocated by mechanisms other than peer review.

Since FY 1996, the Office of Science and Technology Policy (OSTP) and the Office of Management and Budget (OMB) “have jointly provided annual direction to agencies encouraging them to emphasize the funding of peer-reviewed research over non-peer-reviewed research” (GAO 1994:4), but there is no direct way of judging whether or not this directive has had any impact. The Clinton Administration's FY 2001 budget proposal reports that \$26 billion of the nearly \$83 billion

---

<sup>19</sup> See Intersociety Working Group (1999:68) for expenditures at colleges and universities.

total federal R&D (31.4%) was peer reviewed in FY 2000, and that this share is proposed to increase to 33% in FY 2001. This level of peer reviewed R&D corresponds to about 10% of national expenditures. Again assuming that no private funds are peer reviewed,<sup>20</sup> this rate is double what Brooks estimated, despite the nearly two-fold increase in the relative share of private funds in the national account. There has thus likely been a significant increase in the share of federal R&D funds that are subject to peer review.<sup>21</sup>

Such changes, however, may occur as a side effect of the changing composition of total R&D, rather than as a first-order effect of emphasizing peer review. For example, in the early 1990s, the composition of federal R&D spending was approximately 60% defense, 40% civilian. Currently, the composition is closer to parity between defense and civilian accounts, and as civilian R&D is more likely to be peer reviewed than defense R&D, this relative transfer can account for some of the growth in the share of peer reviewed research. Within the civilian account, the recent growth of NIH's share has expanded the influence of peer review as well.

Nevertheless, some recently established federal funding programs tout peer review as an important and novel component:

- In 1991, the U.S. Department of Agriculture created the National Research Initiative as a competitive, peer reviewed grant program to complement its intramural research and its formula funding to land-grant colleges. The National Academy of Sciences has praised the program for producing high-quality science, but Congress has appropriated only about one quarter of its annual \$500 million authorization and the number of applications has fallen 25% off of its 1994 peak (Southwick 2000).
- In 1995, the Environmental Protection Agency created the Science to Achieve Results (STAR) program, which has been funded at \$100 million. STAR uses peer review to allocate funds among responses to requests for proposals in EPA's mission-related research, and highly rated proposals are "subjected to a programmatic review within EPA to ensure a balanced research portfolio" (GAO 1999).
- The Advanced Technology Program (ATP) of the Department of Commerce makes use of independent external review as well. ATP reviewers rank proposals based on scientific/technical merit and on the potential for broad-based eco-

---

<sup>20</sup> Sapolsky (1990:95n29) suggests that NIH found a model for peer review in the private Rockefeller Fund, which began the practice in 1930 (but Rockefeller's reasons for doing so are obscure). Today, a few private foundations use peer review to allocate funds, but review by program officers and boards seems to be the norm (Guston 2000c). Such funds, however, are still a small fraction of private industry's support of R&D.

<sup>21</sup> This increase has not resulted in a concomitant increase the number of grant review committees which, according to reports by the General Services Administration, have actually shown a decrease in numbers over the last five years.

conomic impacts. Since it received its first appropriations in 1990, ATP has funded more than \$1.5 billion in R&D (see Hill 1998).

### **Expansion of Peer Review in the Evaluation of Research Programs**

The jurisdiction of peer review may be further increasing among agencies that sponsor research with the increasing importance of the Government Performance and Results Act (GPRA), which requires federal agencies to engage in program and performance assessment. Because the results of research funding are hard to quantify, research agencies rely on peer assessments (either prospectively or retrospectively) to justify and evaluate their performance (NAS 1999). The National Science and Technology Council (NSTC 1996) finds that “[f]or evaluating current programs in *individual agencies*, merit review based on peer evaluation will continue to be the primary vehicle for assessing the excellence and conduct of science at the cutting edge” (emphasis in the original). Such review can occur not only prospectively, as with current peer review for funding allocations, but “[a] form of merit review with peer evaluation can also be used for retrospective evaluation of an agency’s fundamental science program or programs.” NSTC recommends that assessors performing this evaluation should include “input from stakeholders, next-stage users, and/or customers who will use or have a stake in the results of the research being done,” in addition to those with “relevant scientific expertise and experience in the type of research being done.”<sup>22</sup>

### **Expansion of Peer Review in the Evaluation of Knowledge Inputs to Regulation**

Congress has recently sought to expand the application of peer review to the knowledge inputs to regulatory policy making. Bills introduced in the 106<sup>th</sup> Congress (1999-2000) include: S. 746, to peer review cost-benefit and risk analyses of major rules, among other purposes; H.R. 574, to peer review all regulations supported by scientific data; and H.R. 2639, to peer review the data used in standards promulgated by the Occupational Safety and Health Administration.

Many federal agencies, however, already practice forms of peer review in their regulatory, evaluative, or assessment missions (Guston 2000b; 2000c; Jasanoff 1990; Smith 1992).<sup>23</sup> Some of these mechanisms, such as the Science Advisory Board of the Environmental Protection Agency (EPA), are decades old. Others, such as the Board of Scientific Counselors of the National Toxicology Program (NTP), are recent innovations.

---

<sup>22</sup> NSTC’s recommendation for this kind of input seems to foreshadow what NIH has done in establishing a Council for Public Representation to advise the NIH director.

<sup>23</sup> The General Services Administration reports that approximately 25% of all federal advisory committees are “scientific/technical” in nature (excluding grant review committees); this percentage has held roughly constant since GSA began the categorization for fiscal year 1985.

#### 4 Public Research, Innovation and Technology Policies in the US

EPA has made the most intensive effort to expand the application of peer review to the use of science in its own decision making. A set of documents published over the last decade has:

- Emphasized the importance of external peer review for EPA scientific and technical products, contact between EPA and external scientists, and the use of the best science in decisions – thus setting the agenda for regulatory peer review (EPA 1992);
- Attempted to set standard operating procedures among various EPA divisions by creating a Science Policy Council (SPC) to “expand and improve peer review in all EPA offices” (Browner 1994);
- Articulated the new policy for “peer review and peer involvement” in EPA (1997)<sup>24</sup> that anticipates the peer review of “major work products that are primarily scientific and technical in nature and may contribute to the basis for policy or regulatory decisions;” and
- Primed EPA staff and managers on the “organization and conduct of peer review,” (EPA 1998), including specific criteria of when to apply and not apply it.

NTP is an agency of the Department of Health and Human Services, created in 1978 with the mission to “evaluate agents of public health concern by developing and applying the tools of modern toxicology and molecular biology” (NTP 1999:2).

NTP uses a Board of Scientific Counselors (BSC) to provide peer review for a number of agency activities, including oversight of research conducted in NTP centers and review of nominations for substances to be included in the congressionally mandated *Report on Carcinogens*.

In this latter role, BSC demonstrates another aspect of the expanding jurisdiction of peer review: the requirement that all information considered in the decision to list (or delist) a substance in the *Report on Carcinogens* be from the publicly available, peer reviewed literature. Further research intends to document the extent to which other federal agencies apply such a stricture in their regulatory, evaluative, or assessment missions (but see below as well).

---

<sup>24</sup> A more complete chronology exists in Powell (1999), which also describes a 1996 GAO report that found uneven progress on implementing peer review and an internal study conducted by EPA’s Office of Research and Development that reported few deviations from peer review requirements except in the area of models. ). One of SPC’s early products was “Guidance for Conducting External Peer Review of Environmental Regulatory Models” (Sussman 1994).

### Expansion of Peer Review in State Policy

There is no systematic information about the use of peer review in the states, but ad hoc information suggests that it is taking root and expanding there, too, in both roles of allocating funding and evaluating knowledge inputs.

State agencies support modest amounts of research closely related to missions in, for example, environmental protection, health and human services, or housing, planning and development. They also support R&D programs aimed specifically at economic development. NSF (1999) reports that states fund about \$3 billion in R&D annually.<sup>25</sup>

Many state agencies provide small grants and contracts to academic and other researchers in pursuit of their missions. The New Jersey Department of Environmental Protection, for example, is one of the few such departments in the nation with a separate and highly professionalized Division of Science, Research, and Technology (DSRT). For its extramural research, DSRT conducts a peer review, using three reviewers, two of whom are outside DSRT. The reasons for not using external review include the problem of competitors for the funding reviewing proposals and the problem of quicker turnaround needed by the agency than external reviewers normally provide.<sup>26</sup>

The State Science and Technology Institute (SSTI) is a clearinghouse for information about R&D at the state level. Searching its summaries of state programs in R&D for economic development for “peer review” yielded hits at eleven states (AK, AR, CA, CO, KS, MI, ND, NJ, NY, OK, PA) and the following details:<sup>27</sup>

- Some states use visiting committees to review major investments in R&D centers; such review may occur annually or biennially and is important in decisions to renew or extend funding.
- More states use peer review, in either an ad hoc or panel fashion, for allocating small grants.
- For technology programs, states review both technical elements of proposals and business plans, as ATP does.

---

<sup>25</sup> This figure presumably encompasses both types of funding and is likely an underestimate.

<sup>26</sup> In his proposal for post-war federal support of research, Vannevar Bush implicitly compared the federal government to state governments, finding that the latter lack the institutional means and talent to provide the necessary administrative support for research funding. Many states have improved their capacities since 1945, but it is not clear whether the new capacity is sufficient (See Guston 1996a).

<sup>27</sup> See [www.ssti.org/states.htm](http://www.ssti.org/states.htm). Minnesota, which did not appear in the SSTI search, has mandated peer review in the authorizing statute for Minnesota Technology, Inc. (MN Statutes 116O.071, subd. 2).

#### *4 Public Research, Innovation and Technology Policies in the US*

It is possible that many of these peer review activities did not commence until after the programs were established. In New Jersey, for example, several programs initially implemented in a non-competitive fashion were replaced by competitive, peer-reviewed programs over the first decade of their operation.

California has recently offered \$300 million in matching funds to the University of California (UC) system to establish new collaborative research centers. The choice of the centers will depend on the results of two rounds of peer review: the first stage, conducted by technical experts chosen by the UC system; and the second stage by technical experts selected by the governor from a slate proposed by UC.

There is similarly no systematic information about the use of peer review in state regulatory decisions, although there is evidence of interest among the states in science in policy (Andersen 2000; CGS 1999). California has implemented this interest the furthest. In 1997, Governor Pete Wilson signed SB 1320, which contained a peer review mandate.<sup>28</sup> Also in 1997, a Risk Assessment Advisory Committee issued a final report that asked for peer review to be applied more consistently throughout the state's Environmental Protection Agency (Cal/EPA). Cal/EPA (1998) has begun to implement the request in its strategic planning and the publication of policy and principles for peer review. California's Proposition 65 also mandates a peer function to list chemicals shown to cause cancer, birth defects, or reproductive harm.

There are likely a wide array of uses of peer advisory committees in support of policy or regulatory decision making in other states, including:

- The Michigan Environmental Science Board, established by Governor John Engler by executive order in 1992, which answers referrals from the Governor, who asks specific questions regarding, for example, the review of environmental impact statements or proposed environmental standards for permits or operating licenses. A subcommittee of the Board, with guests added as needed for their expertise, then provide answers.<sup>29</sup>
- North Carolina's Scientific Advisory Board on Toxic Air Pollutants, composed of five scientists who review new or revised acceptable ambient level guidelines for air toxics and who conduct risk assessments.<sup>30</sup> A broader Environmental Management Commission adds economic and feasibility concerns, and an Environmental Review Commission serves as gatekeeper to the legislative process.

#### **Expansion of Peer Review in the Evaluation of Courtroom Expertise**

---

<sup>28</sup> See [www.calepa.ca.gov/publications/factsheets/1997/prop65fs.htm](http://www.calepa.ca.gov/publications/factsheets/1997/prop65fs.htm).

<sup>29</sup> See [www.mesb.org](http://www.mesb.org).

<sup>30</sup> See [daq.state.nc.us/offices/technical/toxics/risk/](http://daq.state.nc.us/offices/technical/toxics/risk/).

#### 4 Public Research, Innovation and Technology Policies in the US

In its 1993 *Daubert v. Merrell Dow* decision, the Supreme Court articulated an important but not dispositive role for peer review in the certification by judges of expert witnesses. *Daubert* also prompted judges to appoint their own experts to re-view scientific findings and offer analysis not based on advocacy. The Court elaborated its view in such subsequent decisions as *GE v. Joiner* and *Kumho v. Carmichael*.<sup>31</sup>

In *Daubert*, the Supreme Court ruled that the Federal Rules of Evidence, and not the 1923 *Frye* rule, govern the admission of expert testimony.<sup>32</sup> Passed by Congress in 1975, the Federal Rules hold that all expert testimony that is “relevant and reliable” is admissible, and they place the responsibility of determining relevance and reliability in the hands of the trial judge – whom the Court felt was competent to make that judgment.<sup>33</sup> The Court also identified four points that judges could take into account when determining admissibility: 1) Is the theory or technique testable? 2) Has the theory or technique been subjected to peer review and publication?<sup>34</sup> 3) Is the error rate known or potentially known? and 4) Is the theory or technique generally accepted within the particular scientific community? The Court specified that judges should consider these questions for the principles and methods of experts, and not their conclusions. It further directed that this list was neither definitive nor dispositive.

*General Electric v. Joiner* was the first case implementing the *Daubert* decision that percolated back to the Supreme Court. Associate Justice Breyer separately concurred with Chief Justice Rehnquist’s opinion in *Joiner*, citing several of the *amici* that judges’ lack of scientific training does not relieve them of the role of gatekeeper that *Daubert* casts them in. Breyer therefore concluded that judges should make greater use of court-appointed experts, as the Federal Rules allow them to do.

A high profile instance of court-appointed experts occurred in the issue of systemic illness from silicone breast implants.<sup>35</sup> Judge Sam C. Pointer, Jr. appointed a panel of four scientists to assist him in a class action suit. After reviewing the published research and questioning the experts offered by both sides, the panel found no clear

---

<sup>31</sup> A helpful summary of the cases appears in Berger (2000).

<sup>32</sup> In its 1923 *Frye* decision, the Supreme Court found the use of a primitive lie detector to be inadmissible because the machine did not work on a principle “generally accepted” within the relevant community.

<sup>33</sup> The Federal Rules of Evidence hold sway in all Federal courts, and most states have adopted them voluntarily.

<sup>34</sup> In their *amicus* brief, *The New England Journal of Medicine* and *The Annals of Internal Medicine* argued that peer reviewed publication should have a stronger role in determining what information is admissible, as peer review is the mark of “good science.” Other briefs, for example that submitted by Chubin, Hackett, Ozoroff and Clapp, disputed this gatekeeper role for peer review, arguing that peer review is more a tool of editing than a guarantor of reliability.

<sup>35</sup> Breyer (2000; 1998) speaks favorably of such a solution elsewhere.

evidence that silicone breast implants cause immune disorders (Kaiser 1998). Although the panel was widely heralded as bringing clarity to the issue, some disputed the idea that the experts were, in fact, independent, and delimiting the evidence to published research may have pre-ordained the conclusion in any event.<sup>36</sup>

## **Conclusion**

A review of the use of peer review by the U.S. Federal government suggests that its role and jurisdiction has continued to expand. Not only is a greater share of federally funded R&D (and national R&D) peer reviewed than in the past, but peer review is taking hold as a means of evaluating larger aspects of the R&D system and of producing “relevant and reliable” knowledge in the regulatory process, in the Federal court system, and even in R&D funding and regulation by states.

The expansion of peer review in this manner is not, however, easy to evaluate. Functions of peer review beyond the allocation of funds are based on an analogy to the success of peer review there; yet many criticisms of funding peer review exist, and the analogy between it and regulatory peer review is not exact (Jasanoff 1990; Powell 1999). Moreover, despite the attempts of many in Congress to quell controversy in regulatory science by adding peer review, the addition of peer review in funding programs does not seem to serve the same purpose, as several of the new peer reviewed funding programs remain politically vulnerable.

The increased instance of peer review is reason enough for increased scholarly attention: Can we further quantify and qualify the ways in which the Federal government invokes peer review? Can we specify more completely what political and policy problems are supposed to be resolved by peer review? Can we understand in both a more comprehensive and more detailed fashion the relationship between the processes of peer review and the supposed outputs of relevance, reliability, quality, and consensus? The answers to these questions may lead to a more complete and nuanced understanding of the relationship between politics and science.

## **Acknowledgements**

The research included in this paper was supported in part by a contract from American Chemistry Council (formerly, Chemical Manufacturers Association). The conclusions are solely the author’s and in no way represent the views of ACC. The author thanks Margaret Brennan, Christine Krupka, Pamela LeBeaux, Elizabeth Nasi, and Veronica Rowan for their research assistance. The author also thanks Bernard D. Goldstein, Lewis M. Branscomb, Neal F. Johnson, Diana Hicks, Caron Chess, and Clinton Andrews for their advice on the report on which this paper is partially based.

---

<sup>36</sup> To help support judges in finding experts, the American Association for the Advancement of Science (AAAS) started the Court Appointed Scientific Experts (CASE) project to will help identify, contact, and vet potential experts. See AAAS (1998), Kieman (1999), and Runkle 1999).

## References

- American Association for the Advancement of Science (AAAS). 1998. "Court appointed scientific experts: A demonstration project of the AAAS." Available at [www.aaas.org/spp/case/case.htm](http://www.aaas.org/spp/case/case.htm).
- Andersen, Glen. 2000. "Science and politics: A balancing act for environmental health policy." *State Health Notes* (28 February):3-6.
- Berger, Margaret A. 2000. "Expert testimony: The Supreme Court's rules." *Issues in Science and Technology* 16(4):57-63.
- Brainard, Jeffrey and Southwick, Ron. 2000. "Congress gives colleges a billion-dollar bonanza." *The Chronicle of Higher Education* (28 July):A29-31.
- Branscomb, Lewis M. 1982. "Industry evaluation of research." Pp. 68-81 in *Quality in Science*, edited by M.C. La Follette. Cambridge: MIT Press.
- Braun, Dietmar. 1993. Who governs intermediary organizations? Principal-agent relations in research policy-making. *Journal of Public Policy* 13 (2): 135-62.
- Breyer, Stephen G. 2000. "Science in the courtroom." *Issues in Science and Technology* 16(4):52-56.
- \_\_\_\_\_. 1998. "The interdependence of science and law." *Science* 280(5363):537-38.
- Brooks, Harvey. 1982. "Science indicators and science priorities." Pp. 1-32 in *Quality in Science*, edited by M.C. La Follette. Cambridge: MIT Press.
- \_\_\_\_\_. 1968.
- Browner, Carol M. 1994. Memorandum on Peer Review to Assistant Administrators, General Counsel, Inspector General, Associate Administrators, Regional Administrators, and Staff Office Directors. 7 June. Mimeo. Available at <http://www.epa.gov/ordntrnt/ord/spc/memo0607.htm>.
- Cal/EPA. 1998, March. "Policy and Guiding Principles for External Scientific Peer Review."
- Carnegie Commission on Science, Technology, and Government. *Science and Technology in Judicial Decision Making: Creating Opportunities and Meeting Challenges*. New York: Carnegie Commission.
- Caswill, Chris. 1998. Social science policy: Challenges, interactions, principals, agents. *Science and Public Policy* 25(5)286-86.

#### *4 Public Research, Innovation and Technology Policies in the US*

- Chubin, Daryl and Hackett, Edward. 1990. *Peerless Science: Peer Review and US Science Policy*. Albany: SUNY Press.
- Council of State Governments (CSG). 1999. *A State Official's Guide to Sound Science*. Lexington, KY: CSG.
- England, J. Merton. 1982. *A Patron for Pure Science: The National Science Foundation's Formative Years, 1945-57*. Washington, DC: National Science Foundation.
- Environmental Protection Agency (EPA). 1998. *Science Policy Council Peer Review Handbook*. Office of Science Policy. EPA 100-B-98-00. Washington, DC: EPA.
- \_\_\_\_\_. 1997. "Peer Review and Peer Involvement at the U.S. Environmental Protection Agency." Available at <http://www.epa.gov/ordntrnt/ord/spc/perevmem.htm>.
- \_\_\_\_\_. 1992. *Safeguarding the Future: Credible Science, Credible Decisions*. EPA/600/9-91/050. Washington, DC: EPA.
- General Accounting Office (GAO). 1999. *Federal Research: Peer Review Practices at Federal Science Agencies Vary*, GAO/RCED-99-99.
- \_\_\_\_\_. 1994. *Peer Review: Reforms Needed to Ensure Fairness in Federal Agency Grant Selection*, GAO/PEMD-94-1.
- 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.
- Guston, David H. 2000a. *Between Politics and Science: Assuring the Integrity and Productivity of Research*. New York: Cambridge University Press.
- \_\_\_\_\_. 2000b. "Regulatory peer review." Paper presented at the annual meeting of the American Political Science Association, Washington, DC, 2 September.
- \_\_\_\_\_. 2000c. *What is Peer Review?* Report submitted to the American Chemistry Council, May.
- \_\_\_\_\_. 1999. "Constructing Consensus in Food Safety Science: The National Toxicology Program's Report on Carcinogens." Paper presented at the annual meeting of the American Association for the Advancement of Science, Anaheim, CA, 22 January.
- \_\_\_\_\_. 1996a. "New Technology Role for the States." *Forum for Applied Research and Public Policy* 11(3):38-44.

#### 4 Public Research, Innovation and Technology Policies in the US

- \_\_\_\_\_. 1996b. Principal-agent theory and the structure of science policy. *Science and Public Policy* 23 (4): 229-40.
- Hill, Christopher T. 1998. "The Advanced Technology Program: Opportunities for enhancement." Pp. 143-73 in Lewis M. Branscomb and James H. Keller, eds., *Investing in Innovation: Creating a Research and Innovation Policy That Works*. Cambridge: MIT Press.
- Intersociety Working Group. 1999. *AAAS Report XXIV: Research & Development, FY 2000*. Washington, DC: American Association for the Advancement of Science.
- Jasanoff, Sheila. 1990. *The Fifth Branch: Science Advisers as Policy Makers*. Cambridge: Harvard University Press.
- Kaiser, Jocelyn. "Scientific panel clears breast implants." *Science* (11 December 1998):1963-65.
- Kiernan, Vincent. "Project will supply 'independent' experts for suits involving scientific issues." *The Chronicle of Higher Education* (11 June 1999):A20-21.
- Kim, Kyung-Man. 1994. *Explaining Scientific Consensus: The Case of Mendelian Genetics*. New York: The Guilford Press.
- Latour, Bruno. 1987. *Science in Action: How to Follow Scientists and Engineers Through Society*. Cambridge: Harvard University Press.
- National Academy of Sciences (NAS). 1999. *Evaluating Federal Research Programs: Research on the Government Performance and Results Act*. Committee on Science, Engineering, and Public Policy. Washington, DC: National Academy Press.
- National Science Foundation (NSF). 1999. *What Is the State Government Role in the R&D Enterprise?* Division of Science Resources Studies. 99-348. Arlington, VA: NSF.
- National Science and Technology Council (NSTC). 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 Toxicology Program (NTP). 1999. "[What is the] *Report on Carcinogens?*" Research Triangle Park: NTP.
- Powell, Mark. 1999. *Science at EPA: Information in the Regulatory Process*. Baltimore: The Johns Hopkins University Press.
- Rip, Arie. 1994. The republic of science in the 1990s. *Higher Education* 28: 3-23.

#### *4 Public Research, Innovation and Technology Policies in the US*

- Roy, Rustum. 1985. "Funding science: The *real* defects of peer review and an alternative to it." *Science, Technology, & Human Values* 10(3):73-81.
- Runkle, Deborah. 1999. "Court appointed scientific experts: A demonstration project of the American Association for the Advancement of Science." *Federal Discovery News* (December):3-4.
- Sapolsky, Harvey. 1990. *Science and the Navy: The History of the Office of Naval Research*. Princeton: Princeton University Press.
- Savage, James D. 1999. *Funding Science in America: Congress, Universities, and the Politics of the Academic Pork Barrel*. New York: Cambridge University Press.
- Smith, Bruce L.R. 1992. *The Advisers: Scientists in the Policy Process*. Washington, DC: The Brookings Institution.
- \_\_\_\_\_. 1990. *American Science Policy Since World War II*. Washington, DC: The Brookings Institution.
- Southwick, Ron. 2000. "Is an Agriculture-Research Effort Dying on the Vine?" *The Chronicle of Higher Education*(18 August):A23-24.
- Sussman, Robert M. 1994. Memorandum on Establishment of Agency-wide Guidance for Conducting External Peer Review of Environmental Regulatory Models to Assistant Administrators, General Counsel, Inspector General, Associate Administrators, Regional Administrators, Deputy Assistant Administrators, and Deputy Regional Administrators (with attachment). Available <http://www.epa.gov/ord/spc/regmod1.htm>.
- van der Meulen, Barend. 1998. Science policies as principal-agent games: Institutionalization and path dependency in the relation between government and science. *Research Policy* 27 (4): 397-414.

## Discussion of David Guston's paper

**Peter Blair** (Sigma Xi, USA). What about the good old boy network of peer review?

**David Guston.** We don't have much information on changes in the peer review process. The literature does suggest a modest old boy network and gender effect in evaluation of proposals. There seems to be an indication that panel review is preferable to ad hoc review.

**Irwin Feller** (Pennsylvania State University, USA). We should reconcile the rise of earmarking of billions of dollars to the rise of peer review. The NSF EPSCoR program is peer reviewed, but it is set aside. Another example is the T21 Transportation bill. The NSF and NIH models are symbolic of peer review.

**David Guston.** We need to figure out more where numbers are coming from. Symbolism of peer review is important, but it does solve the problem of who to give the money to. It is a fair option when political controversy is too high for distributing resources in other manners. If the political environment changes, then there may be other ways to solve the problem. For more of the controversial issues such as EPA research, peer review may resolve controversies associated with these types of activities.

**Peter Blair.** Scientists may not be best people to solve problem of evaluating the societal issues of projects.

**David Guston.** NSF is just at the beginning stage of evaluating using a two criteria model. The NIH attempted to stay in touch with the social relevance by having a committee structure for the technical/scientific merit and another committee comprised of more public people to look at the broader societal issues. This committee may also have lay people who have some technical credentials but who may not be scientists, etc.

**Stefan Kuhlmann** (Fraunhofer Institute for Systems and Innovations Research). ATP has a technical review panel and another review panel so that even if the scientists love the "whizbang" technology, this can be balanced by an educated layperson review panel that looks at its overall impacts. These two panels have separate review processes that are then brought together. A project can lose if it doesn't meet both areas.

**David Guston.** The NSF review forms do not include an area for societal impacts.

#### *4 Public Research, Innovation and Technology Policies in the US*

**Maryelle Kelley** (National Institute of Standards and Technology, USA). Define what you mean by peers. In European programs, more and more assessment panels are represented by the public as well as scientists. This reflects a growing trend to fund generally societal-beneficial programs, but it is difficult to find appropriate peers in this realm.

**David Guston.** My paper doesn't deal with who these people should be but it is not difficult to address. It may answer the question of what the jurisdiction is by defining who the peers are. Evaluative committees are set up to judge proposals in some service programs which is not very different from what's going on in science review. Both are based on communal merit review. This can expand into societal merit review.

## Commentary on the Session and Additional Discussion

**Arie Rip** (Twente University, The Netherlands). The title of this section suggests that a framework could be an articulation of a theory. Here are my comments on the papers:

Kelley's paper creates a picture of the ATP program. The program has been pushed toward specific questions to justify its existence. The pressure to show success has been important, but also shows some very interesting findings including the problem of skewed distributions in the capabilities of firms to do something meaningful with the money received.

Dietz's paper could be linked with other work such as SRI studies to give more insight into its results. Another suggestion is to look at larger populations and possibly case studies as these can convince sponsors/critics by providing more details about the program. It is not enough to have just a few interesting case studies. Using multi-modal evaluation methods is important.

Guston's paper looks at the expansion of peer review into a new jurisdiction. Peer review can be mobilized for certain resources but it's difficult to get a handle on what peer review does for you as an evaluator. Why is peer review expanding in the US? The shift from responsibility for decisions (by courts, regulatory environment) has been seen in Europe less visibly because Europe has never been very clear about responsibilities for decisions. The US starts in a different vein. It does imply that some things happening in Europe could be interesting in US. For instance, as soon as a proposal is made, it is immediately fragmented. There is a danger in evaluation, though, by copying what others are doing abroad.

**Barry Bozeman** (Georgia Institute of Technology, USA). On several occasions, I have conducted reviews for the EPSCoR program. The program is designed to give money to the research "have-nots." There is recognition of social and economic impacts of research programs. He has been invited to review proposals in other areas but not in his area of specialty. What is EPSCoR's peer review process?

**Irwin Feller**. Because EPSCoR is a NSF program, peer review is a way to fundamentally legitimize the program within NSF because it is not a very popular program. Peer review should also select the best programs for funding. This holds true for the NSF. EPSCoR has spread to 6-7 other agencies. The objective of other agencies in this program is not necessarily the same, but is directed by Congress.

**David Guston**. This is reflective of the way these agencies operate.

**Liam O'Sullivan** (European Commission). In reference to Kelley's account of NIST, the focus on control groups has an appeal for economic evaluation, but there's a difficulty with control groups. He can't think of a European parallel where there is a significant focus on outcomes. Notions of additionality that have been developed in EU may be useful in the US. A project has a different quality (not just whether the project would have taken place without funding). Projects have other values such as training and dissemination of results yet the ATP seems to be focused only on the direct economic benefits without considering other impacts.

**Maryellen Kelley**. After the mad cow disease situation, there was a feeling that scientific advice itself should be put out to peer review. Many peer reviewers say they are not sufficiently equipped to evaluate socio-economic impacts.

**Terttu Luukkonen** (VTT Technology Group, Finland). Looking at socio-economic impacts can also have negative impacts. People use phrases from other projects just to get funded. There is a type of hypocrisy in evaluating proposals.

**Maryellen Kelley**. Additionality of behavior is something that I would like to push because product development is perhaps too far afield. The importance of collaboration and willingness of firms to share information and form partnerships are types of additionalities that they should consider more in-depth. Yet they have the problem of what to look at. For the R&D tax credit question, the real answer is that they are not really supporting research with the credit, but are instead supporting pre-competition, which has a benefit to other firms.

## 5. Evaluation of Large-Scale Programs in Europe

Session Chair: Luke Georghiou (University of Manchester, UK)

---

### Evaluation of the Brite/Euram Program

**Laurent Bach, Marc-Jacques Ledoux, Mireille Matt**

Bureau d'Economie Théorique et Appliquée (B.E.T.A.), L. Pasteur University of Strasbourg, France

Email: bach@cournot.u-strasbg.fr

---

#### Introduction

The article deals with the experience of the BETA group in the evaluation of large RTD programmes, and notably the EU RTD ones (see Table 1). This experience mainly relies on the design and the use of an original evaluation method, which aims at evaluating the economic effects of such programmes at a micro level (i.e. the participants to the programme) by means of direct interviews of participants involved.

We will first describe the so-called BETA methodology, putting the emphasis on the definition of the economic effects and the way to quantify them. Then the types of results that can be obtained will be showed and illustrated by the largest study performed by BETA about the EU RTD programme, that was focused on the Brite-Euram programme. Finally we conclude by discussing the relevance of the Beta method as well as the lessons learned from the evaluation works.

#### Part 1. Methodology of evaluation

##### *Overview*

The BETA approach is basically usable for the evaluation of RTD programmes partially or wholly funded by public money. In this perspective, RTD programmes are the ones which exhibit the following features:

- a programme generally includes different projects
- there is an explicit agreement about the objective of each project, expressed in technical and possibly economic terms
- a programme has a limited duration
- firms are involved, possibly in cooperation with university or research labs from other institutions (thereafter called participants or partners); and
- obviously a R&D activity has to be performed by (at least one) participants.

<b>PROGRAMMES EVALUATED BY BETA TEAM</b>	
EUROPEAN SPACE AGENCY	1980
EUROPEAN SPACE AGENCY	1988
EUROPEAN SPACE AGENCY - TT	1991, 1996, 2000
EU / BRITE EURAM	1993-1995
EU / ESPRIT-HPCN	1997
<b>OTHERS:</b>	
DENMARK (ESA Prog.)	1987
CANADA (ESA Prog.) with HEC Montreal)	1990 & 1994
Private Co - France	1992
ANVAR - France	1995
MATERIALS IRELAND -Ireland	1995
PETROBRAS Co - Brazil (with DPCT Unicamp)	1999
BRAZILIAN Space Prog. (DPCT Unicamp)	in progress

**Table 1**<sup>1</sup>

The evaluation is limited to the participants to the programme: what is evaluated is the economic impact (or the economic effects) generated by and affecting these participants.

The approach is based on a microeconomic approach: economic effects are identified, evaluated in monetary terms at participants' level, and then aggregated. From an empirical point of view, it means that in most cases representative sample of participants have to be built up and investigated, and confidentiality of information at participant's level has to be protected. Information about the effects are gathered through direct interviews of participants.

The evaluation has two main goals: providing a minimal estimation of the effects on one hand, and on the other allowing to better understand how the effects are generated and more generally how innovation processes are generated by large RD programmes and are creating economic value.

---

<sup>1</sup> The results of the Brite-Euram study presented further have been published by the EC in [B.E.T.A., 1993]. Results of other studies are provided and commented in [BACH et al. 2000, 1997, 1995, 1994, 1993, 1992] and [FURTADO et al., 1999] or could be found in the study reports (list available at BETA).

## *5 Evaluation of Large-Scale Programs in Europe*

### ***Scope of the evaluation: definition of direct and indirect effects***

Two types of effects are distinguished by BETA: direct and indirect effects.

#### *Direct effects*

The direct effects are the effects that are directly related to the objectives of the research projects, as they were defined at the beginning of each of these projects. For instance, if the objective is to develop a new product or a new family of products, the sales of such products are considered as direct effects; correspondingly, if the objective is to develop a new process, the economic effects of the use of this new process are considered as direct effects. This rule is not modified in the case of more fundamental research-orientated projects: direct effects are related to the application of the new scientific knowledge or the new technologies in the field foreseen at the beginning of the projects; only the range of possible direct effects may be enlarged in these cases, since the fields of application may be broadly defined.

#### *Indirect effects*

In comparison to direct effects, indirect effects are those which go beyond the scope of the objectives of the projects; generally speaking, indirect effects are derived from the use of what has been learned during the execution of the project, in participant's activities which are not directly related to the objective of the project. All types of learning leading to the creation of all types of knowledge are taken into account: technological, organisational, networking, management, industrial, etc. This is probably the main feature of the approach since it provides a considerably detailed view of how a RD activity performed in the framework of a public programme affects the learning processes of the participants.

Indirect effects have been broken down into four sub-categories: technological effects, commercial effects, organisation and method effects, work factor effects.

Technological effects: These effects concern the transfer of technology from the project to other activities of the participant. Following definitions proposed by the "evolutionary economics" and "knowledge economics", the term "technology" here encompasses artefacts (products, systems, materials, processes...) as well as codified, tacit, scientific, technological, etc... knowledge (apart from methods, see Organization and Method Effects). What is transferred can therefore be of a very diverse nature, from scientific expertise to worker's know-how, including technology laid down as a blueprint, new theories or "tricks of the trade"; this broad approach is one of the originalities of BETA methodology. The transfers lead to the design of new or improved products, processes or services that allow the participant to achieve new sales, to protect existing market shares, or to obtain new research contracts.

Commercial effects: Commercial effects basically take the form of increased economic activities (sales of products and services or new research projects) that do not incorporate significant technological innovation coming from the project itself. This

## *5 Evaluation of Large-Scale Programs in Europe*

can be achieved in two ways. First the network effects refer to the impact of projects on the cooperation among economic actors (firms, research centres, universities). Some of these effects concern the establishment of business links between participants of the same consortium, which leads to the continuation of commercial or technical collaboration after completion of the project. The same types of cooperation could also be set up between participant(s) and organisms or firms not involved in the project, for instance with a supplier of another participant, or thanks to a conference or workshop organized by the public organization in charge of the management of the evaluated programme. Second, by working on behalf of a given public programme, participants sometimes acquire a quality label, which is afterwards used as a marketing tool.

Organization and Method Effects: Organisation and method effects (subsequently referred to as O&M effects) occur when experience gained through the project allows the participant to modify its internal organization and/or to apply new methods in project management, quality management, industrial accounting.... .

Competence & training (or “work factor”) effects: These last indirect effects are of a different nature to the first three. They tend to describe the impact of the project on the “human capital” of the participant. Each of the participating organizations masters a certain range of competences related to more or less diversified scientific and technological fields, which form what has been called the “critical mass” or the “knowledge base” of this organisation. The impact of the project on this “critical mass” constitutes the work factor effect. In other words, the aim is to differentiate between routine work and innovative work that really makes the technological level of the participant increase or diversify.

### *Quantification of the economic effects*

Direct effects and most of the indirect effects are expressed in terms of added value generated by sales and cost reductions that have been achieved thanks to the knowledge gained by the participants during the evaluated programme. Only gross effects are quantified: additional costs of transfers, industrialization, marketing and s.f. are not deducted (they anyway are impossible to measure of each case of effect, because of the size of our sample). Cash flows or income are not calculated, which means that the results provided by such a study cannot be directly used for classical financial analysis of rate of return or profitability.

Only real sales are quantified, and not the size of the markets on which the product/service could eventually be sold. When the transfer of technology or method, or the commercial effects are only partly influencing sales or cost reductions, the value of the corresponding effect only amounts to a share of those sales or cost reductions. This share is in proportion to the influence on those sales or cost reductions of work performed within the framework of the evaluated project (“fatherhood coefficients” are thus used). When such an evaluation is complex, a two-step process is used: first, the influence of one parameter is evaluated (influence of technological aspect, commercial aspect etc, following the logic of the indirect effects classifica-

## 5 Evaluation of Large-Scale Programs in Europe

tion), then the evaluation of the specific influence of the evaluated project on this parameter is evaluated.

There are two exceptions to the quantification method described above in terms of sales or costs reductions leading to added value increase. Firstly, in the case of patents that are not protecting existing product or process, the minimum estimation of the value of indirect effects was provided by the amount spent by the participant to register and keep patent “alive”. Secondly, in the case of quantification of the work factor, a rough estimation based on the use of a proxy value is adopted by the BETA group. The method consists of: a) isolating people who contribute to the technological capacity of the participant by the increase in their competences (which may enlarge or diversify the “critical mass” of the participant); b) evaluating the time spent by these people in truly innovative activity; c) for reasons of homogeneity, quantifying the effect in monetary terms by taking into account the average cost (including overheads) of these engineers or technicians over the time period estimated at the previous step.

TYPE OF EFFECT	QUANTIFICATION
DIRECT EFFECTS	sales / cost reductions
INDIRECT EFFECTS	
<i>Technological</i> Transfer of product-related techno Transfer of process-related techno Transfer of service-related techno Patents	sales / new research contracts cost reduction / new research contracts sales / new research contracts cost of patenting (proxy value)
<i>Commercial</i> Network effect Reputation effect	sales / cost reductions / new research contracts
<i>O&amp;M</i> Project management Other methods Organization	cost reductions cost reduction cost reduction / sales / new research contracts
<hr style="border-top: 1px dashed black;"/> <i>Competence &amp; training (work factor)</i>	monetary equivalent of man-hours(proxy value)

**Table 2**

### *The sampling of projects and participants*

The methodology is based on a sample of projects that should be representative of the full project population. To tackle this difficulty it is necessary to choose a set of objective parameters which can be obtained independently of the results and which also correspond to some criteria operating in the definition or in the objectives of the R&D programme (for instance size of the firm, proportion of universities, or nationality of participants).

***The gathering of information***

The BETA methodology is based on direct interviews with managers of the participants to the evaluated programme. Each interview is made by two members of the BETA team. The managers interviewed are those responsible for the project within the participant; in many cases, especially in small and medium-sized firms, this manager is the DG of the firm or the Technical, Engineering or R&D Manager. Prior to each interview, the BETA team sends some information about the study and the methodology of evaluation to the managers. The data are kept strictly confidential by BETA.

***The meaning of results provided in terms of ratios***

In addition to the evaluation of direct and indirect effects in real value, some results are summarized in the form of ratios or coefficients. These figures represent the ratio between the amount of a certain type of economic effect generated by a given set of participants and the total payment made by the funding public body to the corresponding set of participants (i.e. the contribution of the public body to the participants' R&D budgets). It means that on average, for the sample of participants studied, every 100 units financed by the public body result in a minimum amount of added value generated by the participants equal to 100 times the coefficient.

It is important to emphasize that BETA ratios are directly comparable neither to classical CBA ratios nor to investment appraisal ratios such as ROI, for the following reasons:

- as it has been said, all costs are not taken into account (for instance investment made by the firms to transfer and adapt technology derived from the evaluated programme);
- there is no consumer surplus added;
- economic effects are expressed in added value generated by the participants and not in sales, cash flows or net income.

It is obvious that the last type of indirect effect (*work factor effects*) is of a different type from the three other ones, because of its very nature and of the way used to quantify it. Following the standard way of showing the results, all effects are added in order to compute ratios, which does not really matter since the ratios are not CBA or Investment appraisal ratios. But the reader can easily separate the work factor figure from the others.

***The principle of minimum estimate***

The measure of economic effects performed by the BETA group must be considered as the minimum estimates of these effects, for the following reasons. All estimations of figures provided by participants (rate of added value, coefficient of influence of the project, time spent on innovative activity, probability of achievement) are systematically minimized. That is to say that these estimations are expressed most of the time as a range from which only the lower boundary is used for compu-

## *5 Evaluation of Large-Scale Programs in Europe*

tation. Some effects cannot be measured, for instance because the influence of the project exists but is not separable from other factors. In spite of the time spent in interviewing people, some cases may also escape the interviewers, for instance when the technical aspects are very complex or when the firm has forgotten. Finally, in spite of the guarantee of confidentiality provided to all participants interviewed, they may still be reluctant to divulge very strategic information.

More generally, it should be underlined that the BETA group methodology aims to assess only the economic effects for the participant and not the effects benefiting the rest of the economy. These are mid- or long-term effects, which may consist of diffusion of technology (through imitation, technology transfers, staff mobility, ...) or increase in consumer satisfaction. Thus the effects measured by the BETA group are only a sub-set of the global economic effects of the projects.

### *The time period covered by the evaluation*

For each project studied, the time period covered by the evaluation obviously starts with the project, since some effects may appear at the very beginning of the research work. For the end of the time period, two years forecast are generally added.

## **Part 2. The evaluation of EU BRITE EURAM programme**

In this part, we follow two objectives: showing the type of results that can be obtained from evaluation of the BETA-type, and illustrate this point by the results of the largest study made by the BETA for the EU, that is the so-called Brite-Euram study.

### *The EU BRITE EURAM programme*

Among other European R&D cooperative programmes such as Esprit or Race, the Brite-Euram programme has been set up by ex-EEC since the First Framework Programme (1984-1987) and has been going on for almost fifteen years through the following Framework Programmes (although under different names). The first Brite (for Basic Research in Industrial Technologies for Europe) programme was set up in 1985 to develop applications of new technologies and materials in industrial sectors. In 1986, it was followed by Euram (European Research on Advanced Materials) programme, aimed at stimulating the development of new materials. Approximately 300 R&D projects were selected from both programmes up to 1989. In 1989, the two programmes were grouped together in the Brite-Euram I programme (1989-1992), for which 368 new projects were added. The projects at stake are cooperative; they are proposed by industry, and selected by EU which generally funds 50 % of the cost for firms, and 100 % for universities. 3 to 10 partners are associated in each project, one of those being "prime" or leader partner. In brief, all projects: a) are transnational (involving at least two independent partners from different EU Member States), b) are focused on R&D, c) last 3 to 4 years, d) involve at least one university lab or one research centre, e) cost in average 1 to 2,5 MECU.

## *5 Evaluation of Large-Scale Programs in Europe*

The main BETA study is based on an evaluation of a representative sample including 176 participants involved in 50 projects in the EU BRITE, EURAM and BRITE EURAM 1 programmes (BETA, 1993). Three criteria have been taken into account: the nationality of participants, the type of programmes (Euram, Brite or Brite-Euram I), and the type of partners (big firms, SMEs, universities and research centres). This study has been followed by two other ones, focusing on SMEs, “small countries” and technological transfers.

### ***Main results***

The 176 contractants of the sample received 39.4 MECU (all results are expressed in ECU base 1991). 611 economic effects have been measured while in addition approximately about 300 have been identified without being quantified (mainly due to lack of information). The table 3 summarizes the results computed for the full sample.

### *Direct effects*

522.5 MECU will have been directly generated at the end of 1995 (413.3 MECU until the end of 1993). The corresponding ratios direct effects/EU funding amount to 13.3 and 10.5 respectively. 54 cases of direct effects have been measured<sup>2</sup>, and were observed in 37 firms out of 113, and in 2 Research Centres which should not generate such direct effects as they are non-profit-making organizations. These 39 contractants were involved in 22 projects out of 50. The distribution of these direct effects in terms of ECU is wide, the smallest amounting to 12 190 ECU and the biggest to 250 MECU.

It should be stressed that the existence of such direct effect put in question the concept of “pre-competitiveness” which is one of the criteria used by EU to select the B/E project. Research carried out in the framework of B/E projects is not pre-competitive since it directly generates commercialisation of products or processes. However, when one gives up linear conception of innovation in which basic research roughly corresponds to pre-competitive research, it is impossible to say what is pre-competitive and what is not<sup>3</sup>.

---

<sup>2</sup> Two negative effects have also been evaluated.

<sup>3</sup> Note that new definitions of the concept of pre-competitiveness have appeared both in the US and the EC. They mainly introduce the technological risk, but are hardly practical.

Number of projects surveyed	50
Number of participants	176
Number of measured economic effects	632
Ratio of direct effects/EEC fundings	13.3
Total direct effects in MECU 91	522.5
Ratio of indirect effects/EEC fundings	4.1
Total indirect effects in MECU 91	160.8
Technological	76.5 (47.6 %)
Commercial	16.5 (10.3 %)
Organisation & Methods	18.6 (11.6 %)
Competence and training	49.2 (30.6 %)

**Table 3**

*Indirect effects*

160.8 MECU will have been indirectly generated at the end of 1995, which leads to a ratio of 4.1 (132.2 MECU at the end of 1993, ratio = 3.4). 555 indirect economic effects were measured. Most of the contractants, 155 out of 176 (88 %), have generated indirect economic effects. As for the direct effects, the distribution is wide, the smallest amounting to about 1 000 ECU and the largest reaching 20 MECU. They are in general much smaller than the direct effects and follow a decreasing distribution.

The indirect effects evaluated in terms of added value on sales and of reduction of costs represent almost 70 % of the total, the remaining 30 % being attributed to increases in competence and training that are evaluated through proxy values.

The technological indirect effects (47.6 %) are divided into four categories: product transfer (43.4 %), process transfer (53.4 %), service transfer (1.5 %) and portfolio of unused patents (1.7 %). The domination of the process transfer is consistent with the definition of the research programmes which are more process than product oriented. On average very few patents were drawn from these research projects and only 10 patents without market applications were still held by the contractants. The commercial indirect effects (10.3 %) are equally spread between network (51.7 %) and reputation (48.5 %) effects. It is less than could have been expected when it is considered that one of the main objectives of these research programmes was to create a network of industries throughout Europe. However, more fine analysis show that the main beneficiaries of network effects are the SMEs and the “small” countries (Greece, Ireland and Portugal), for which the percentage of commercial effects amount to 33 % and 62 % respectively.

The organization and method indirect effects are mainly due to the organization transfers 85.6 % versus 8.1 % for method transfers and 6.3 % in management cost reductions. It should be stressed that for most of the cases, the transfers of method result in an improvement in the quality of the final products, in other words 1.5

## 5 Evaluation of Large-Scale Programs in Europe

MECU of indirect effects are due to a better quality of products. This was not included in the objectives of the research projects.

49.2 MECU correspond to the minimum evaluation of the gain in competence (86.5 %) and training (13.5 %) of the partners belonging to the sample, due to their participation in the EURAM, BRITE and BRITE-EURAM I programmes.

### *Direct versus indirect effects*

In this section, we try to see whether there is a correlation between direct and indirect effects, i.e. whether the generation of direct and indirect effects are simultaneously or mutually exclusive.

39 firms and research centres were the champions in generating direct effects. The question arises : are they better than the others at generating indirect effects ? The answer is “yes” for the big firms, “no” for the SMEs and “no difference” for the research centres (but the statistics are not significant for these centres). In addition, the direct effects are accompanied by large technological indirect effects (technological transfers).

	39 contractants with direct effects			107 contractants (no universities) without direct effects		
	Ratio of indirect effect	Number of contractants	% technological indirect effects	Ratio of indirect effect	Number of contractants	% technological indirect effects
Big	9.2	25	69 %	3.1	50	39 %
SME	1.4	12	11 %	3.5	26	19 %
Research centres	2.9	2	-	3.0	31	-

**Table 4**

Differences in access to human, technical or financial resources between large and small firms obviously explain this difference in behaviour. While large firms can economically exploit results of RD projects in different ways, SMEs are very often compelled to concentrate their efforts on one type of results.

Note that this strong link between direct and indirect effects within the big firms is not shared with the other members of the consortium. If the 50 consortia are divided into two groups: group one, consortia containing at least one participant with at least one direct effect; group two, consortia without any direct effects, the first group shows only a small advantage over the second (ratio of 4.6 versus 3.5).

## 5 Evaluation of Large-Scale Programs in Europe

### *Economic effects and technological or scientific success and failure*

The technological or scientific success or failure of a project has been evaluated by the BETA team and not by the contractants. Of the 50 projects in the sample, one was abandoned six months after it started, 38 were a success and 12 a failure. We consider that a project is an economic success when it generates economic effects.

Parameters	Success Technological/ scientific	Failure Technological/ scientific
Number of contractants Total EEC funding MECU 91	136 30.1	40 9.2
Total Direct effects MECU 91 Ratio Direct effects/ EEC money	516.6 17.1	5.9 0.6
Total Indirect effects MECU 91 Ratio Indirect effects/EEC money	145.3 4.8	15.5 1.7
Technological	50.3 %	22.5 %
Commercial	10.1 %	12.2 %
Organisational and method	11.3 %	14.4 %
Competence and training	28.4 %	50.9 %

**Table 5**

There are no technological/scientific successes with only direct economic effects or without any economic effects. There is only one case of total failure, the 11 other technological/scientific failures generate at least indirect economic effects and, even, in two cases direct and indirect economic effects together (Table 5). The contractants (firms + research centres + universities) belonging to technologically or scientifically successful projects obviously generate many more economic effects than the contractants belonging to technologically/ scientifically failed projects (see table above). However it is interesting to note that this last group generates 5.9 MECU of direct effects (ratio 0.64) and 15.5 MECU of indirect effects (ratio 1.68).

This result clearly shows that R&D which do not directly leads to commercialised product or process could be profitable for the firms, even in the short term. From a methodological point of view, it also points out the interest of tracking the so-called indirect effects and not only direct effects.

### ***Detailed results***

According to the programme evaluated, it is possible to cross different qualitative features of the participants and of the projects with the quantitative results, in order to enrich the analysis and providing a finer understanding of the innovation processes triggered by the RTD programme. A lot of the results drawn from the studies are in line with the features of modern innovation processes, already mention for instance in the seminal works from [KLINE S.J. - ROSENBERG N., 1986] or [ROTHWELL R., 1994]. Basically, three type of characteristics can be distinguished: characteristics of participants (for instance nature, size, nationality, etc),

## 5 Evaluation of Large-Scale Programs in Europe

characteristics of the project (for instance cost or duration) and characteristics of the involvement of the participants in the project (for instance type of research performed, level of responsibility, money received, etc). Moreover, when projects are collaborative ones, it also very interesting to try to analyse the combination of participants' intrinsic characteristics or involvement characteristics, in other words to take into account the structure of the network set up for each project. Some examples of such analysis are provided below, again taken from the Brite-Euram study.

### *Nature of the participants: size of the firms*

It is obvious that SMEs cannot compete with the big firms in terms of direct and indirect effects (Table 6). However that does not mean that the situation is catastrophic for the SMEs; they managed to generate 17.9 (direct) + 24.9 (indirect) MECU from 9.0 MECU received from the EEC (ratio = 4.75). Comparatively to big firms, SMEs show more work factor effects, less O&M and above all technological effects, while their only strong position is the amount of commercial indirect effects. They have fully used their membership of a new network to generate economic effects up to 8.3 MECU, about 19 % of the total effects. We can also notice that SMEs generate proportionally more product transfer and less process transfer than big firms, and beneficiate more than those latter form experience in project management.

Parameters	Large firms	SMEs	Reserach centers	University labs
Number of partners	75	38	33	30
Ratio : direct effects/EC funds	25.3	2	0.1	0
Total direct effects **	503.7	17.9	1.0	0
Ratio : indirect effects/EC funds	5.4	2.7	3.0	2.0
Total indirect effects **	108.5	24.9	19.8	7.6
Technological	59%	17%	28%	32%
Commercial	5%	33%	11%	9%
Organization & methods	14%	5%	10%	3%
Competence & training	22%	45%	51%	56%

\* staff less than 500

\*\* in 1991 MECU

**Table 6**

Their main weakness is due to their small size, in absolute terms, both in critical mass of competence and in financial terms. Two signs among many others, extracted from the results, well illustrate this handicap.

When a firm generates direct effects, it is an indicator of its competence and its ability to transform a research project into a market product. This firm should generate more indirect effects (as is largely confirmed by the performances of the big firms) and among these indirect effects, a large part should be technological transfers (product or process). To generate these indirect effects they do not need large extra investments, but a large spectrum of competences. The SMEs do not have these competences and the champions at generating direct effects are not able to generate even the same amount of indirect effects as the other SMEs (ratio 1.4 versus 3.5). In

## *5 Evaluation of Large-Scale Programs in Europe*

other words, SMEs seem to be able to generate direct or indirect effects, but not both together.

The critical lack of financial size is well illustrated by the difference between big and SME in terms of direct effects (25.3 versus 2.0). For a big firm, in most cases the extra investment necessary to put the product of the research on the market was very small when compared to the total turn-over (generally well below 1 %); for the SME, even for extra investment equal to the amount of money previously invested in the research project (generally in a ratio 1 to 1 and not 1 to 10 or 1 to 100 as is often put forward), the barrier is too high and the technologically successful project is abandoned.

The sample (38 SMEs) of the study completed in 1993 is too small to further explore the analytical situation of these small firms. An extension of the study to 64 other SMEs carried out by BETA has firstly allowed a new classification-definition, pertinent to the analysis of economic effects generated by the SMEs participating in the BRITE-EURAM programmes and probably also pertinent to any other R & D programme. The SMEs can be divided into four families according to their degree of independence, while keeping as a basic definition less than 500 employees and an annual turnover of less than 40 million Ecu: the SME subsidiaries of private industrial groups with more than 30 % of their capital in the hands of a big industrial group, the SMEs grouped in an "holding", the "true" SMEs in which less than 30 % of the capital belongs to an industrial group and finally the nationalized SMEs. The study shows that only the private subsidiaries SME, or, to a lesser extent, those grouped in a "holding", can generate direct economic effects from the R & D programmes since they have access to financial and human resources of large enough size, either from their parent company or thanks to their holding. The blunt conclusion regarding the "true" SMEs tends to confirm that this form of firm does not have the dimensions necessary to directly use research to innovate. However, the table is not totally negative since the "true" SMEs generate some indirect effects that they were able to obtain by applying some of the results of their participation in BRITE-EURAM to products that already existed. Some varying performances are observed according to different criteria, such as nationality, but these are really minor when compared to the criterion of independence already described.

### *Nature of the participants: research centres and universities*

The economic effects generated by research centres and universities were one of the surprises in this study (Table 6). These two kinds of organization should not generate direct effects because of their status in the programmes. They are considered as non-profit-making organizations and should work at marginal cost. If the universities never sold anything, this was not the case for all of the research centres. Two of these centres have raised 1 MECU of direct effects.

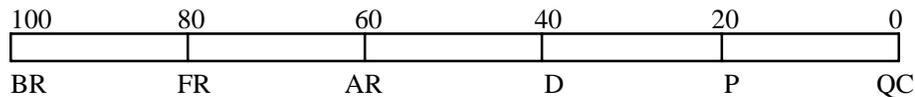
If the work force effects are left aside, one can observe that the rest of the indirect effects reach 9.6 MECU for research centres and 3.3 MECU for universities, mainly through new research contracts from industry, national governments and the EEC. The increase in competence and the training effect equal 10.2 MECU for the re-

## 5 Evaluation of Large-Scale Programs in Europe

search centres and 4.3 MECU for the universities. These two figures strongly underestimate the real indirect effects that will be generated in the future. Indeed, the increases in competence are considered by the participants as very important, mainly because of the discovery of new fields of research directly connected to the industry with many future contracts in good perspective, and also because of the fact that the people who have acquired these new competences stay in the laboratory for a very long period (on average more than 10 years) and hold high hierarchical positions.

### *Nature of the involvement of each participant: effect of the nature of the research and upstream participation vs. downstream participation*

During the interviews each contractant was asked to position his own research programme in the project on a scale which determines a typology of the research. It should be stressed that this scale does not imply that the linear model of innovation is recognized. The scale was designed as follows:



- BR basic research without any predictable field of industrial application; (e.g. : how old is a star ?).
- FR fundamental research with a possible field of industrial application ; (e.g.: mechanism of reaction in chemistry).
- AR applied research directly devoted to an industrial problem ; (e.g. : modelling of a multiphasic flow in a chemical reactor).
- D development (e.g.: building of a prototype).
- P process (e.g.: improvement and uses of a software for design).
- QC quality control.

In the same project, the individual contractants can be located in different places on the scale; for instance a university lab can be at 60 (AR), a big firm at 80 (FR), a research centre at 50 (between AR and D) and an SME at 20 (P).

The distribution of the 176 positions determined by the participants is centred around 60, showing that the three programmes are principally applied research oriented. On the basis of this typology, three types of analysis have been carried out.

The classification of partners using the research scale presented above was first used at the partner level. Accordingly, the contractants were divided into two groups: upstream where they estimate their location to be above 59 (including Applied Research) and downstream for those located below 60. The effects generated by these two groups are very different (Table 7). The number of measured effects per contractant is slightly higher for downstream research (3.9 vs 3.5). But the most interesting point is the large difference between the two groups when direct and indirect effects are compared. Upstream research generates twice the amount of indirect effects as downstream research (5.0 vs. 2.8) while the situation is com-

## 5 Evaluation of Large-Scale Programs in Europe

pletely reversed for the direct effects (8.5 vs. 21.0). Downstream research, generally close to the process, is obviously generating direct sales but few indirect effects because it is less generic and more specialized, contrary to upstream research. Another important difference which should be stressed is the very small proportion of technological indirect effects generated by downstream research (30 %). This confirms that such research is not generic.

But far more interesting are analysis made at the consortium level. By fact, on one hand the two situations (upstream and downstream) co-exist within a lot of consortia, and on the other hand, the breakdown of research activities between partners differ from one consortium to another. We then have tried to investigate i) whether the involvement of one partner in fundamental research has an impact on the effects of its partners, and ii) whether the breakdown of research activities has an impact on the observed effects of the different partners.

Parameters	Upstream	Downstream
Number of contractants *	104	68
Number of measured effects	366	262
Average effect per contractant	3.5	3.9
Total Direct effects in MECU 91	202.8	319.7
Ratio Direct effects/EEC funding	8.5	21.0
Total Indirect effects in MECU 91	118.5	42.3
Ratio Indirect effects/EEC funding	5.0	2.8
Technological	54 %	30 %
Commercial	11.5 %	7 %
Organisational and method	11 %	13 %
Competence and training	23.5 %	50 %

\* Four of them were not ranked.

**Table 7**

### *Structure of project: projects with fundamental research vs. projects without fundamental research*

Among the 50 projects, five consortia do not associate fundamental or basic research (100 to 70 on the research scale explained above) to the rest of the innovation process. The performances of the 19 participants included in these five consortia are far less successful both for direct and indirect effects as shown in the Table 8.

## 5 Evaluation of Large-Scale Programs in Europe

Parameters	Associated with F.R.	Not associated with F.R.
Number of consortia	45	5
Number of participants	157	19
Total Direct effects in MECU 91	505	17.5
Ratio Direct effects/EEC funding	14.8	3.3
Total Indirect effects in MECU 91	149.7	11.1
Ratio Indirect effects/EEC funding	4.4	2.1
Technological	50.5 %	8.9 %
Commercial	10.2 %	11.6 %
Organisational and method	12.2 %	3.1 %
Competence and training	27.2 %	76.3 %

**Table 8**

The nature of the indirect effects is dramatically different from one group to the other. The first group, where fundamental research is associated with the full innovation process, generates a large amount of technological transfers (50.5%), while the second group without fundamental research, generates mostly competence and training effects (76.3%). It should be underlined that fundamental research is not only the prerogative of the University; many other participants declared that a part of their contribution to the project included fundamental research.

These results are consistent with the conception of innovation as an interactive (Kline-Rosenberg type) rather than a linear process. Whatever is the objective of the research effected in a project (on new family of materials, on new application of existing materials, on enhancement of performance of existing process) the association of fundamental research seems to be required in order to generate economic effects. Correspondingly, basic or fundamental research cannot be dissociated (physically or practically) from the rest of the process if radical or incremental innovation is to succeed. As the good fundamental research university labs are nowadays largely financed by industry, even the possibility of having fundamental research outside the competitive organization is not feasible. These results may be regarded as in line with works explaining the new modes of relations between research and industry, putting for instance the emphasis on the problem solving focused attitude of research-industry collaborations ([GIBBONS M. et al., 1994]). It is thus interesting to mention that the scientific discipline and technological fields which were at the origin of the biggest amount of effects were related to applied physics, mathematics models and simulation, showing that this was really the capability of solving problem with the means of simulation tools based on strong scientific knowledge and methods which were crucial.

### *Structure of the project: breakdown of research activities*

The consortia have been split in three groups: techno success/direct and indirect effects, techno success/indirect effects only, techno failure/direct or indirect effects. We then have analysed the breakdown of research activities among the partners

## *5 Evaluation of Large-Scale Programs in Europe*

involved in each of those three groups. For this purpose, we used the classification presented above and analysed the range of research activities covered by the different projects (for instance, project A encompasses partners involved at 80, 65, 60 and 35 level thus the range is 80-35; project B encompasses partners all involved at 60 or 70 level then the range is 70-60, and s.f.). Results suggest that the most successful projects are those in which fundamental research is most widely associated with the other steps of innovation. 65 % of projects from group 1 (technological success with direct and indirect effects) encompass partners distributed over 40 or more units on the research scale (typically partners ranging from 80 to 40), against only 35 % for projects of group 2 and group 3 (technological success with only indirect effects, and technological failure with indirect effects, respectively).

### *Structure of the project: the role of universities*

A significant consortium effect has been identified in the analysis of the sample. The presence of a university (or better, a fundamental research institution such as CNRS, Max Planck, etc.) in a consortium has a very positive action on the generation of economic effects (Table 9). The positive effect is especially efficient for the direct effects but can also be observed for the indirect effects. On this last point it should be stressed that the main effect of the presence of a university in a consortium is to accelerate the process of generation of indirect economic effects. The influence of universities is also observed outside the firms, in the research centres but to a lesser extent.

Very strong qualitative differences are also observed in the spread of the indirect effects when universities are or are not associated with firms and research centres. Almost all the categories are affected. The presence of universities favours all technological effects in general, product transfers instead of process transfers, network effects instead of reputation effects, organization effects instead of method effects and even, to a lesser extent, competence instead of training.

It must be stressed that this positive impact of the involvement of universities on the generation of effects should not be mixed up with the importance of fundamental research described above. By fact, all universities do not systematically act at the fundamental research level; conversely, a lot of firms are involved in fundamental research work within the framework of the B/E project.

## 5 Evaluation of Large-Scale Programs in Europe

Parameters	Firms with universities	Firms without	Firms and research centres with universities	Firms and research centres without
Number of partners	49	64	60	86
Total Direct effects in MECU 91	312.1	209.5	312.2	210.4
Ratio Direct effects/EEC funding	24.4	13.0	20.4	10.3
Total Indirect effects before January 94	63.3	43.1	69.3	55.7
Ratio Indirect effects	5.0	2.7	4.5	2.7
Total Indirect effects before January 96	65.1	68.5	71.4	81.8
Ratio Indirect effects	5.1	4.2	4.7	4.0

**Table 9**

### *Structure of the projects: integration, specialization and user-producer interaction*

Three types of activities have been defined, according to the possible role of the partners within the project. These latter are thus referred as Producer (P) - of materials, equipment, instrumentation, ...; User (U) - of those materials, equipment, instrumentation, ...; and Researcher or Tester (T). Note that a single firm can run different activities within the same projects. The economic success of the integrated firms is at least four times greater for direct effects and almost three times greater for indirect effects, than that of the non-integrated firms. The group T contains mainly universities and research centres and therefore cannot be compared in economic terms. The results show that it is easier to market research results directly or indirectly, if all the competences are under the same roof.

Another interesting point is to check if a Producer and a User firm (two non-integrated partners at least) are bound together through a consortium, a “consortium effect” such as the one classically referred to as “user-producer interaction” (Lundvall B.-A., 1988) can replace the integration effect. Then the contractors have been divided into three groups according to the nature of their partnership collaboration :

- vertically integrated firms (producers + end-users + suppliers - PU+PUT)
- partners who are complementarily associated within the project (end-users and producers U+P associated)
- partners who are not complementarily associated (suppliers alone, producers alone or end-users alone-U+P not associated)

The analysis of the results shows that there is a slight consortium effect on the associated firms; but, this never reaches the efficiency of the integration effect (Table 10).

## 5 Evaluation of Large-Scale Programs in Europe

Parameters	INTEGRATION	CONSORTIUM	INDEPENDENT
	PU integrated	U+P associated	U+P not associated
Number of contractants	26	74	25
Total Direct effects in MECU 91	289.1	189.5	37.3
<b>Ratio Direct effects/EEC funding</b>	<b>40.2</b>	<b>10.7</b>	<b>5.4</b>
Total Indirect effects in MECU 91	58.7	67.3	15.3
<b>Ratio Indirect effects/EEC funding</b>	<b>8.2</b>	<b>3.8</b>	<b>2.2</b>

**Table 10**

However, this consortia effect which is to be linked to “user-producer” type of interaction is more effective when partners have already worked together before the B/E project. This underlines the importance of long-term relations in user-producer interaction ([following for instance works from LUNDVALL B.-A., 1988] or [VON HIPPEL E., 1988]).

### Conclusion

The evaluation proposed here has clearly shown the impact of R&D cooperative projects on the industry. It is important to underline that this impact is generated both directly and indirectly through many different ways within the partners involved in the projects. This shows the difficulty of the evaluation of success or failure of a programme, since it reveals that any evaluation limited to direct effects, or main partners, or short-term effects, or limited type of effects can be misleading.

The detailed analysis of the large data bases set up by BETA also highlight the impact of the nature of the participants and of the organisation structure of the analysed R&D projects on the amount of observed effects. For instance, the effects depends heavily on factors such as the participation of a university lab, the participation of at least one partner involved in a rather fundamental research work, or the diversity of research tasks over a scale ranging from fundamental research to industrialisation work. Other studies performed by BETA about the space programmes conducted by the European Space Agency ([BETA, 1980, 1988]) have also shown that the amount and the profile of indirect effects of a firm are largely correlated to both its size and its position in the industrial network set up for the purpose of the R&D projects (see [Zuscovitch E.- Cohen G., 19]). Another aspect illustrated (if not revealed in some cases) by the BETA method is the importance of a coherence between the RTD programmes and the industrial context in which they take place (see second Brite-Euram, Material Ireland or Procap studies for instance). The conclusion is quite obvious: when public authorities want to promote innovation through public/private R&D collaboration, the success or the failure of the public policy is much more determined by the way how the programme is design (including the networks set up or used and the articulation between the programme and the industrial context in which it is performed) and implemented than by the amount of public money spent.

## 5 Evaluation of Large-Scale Programs in Europe

From a methodological point of view, BETA method only provides partial answers to some of the classical problems encountered by all evaluation methods. These classical problems may be broken down into four categories (following in part [Bach L. and Georghiou L., 1998]).

The first is related to the difficulty to identify and assess the *propagation* of the effects through space and time, that is in the long term, from participants to the programme to non participants, from the micro level to the meso and macro ones. This propagation or diffusion phenomena make difficult the assessment of “net effects” (as opposed to “gross effects” that would be limited to participants, short term or micro level) as well as any sampling and extrapolation attempts. On this point, the BETA approach is unambiguous: such an evaluation is out of its scope, and it only allows to provide some hints (from the very beginning of the propagation that starts at the participant level). In other words, the BETA approach as it stands now does not allow an evaluation of the impact of the RD programme on the “rest of the economy”.

The second series of problems concerns the *separability* between factors that lead to innovation. A first dimension of this is the tendency to “attribute” all the benefits to the evaluated project, as regards other projects run in parallel for instance within a firm (then evaluating only one project is a cause or symptom of “project fallacy”). A second dimension is the complementarity of assets (in the form of knowledge and competencies and/or individuals, such as marketing capabilities, management skills, “product champion”, “gate keepers” ...) that are more and more required in order to innovate. Here again, the influence of the R&D work on which the evaluation focused tends to be overestimated. A third dimension is inherent to collaborative projects: this is the difficulty (or impossibility) to split between the respective influence of each associated partner. The BETA answer to the first dimension of the separability problem is the use of “fatherhood coefficient”, which are assessed by the managers interviewed with the support of the interviewers. This is especially the case when the two-step evaluation process is used (see above; the first step being related to the separability between assets, the second with the separability between projects). In most of the circumstances, different managers are interviewed together, who are in charge of different departments (research, marketing, finance etc) and are from different educational background. Long practice of interviews clearly shows that managers agree on a minimal value of the fatherhood coefficients. Moreover, a conservative estimation is used, since the value of the coefficient is not evaluated in the form of an exact value but in term of a range of which only the lower bound is used for the calculation of the effect. But anyway, subjectivity remains inevitable in this exercise. As regards the collaborative dimension of the separability problem, the fact that the BETA covers all participants of projects evaluated allow to make cross-comparisons which undoubtedly diminish the risk of double-counting.

Taking into account the *additionality* related problems is the third difficulty encountered in all evaluation, obviously related with the separability problems. Recent works have emphasized the different perspectives that can be adopted when dealing with this problem: input additionality, output additionality ad behavioural additionality. Two questions are at stake: what difference does the evaluated

## 5 Evaluation of Large-Scale Programs in Europe

ality. Two questions are at stake: what difference does the evaluated programme make ? and if it does, could it justify the public intervention ? and a third one sometimes follows : Or would it have been better to invest in another programme ? (what is being translated into the “opportunity cost question” in standard economics vocabulary). These problems can be addressed at partners or projects level, or at programme level, or even at policy level. Since the BETA approach basically is a micro one, it can only provide some partial information or impressions that could be used to answer the question at higher levels (programme, policy). Conversely at micro level, there are probably some methodological enhancements to be made to the BETA approach, for instance by refining the “alternative reference scenario” (in the BETA approach, no alternative scenario is explicitly asked to the managers interviewed, in other words the “null hypothesis” or the “absence of the evaluated project” is the reference situation), or by including some “control organisation”, for instance in the EC case those who were ranked just after the “cut” of the EC choice among the proposals.

A last series of problem arises when there is a need or willingness to an *institutionalisation* of the evaluation. Typical concerns are the involvement of multiple stakeholders, the timing of the evaluation as compared to the timing of the programme design and implementation, the resources required for the evaluation, the selection of the evaluators, the reproducibility of the evaluation exercise, the data collection approaches, etc. Some remarks can be made about the BETA approach in this context. i) this is a time and resources consuming approach (interviews, data collection and analysis work amount to approx. 1 man-week per partner), then it cannot be used for massive or exhaustive coverage of programmes. ii) this leads to a second question, which is the sampling phase and the choice of the criteria used for this purpose; a good knowledge of the programme itself and of the objective of the evaluation work is required. iii) experience with different university labs with which the BETA is collaborating and studies performed by other organizations on the same methodological basis (see for instance [GARCIA A., 1996]) have shown that the transfer and the appropriability of the method is not immediate and obvious, especially because it requires a specific practice of interviews. iv) one practical and crucial aspect is the confidentiality of the information gathered by the BETA throughout the interviews, and even the confidentiality of the sample of participants. This is a very good argument to put forward the managers interviewed and make them give important if not strategic information. It is not sure that such confidentiality could be maintained with recurrent and regular evaluation campaigns based on interviews. v) one aspect of the institutionalisation is the very use of the result by the decision-makers of the institution. Results presented by the BETA have sometimes been subject to confusion and misinterpretation, when exhibiting in a two simplistic way the quantified result and the “x euros return to 1 euro invested by state” ratios. As well as for (many ?) other evaluation methods, users of the results must probably be in some way “educated” in order to exploit those results and help the evaluators to enhance the relevance and the quality of their approaches, in an endo-formative or pro-active way affecting not only the programme itself but also its evaluation.

On the other hand, the experience gained by using the BETA method also shows that it is adaptable to a wide variety of RTD programmes, provided that they fulfil the conditions given earlier. In each case, the BETA approach allows a fine understanding of the innovation processes triggered by the RD programmes (especially when cooperative projects are at stake), with qualitative dimensions that may help a pro-active use of the evaluation. Another originality is that the BETA approach may be regarded as a “mixed approach” between i) case studies and statistical analysis, and ii) between main-stream / cost-benefit analysis and more “heterodox” approach attempting to account for phenomenon such as networking (internal and external to participants), the creation and change of technical and organizational routines, the creation and change of learning processes, etc. All those phenomena reflect the impact of one given programme not in terms of final or intermediate outputs, but in term of various changes affecting the participants to the programme, i.e. affecting the capability of the participants. This is probably because the BETA approach basically consists in examining the development, codification, combination, acquisition, storing, sharing, exploitation, etc of all forms of knowledge by the participants; and it is more and more acknowledged that all these forms of knowledge processing are at the core of all innovations, giving birth to competencies, skills and value. However, theoretical and methodological work remain to be done to better fit the emerging concepts and tools of the knowledge economics with the ones that have been used by the BETA approach for some years.

### **Bibliography**

- B.E.T.A., 1993, “Economic evaluation of the effects of the Brite-Euram programmes on the European Industry”, Final Report, EUR 15171, CEC Luxembourg.
- BACH L. and GEORGHIOU L., The Nature and Scope of RTD Impact Measurement , A discussion paper for the *International Workshop on “Measurement of RTD Results/Impact”*, Brussels, 28-29 May 1998 A discussion paper for the International Workshop on “Measurement of RTD Results/Impact”, Brussels, 28-29 May 1998
- BACH L., COHENDET P., SCHENK E., 2000, “Technological transfers from the European Space programmes: a dynamic view and a comparison with other R&D projects”, to be published in *International Journal of Technology Transfer*.
- BACH et al., “Measuring the economic effects of large scale R&D programmes : experience drawn from an evaluation of space programme”, 1997, in *The Strategic Management of research and technology. Evaluation of programmes*, M. Callon, P. Laredo et P. Mustar (eds), Economica International, Paris.
- BACH L., COHENDET P., LEDOUX M.J., 1995, “The evaluation of big R&D projects : A comparison between the case of the Brite-EuRam projects and the case of the space projects”., in *International Journal of Technology Management*, Special Issue on Evaluation of Research and Innovation, Vol.10, Issue 4/5/6.
- BACH L., CONDÉ-MOLIST N., LEDOUX M.J., MATT M., SCHÆFFER V., 1995, “Evaluation of the economic effects of Brite-Euram Programmes on the European industry”, *Scientometrics*, Vol.34, n°3, pp.325-349.

- BACH L., LEDOUX M.J., MAGNAVAL R., PERO H., 1994, "The Economic Effects of public/private cooperative research", *Futures*, Vol.26, n°8.
- BACH, LAMBERT G., 1993, "Evaluation of the Economic Effects of large R&D Programmes : the Case of the European Space Programme", *Research Evaluation*, Vol.2, n°1, april.
- BACH L. et al., 1992, "Measuring and Managing Spinoffs : The Case of the Spinoffs Generated by ESA Programs", in J.S. Greenberg, H.R.Hertzfeld (eds), *Space Economics*, Progress in Astronautics & Aeronautics, vol. 144, AIAA, New York.
- BUISSERET J. et al., 1995, "What difference does it make ? Additionality in the public support of R&D in large firms", *International Journal of Technology Management*, 10 (4-6).
- FURTADO F., SUSLICK S., PEREIRA N.M., DE FREITAS A. G., BACH L., 1995, "Economic Evaluation of Large Technological Programmes : The case of Petrobras's Deeptwater Programme in Brazil – Procap 1000", *Research Evaluation*, Vol.8, n°3, december.
- GARCIA A., AMESSE F., SILVA M., (1996), "The indirect economic effects of Ecopetrol's contracting strategy for informatics development", *Technovation*, vol. 16, n°9.
- GIBBONS M. et al., 1994, *The New Production of Knowledge*, SAGE Publication
- KLINE S.J. - ROSENBERG N., 1986, "Innovation : An Overview", in LANDAU R. - ROSENBERG N. (ed.), "The Positive Sum Strategy", Academy of Engineering Press (pp. 275-305).
- LUNDEVALL B.-A., 1988, "Innovation as an Interactive Process : From User-Producer Interaction to the National System of Innovation", in DOSI G. et al., "Technical Change and Economic Theory", Pinter Publishers, London.
- OECD, (1997), "Policy Evaluation in Innovation and technology - Towards Best Practices", Proceedings of the Conference organised by OECD's Diectorate of Science, Technology and Industry, June.
- VON HIPPEL E. (1988), "The Source of Innovation ", Oxford University Press.
- ROTHWELL R. (1994), " Industrial Innovation : Success, Strategy, Trends ", in *The Handbook of Industrial Innovation*, Edt M. Dodgson - R. Rothwell, Brookfield, US.

---

## Discussion of Lauren Bach's paper

**Arie Rip** (University of Twente, The Netherlands). Was the evaluation planned as a whole or were parts added over time? The approach used was developed some-time ago. Has it changed since then?

**Laurent Bach.** The main methodological development took place for this study. Data were collected initially only on the indirect effects. During the 1980s other effects were discovered, and thus the methodology was broadened. The critical mass is what brings about new knowledge. When new programs are evaluated we find new effects that we did not pick up previously.

## *5 Evaluation of Large-Scale Programs in Europe*

**Maryellen Kelley** (National Institute of Standards and Technology, USA). We heard from Liam O’Sullivan about the emphasis on behavioral additionality in the EU plans. Do indirect effects fall into that category? Are indirect effects unintended effects or were they goals of the program?

**Laurent Bach.** This is a matter of vocabulary. We took direct effects to be those having to do with the technical goals of the program. But if you examine the EU framework, you see that there are other objectives. But the objectives of our project at the direct effect level are better defined.

**Maryellen Kelley.** Is behavioral additionality in this case congruent with technological connections/effects? Did you see that with technological success also came behavioral additionality?

**Laurent Bach.** All of the research projects were collaborative. We didn’t see a direct correlation between qualitative effects and economic effects.

**Ken Guy** (Wise Guys, UK). When I look at this, I want to know more detail of ascribing values and attributing problems.

**Laurent Bach.** This attribution and ascription was done through interviews with program managers in the firms. We cannot attribute all of the outcomes to the R&D projects, so we asked managers what share of the credit they felt was attributable to the R&D. We asked both technical and financial managers for their estimates.

**Ken Guy.** I have a problem in doing this in my own work. When I ask for a margin of error I get a very [too] wide one.

**Laurent Bach.** We did not use a probabilistic system. We asked instead for a range and we always took the minimum of that range.

**Ken Guy.** Still that gives a margin of error. You still end up with a figure that must have error margins. When data such as these are available publicly they can be misused and that is very dangerous.

---

## Evaluation of the EU's Fourth Framework Programme in Finland

**Terttu Luukkonen**

VTT Group for Technology Studies, Finland

Email: terttu.luukkonen@vtt.fi

---

### Introduction

The first Finnish study of the national impacts of participation in the EU Framework Programme was carried out in 1996-97 and it concerned the Second and Third Framework Programmes (Luukkonen & Niskanen, 1998). It was started soon after Finland joined the European Community as a full member in 1995. Finnish organisations had been able to participate in the Framework Programme earlier, since 1987, either on a project or on a programme basis. Because of many difficulties and delays involved in the atypical procedure, Finns did not participate very frequently until the country's full membership, which coincided with the start of the Fourth Framework Programme.<sup>1</sup> Participation rates in the Fourth Framework Programme became fourfold compared with those in the Third far exceeding the average growth of the Framework Programme (the inclusion of new programmes and actual growth) (see Luukkonen and Hälikkä, 2000).

The evaluation of the Fourth Framework Programme was a continuation of evaluative studies concerning the Second and the Third and not only repeated, but further elaborated the approaches used earlier. The evaluation of the Fourth Framework Programme, like that concerning the Second and Third, consisted of a series of studies using different approaches. Surveys took their model from impact studies which were carried out in older EC member countries in the early 90s concerned the impacts of the Second Framework Programme (see e.g. Reger and Kuhlmann; 1995, Laredo, 1995; Georghiou et al, 1993).

### Institutional setting

The evaluative studies of the national impacts on participation in the Framework Programmes have been carried out at a contract research institute, the Group for Technology Studies at the Technical Research Centre of Finland (VTT). The first evaluative study was done on the initiative of the researchers - mainly for pure intellectual curiosity. Later on their research interests included an attempt at understanding the role of public programmes in the promotion of technological change and EU

---

<sup>1</sup> There was a short period of one year, 1994, when Finland was a member of the European Economic Area (EEA). This allowed Finnish organisations to participate in the Framework Programme on an equal basis but did not allow Finland's participation in decision-making. Because membership in EEA lasted only a very short period, it is difficult to detect its influence on participation rates.

RTD policies in general. In the beginning, the studies could not be regarded as evaluation, but rather as research activities in the domain of R&D policies and technological change.

By providing useful information, the studies created an information demand among the national authorities responsible for the national policy on EU RTD activities. They became to be regarded as part of an evaluation process, the purpose of which was to provide an assessment of the benefits gained from participation. The Finnish authorities had actively promoted participation by Finnish organisations in the Framework programmes and wanted to find out whether this policy was successful in its impacts and worthwhile continuing.<sup>2</sup>

### **Information needs and uses**

Finnish authorities have been interested in obtaining up-to-date information about the participation rates of Finnish organisations in the Framework Programme and about the monies obtained. This is the basic level of assessing the 'juste retour', that is, whether in money terms, member countries get out as much as they contribute. With regard to participation rates, the Commission sources, such as the Cordis database, produced data that was neither sufficient nor timely. There were many gaps in information about participating organisations and summaries based on such data were incomplete and their representativeness was not known. There was also an interest in learning about the acceptance rates by Finnish organisations in the application process, information not obtainable from any previous sources. Another knowledge interest concerned an assessment of the success of the policy to promote EU research collaboration, as referred to above. Such information could not be inferred from the studies carried out in other countries.

The findings of the studies concerning Framework Programme participation have been used in considerations concerning the national standpoints on EU RTD policy. According to the knowledge of the author of this paper, the studies have not changed any national standpoints noticeably. This is at least partly because the studies did not bring about any fundamentally critical information, such as a very negative analysis of the utility of participating in the EU programmes, but confirmed the adopted policies. It is also to be noted that all the studies on the Fourth Framework Programme have not yet been completed and all the knowledge inputs have not yet been disseminated sufficiently in Finland. Also some standpoints on the development of the Framework Programme in general, which run counter to the prevailing notions such as those suggested by Luukkonen (2000b), have not had, at least yet, influence.

Another important use of the study findings has been in information dissemination and information campaigns about Framework Programme participation. Finnish national authorities actively promote the participation of Finnish organisations and

---

<sup>2</sup> This exemplifies the fact that 'evaluation' is a socially defined and negotiated activity. A study containing evaluative data can be regarded as an evaluation or ignored in such a context.

## *5 Evaluation of Large-Scale Programs in Europe*

at the time of the opening of an application round, arrange, as in other EU Member countries, national dissemination days on the specific research programmes. They also publish a newsletter and provide other information on demand and together with representatives of major research organisations. The studies have provided information about different aspects of participation such as factors contributing to success in project participation, the benefits of being a co-ordinator etc. Also general information about participation activity by different types of organisation has attracted interest in this context.

Finnish authorities have also used the findings in their discussions on EU RTD policy with representatives from other countries. This information has particularly been used when relating Finnish experiences to representatives of the pre-accession countries.

### **Organisation**

All the studies including that on the Fourth Framework Programme have been carried out in close interaction between the national authorities responsible for Finland's EU R&D policy and the researchers. Close interaction between the researchers and national authorities has been possible particularly through the steering groups. All major research projects at the VTT Group for Technology Studies have steering groups consisting of representatives of major funding agencies of the studies, other importance user and reference groups as well as other researchers in the area. This was therefore not a special arrangement for the studies on EU participation. In this case, the steering groups have included representatives from the Ministry of Trade and Industry, Ministry of Education, Cabinet level Science and Technology Policy Council of Finland, National Technology Agency (Tekes), the Academy of Finland, the national intergovernmental committee for EU RTD policy, and industry. Major research plans have been discussed in the steering groups and they have received preliminary data at an early stage. Exchanges in the groups have enhanced the dissemination of the research findings to the major users. Findings have also been presented at national seminars on EU RTD policy and in informal discussions.

### **Studies carried out and ongoing**

#### *1. Database on participation*

Like the studies on the Second and Third Framework Programme, that on the Fourth started with the creation of a database of Finnish participants in the Framework Programme. Basic information was obtained from Cordis and the Commission and it was complemented with data obtained from the national delegates and major Finnish research organisations.

The results were published in Niskanen et al., (1998) and in Luukkonen et al., (1999). They showed that, besides a sizeable increase in participation, with the exception of a few programmes, Finns had much higher than average acceptance rates. In money terms, they got more money than they contributed, and compared to their

## *5 Evaluation of Large-Scale Programs in Europe*

share of the EU budget, they participated much more than on average. However, because of the high R&D intensity of the country, their participation rates were below the EU average when R&D personnel numbers were taken into account indicating a potential for further growth in participation rates.

From the Fifth Framework Programme onwards, this kind of data will be collected by the National Technology Agency (Tekes) on a real-time basis from the national contact points in different special programmes, and the first such data are already available. The collected databases allow national authorities to monitor national participation rates in different special programmes, acceptance rates in the proposal phase, and the different roles Finnish organisations have in the Framework projects.

### *2. Survey with participants in the Fourth Framework*

A comprehensive survey with participants in the so-called shared-cost activities about participation experiences was an important part of the studies. This survey was addressed to 955 participants and the response rate was 70% (Luukkonen and Hälikkä, 2000). In addition to the participants' assessments on their participation objectives and achievements, and impacts of participation, these data included information about collaboration networks and their additionality. The major findings have been reported in Luukkonen and Hälikkä (2000). This paper will only refer to a few of its findings.

### *3. Study about the strategies of Finnish companies in EU RTD participation*

This study was based on interviews with technology or R&D directors of major Finnish companies and directors of SMEs about the extent to which EU projects furthered their technology and business interests. The study also explored the pre-competitiveness of the EU projects and problems related to intellectual property rights. They have been reported in a paper which has not yet been published (Luukkonen 2000c).

### *4. Study of the impacts of EU research collaboration on university research*

This study focuses on university departments. It analyses the quality of EU projects, the influence of EU projects on the allocation of research funds to departments nationally, on industry-university links, and on the internationalisation of Finnish science (Niskanen, 2000). The data are based on surveys and interviews with researchers who have participated and also with those who have not participated in EU research programmes. The study will be completed by autumn 2001.

## **Some findings**

This report will not even attempt to summarise all major findings of the evaluation studies obtained so far. It will instead highlight a few issues which have been explored in the studies and which relate to more general issues about the role of the Framework Programme.

## 5 Evaluation of Large-Scale Programs in Europe

### 1. Additionality

The evaluation of public R&D programmes within the framework of additionality has been a practice in Europe since the well-known UK Alvey programme evaluation (1984-1990) (Quintas and Guy, 1995), which developed and refined many evaluation tools used later on in evaluation in other European countries and at the EU level. The concept of additionality was further elaborated within the EC MONITOR-SPEAR programme studies in the late 80s (Georghiou, 1994). At the moment, additionality is a major general framework offered for the evaluation of RTD programmes within the European Union. The Finnish studies on the Framework Programme have examined current notions of additionality and their usefulness in the evaluation of public RTD programmes (see Luukkonen, 1999; Luukkonen, 2000a).

Additionality is the incentive effect of public support of R&D activities in companies in particular. It is originally based on the notion of market failure, which the corrective measures by the government are expected to deal with (Luukkonen, 2000a). Considerations of additionality reflect a fear that public support would substitute for indigenous R&D investments made by the companies themselves, and would thus make the support superfluous. In practice, additionality is not easy to measure.<sup>3</sup>

In order to explore the usefulness of the concept of additionality in evaluation in the Finnish studies, it was juxtaposed with the strategic value of the project, and interpreted within a general framework of market failure thinking, originating from neo-classical economics (Table 1). It means that attention was paid to the role of public funding as a correction of the supposed market failure. High additionality means that the research would not have been done at all without EU funding and low additionality that the research was done differently. No additionality, that the research would have been done anyway, comprised only 3 % of the cases, and in Table 1 has been combined with low additionality.

**Table 1. Additionality and strategic value**

Strategic value	Additionality	
	High	Low
Great	1 Ideal	2 Substitution
Potential	3 Potential	4 Marginal
Marginal	5 Trivial	6 Truly marginal

<sup>3</sup> In the surveys conducted in Finland, additionality was measured by the routine questions used in this connection and the model was adopted from previous impacts studies on the Framework Programme (see, e.g. Georghiou et al, 1993). These questions included the following: whether the research would have been done anyway without EU funding, whether EU funding enabled the project to be conducted faster, in larger scale or differently, or whether the project would not have been carried out at all without it.

## *5 Evaluation of Large-Scale Programs in Europe*

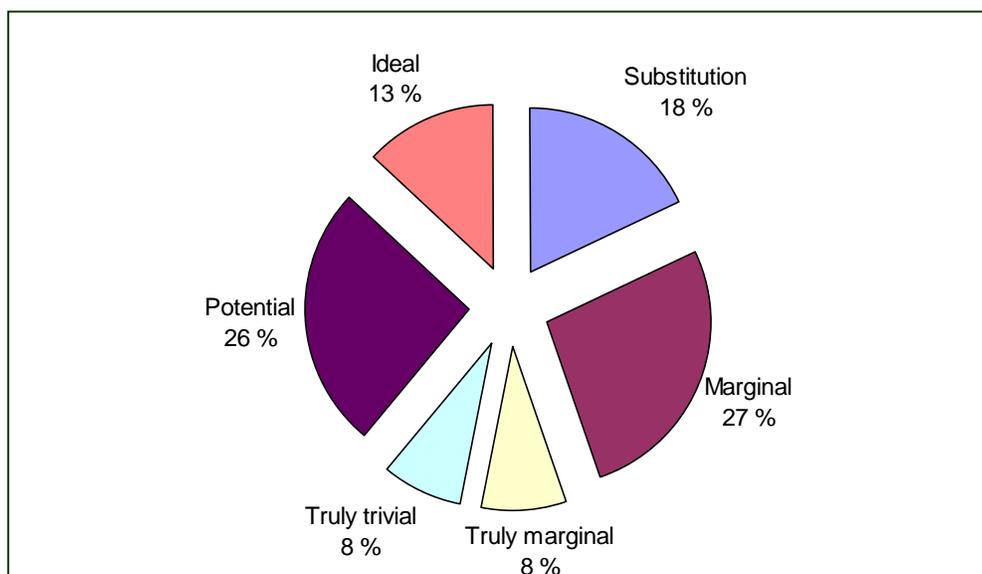
In order to illustrate the point, the names for the categories are to some extent an exaggeration. The first category, coined 'ideal', indicates strategically important R&D which would not have been carried out without government funding for various reasons (uncertainty, risks, expenses etc.). It is called 'ideal' because it is often assumed that, in an ideal case, a government programme has a high additionality in advancing strategically important endeavours. 'Substitution' in category 2 is strategically important R&D which the firm would have done in any case, but when government money was available, it utilised it. This is the case which should be avoided in public funded. Category 3, 'potential', indicates research that may be of future strategic importance, often longer-term and risky research, but its importance is not yet known. Without EU funding, the company would not have carried out the project. Interviews with technology directors of Finnish companies have amply illustrated the importance of public funding in enhancing decision-making in companies in situations in which the project outcomes are uncertain (Luukkonen 2000c). Category 4 denotes 'marginal' R&D, which here means, like in the 'potential' case, longer-term and riskier research, but which would have been carried out somehow even without the public support. 'Truly trivial' in category 5 indicates research which is non-essential and which companies would not have done if government funding had not been available. The last category, 'truly marginal', category 6, denotes non-essential, non-important R&D which would have been carried out anyway, perhaps to search for new potential avenues for development, R&D, likely also not very expensive R&D.

When firm data from the survey to the Finnish companies in the Fourth Framework Programme (Luukkonen and Hälikkä, 2000) are put in the various categories of Table 1, the relative sizes of each class are indicated by the pie chart in Figure 1.

The ideal category in Figure 1 is quite small, only 13 %. Even together with the category of 'potential', both seemingly good cases, the percentage is only 39%, well below half the cases. If considered in a strict sense of additionality, this finding seems to point to a poor result: only 39 % proved to have additionality.

These data can, however, be considered in a different framework. Evolutionary perspectives on technological change have highlighted the importance of knowledge flows and interactions within the innovation system (see e.g. Metcalfe, 1995). Collaboration among different types of organisation is important because it will expand the knowledge frontiers of current technological know-how and further enhance the knowledge base of the companies involved in collaboration. Changes in the ways in which companies carry out R&D, for example, their collaborative behaviour, are important for furthering technological change. The dimension of low additionality in the above Table 1 and Figure 1 could be interpreted to represent behavioural additionality, a term coined by Geoghiou (1994). This class is based on survey questions which enquired alternatives of doing the research project faster, on larger scale or with different objectives, in sum, differently. Doing research differently is closely related to doing research in a collaborative consortium, which offers new opportunities for knowledge flows and expertise. If we interpret 'low additionality' as 'behavioural additionality', we can come up with the following (Table 2).

**Figure 1. Finnish companies and the additionality of the Fourth Framework Programme projects**



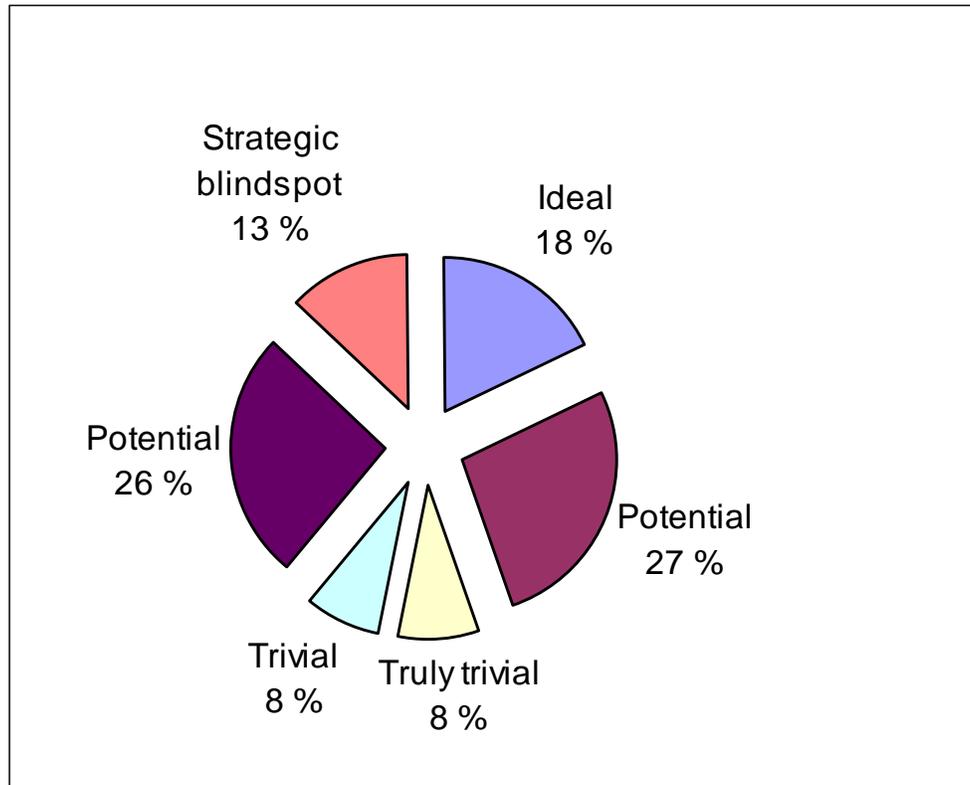
**Table 2 Additionality reinterpreted**

Strategic value	Additionality		
	Input	Behavioural	No additionality
Great	Strategic blindspot	Ideal	Substitution
Potential	Potential	Potential	Marginal
Marginal	Trivial	Truly trivial	Truly marginal

Input additionality in Table 2 refers to Georghiou's classification of additionality (1994) and means the extent to which a subsidy is reflected in increased R&D expenditure by a firm. This concept seems to draw on the market failure approach, and Table 1 is in fact based on a mixture of the above-mentioned neo-classical and evolutionary approaches. Low additionality in Table 2 is divided into two classes, behavioural additionality (doing the research differently) and no additionality. These two classes were, however, merged together in Figure 2, since 'no additionality' was only 3% of all the cases.

According to the evolutionary perspectives, behavioural additionality in projects of high strategic importance represents the 'ideal' case, while input additionality in projects of great strategic importance represents 'a strategic blindspot' for the company, a failure to fund important projects with the company's own resources. There are two classes of 'potential': research of potential future importance that has input additionality and research of potential, future importance that has behavioural additionality. All four classes can be regarded as representing success in policy, and the interpretation of Figure 2 is quite different from the one according to Figure 1: as many as 84% of the cases would represent a good outcome, quite contrary to the purely neo-classical interpretation provided above.

**Figure 2. Additionality reinterpreted**



We can conclude that the framework in which additionality is interpreted is highly important for the evaluative outcome. The way in which the 'ideal' case is defined varies in the two examples. An overall interpretation of additionality also varies. The above interpretation in the combined evolutionary and market failure framework (Table 2) has a benefit of providing an improved elaboration of different types of additionality. By drawing attention to behavioural additionality, it highlights that additionality is not dichotomous. It may, nevertheless, be claimed that it is probably too generous and exaggerates the success of the programme under examination. In collaborative R&D programmes, it is easy to achieve some additionality. In such programmes, it is practically self-evident that things are done differently than would otherwise be the case. For example, when obtaining EU money, a company will have partners from other European countries, and quite likely partners with expertise to which the company might have difficulties in having access without the EU project; the consortium is in most cases larger than in national projects, and more areas of expertise will be represented. This was found in the interviews with company technology directors (Luukkonen, 2000a; Luukkonen, 2000c). According to these data, for the companies, the choice was often among a national collaborative R&D and EU collaborative project. The national collaborative programmes have similar behavioural effects compared with the European ones. The latter are typically larger, have better, in many cases the best possible, European expertise and

## *5 Evaluation of Large-Scale Programs in Europe*

partners from other European countries. It is thus easy to achieve behavioural additionality in a collaborative R&D project, be it national or European.

Yet, our data have shown that the overall percentage of behavioural additionality is not greater than a little over 50 % and that the proportion of input additionality is also great. Additionality is thus not a simple matter and needs further examination.

### *2. Networks*

#### *2.1. Cross-sector collaboration*

The original model for the Framework Programme entailed collaboration among companies and between companies and public sector research institutes. These relationships are important because they offer an important avenue for interaction among knowledge-exploiting organisations (companies) and knowledge-producing institutions (universities and research centres) around research agendas of common interest. It is not a question of simply diffusing research findings to practice or of a linear process of information relay, but of reframing or redefining research agendas.

The extent to which it has succeeded in this task has been disputed. For example, Peterson and Sharp (1998) have claimed that the Framework Programme has not succeeded very well in fostering university - company or public sector research institute - company relationships. Peterson and Sharp (1998) base their claim of few cross-sector links on indirect and insufficient evidence, and this matter has not really been examined (cf. The Second European Report on S&T Indicators - Report, 1997). The Finnish survey explored this question and showed that in the Fourth Framework Programme, as many as 64 % of the projects with Finnish participants involved company and university or company and research centre collaboration. When looked from the point of view of companies, over 80% of this sector's participations in EU projects entailed partners from either a university or a research centre: 70% of companies collaborated with a university and 75% with a research centre in their EU project. These are much higher figures than any estimates of R&D collaboration between these different types of organisation in other contexts. These figures are higher than those obtained in the Second Community Innovation Survey for Finland (54% and 42% respectively) (see Eurostat, for example, through New Cronos database).<sup>4</sup> This would mean that the EU Framework Programme effectively promotes cross-sector collaboration. According to the Finnish studies, they also attract participation by researchers who have earlier experiences of collaboration with companies and therefore are more inclined to do so in an EU project.

#### *2.2. Interfirm networks*

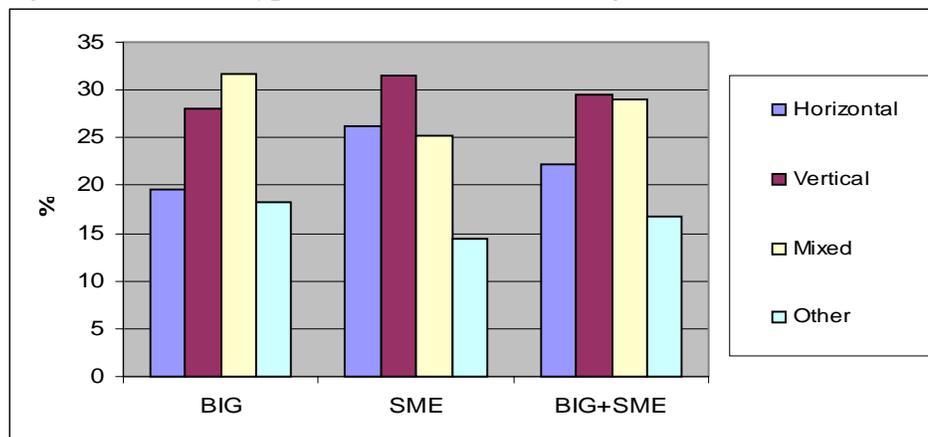
The Finnish survey provided information for the analysis of different patterns of interfirm collaboration. Research literature on interfirm collaboration usually de-

---

<sup>4</sup> There is reason to believe that in their answers to the Community Innovation Survey, Finnish firms have defined innovations in a way similar to R&D and therefore these figures are comparable (see Luukkonen and Hälikkä, 2000).

finds two basic types: that between partners at the same level in the production process, such as collaboration between competing companies, that is *horizontal* collaboration, and that which involves co-operation throughout the chain of production for particular products, for example, by a company with its clients or subcontractor firms, that is, *vertical* collaboration (Arnold, Guy, Dodgson, 1992; Dodgson, 1994). Collaboration within the vertical supply chain and in particular, with customers, has been found to be the most important source of information in the innovation process (Palmberg et al., 2000) or according to some findings, as important as public research institutes (Arundel and Geuna, 2000). The Finnish impact study looked at the types of interfirm collaboration in EU projects and also found a third type, a *mixed* type, which involved collaboration with both competitors and suppliers or subcontractors.

**Figure 3. Different types of interfirm networking**



The mixed type turned out to be as important as the vertical one (see Figure 3). It was found to be particularly prevalent in information technology and telecommunication projects and fields (Luukkonen and Hälikkä, 2000). These are the fields that are most high-tech and where technological change is most rapid. It seems that in new technological areas, there is a need to reach out to a wider knowledge base than before. Increasingly complex technologies develop in heterogeneous collaboration networks, and purely vertical or horizontal collaboration is no longer sufficient to describe the collaboration patterns prevalent in these areas.

It was further found out that the mixed collaboration networks had higher input additionality than other network types (Luukkonen and Hälikkä, 2000). This may be explained by the fact that complex networks are more difficult to organise without the EU collaboration framework and funding. It was found out that in the EU Framework Programme, it is not only a question of funding as such, but also of a ready-made legal collaboration frame, the existence of which facilitates the creation of cross-country and cross-institutional projects (Luukkonen and Niskanen, 1998). The sharing of intellectual property rights has been predefined, which means that companies do not have to start lengthy discussions about the matter before embark-

## 5 Evaluation of Large-Scale Programs in Europe

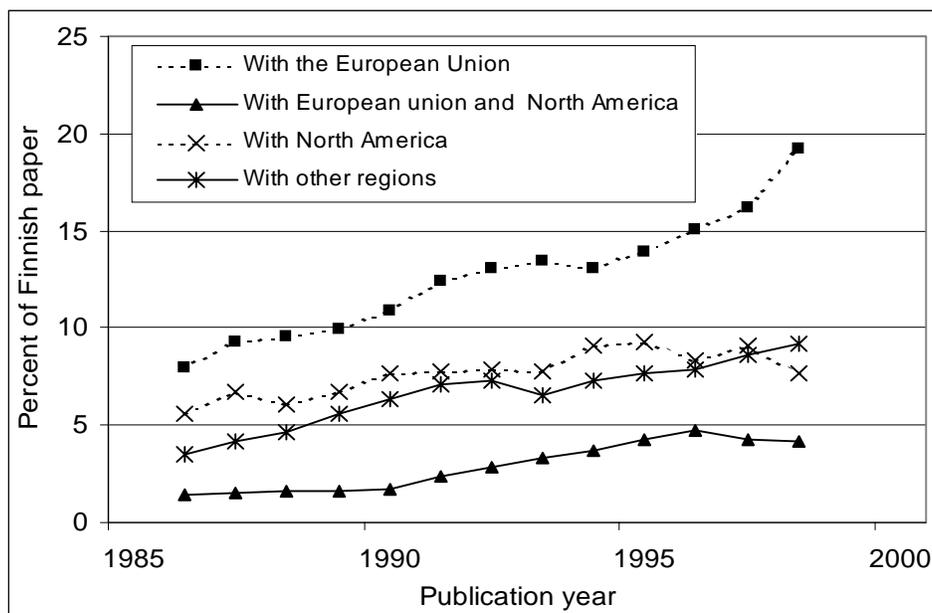
ing upon collaboration. The benefit of a legal collaboration framework may be the greater, the more complicated the network structure.

### 2.3. Cross-country collaboration networks

That the Framework Programme has advanced inter-European collaboration has been widely acknowledged. It is after all a basic requirement that the projects have at least two partners (legal entities) independent of each other, and established in two different Member States, or one Member State and one Associated State.

A bibliometric study of Finnish science showed that inter-EU collaboration has grown much more than other international collaboration relationships as measured by bibliometric data on coauthorship (Persson et al., 2000).

**Figure 4. Percent of Finnish papers coauthored with different country groups. Based on the SCI (Persson et al., 2000)**



It is likely that many factors have contributed to this development and international collaboration is a growing tendency. Still, we may presume that the Framework Programme has had the greatest impact on this, especially since the tendency is increasing towards the end of the 90s. The Framework Programme is practically the only European programme that grants substantial sums of money to fund actual research via collaborative projects. Other European programmes are networking programmes (COST), a mechanism and a status to create collaboration (Eureka), grant research money only on a small scale (European Science Foundation), or are research institutes where joint research is carried out (CERN, EMBL, ESA).

## *5 Evaluation of Large-Scale Programs in Europe*

Because international collaboration is a growing tendency, we have so far no reason to presume that it would have made collaboration relations too one-sided, as is sometimes feared in discussions in Finland.

### *3. Precompetitiveness*

The original model for collaboration, adopted for the Framework Programme, was that the research was to be generic, it was to be carried out jointly by European firms and research organisations (universities and research institutes), and it was to be precompetitive. By it was meant that the research to be carried out was a few years removed from the market phase.

Given the fact that in recent years, programme goals have emphasised market orientation and implementation of knowledge to economic applications, one of the research questions in the Finnish studies has been whether the above evolution has led to near-market research at the cost of precompetitiveness.

This question was explored in the interview study done with the technology and R&D directors of large companies and the directors of SMEs (Luukkonen, 2000c). There were 49 interviews in 41 companies. It was first found out that 90 % of the companies had precompetitive projects in the Framework Programme, while 20% had near-market projects. The overlap expresses that some firms had both types of EU project. Precompetitiveness was the major type and this was true independent of whether the company sought research collaboration to advance technological know-how of the company or to advance commercial objectives. Since such a great majority of the firms had precompetitive projects, we can assume that there have not been great changes in this respect over time.

The fact that the Framework Programme continues to be mainly precompetitive is due to the special circumstances and the contract principles of the Framework Programme, particularly the requirement that partners share the intellectual property rights amongst themselves. This can be a facilitating factor in the creation of consortia, but at the same time, restrict the types of project which firms can bring to the European collaboration. Doing near-market research in an EU project seems very difficult in practice because of reasons related to confidentiality, intellectual property rights and related issues. The slowness of the decision-making process on the funding of the projects may also be more irksome for near-market research.

In spite of the fact that the large majority of EU R&D collaboration involves precompetitive projects, around one third of the companies reported commercial utilisation of the results quite soon after the project or even during its lifetime. The explanation for this potential conflict is the possibility that firms can carry out parallel projects internally, in which they utilise the findings obtained in the EU project.

### **Conclusions**

The above results have shown that the largest group of projects consists of those of potential future importance for the company. They represent longer-term and riskier

## *5 Evaluation of Large-Scale Programs in Europe*

projects. The public co-funding lowers the threshold for the decision to embark upon such projects.

EU projects can have commercial or knowledge-enhancing objectives and motivation, but in spite of this matter, the large majority of them are still precompetitive. There is not contradiction in a project being precompetitive and still advancing commercial objectives and helping creating commercial outcomes. It was found out that companies often carry out parallel commercial projects and these can draw on the precompetitive EU project. EU collaboration is in fact more effective in its outcomes, if the company has other, parallel projects in-house.

The Framework Programme is still largely precompetitive because the contract model, especially the rules concerning sharing intellectual property rights amongst the partners in a consortium, restrict the types of project possible in this framework. Most near-market research is quite confidential and companies cannot carry out this kind of projects in it.

The Framework Programme has been noted to enhance different types of networking: as can be expected, cross-country networking, but also cross-sector collaboration and complex interfirm networks. EU money plays a role in bringing the partners together and in enhancing collaboration. The Framework Programme is not, however, only a funding arrangement. It also important in providing a framework within which companies find it easier to make collaboration contracts. The model contract facilitates the start of a collaborative project. Companies do not have to start negotiating the division of intellectual property rights from the scratch. It is ironical that, while the contract model restricts the type of research that can be carried out within this framework, it facilitates making the contract on research that can be carried out in it.

### **References**

Arnold, Erik, Guy, Ken, and Mark Dodgson, *Linking for Success: Making the Most of Collaborative R&D*. Report to NEDC Electronics Industry Sector Group, September 1992.

Arundel, Anthony and Aldo Geuna, *Does Localisation Matter for Knowledge Transfer among Public Institutes, Universities and Firms?* Paper presented at the 8<sup>th</sup> Schumpeter Conference "Change, Development and Transformation", University of Manchester, 28 June - 1 July, 2000.

Dodgson, Mark, *Technological Collaboration and Innovation*, in: Dodgson, Mark and Roy Rothwell (eds.), *The Handbook of Industrial Innovation*, Edward Elgar, Adershot, 1994, pp- 285-292.

Georghiou, L., H. Cameron, J. A. Stein, M. Neveda, M. Janes, J. Yates, M. Piper, M. Boden, and J. Senker, 1993, *The Impact of European Community Policies for Research and Technological Development upon Science and Technology in the United Kingdom*. London, HMSO.

## *5 Evaluation of Large-Scale Programs in Europe*

Georghiou, L., 1994, *Impact of the Framework Programme on European Industry*, EUR 15907 EN (Office for Official Publications of the European Communities, Luxembourg).

Larédo, Philippe, *The Impact of Community Research Programmes in France*. Final Report prepared for the European Commission. Ecole des Mines, Paris, 1995.

Luukkonen, Terttu, *The Difficulties in Assessing the Impacts of EU Framework Programmes*. *Research Policy*, Vol. 27, 1998, 599-610.

Luukkonen, Terttu, *Additionality of Publicly-Funded RTD Programmes*, The IPTS Report, No. 40, December 1999, 10-15.

Luukkonen, Terttu, *Additionality of EU Framework Programmes*, *Research Policy*, Vol. 29, 2000, 711-724, 2000a.

Luukkonen, Terttu, *Old and New Strategic Roles for the EU Framework Programme*. paper presented at Europolis Workshop in Maastricht, 13-14 April, 2000, 2000b.

Luukkonen, Terttu, *Technology and Market Orientation in Company Participation in the EU Framework Programme*, paper submitted to *Research Policy*, 2000c.

Luukkonen, Terttu and Sasu Hälikkä, *Knowledge Creation and Knowledge Diffusion Networks - Impacts in Finland of the EU's Fourth Framework Programme for Research and Development*. Tekes, Publications of the Finnish Secretariat for EU R&D 1/2000, Helsinki 2000.

Luukkonen, Terttu and Pirjo Niskanen, *Learning Through Collaboration - Finnish Participation in EU Framework Programmes*. VTT Group for Technology Studies, Helsinki, 1998.

Luukkonen, Terttu, Sasu Hälikkä, Pirjo Niskanen, Riikka Eela, *Finnish Participation in the Fourth Framework Programme*, Tekes, Publication of the Tekes International Co-operation 4/1999, Helsinki 1999.

Metcalf, Stan, *The Economic Foundations of Technology Policy: Equilibrium and Evolutionary Perspectives*. In: *Handbook of the Economics of Innovation and Technological Change*, ed. by Paul Stoneman. Oxford: Blackwell Publishers, Ltd., 1995, 409-512.

Niskanen, Pirjo, *EU Research Collaboration and Finnish Universities*. Paper presented at 4S/EASST Conference 2000, Vienna, 27-30 September 2000.

Niskanen, Pirjo, Riikka Eela, Sasu Hälikkä, Terttu Luukkonen, *Suomalaiset EU:n tutkimuksen neljännessä puiteohjelmassa*. Tekes, Kansainvälisten verkostojen raportti 3/1998, Helsinki 1998.

## *5 Evaluation of Large-Scale Programs in Europe*

Palmberg, Christopher, Petri Niininen, Hannes Toivanen and Tanja Wahlberg, Industrial Innovation in Finland, The Results of the Sfinno Project, Group for Technology Studies, Working Papers, 47/2000, Espoo, 2000.

Persson, Olle, Terttu Luukkonen, and Sasu Hälikkä, A Bibliometric Study of Finnish Science. VTT Group for Technology Studies, Working Papers 48/2000, Espoo 2000.

Peterson, John and Margaret Sharp, Technology Policy in the European Union. Macmillan Press, Houndmills and London, 1998.

Quintas, P. and K. Guy, Collaborative, 1995, Pre-competitive R & D and the Firm. Research Policy, 24, 325-348.

Reger, Guido and Stefan Kuhlmann, European Technology Policy in Germany - The Impact of European Community Policies upon Science and Technology in Germany, Physica-Verlag, Heidelberg, 1995.

The Second European Report on S&T Indicators 1997. European Commission. Luxembourg: Office for Official Publications of the European Communities, Luxembourg, 1997.

---

### **Discussion of Terttu Luukkonen's paper**

**Nick Vonortas** (George Washington University, USA) Networks in Europe are being brought up frequently in the discussion and this is one of the foundations of the framework program. Outside bibliometrics, how much rigorous work looking at networks (e.g., graph theory, maps etc.) has been conducted?

**Terttu Luukkonen.** One source is EU cooperative programs, OECD projects going on well. Recently there is an ongoing OECD project that uses the EU database with the project data of the Framework Programme. Unfortunately the data is incomplete, and it has to be cleaned. Results are not yet available. There has to be done more work on than topic.

**Luke Georghiou** (University of Manchester, UK). The Center for Sociology of Innovation (Mainz) characterized networks based not only on who is in them but on what they are doing.

**Maryellen Kelley** (National Institute of Standards and Technology, USA) The complete network is very difficult to measure. The requirements are dense and the task is enormous.

## The Evaluation of Nanotechnology Competence Centers in Germany

**Susanne Bühner**

Fraunhofer Institute for Systems and Innovation Research, Karlsruhe, Germany

Email: sub@isi.fhg.de

---

### Background

Since October 1998 the Federal Ministry of Education and Science (BMBF) has been promoting six competence centers (CC) for nanotechnology<sup>1</sup> which were selected on a competitive basis. It is important to mention that the CC are “virtual” ones, i.e. they are not concentrated in a small/close region.

The six nanotechnology competence centers get funding for their infrastructure respectively coordination efforts; in addition, the Ministry offers project grants. However, the invitations to tender are public (i.e. all interested German researchers are allowed to submit proposals) and the grants are awarded only after an ex-ante evaluation has been performed by a group of external experts.

The BMBF has created the CC in order to support the activities in the nanotechnology sector, especially for industrial application. The functions of the competence centers are mainly public relation, education and training, creation of an economically attractive environment and the counseling of mainly industrial prospects in the corresponding field of nanotechnology. Apart from these tasks, a first coordination of R&D activities and projects as well as consulting of applicants in the centers is expected.

The scheme has the following main objectives:

- Optimal transfer of (nanotechnological) knowledge into marketable products, processes and services.
- Setting up a competence profile in the selected technology field which makes the location attractive and well-known, at the national as well as at the international level.

---

<sup>1</sup> The promoted competence centers are (1) ultrathin functional layers (coordinator located in Dresden), (2) NanOp-application of nanostructures in optoelectronics (coordinator located in Berlin), (3) Construction and application of lateral nanostructures (coordinator located in Aachen), (4) Nanotechnology: Functionality by means of Chemistry (coordinators located in Kaiserslautern, Saarbrücken) (5) Ultraprecise surface manufacturing (coordinator located in Braunschweig), (6) Nanoanalytics (coordinators located in Münster, Hamburg and München).

### **The Evaluation (Objectives, Design, Methods)**

The objectives of the evaluation are (1) the assessment of the *impact of the scheme* illustrated by the six promoted nanotechnology competence centers, (2) the *stimulation of learning experiences and optimization processes* through the intensive exchange among all parties involved.

With regard to the schedule, the study is divided into three phases: (1) preparatory phase (about 6 months), monitoring phase (18 months), evaluation phase (6 months).

During the *preparatory phase*, the ancillary conditions are investigated, including the scientific and technological competences of the actors involved, the aims and strategies of the centers and the organisational models developed in order to achieve these aims. The *monitoring phase* is characterised by a detailed analysis of the structures of the (internal and external) cooperation and communication and a comparison with alternative models of competence centers — at the national and also the international level. Finally, the *evaluation phase* is devoted to the analysis of a number of quantitative and qualitative indicators referring to nine performance dimensions (scientific and technological rating, competence with regard to collaboration, interdisciplinarity, efforts to influence the (business) environment, education and training, public relations, norms and standards, long-term perspectives of the centers).

These indicators are applied, taking into consideration the technological and economic circumstances under which the competence centers operate. This means that every competence centre will get own weighting factors of the performance dimensions mentioned, for example center A a stronger weight for the scientific performance than center B, which focuses more on the educational side of the scheme.

### **Preliminary Results**

The evaluation is challenged by the two roles the evaluators have to play, that of a critical friend as well as an objective assessor. This role conflict is not easy to communicate within the nanotechnology scene. A further problem is linked to the particularities of the scheme: because project grants are not automatically involved, it is rather difficult for the persons at the top of these centers (the speakers of the competence centers or the coordinators) to make the benefits of the participation clear to the personnel involved.

---

## **Discussion of Susanne Bühler's paper**

**Irwin Feller** (Pennsylvania State University, USA) If the objective is to promote technology transfer into products/processes, are there objective assessment about how far nanotechnology is away from marketable products?

## *5 Evaluation of Large-Scale Programs in Europe*

**Susanne Bühler.** There was an assessment in the 1990s – a study of German engineers that estimated a large potential for nanotechnology.

**Arie Ripp** (Twente University, The Netherlands). Why has the Ministry chosen to support competence centers instead of Research Centers- is this a general trend? And, what are the stakes of evaluation within the decision context?

**Susanne Bühler.** Some of the competence centers may not survive. There are other competence centers internationally that can be used for comparison (e.g. Technopolis did a study on the success of Competence Centers).

**Stefan Kuhlmann** (Fraunhofer Institute for Systems and Innovations Research, Germany). There is currently lots of research funding in nanotechnology. The Ministry will support clusters of research institutions in the areas of competence and education. These institutions are needed to stimulate and speed up innovative activities in nanotechnology. Additional funding is supplied on top of the research funding to do cooperative activities. There is a high risk of free rider effects connected to this kind of funding.

**Joe Cullen** (Tavistock Institute, UK). For alternative competence centers, is there a benchmarking exercise where competencies must be mapped? What are the standards for comparison?

**Susanne Bühler.** There are success factors models that are supposed to be sustainable without program funding. The Benchmarking is not the most important task but, the identification of successful Centers. For comparison there is a center that was not selected. Thus, there is a model, in this case "Regional Centers", for comparison.

**John Barber** (Department of Trade and Industry, UK). It is not uncommon for governments to underestimate the time required to commercialize an innovation. So, the way you look at centers must account for this. There are barriers to use that lie in demonstrations that the technology will work. When program is set up, there must be a clear rationale for the barriers, and this should be used for deciding whether to establish a Research or a Competence Center.

**Maryellen Kelley** (National Institute of Standards and Technology, USA). Susanne, in your presentation you showed a slide that said that the centers did not think that standardization was very important. Do you know a reason for this assessment? Do you think that in a new field measures are important?

**Susanne Bühler.** From the Centers point of view it is not important so much for the evaluation but very important for centers in general, and they know that. The same holds e.g. for interdisciplinarity

**Maryellen Kelley.** Standards for instrumentation are often major obstacles to advances. If the center can contribute to this problem, it is a big deal.

**Susanne Bühler.** One of the competence centers focuses on standards in particular and thus for them it is very important.

**John Barber.** Ideally, evaluation should start before the program begins.

---

## **Commentary on the Session and Additional Discussion**

**Louis Tornatsky**

Southern Technology Council, USA

My comments on each of the three paper presentations are as follows:

For Laurent Bach: The BRITE project and the Fourth Framework programs are ideal for doing ATP-like evaluations and ATP might serve as a useful comparison. Additionally, quantitative economic impact studies end up in the public domain, but it is not clear always just how real they are. I also suggest using ordinal measures for participants which can be used to assess best practice. Then the projects could be evaluated in a more qualitative way.

For Terttu Luukkonen: I would liked to have heard more about the Fourth Framework programs in general. However, I did like your multi-modal approach.

For Susanne Bühler: The nanotechnology evaluation is much along the lines of the NSF Industry-University Cooperative Research Centers program. A person month of evaluation time of a faculty member from the same university is used in assessing this program. That person gets the inside “lore.” Those evaluators also meet twice per year and they then play the “critical friend” role of working with the centers for the purpose of program improvement.

**Luke Georghiou** (University of Manchester, UK). Just to remind the US colleagues, the European Framework Programs are not “a” program but a collection of programs with different audiences and objectives.

**Terttu Luukkonen** (VTT Technology Group, Finland). The EU framework programs are different from national programs. The companies have to choose what [EU or national programs] they want to participate in. The EU requires more sharing of information with other companies than do some national programs. Those programmes are specific cross border programmes.

**Irwin Feller** (Pensylvania State University, USA). I am fascinated with the Framework Program. At the policy level, have policymakers gone to ground zero to think about what should be done? Twenty years ago, the aim of the EU was to force collaboration across member states. But today that is not the problem that it was then. In fact, there are many global (let alone European) links between companies. What is the current rationale in political, economic, and social terms for the program?

## *5 Evaluation of Large-Scale Programs in Europe*

**John Barber** (Department of Trade and Industry, UK). The Framework programs are set up under protracted negotiation where there is a lot of political compromise. We are unable to go back to ground zero.

**Terttu Luukkonen.** To respond to Irwin Feller's question, in interviews conducted with people in telecommunications, they say the Framework Programs are too slow to be useful in the R&D cycle. As a result, companies create their own alliances which are most important and are time dependent. But even if companies do have cross border branches they do not necessarily have many cross border R&D activities ongoing.

**Joe Cullen** (Tavistock Institute, UK). In programs that are funded under the framework, organizations only get 50 percent of the cost of R&D (for non academic programs). Does this inhibit participation?

## 6. Evaluation of Regionally-Based Science and Technology Programs

---

### Evaluating Manufacturing Extension Services in the United States: Experiences and Insights

**Philip Shapira**

School of Public Policy, Georgia Institute of Technology

Atlanta, GA 30332-0345, USA

Email: ps25@prism.gatech.edu

---

#### Introduction<sup>1</sup>

Increased attention has been focused in the United States over the last few decades to the promotion of business and economic development through technology-based business assistance services (Coburn and Bergland 1995). The emergence of the US Manufacturing Extension Partnership (MEP) is a prominent example of this trend. At the start of the 1990s, the U.S. infrastructure for manufacturing extension services was patchy, comprising a handful of individual state programs, a few embryonic federal manufacturing technology centers, and a series of other rather uncoordinated federal and state technology transfer efforts. However, by 1999, more than sixty manufacturing technology centers or programs were operational in all fifty states, under the aegis of the MEP.

The MEP provides manufacturing extension services to small and medium-sized manufacturers and, in so doing, aims to upgrade the performance and competitiveness of these firms. The approach to technology is pragmatic, with an emphasis on best practice, known, and commercially tested techniques and methods. The kinds of services offered include information dissemination, assessment, problem solving, referral, and training. Typical customers are often conservative and risk-averse when it comes to technology, may at times be in need of basic training and workforce skills upgrading, yet have existing products and customers, and cash flow. The MEP is a collaborative initiative between federal and state governments that also involves non-profit organizations, academic institutions, and industry groups (see Table 1). The MEP's federal sponsor is the

---

<sup>1</sup> An earlier version of this paper was produced for the 1999-2000 Project on the Assessment of Business Incubators as Economic Development Tools in Developing Countries, sponsored by the United Nations Industrial Development Organization, Vienna. This latest draft has benefited from comments from Jan Youtie and Ken Voytek. The author remains responsible for the opinions and judgments contained in the paper.

## *6 Evaluation of Regionally-Based S&T Programs*

National Institute of Standards and Technology (NIST), within the U.S. Department of Commerce. While the federal sponsor, NIST, performs an ongoing funding and coordinating role, the states are also major funders of manufacturing extension services. In so doing, they bring into play state-level concerns about business development, job creation, state economic development, and local development (particularly in peripheral areas within their jurisdictions).

One of the interesting features of the U.S. manufacturing extension system is the considerable effort placed on performance measurement and evaluation. The federal sponsor has established ongoing performance monitoring and tracking systems; there is external review of center performance; and, as discussed later in this paper, about 30 evaluation studies of the program have been produced by a variety of authors over the last five years or so. Several factors help to account not only for the focus on evaluation within the MEP but also its form. In some areas of technology policy endeavor (for example, technology incubators or technology parks), states and local agencies develop their own programs with little consistent federal guidance, even though although they may draw on federal funds. States and localities vary greatly in how they evaluate technology programs within their jurisdictions. While robust evaluations do occur, typically states and localities are satisfied with simple, program self-reported, activity counts (see Melkers and Cozzens 1996). In contrast, the U.S. manufacturing extension system has evolved as a national federal-state partnership rather than a series of centers under state and local sponsorship operated individually or in association with area institutions as in the case of technology incubators. The role of the federal government as an ongoing stakeholder in manufacturing extension service provision encourages more consistent, if not increased, attention to be paid to evaluation. Manufacturing extension's national framework has allowed financial and intellectual resources to be invested on a system-wide basis, allowing economies of scale and learning in evaluation methodologies. Finally, the most substantial period of growth for manufacturing extension occurred in the mid-1990s during a period of heightened interest to performance measurement and "reinvented government" (Shapira, 2001). Again, this has encouraged attention to evaluation and review.

The basic categories of information that MEP evaluation seeks to develop are not dissimilar from those of other technology programs. There is a need to track public and private resource inputs and the particular services offered to individual firms. It is necessary to measure intermediate outcomes associated with services, track firms over time to assess longer-term effects, assess the performance of customers against non-assisted controls, identify regional and broader industrial impacts, and to provide an overall assessment of benefits, costs, and net returns from program intervention. Like other program areas, manufacturing extension faces considerable challenges in providing this information. For example, there are issues related to the period of time after program intervention that should elapse before full measurement of impacts can occur and to the diversity of hard and soft impacts that more complete evaluations should seek to measure. Similarly, evaluators also have to confront the problem of indirect links between program intervention and desired

## *6 Evaluation of Regionally-Based S&T Programs*

program outcomes. Manufacturing extension intervention occurs at initial and intermediate steps, for example by providing information or making technical resources available. Desired program outcomes, such as increased sales or business formations, will usually require further steps and investments by customers which are typically removed from program control and which may be affected by other external factors. The degree of attribution of subsequent downstream outcomes to program intervention is thus a shared challenge for evaluators. Evaluators face challenges of developing counter-factual evidence and controls or benchmarks. Ideally, program evaluation design should incorporate elements that can consider counter-factual evidence and arguments, i.e. what would have happened without the investment of program funds? Additionally, evaluators for each program have to address differences in stakeholder perspectives about what particular measures should receive most weight.

This paper reviews the justification and development of manufacturing extension services in the United States, then examines experience with evaluation with the aim of offering insights that might have broader relevance. In so doing, there is no suggestion that manufacturing extension has developed a “fault-free” system of evaluation. Far from it, for as we will see – notwithstanding the resources invested in manufacturing extension evaluation – many unresolved issues and challenges are evident. Nonetheless (or possibly because of this), examining the manufacturing extension case is useful because of the many points of comparison and differentiation it can offer.

### **US Manufacturing Extension: Justifications for Policy and Programmatic Intervention**

Manufacturing extension services seek to upgrade technology, business practice, and performance in industrial enterprises and industrial communities. In the United States, manufacturing extension services focus mainly on industrial companies with fewer than 500 employees. There are some 400,000 of these small and mid-sized manufacturers, forming an important and integral part of America’s industrial base. Many of these smaller firms find it difficult to introduce modern manufacturing technologies and methods. There are market failures on both the demand and supply side. Smaller firms frequently lack information, expertise, time, money, and confidence to upgrade their manufacturing operations, resulting in under-investment in more productive technologies and missed opportunities to improve product performance, workforce training, quality, and waste reduction. At the same time, private consultants, equipment vendors, universities, and other assistance sources often overlook or cannot economically serve the needs of smaller firms; potential suppliers of information and assistance also face learning costs, may lack expertise, or face other barriers in promoting the diffusion of rewarding technologies. System-level factors, such as the lack of standardization, regulatory impediments, weaknesses in financial mechanisms, and poorly organized inter-firm relationships, can also constrain the pace of technological diffusion and investment.

## *6 Evaluation of Regionally-Based S&T Programs*

Federal and state policymakers in the U.S. have sought to address the modernization problems of small and mid-sized manufacturers by organizing manufacturing extension programs. These programs promote the diffusion and deployment of new technologies and improved business practices among industrial firms. Approaches to manufacturing extension in the U.S. have evolved from initial single source providers of technical assistance to current networks which aim to coordinate public and private service providers to better meet the needs of smaller firms, regions, and industries. U.S. manufacturing extension programs typically employ industrially experienced field personnel who work directly with firms to identify needs, broker resources, and develop appropriate assistance projects. A variety of services are offered, including information provision, technology demonstration, training, and referrals. Given the economy-wide benefits of accelerating the deployment of technology and the difficulties many companies and industries face in independently implementing technological upgrades, manufacturing extension services in the U.S. are often viewed as examples of how collective public action in partnership with the private sector can make markets and economic development processes more effective.

In addition to the “pure” market failure rationalizations for manufacturing extension services, it is evident that proponents of these services also justify policy intervention on strategic and competitive grounds. A primary motivating factor underlying the initial federal involvement in manufacturing extension in the late 1980s was concern about poor industrial performance in the United States in the face of intense foreign competition. Similarly, at the sub-national level, state support for manufacturing extension services hinges greatly on how it performs and is perceived as an economic development and job strengthening strategy. In this sense, manufacturing extension is promoted as a strategic intervention with a combination of economic, technological, industrial, developmental, and political rationales. Manufacturing extension’s performance (and continued funding) is thus judged not solely by whether extension services remedy market failures, but also in terms of whether it maintains business support, promotes technological success stories, is associated with job creation, and is politically sustainable at both federal and state levels.<sup>2</sup>

### **Development of US Manufacturing Extension Services**

The provision of technology and related business assistance to small and mid-size firms in the United States is not an entirely new activity. In the 1950s and 1960s, state industrial extension and technology assistance programs were established in several states, including Georgia, North Carolina, and Pennsylvania. These early

---

<sup>2</sup> While in theory strategic motivations may lead to an “oversupply” of public manufacturing extension services, it turns out that in the U.S. the opposite (an “undersupply”) is at least as likely. In several instances, incoming state governors ideologically opposed to many forms of public intervention have reduced or eliminated support for manufacturing extension, despite coherent arguments about market failure needs for services in their jurisdictions.

## *6 Evaluation of Regionally-Based S&T Programs*

programs diffused technical information and used professional engineers and other technical specialists to help local firms improve their use of technology. By 1990, manufacturing extension and technology transfer programs had been established in 28 states. However, of these only about one dozen states operated field service networks using industrial experienced staff able to work on-site with firms - a factor critical in being able to address specific manufacturing problems on the shop floor of small firms.

The federal government started its own direct support of industrial extension programs in the late 1980s and early 1990s. Under the auspices of the 1988 Omnibus Trade and Competitiveness Act, the National Institute of Standards and Technology set up a handful of manufacturing technology centers, working in an initial collaboration with selected states (U.S. Congress, 1990; National Research Council, 1993). A dramatic expansion of federal sponsorship came with the Technology Reinvestment Program, first announced in 1993 and implemented in 1994 and 1995 (Shapira 1998). Although this multi-billion dollar program was targeted towards the post-cold war conversion and restructuring of America's defense-industrial base, it did make significant funds available for the upgrading and deployment of technology in civilian industries, including small and mid-sized manufacturers. The National Institute of Standards and Technology was allocated a share of these resources to increase the number of manufacturing technology centers and to organize, through what was then the embryonic Manufacturing Extension Partnership, a collaboratively-delivered set of industrial extension services to manufacturers throughout the country. Funding was awarded through a competitive review process and states and other applicants generally had to match one-half of proposed costs with their own or private funds. The Technology Reinvestment Program had a major impact on the structure and character of the MEP. Funding was awarded through a competitive review process and states and other applicants normally had to match one-half of proposed costs with their own or private funds. Additionally, applicants were guided to form partnerships of service providers. MEP proposals were judged in terms of the number, diversity, and skills of constituent service providers, geographic scope and coverage, cohesiveness, organization, and management structure.

The funds made available by the Technology Reinvestment Project allowed a rapid ramp-up of the MEP program, to achieve near national coverage. By the mid-to-late 1990s, Technology Reinvestment Funds were phased out, and the federal share of the MEP was incorporated entirely into the civilian budget of the US Department of Commerce. By the end of the 1990s, the MEP had grown into a national network of more than 60 MEP centers in all 50 states.

### **Structure and Operation of the MEP**

Although there have been changes in sources of federal funds for the MEP, there has been consistency in three key program design elements. Mostly, these design elements were established as part of the rules and procedures of the 1988 Omnibus

## *6 Evaluation of Regionally-Based S&T Programs*

Trade and Competitiveness Act. First, all proposals for MEP centers are reviewed through a competitive process, typically involving independent review panels. Second, MEP center proposers are required to provide state matches to federal funds, and must also develop strategic and operational plans to generate private fee income. Third, all MEP centers are subject to periodic external reviews, with continued federal funding based on performance and strategic orientation.

However, there have been two important departures from the early program design. First, Congress originally envisaged that NIST's manufacturing centers would transfer advanced cutting-edge technology developed under federal sponsorship to small firms. But MEP staff soon realized that small companies mostly need help with more pragmatic and commercially proven technologies; these firms often also needed assistance with manufacturing operations, workforce training, business management, finance, and marketing to get the most from existing and newly introduced technologies. Most MEP centers now address customers' training and business needs as well as promote technology. In general, centers have found that staff and consultants with private-sector industrial experience are better able than laboratory researchers to deliver such services.

Second, Congress initially expected that MEP centers could become self-sustaining after 6 years of operation, without ongoing federal funds through the MEP program. The legislation set out a schedule for sustainability, with federal funds ramping down from one-half after the third year to zero after year six. This was known as the "sunset" clause. In practice, it was realized soon after the MEP program got underway that a regime full self-sufficiency was too severe, and would work against desired public purpose objectives. To generate high levels of private fee revenue to replace federal funds, MEP centers would need to narrow their range of service offerings and go "up-market" – serving larger firms willing and able to pay for market-priced consultancy services. Less assistance would be available for small and medium sized firms – the primary target group of policy. Additionally, in many states, the availability of state funds is contingent on federal funds. Thus, if federal funds were to be withdrawn, it is likely that a number of states would end their funding too. The goal of a national system would no longer be met, and service provision would return to the patchy state coverage existing prior to the development of the MEP. After several years trying to change the sunset clause, MEP proponents succeeded in gaining language in the Technology Administration Act of 1998 that allowed continue federal funding past the center six-year mark. This allows federal funding to be continued after year six at one-third of an MEP center's budget, subject to performance and satisfactory external reviews every two years. Although individual center funding sources will vary, on average it is now anticipated that typical MEP center budgets will be comprised of one-third federal funds, one-third state funds, and one-third private fee revenues.

In its current structure and operation, the MEP incorporates most of the principles articulated in recent government reform proposals in the United States. First, the program seeks a cooperative relationship between the public and private sectors. The private sector is involved not just as a recipient, but also as a service partner and an

## *6 Evaluation of Regionally-Based S&T Programs*

advisor. Second, the program is decentralized and flexible, with individual centers able to develop strategies and program services which are appropriate to state and local conditions. Third, the MEP seeks not to duplicate existing resources. Rather than provide services directly from the federal level, MEP awards are designed to get existing service providers, whether they be consulting firms, non-profit organizations, academic institutions, public agencies or trade associations, to cooperate and coordinate in their efforts to assist local manufacturers.

It is important to understand the hybrid organizational nature of the MEP. The MEP program is national, in the sense that there are centers in all U.S. states. But the MEP is not a purely federal effort, in contrast to the U.S. Advanced Technology Partnership – a federal program that promotes national consortia of private companies, laboratories, and other institutions to develop and commercialize leading-edge technologies; nor is the MEP a purely local effort, as is the case with state technology initiatives like Pennsylvania's Ben Franklin partnership or the Georgia Research Alliance – programs that are concerned only with technology development in individual states, without regard to national impact. Rather, the MEP embodies a partnership between different levels of government – with the relationships between these levels undergoing change over time.

MEP centers are typically structured either as separate non-profit corporations or as part of other organizations, such as universities, state agencies, technology centers, or economic development groups. Individual MEP centers typically operate with multiple field locations: latest estimates indicate that the 60 MEP centers operate more than 400 MEP field offices. Additionally, individual MEP centers operate with and through local networks of associated public and private service providers. The MEP partnership network is comprised of about 2,500 affiliated public and private organizations that deliver or support the delivery of services to small and medium-sized manufacturers. Centers usually have governing or advisory boards that include local public and private sector representatives. NIST, the federal sponsor, works with individual centers and with states in program management and development and in backing joint working groups, staff training, information and communications systems, common tools, and performance measurement and evaluation.

For fiscal year 1999, MEP's federal funding was \$98 million, of which about \$88 million went to direct center support. State matching funds to centers added a further \$101 million, while private fee revenues reached about \$81 million. Total MEP center revenues from all sources thus equaled \$270 million, with the following revenue distribution for a typical center: 33% federal, 37% state, and 30% private fees. For the MEP system as a whole (adding back in the \$10 million of NIST central costs plus an imputed \$5 million in state central administrative costs), the total system cost was around \$285 million. There are some important points to observe about these budgetary numbers. First, public funding in the U.S. for manufacturing extension has increased by an order of magnitude over the last ten years. Federal funding in fiscal 1999 was 16 times greater than the \$6 million in federal funds allocated a decade earlier, in fiscal 1989. State matching funds have

## *6 Evaluation of Regionally-Based S&T Programs*

increased by about 8 times over the same period. Second, although federal MEP funding peaked at \$112 million in fiscal 1997, it now seems to have stabilized at around the \$100 million mark. Third, private fee revenues have increased rapidly in recent years. For example, in fiscal 1997, about \$43 million in private fees were generated by MEP centers. The current level of fee generation has almost doubled in two years, and is likely to increase somewhat further in future years. The goal of one-third revenues from private sources, to match equivalent thirds from federal and state funds respectively, thus looks feasible. Finally, while these reported numbers capture the budget trends for the MEP program, there are other federal and state resources allocated to services that have some similarities to manufacturing extension that are not counted here in terms of dollar budgets. These include such programs as federal laboratory technology transfer services, NASA regional centers for technology transfer, DoD manufacturing technology programs, and various state and local business assistance programs.<sup>3</sup>

### **MEP Center Services**

In 1999, 23,100 firms were assisted by the MEP through assessments, technical assistance projects, information workshops, training, and other services. This was up from about 19,000 assisted firms in 1996.<sup>4</sup> For 2000, the MEP aims to serve 25,000 firms. Although the MEP does serve some larger firms, the program primarily serves the target group of companies with fewer than 500 employees. In 1999, 64 percent of assisted companies had fewer than 100 employees, while a further 30 percent employed between 100 and 499 employees. About 60 percent of the companies served annually are new clients, with the balance comprised of repeat customers.

About 68 percent of all MEP activities with firms fall within the category of technical assistance projects. These projects are usually focused towards specific problems or opportunities within firms. A further 10 percent of activities are categorized as assessments – these activities include initial assessments and strategic reviews at companies to determine priorities and needs for action. Training activities, for managers and employees, comprise the balance (22 percent) of MEP activities. By substantive area, MEP activities tend to focus on process techniques and “soft” business and manufacturing practices rather than “hard” new technologies such as factory automation. The top five substantive areas of assistance (by percent of activities) are process improvement (17 percent), quality (14 percent), business systems and management (11 percent), human resources (11 percent), and plant layout and shop floor organization (9 percent). Environmental and market development activities each comprised a further 8 percent. At the other end of the

---

<sup>3</sup> Although not counted in the MEP’s direct budget numbers, many affiliates of these other programs are counted as service partners in the MEP’s estimate of its national network.

<sup>4</sup> The total volume of inquiries is somewhat greater than the actual number of firms assisted, although there is no national data on this measure.

## *6 Evaluation of Regionally-Based S&T Programs*

spectrum, automation projects and CAD, CAM, or CAE assistance comprised just two percent of activities.<sup>5</sup>

The typical MEP center has a budget of about \$4.5 million a year, of which about 30 percent is derived from fee revenues, employs about thirty-five professional and technical staff and uses an additional ten consultants each quarter. However, there are wide variations in individual budgets, fee revenue generation, staffing, market areas, and clients served. Smaller MEP centers, usually serving dispersed rural areas, may have as few as 1,300 to 1,500 enterprises in their target market areas, employ perhaps one-half a dozen staff members, and serve under 80 firms a year. Larger centers in urbanized locations or which serve large states may have more than 15,000 firms in their target markets, employ 60 or more staff, and assist upwards of 1,200 firms a year. Annual budgets vary accordingly, from several hundred thousand dollars for small MEP centers to upwards of eight to ten million dollars for the largest centers. The proportion of fee revenues also varies, although this is influenced by other factors besides size. These other factors include the center's aggressiveness and strategies in seeking fee revenue, the level of state support (lower levels of state funding usually promote higher fee seeking activities), state policy (some states view manufacturing extension as a public service mission that should not be driven by high fee revenue goals), and market sophistication (firms in urban areas and in higher value-added industrial sectors may be more willing to pay). Depending on such factors, fee revenues for individual centers can range from around 20 percent to more than 40 percent of total budget.

The decentralized and flexible structure of the MEP allows individual centers develop strategies and services appropriate to state and local conditions. For example, the Michigan Manufacturing Technology Center specializes in working with companies in the state's automotive, machine tool and office furniture industries. The Chicago Manufacturing Center has developed resources to address the environmental problems facing the city's many small metal finishers. In Georgia, there has been a focus on quality and lean manufacturing methods, aimed at upgrading the state's many routinized manufacturing facilities.

In addition to individual center services, groups of centers collaborate with one other, the NIST national program, and other organizations to implement shared tools and service offerings to firms. An early example is the Performance Benchmarking Service (PBS), which allows manufacturers to benchmark quality, productivity, other shop-floor measures, and corporate performance with those of best practice firms in relevant sectors. PBS was initially developed by the Michigan Manufacturing Technology Center, received further sponsorship by the MEP national program, and is now used for client benchmarking by several MEP centers, with the coordination of the Michigan center. Another initiative, known as Supply America, combines the resources of individual centers to upgrade geographically dispersed small firm supply chains of larger manufacturers. The aim to offer standardized, easy access MEP services that can across state lines. Other

---

<sup>5</sup> Data in this section is drawn from the MEP National Reporting System.

## *6 Evaluation of Regionally-Based S&T Programs*

collaborative service thrusts sponsored by the national MEP include quality, ecological manufacturing, and, now receiving greatest attention, lean manufacturing and electronic commerce (“eBusiness”). In lean manufacturing, the MEP has developed personnel training programs and a suite of products that field staff can offer to companies. For eBusiness, the MEP has developed awareness and education materials, eBusiness assessment, planning and implementation tools, and an online eBusiness solutions center.

### **The Evaluation of the MEP**

Manufacturing extension services operated through the U.S. MEP are subject to a variety of performance measurement strategies and evaluation approaches. Among the sponsors of measurement and evaluation studies are NIST, federal oversight agencies, state governments, individual centers, and research and policy foundations. The performers of studies include program staff at federal and state levels, external auditors, university researchers, and private consultants. The following section describes the broad types of performance measurement and evaluation activities undertaken within the manufacturing extension field.

At NIST, performance measurement focuses on tracking such indicators as clients served, types of service activities, costs per unit of service delivered, client satisfaction, and client actions taken. Center inputs are captured by MEP management information systems, to which centers submit regular standardized reports. The principal performance measurement tool to capture intermediate business outcomes is a national post-project survey conducted up to a year after MEP projects have closed.<sup>6</sup> The survey is implemented by telephone by an outside survey house and probes quantitative impacts such as sales, cost savings, jobs, capital investment, and productivity improvements that result from MEP projects. In the first quarter of 1999, NIST reported impacts of \$435 million in added sales, \$42 million in cost savings, \$171 million in capital investment, and \$22 million in capital savings, as a result of MEP projects.<sup>7</sup> Results from this survey are reported by NIST against goals established for the MEP program under the U.S. Government Performance and Review Act (GPRA).<sup>8</sup> On an annualized basis, NIST MEP

---

<sup>6</sup> The NIST post-project survey focuses on collecting intermediate and quantitative impacts such as sales, cost savings, jobs, capital investment, and productivity improvements. The questionnaire is administered by telephone to all MEP project customers. The survey questionnaire and implementation procedure has been subject to a great deal of attention by system stakeholders and has recently been revamped. The Census Bureau formerly administered the survey, but it is now administered by an outside private survey house.

<sup>7</sup> These are aggregates of customer self-reports in the post-project survey and are not controlled or audited.

<sup>8</sup> GPRA requires federal agencies to develop measurable goals and conduct annual performance assessments. Goals for the NIST MEP are contained in the Technology Administration’s FY 2000 Performance Plan.

## *6 Evaluation of Regionally-Based S&T Programs*

significantly exceeded its sales goals (by a factor of 3), and has met its designated cost savings and capital investment goals.

In addition to the post-project survey, NIST has also supported a series of other evaluation studies. Logic-based case studies have been conducted to better understand high impact projects and the transformation of firms through MEP intervention (Cosmos 1997). Controlled studies have combined program and census data to ascertain and control for pre-project characteristics of customers and post-project outcomes (e.g. Jarmin 1997, 1998, 1999). Pilot assessments of inter-firm networking have been supported (such as??). Furthermore NIST has supported several state-of-the art workshops on the evaluation of industrial modernization and organizes an evaluation working-group to bring together those working on evaluation within the MEP system to review procedures and results.

The 1988 Omnibus Trade Act required external reviews of manufacturing technology centers by outside reviewers. This is a very important element in the portfolio of MEP evaluation measures. Outside review teams are constructed for each center by NIST and typically include program managers from other states, business representatives, other independent experts, and NIST staff. Each center has submits a report including an overview of its performance, strategies, and financial position to the review team. As amended, federal law requires satisfactory review of center performance every two years to ensure continued federal funding. Recently, NIST has established a performance review system based on the Baldrige quality criteria. Centers are reviewed under seven groups of criteria: 1) leadership, 2) planning, 3) customer knowledge and relationships, 4) performance information and analysis, 5) center staff workforce practices and workforce environment, 6) process management, and 7) performance measures. Centers that are judged to be weak in some or all of these areas may be subject to action recommendations by NIST as a condition of further federal funding. Recommendations may range from requirements to revise strategic plans to substantive organizational change.

In addition to NIST's reviews of individual centers, external oversight reviews of the whole MEP system have been conducted by oversight agencies, for example by the U.S. General Accounting Office (GAO 1995). The U.S. Government Performance and Review Act (GPRA) also requires federal agencies to develop measurable goals and conduct annual performance assessments. Goals for NIST MEP are contained in the Technology Administration's FY 2000 Performance Plan (U.S. Department of Commerce 1999).

At the state level, additional evaluation norms and procedures may be imposed on MEP centers, although with much variation. A few states have required independent evaluations of their programs. One example is New York, which has sponsored its own external evaluations conducted by outside consultants (see Nexus Associates 1996). Pennsylvania has also sponsored robust evaluations of its Industrial Resource Centers, which hold MEP franchises too (Nexus Associates 1999). However, to date many states have been satisfied with simple counts of program activities (e.g. number of firms served), and simple testimonials as to job

creation. This may change as states now increasingly seek improved documentation of program performance.

Individual centers have established their own procedures to evaluate customer satisfaction and program impact. The methods used vary greatly in terms of sophistication, metrics, and robustness. Some centers do very little, leaning instead on NIST's national evaluations. A few centers sponsor extensive efforts. For example, the Georgia MEP has a distinct evaluation element that has conducted a series of evaluative studies including controlled surveys, cost-benefit analyses, and other special studies (see Shapira and Youtie 1998). The Michigan MEP has used a comprehensive benchmarking protocol, known as the Performance Benchmarking Service, to undertake controlled studies of program impact (Luria and Wiarda 1996).

### **Findings from Evaluation Studies**

Beginning in the early 1990s, the evaluation procedures described above have produced steady enlarged the body of knowledge and methodological experience about the impacts and effects of manufacturing extension services. Mostly, evaluation studies have resulted in publicly available reports produced by evaluation performers for sponsoring bodies. A number of these studies have also been published in peer-reviewed journals.

The largest of the evaluation efforts - the NIST post-project survey - has to date primarily been used to report gross program results to the U.S. Congress and other interested parties. NIST reports the number and type of projects conducted, characteristics of the firms assisted, and aggregated impacts attributed to MEP projects for factors such as sales, cost savings, capital investment leveraged, and jobs (MEP 1997). In addition to the ongoing NIST national survey, more than 30 other separate empirical evaluation studies of manufacturing extension service impacts have been conducted between 1994 and 1999. A variety of methods have been used in these studies, with some studies using more than one method. While the most common method was to survey customers, it is worth noting that the next most common method was a survey that also used a comparison group. Case studies were the third most frequently used approach. (See Table 2.) A summary review of these studies is attached in the Appendix to this paper.<sup>9</sup>

Those studies that use customer surveys report mostly positive results. Manufacturing extension customers are generally satisfied with services and indicate overall business performance is improved (GAO 1995, Ellis 1998). Although, Swamidass (1994) finds that the MEP is only one of multiple sources of

---

<sup>9</sup> The author would appreciate references to any manufacturing extension evaluation studies that are not included in the appendix table.

## *6 Evaluation of Regionally-Based S&T Programs*

assistance,<sup>10</sup> and may not be most important, another study indicates that over two-thirds of customers act on the recommendations provided by program staff (Youtie and Shapira 1997). Several studies confirm that firms find it hard to report precise dollar impacts, but those who do report indicate increased sales, cost savings, and job impacts (Oldsman 1996; MEP 1997, 1998). Improvements in information and technology access are also reported (Youtie and Shapira 1997).

However, although most customers of manufacturing extension services are satisfied with the services received and do tend to pursue follow-up actions, the distribution of subsequent benefits is skewed. For example, Oldsman (1996) finds that a few firms generate large impacts from manufacturing extension projects; many firms report negligible impacts. Shapira and Youtie (1998) find product development projects most likely to generate new sales and jobs. Quality projects have the lowest reported quantitative dollar impacts. In part, reporting of negligible impacts may reflect survey questionnaire deficiencies in capturing the full range of hard and soft effects. But it may also reflect the fact that many manufacturing extension projects are of small scale and, by themselves, cannot be expected to have dramatic effects. Survey timing is also an important factor. MEP customers overestimate benefits and underestimate costs immediately after project completion, compared with follow-up interviews with the same customers one year later (Youtie and Shapira 1997).

Case studies have proven valuable in further probing the reasons for differential effects of program services, particularly in understanding the factors that lead to high impact projects (see Cosmos 1997; Youtie 1997). For example, parallel changes in management personnel and strategy often seem to be associated with firms pursuing more intensive engagements with manufacturing extension services.

When control group surveys are implemented, the results that emerge from the portfolio of such studies show more mixed results than surveys without comparison groups. This is not surprising. When controls are put in place, more complete allowance is made to discount the effects of changes that might well have occurred without program intervention. One controlled study of manufacturing extension customers and non-customers in by Shapira and Rephann (1996) in West Virginia found that customers adopt some (but not all) technologies at higher rates than firms not associated with manufacturing extension programs and are more receptive to new technology. Similarly, Luria and Wiarda (1996) found that MEP customers improve faster than non-MEP customers in adopting most technologies, except information technologies (Luria and Wiarda 1996). Two other controlled studies have found that customers have faster growth in value-added per employee (Jarmin 1999; Shapira and Youtie 1998) than non-assisted firms. However, Luria (1997) finds that MEP customers improve over non-customers in growth in sales, jobs, and certain process improvements, but there is no distinguishable improvement in

---

<sup>10</sup> The question phrasing in the Swamidass study may have lowered respondent reports of their use of MEP-like services. MEP centers are often not recognized as “government”, although this was the category used in this particular study.

## *6 Evaluation of Regionally-Based S&T Programs*

wages, profits, and productivity. Oldsman (1996) finds that the average customer added fewer workers than similar non-customers. Poorer performance in job growth is not necessarily a weakness, since one of the major aims of manufacturing extension is to enhance productivity. Where productivity is increased, direct employment (especially of shop-floor workers) may be reduced as efficiency is increased. But it does suggest that not all desired policy outcomes can be achieved simultaneously.

Another small group of studies has examined the costs and benefits of manufacturing extension services from public perspectives, rather than those just of customers. Here, the findings are moderately positive. Cost-benefit analyses by Shapira and Youtie (1995) and by Nexus (1996) show the net public and private benefits of manufacturing extension services exceed costs by a moderate ratio, while a Pennsylvania study (Nexus 1999) reports strongly positive net public benefits. Thompson (1998) finds the taxpayer payback to a state varies from just under break-even to positive.

### **Use of Evaluation Studies and Outstanding Issues**

For programs providing firm level services, be it through technical assistance to existing firms or support for new business formation, the manufacturing extension case exemplifies the importance and value of understanding who is being served, what services are being provided, what are the public and private costs, and what results are being generated. Evaluation at the firm level requires the development and maintenance of information systems that can accurately track program expenditures and income, staffing, customer profiles, and service interventions. Manufacturing extension services have also demonstrated the usefulness of implementing customer valuation procedures that measure satisfaction with services received and which can also track immediate actions taken and the value placed on service.

We have seen that evaluators have drawn on manufacturing extension service tracking databases to undertake impact assessments that seek to measure effects over time on such factors as investment, productivity and value-added per worker, wages, jobs, sales, quality and manufacturing operations. Survey-based tools have been most commonly used to identify service impacts, although evaluators have also used structured case studies, performance benchmarking tools, and other methods, and more formal econometric tools and analysis to explore the effects. In several instances, longer-term service impact evaluations have been designed to include control groups of non-assisted customers, to try to measure the “additionality” generated by program intervention.

The evaluation of firm level services in manufacturing extension has been aided by the elaboration of program logic models that delineate program inputs, work processes, intended intermediate outcomes, and anticipated technological, business, and economic development consequences. Information systems, procedures, and tools have been developed to obtain measures at successive steps. However, one weakness (or at least a trade-off from decentralized innovation) in information

## *6 Evaluation of Regionally-Based S&T Programs*

systems is that centers implement their own systems. There is no national standard or system, although NIST seeks to develop standardization in information systems through a new 13-center pilot integration project.

In addition to evaluating firm level services, the manufacturing extension system has constructed procedures to inform and implement external reviews of manufacturing assistance centers and associated service providers. Periodic independent performance assessments by outside panels provide expert advice and credibility to this process, drawing on center management reports, accumulated evaluations of firm services, and on-site center visits. Evaluation has also sought to contribute to system-wide assessments of manufacturing assistance policies and to inform strategic decision-making at regional and national levels. This has required linking results from program evaluations to ongoing assessments of firm and industry problems and directions and benchmarks of how firms are using technologies and techniques. The aim has been to promote learning about unmet needs, new opportunities, and interactions with other private and public actions.

The information provided by these varied evaluation studies and methods has served to justify and highlight the positive impacts that the program has to national and state funding agencies and stakeholders. In some cases, individual sets of findings have been linked with regional econometric models to estimate broader economic effects. To a lesser degree, evaluation findings have been used to prompt learning within the program and to suggest strategies for improvement. For example, it has been found that larger projects (involving more hours of staff and company time) have greater strategic impacts on companies. Although this seems a commonsense finding, it is important because the norm in the MEP is to conduct relatively small projects with firms. It has also been found that different services lead to different kinds of impacts. While environmental projects are more likely to generate cost savings, product development services (as already noted) are more likely to lead to increased sales and jobs impacts. Program managers indicate that these kinds of findings are useful for orienting program resources and are certainly easier for them to interpret than complex econometric controlled studies (Shapira and Youtie 1998).

However, although the MEP has a better record when it comes to evaluation than most federal programs, a series of issues have arisen. First, different stakeholders have rather different expectations related to evaluation. These differences reflect diverse objectives, contrasts in time horizons, and varied capabilities to consume evaluative information. States, for example, are most interested in jobs (and don't mind if their efforts to support business shift jobs from other states). At the federal level, however, competitiveness and productivity in a national context is stressed. And, while individual centers are concerned with economic measures, their immediate yardsticks may focus more on project counts and fee revenues.

Second, it is apparent that manufacturing assistance services have wide-ranging, often disparate effects that are not fully captured by most current evaluation procedures. In efforts to develop standard impact measures of "visible" results (e.g.

## *6 Evaluation of Regionally-Based S&T Programs*

changes to sales, costs, or jobs), longer-term, specialized, or “softer” impacts of program activities have been de-emphasized. Interestingly, MEP initiatives in training or education receive little systematic evaluation, although they comprise about one-third of program activities. Similarly, services to promote customer-supplier dialogue or inter-industry networks tend to have less tangible effects on the immediate bottom-line of firms than, say, fixing machine problems or saving inventory costs. Hard to measure or intangible activities are given less attention in standard evaluation schemes, meaning that programs have weaker incentives to conduct these activities - even though these activities may ultimately be more important for long-run fundamental upgrading.

Third, even for the more standardized program measures, it has proven quite difficult to reliably measure many program impacts. Many MEP interactions with firms are relatively small. It is difficult to discern effects on jobs or sales or to separate effects attributable to other factors. Indeed, when asked, most firms are unable to ascribe accurate numbers to program impacts on jobs, sales, or other economic variables (they are more likely to check a yes / no box which simply indicates whether an impact is present).

Fourth, the development of appropriate information bases and methods for evaluation remains a problem. At the center level, most MEP centers have complex relationships with service partners, and developing systematic information (and avoiding double counting) is an issue. The robustness of some center methodologies is also problematic. At the national level, the NIST evaluation effort has tended to become highly focused on producing “measurable” economic results for funding sponsors. NIST has also sought to tie evaluation results to center funding reviews. Both of these purposes are relatively narrow uses of evaluation, meaning that more penetrating and exploratory studies that emphasize learning and improvement are under-emphasized.

### **Insights from the Manufacturing Extension Case**

Previous sections of this paper have described the practice and progress of evaluation of U.S. manufacturing extension services. It is now appropriate to probe how this particular set of experiences might be more broadly relevant to other areas of technology policy evaluation. The strongest grounds for comparison are not so much in the performance of particular services, but in how underlying and common challenges of evaluation are addressed. This final section seeks to distill insights about commendable (and less than commendable) evaluation experiences from the manufacturing extension field that have broader application and which can offer points of comparison for the evaluation of technology incubators.

Perhaps the first point to reiterate is that evaluation should be integrated with, and used to inform, public sponsorship, policymaking, and program management in an ongoing way (and not after the fact). This will be aided if evaluation is implemented simultaneously with program funding. Prior to federal government involvement in U.S. manufacturing extension, evaluation of existing state level services was patchy.

## *6 Evaluation of Regionally-Based S&T Programs*

However, with the enlargement of an ongoing federal role and the building of the MEP in the 1990s, evaluation has been rather more consistently built into program and service design. Ideally, as new technology incubators are established or as new funding or programmatic directions are pursued with existing incubators, opportunities will be presented to initiate or strengthen ongoing evaluation systems. It is to be expected that conflicts and tensions will arise, but nonetheless the utility of evaluation will be much enhanced if it is incorporated into planning and program design as early or as soon as possible.

Clear statements of what programs are designed to accomplish clearly also aid evaluation. Statements of desired outcomes should not focus exclusively on economic impacts, but should also seek to assess learning outcomes such as knowledge transfer, skill upgrading, and effects on such elements as strategic management, risk perception, and technological culture. However, the reality is that such statements are not always easily agreed upon. In the U.S. manufacturing extension case, the legislative language is not conclusive (it emphasizes the improvement of “competitiveness” without much further definition), different public and private stakeholders have diverse goals, and local programs have the flexibility to customize program priorities. In many ways, the evaluation process has served as a “negotiating forum” to detail what outcomes should be emphasized through measurement and to adjust outcome priorities as the program has evolved over time. A consensus has emerged on the importance of a few factors such as productivity, jobs, and wages. Tensions remain about the importance of other outcome measures, how to incorporate both hard and soft measures, and the accommodation of differential time periods for varied kinds of impacts. While this experience takes nothing away from the value of upfront statements about program objectives, it suggests that program design needs to establish mechanisms whereby stakeholders can periodically review and refine what outcomes are most important in future periods, so that evaluation measures can match what programs are trying to achieve.

The manufacturing extension case also illustrates that evaluation involves far more than collecting information. Resources need to be allocated to analyze data and report it in ways that are useful to program managers, staff, sponsors, and outside stakeholders. Results (whether good or bad) need to be explored to understand what caused what and why and how much was attributable to the program. Good case studies are often worthwhile here. Findings should be discussed and made available outside of the program, to policy bodies, industry councils and advisory groups. Peer review, including publication in recognized scholarly and professional journals, should be encouraged. Good evaluation should probe and stimulate dialogue, as part of a continuous improvement management approach.

## *6 Evaluation of Regionally-Based S&T Programs*

To achieve these goals, evaluation needs to be carefully organized and adequately funded.<sup>11</sup> Program staff can and should conduct routine data collection and reports. However, external independent evaluators should also be used to conduct special studies, ensure robust methodologies, and to monitor internal processes to ensure credibility.<sup>12</sup> Evaluation resources should also be managed not only to implement regular and timely assessments of key variables, but also to support special studies, methodological innovations, and the development of a network of evaluators. At the same time, caution has to be exercised so as not to overburden customers and programs with too much data collection. Appropriate data protection should always be maintained, but there should be opportunities to make data available to outside researchers.

For manufacturing extension, a mix of evaluative methods has proven necessary and valuable. This is an expected finding that also applies for other technology policies. As a starting point, systematic procedures for tracking program activities are necessary. Knowing who has been served by what kind of service is critical. From this, simple follow-up surveys may provide information on customer valuation and satisfaction. These can be conducted in-house. More comprehensive, longer-term assessments, case studies, and reviews are desirable, preferably involving external evaluators. All methods should be carefully tailored and administered to ensure data collection is feasible, appropriate, and not overburdensome.

The national federal-state-industrial partnership framework found in the MEP highlights the point that attention needs to be given to the relationships between national and regional sponsors and local service centers in conducting evaluations. In many ways, the program is a wonderful example of a performance measurement partnership. It would be inefficient and ineffective if local centers were exclusively charged with evaluation responsibilities. Each local organization would need to implement its own evaluation plan, there would likely be wide differences in commitment and approach, and economies of scale and learning opportunities would be lost. On the other hand, while it is administratively easier for a national or regional body to assume all responsibility for evaluation, a top-down approach would probably be too rigid and standardized, constrain learning, and be hard to implement in a multi-organizational environment. Perhaps the most preferred institutional architecture is a collaborative one, where national, regional, and local levels collaborate in designing, implementing, and improving evaluation systems, and where cross-organizational benchmarking and learning is encouraged.

---

<sup>11</sup> The effort expended in evaluation should be proportional to the resources expended on program services. As a rule of thumb, for a \$5 million annual program in the manufacturing extension field, it is reasonable to allocate 2 - 3 percent of the total budget to evaluation. As the program size diminishes, a slightly higher percentage is justified. As the program size increases, a somewhat lower percent will suffice.

<sup>12</sup> Up to now, most of the robust studies and methodological innovations in the evaluation of manufacturing extension services have come from independent outside researchers (at times supported by NIST funds), rather than from internal NIST program staff.

## *6 Evaluation of Regionally-Based S&T Programs*

Finally, the manufacturing extension case emphasizes that evaluation should be tasked not only with justifying the program to funding sponsors, but also with promoting learning and program improvement. If the latter is to succeed, it is also important that an evaluation community be stimulated and supported (through training, workshops, and electronic interaction) to allow the development of evaluation capabilities, the application of robust and innovative methods, and the active dissemination of findings. This evaluation community should have opportunities and requirements to engage in dialogue about findings, needs, and implications with program managers, sponsors, industry representatives, and policymakers.

**Table 1. Organization and services of manufacturing extension services in the U.S.**

	<b>Manufacturing Extension Services</b>
Objectives	Upgrade technology and business practices in industry to promote competitiveness
Program framework	Nationwide federal-state partnership (Manufacturing Extension Partnership)
Federal sponsor	National Institute for Standards and Technology (MEP), U.S. Department of Commerce
Local hosts	Diverse array of state and local hosts, including non-profit organizations, universities and colleges, state and local agencies.
Primary customers	Existing manufacturing enterprises with fewer than 500 employees (SMEs)
Sponsors	States, nonprofits, universities, other organizations
Scale	60+ manufacturing extension centers in 50 states; 400 offices; 2,600 partner organizations
Staffing	About 2,000 industrially experienced field staff. Ratio of about 350 target SMEs per 1 FTE staff
Services	Information dissemination; technology needs assessment; problem solving; technical advice and guidance; implementation assistance; brokerage and referral; training; demonstration; network promotion.
Revenues	Over \$280m annually in FY 99-00: \$98m federal; \$101m state; and about \$81m from fee revenues
Technological orientation	Pragmatic approach to technology, focusing mainly on best practice, known, commercially available, known technologies.
Trade organization	Modernization Forum
Typical outcome measures	Quantitative: Improvements in productivity, increases in wages, jobs created or saved; Qualitative: Upgrading of technological and strategic competence

**Table 2. Methods used in Empirical Evaluation Studies of Manufacturing Extension Services, 1994-99**

<b>Methods*</b>	<b>Number of studies using method</b>
Customer survey	8
Survey with comparison group	8
Case study	5
Benefit-cost	5
Longitudinal study	3
Simulation model	3
Member survey	1
Center study	4
<b>Total</b>	<b>37</b>

\*Some studies used more than one method.

Source: Analysis of 30 empirical evaluation studies of manufacturing extension services conducted between 1994 and 1999. Updated from Youtie and Shapira (1998).

**Appendix 1. Summary of Manufacturing Extension Impact Studies**

Author/Year	Method	Focus	Main Findings	Comments
MEP (1994)	Center surveys of customer impacts	MEP customers	Benefits per company anticipated by 610 firms responding to MEP center surveys in 1994 included 5.5 jobs added or saved, \$43,000 savings in labor and material costs, and an increase of almost \$370,000 in sales. Benefits exceeded federal costs by 8:1 ratio.	
Swamidass (1994)	Member survey	National Association of Manufacturers members	Only 1% of manufacturers say government is an important source of assistance in technology investment decisions.	Suggests that market penetration of modernization services is low.
GAO (1995)	Survey of MEP manufacturing customers	Nationwide	73% of 389 respondents indicated that their overall business performance had been improved	
Shapira and Youtie, (1995)	Benefit-cost study	Georgia, MEP customers	Combined net public and private economic benefits exceed costs by a ratio of 1.2:1 to 2.7:1	
Luria and Wiarda (1996)	Benchmarking survey, comparison group	Michigan MTC customers; nationwide manufacturers	MEP customers improve faster than comparable firms in a comparison group. However, assisted firms had smaller increases in computer-based technologies.	17 key technology and business performance metrics used; ITI Performance Benchmarking Service dataset
Michigan Manufacturing Technology Center (1996)	Benefit-cost study	Michigan, MTC customers	Combined net public and private economic benefits exceed costs by a ratio of 1.45:1	
Nexus Associates (1996)	Survey of NYMEP customers, comparison group, benefit-cost study	NYMEP customers	NYMEP generated \$30 million to \$110 million of value-added income; 510 to 1920 jobs. Benefit cost ratio of 0.14:1.0 to 0.51:1.0.	Cobb-Douglas Production Function; A priori prediction of high impact oversampling; ITI Performance Benchmarking Service dataset is control group
Oldsman (1996)	Customer Survey, comparison group	New York Industrial Technology Extension Service customers	Total annual cost savings for the 1,300 companies participating in the program between July 1990 and March 1993 is \$30 million. Majority companies said their ability to compete was improved as a result of the program.	The average customer added 5.7% fewer workers than similar, non-participating companies.

## 6 Evaluation of Regionally-Based S&T Programs

Author/Year	Method	Focus	Main Findings	Comments
Shapira and Rephann (1996)	Survey with comparison group, multivariate regression	West Virginia, manufacturing extension customers and non-customers.	Participation in a manufacturing technology assistance program is not yet associated with higher levels of aggregate new technology use, but it is found to associated with adoption of specific technologies and receptivity to new technology investment.	The study's results also confirm the value of training and suggest that a strategy of targeting smaller and medium-sized plants with services focused on multiple clustered locations may be effective in stimulating new technology use among these manufacturers.
Cosmos Corporation, NIST MEP, 1997	Case studies	25 MEP engagements in 13 states	Structured case studies of MEP projects show that program services help smaller manufacturers to modernize their operations, improve quality, and increase profitability through such means as reducing waste, redesigning plant layouts, and improved inventory control and employee training.	
Jarmin (1997)	Longitudinal study, comparison group	Longitudinal Research Databases, 1987-1992, MEP customer data from 8 centers	Manufacturing extension clients had 4-16% higher growth in value-added per worker than non-clients	Standard value-added production function; Controls for self-selection
Kelly (1997)	Case studies of 3 centers	Northern Pennsylvania, Michigan, Minnesota	MEP's focus on one-on-one assistance fails to address problems that limit the diffusion of knowledge and skills in using more advanced technologies.	
Luria (1997)	Performance Benchmarking Service dataset, comparison group	Michigan MTC customers	Customers improved to a greater extent than non-customers in sales growth, employment growth, and adoption of certain process improvements and technologies. However, center customer growth in wage rates, profitability, and labor productivity were not significantly different from that of non-customers.	The author attributes the results to the center's service mix, which attracts companies that are not on a rising productivity path, combined with intense customer price pressures.

## 6 Evaluation of Regionally-Based S&T Programs

<b>Author/Year</b>	<b>Method</b>	<b>Focus</b>	<b>Main Findings</b>	<b>Comments</b>
MEP (1997)	Telephone survey of MEP customers by U.S. Census Bureau	Nationwide, MEP customers	MEP customers report \$110 million increased sales, \$16 million from reduced inventory levels, \$14 million in labor and material savings, 1,576 net jobs created, 1,639 total jobs retained as a direct result of MEP services.	Information provided 9-10 months after project close
Modernization Forum and Nexus Associates (1997), Oldsman (1997)	Survey, comparison group	Manufacturers that used consultants	94% of MEP customers reported improvement in services vs. 77 percent of non customers who worked with consultants	
Shapira and Youtie (1997)	Case studies and analysis of reporting data	6 MEP centers and their partnerships	MEP sponsorship has led to increased service coordination not readily obtained through individual center efforts alone or through demands of state governmental funders. Increased service coordination, in turn, has mostly improved the assistance delivered to firms, though significant expenditure of resources were required to achieve these benefits.	
Welch, Oldsman, Shapira, Youtie, and Lee (1997)	Survey of manufacturing network customers	99 members of 13 separate business networks	The median net benefit of network participation to the firm is \$10,000 (the average was \$224,000)	
Youtie and Shapira, (1997)	Customer survey - longitudinal tracking study	Georgia, MEP customers	68% assisted firms took action, with more than 40% percent reporting reduced costs, 32% improved quality, 28% capital investment	Customers overestimate benefits and underestimate costs close to point of survey, except for small number of high impact projects
Chapman (1998)	Data envelopment analysis of MEP reporting data.	Compares 51 MEP centers using second half of 1996 data.	Centers excel in different areas. (Specifically, MEPs on the frontier in one area may move out of/not be on the frontier in another area).	

## 6 Evaluation of Regionally-Based S&T Programs

Author/Year	Method	Focus	Main Findings	Comments
Ellis (1998)	Surveys of MEP customers	Massachusetts MEP customers	29% MMP customers may not have undertaken changes without MMP assistance. 71% of MMP customers reported some improvement in competitiveness.	
Glasmeyer, Fuellhart, Feller, and Mark (1998)	Survey of 51 manufacturers	Information requirements of plastics industries the Appalachian Regional Commission's counties in Ohio, Pennsylvania, and West Virginia.	Firms most often use traditional information sources because of their credibility and reliability, so MTCs need time to establish a history to demonstrate their effectiveness to firms.	
Jarmin (1998), Jarmin (1999)	Panel, longitudinal study	Longitudinal Research Database, Annual Survey of Manufacturers 1987-1993, MEP customer data from 9 centers	The timing of observed productivity improvements at client plants is consistent with a positive impact of manufacturing extension.	
Kingsley and Klein (1998)	Meta-analysis of 123 case studies	Cases of industrial networks in Europe, North America, and Asia	Network membership can be built with the sponsorship of parent organizations and with public funding, but the networks that generate new business are associated with private sector leadership and at least some private sector funding.	
MEP (1998)	Telephone survey of MEP customers by U.S. Census Bureau	Nationwide, MEP customers	MEP customers report increased sales of nearly \$214 million, \$31 million in inventory savings, \$27 million in labor and material savings, and a \$156 million increase in capital investment as a direct result of MEP services.	Information provided 9-10 months after project close
MEP (1998) (with Nexus Associates)	Simulation model	MEP centers nationally	2/3 of states would end state funding if federal funding were ended; 60-70% of centers would not be able to maintain a focus on affordable, balanced service.	

6 Evaluation of Regionally-Based S&T Programs

Author/Year	Method	Focus	Main Findings	Comments
Oldsman and Heye (1998)	Simulation	Hypothetical metal fabrication firm	Reducing scrap by 2% raises profit margins by 1.2%, but increasing piece price by 2% adds \$200,000 a year.	Authors conclude that manufacturing extension centers should help companies become more distinctive as well as more efficient.
Shapira and Youtie (1998)	Customer survey; project-impact analysis	Georgia, MEP customers	Product development, marketing projects are 60% more likely to lead to sales increases; energy projects are most likely to lead to cost savings; plant layout, environmental projects help companies avoid capital spending. Quality projects do not rate highly anywhere, although they require the largest MEP customer time commitment.	
Shapira and Youtie (1998)	Survey of manufacturers, comparison group	Georgia manufacturers with 10+ employees	The average client plant had a value-added increase of \$366k-\$440k over non-clients	Cobb-Douglas Production function; Controls include use of other public and private sector service providers
Thompson (1998)	Benefit-cost study, simulation	Wisconsin taxpayers	Taxpayer payback ratios of 0.9:1.0 to 3.5:1 from the point of view of the state taxpayer who receives a federal subsidy. However, there is considerable variation in payback ratios by industry and by service type. Increasing sales shows the greatest taxpayer-payback.	
Wilkins (1998)	Center management benchmarking	14 MEP centers	No single measure designates a high or low performing center. Costing rate of \$200-\$400 per hour resulted. Field staff tend to develop more projects than they close. 75% of centers have moved from subsidizing services to generating positive cash flow	

6 Evaluation of Regionally-Based S&T Programs

<b>Author/Year</b>	<b>Method</b>	<b>Focus</b>	<b>Main Findings</b>	<b>Comments</b>
Yin, Merchlinsky, Adams-Kennedy (1998)	Survey and case studies, comparison group	7 pilot centers (receiving \$750,000 over 3 years to establish a manufacturing SBDC) and 7 comparison centers with SBDC relationships but no special funding	Pilot and comparison centers did not differ markedly either in the nature of their partner relationships with SBDC or in the seamlessness of their service delivery.	
Nexus Associates (1999)	Quasi-experimental controlled analysis of customers	SME clients of Industrial Resource Centers (IRCs) in Pennsylvania.	On an annualized basis, IRC clients increased labor productivity by 3.6-5.0 percentage points and output by 1.9-4.0 percentage points more than they would have done without assistance. Productivity gains resulted in an inflation-adjusted \$1.9 billion increase in gross state product between 1988-1997. A benefit-cost analysis finds returns to state investment of 22:1.	Opportunities for even further impacts were identified, through improved targeting and changes in IRC service strategies.

Source: Youtie and Shapira (1998), updated.

## References

- Coburn, C., and D. Berglund (1995). *Partnerships: A compendium of state and federal cooperative technology programs*. Columbus, OH: Batelle Memorial Institute.
- Chapman, R. 1998, July. Using Data Envelopment Analysis to Assess Performance of Manufacturing Extension Centers, NISTIR 6198, Gaithersburg, MD: National Institute of Standards and Technology.
- Ellis, S. (1998). Evaluation of Massachusetts Manufacturing Partnership: Selected Findings. In P. Shapira and J. Youtie, editors. *Manufacturing Modernization: Implications of Evaluation Results for Program Improvement and Policy Development: Proceedings of Fourth Workshop on the Evaluation of Industrial Modernization Programs*, Atlanta, Georgia: Georgia Institute of Technology.
- General Accounting Office (1991). *Technology Transfer: Federal Efforts to Enhance the Competitiveness of Small Manufacturers*, GAO/RCED-92039. Washington, D.C.
- General Accounting Office (1995). *Manufacturing Extension Programs: Manufacturers' Views of Services*, GAO/GGD-950210216BR. Washington, D.C.
- Glasmeier, A., K. Fuellhart, I. Feller, and M. Mark, 1998, May. The Relevance of Firm-Learning Theories to the Design and Evaluation of Manufacturing Modernization Programs. *Economic Development Quarterly*, 12 (2): 107-124.
- Jarmin, R. (1997). Measuring the Impact of Manufacturing Extension. In P. Shapira and J. Youtie, editors. *Manufacturing Modernization: Learning from Evaluation Practices and Results: Proceedings of Third Workshop on the Evaluation of Industrial Modernization Programs*, Atlanta, Georgia: Georgia Institute of Technology.
- Jarmin, R. (1998). Manufacturing Extension and Productivity Dynamics: Preliminary Evidence. In P. Shapira and J. Youtie, editors. *Manufacturing Modernization: Implications of Evaluation Results for Program Improvement and Policy Development: Proceedings of Fourth Workshop on the Evaluation of Industrial Modernization Programs*, Atlanta, Georgia: Georgia Institute of Technology.
- Jarmin, R. 1999. Evaluating the Impact of Manufacturing Extension on Productivity Growth, *Journal of Policy Analysis and Management*. 18 (1), 99-119.
- Kelly, M. (1997). From Mission to Commercial Orientation: Perils and Possibilities for Federal Industrial Technology Policy, *Economic Development Quarterly*, November, Vol. 11, No. 4, pp. 313-328.
- Kingsley, G. and H. Klein (1998). Interfirm Collaboration as a Modernization Strategy: A Survey of Case Studies, *The Journal of Technology Transfer*, Vol. 23, No. 1, Spring, pp.65-74.
- Luria, D. (1997). Toward Lean or Rich? What Performance Benchmarking Tells Us About SME Performance, and Some Implications for Extension Center Services and Mission. In P. Shapira and J. Youtie, editors. *Manufacturing Modernization: Learning from Evaluation Practices and Results:*

## 6 Evaluation of Regionally-Based S&T Programs

- Proceedings of Third Workshop on the Evaluation of Industrial Modernization Programs, Atlanta, Georgia: Georgia Institute of Technology.
- Luria, D. and E. Wiarda (1996). Performance benchmarking and measuring program impacts on customers: lessons from the Midwest Manufacturing Technology Center, *Research Policy*, Vol. 25, No. 2, pp. 233-246.
- Manufacturing Extension Partnership (1994). Making a Difference for America's Manufacturers. Gaithersburg, MD: National Institute of Standards and Technology.
- Manufacturing Extension Partnership (1997). MEP National Data Highlights. Gaithersburg, MD: National Institute of Standards and Technology, December.
- Manufacturing Extension Partnership (1998). Review of Mission and Operations of Regional Centers of the Manufacturing Extension Partnership, Gaithersburg, MD: National Institute of Standards and Technology, February.
- Melkers, J., and S. Cozzens, (1996). Performance measurement in state-science and technology programs. Paper presented at the Annual Meeting of the American Evaluation Association, Atlanta, Georgia.
- Michigan Manufacturing Technology Center (1996). Fifth Year Review. Ann Arbor, MI: Michigan Manufacturing Technology Center.
- Modernization Forum and Nexus Associates (1997). Competition or Collaboration: The Role of Manufacturing Extension Centers in the Private Consulting Market. Working with Consultants: A Tool Kit for Manufacturing Extension Centers, Volume 3. Washington, D.C.: NIST MEP.
- National Research Council (1993). *Learning to Change: Opportunities to Improve the Performance of Smaller Manufacturers*. Commission on Engineering and Technical Systems, Manufacturing Studies Board. Washington, DC: National Academy Press.
- Nexus Associates, Inc. (1996). *Evaluation of the New York Manufacturing Extension Program: Final Report*, prepared for the New York Science and Technology Foundation/ Empire State Development. Belmont, MA, March 18.
- Nexus Associates, Inc. (1999). *The Pennsylvania Industrial Resource Centers: Assessing the Record and Charting the Future*. Report prepared the Ben Franklin/IRC Partnership Board. Belmont, MA, October
- Oldsman, Eric (1996). Does manufacturing extension matter? An evaluation of the Industrial Technology Extension Service in New York. *Research Policy*, Vol. 25, No. 2, pp. 215-232.
- Oldsman, E. 1997, May. Manufacturing extension centers and private consultants: collaboration or competition? *Technovation*. 17 (5), 237-244.
- Shapira, P. and J. Youtie (1995). Assessing GMEA's Economic Impacts: Towards a Benefit-Cost Methodology. Atlanta, Georgia: Georgia Institute of Technology, March (Revised October, 1996).
- Shapira, P. and J. Youtie (1997). Coordinating Manufacturing Extension Services, *The Journal of Technology Transfer*, Vol. 22, No. 1, Spring, pp. 5-10.

## 6 Evaluation of Regionally-Based S&T Programs

- Shapira, P. and J. Youtie (1998). Evaluation Industrial Modernization: Methods, Results, and Insights from the Georgia Manufacturing Extension Alliance" *The Journal of Technology Transfer*, Vol. 23, No. 1, Spring, pp. 17-28.
- Shapira, P. and T. Rephann (1996). The adoption of new technology in West Virginia: Implications for manufacturing modernization policies, *Environment and Planning C: Government and Policy*, 1996, 14, pp. 431-450.
- Shapira, P., (1998). Manufacturing Extension: Performance, Challenges and Policy Issues. In L.M. Branscomb and J. Keller, Eds., *Investing in Innovation*, Cambridge, MA: MIT Press, pp. 250-275.
- Shapira, P., (2001). US Manufacturing Extension Partnerships: Technology Policy Reinvented? Forthcoming, *Research Policy*.
- Swamidass, P. (1994). Technology on the Factory Floor II: Benchmarking Manufacturing Technology Use in the United States, Washington D.C.: The Manufacturing Institute.
- Thompson, C. (1998). Local Politics, National Policy, and the Taxpayer-Payback of Manufacturing Extension, *The Journal of Technology Transfer*, Vol. 23, No. 1, Spring, pp.37-42.
- U.S. Congress, Office of Technology Assessment (1990). *Making Things Better: Competing in Manufacturing*. OTA-ITE-443, Washington, DC: U.S. Government Printing Office.
- U.S. Department of Commerce (1999). FY 2000 Annual Performance Plan (Section on National Institute of Standards and Technology, Office of Technology Administration), Washington, D.C.
- Welch, D., E. Oldsman, P. Shapira, J. Youtie, J. Lee (1997). Net benefits: An assessment of manufacturing business networks and their impacts on member companies. Chapel Hill, NC: USNet and Regional Technology Strategies.
- Wilkins, T. (1998). Benchmarking Manufacturing Extension Centers. In P. Shapira and J. Youtie, editors. *Manufacturing Modernization: Implications of Evaluation Results for Program Improvement and Policy Development: Proceedings of Fourth Workshop on the Evaluation of Industrial Modernization Programs*, Atlanta, Georgia: Georgia Institute of Technology.
- Yin, R., S. Merchlinsky, and K. Adams-Kennedy (1998). Evaluation of MEP-SBDC Partnerships. Bethesda, MD: COSMOS Corporation.
- Youtie J. and P. Shapira (1997). Tracking Customer Progress: A Follow-up Study of Customers of the Georgia Manufacturing Extension Alliance, *Journal of Technology Transfer*, Vol. 22, No. 2, Summer, pp. 43-52.
- Youtie, J. (1997) Toward a cross-case analysis of outcomes of exemplary engagements by MEP centers. In Shapira, P., & Youtie J., eds., *Manufacturing modernization: Learning from evaluation practices and results*. Proceedings of the third workshop on the evaluation of industrial modernization programs. Atlanta, Georgia: School of Public Policy, Georgia Institute of Technology.
- Youtie, J. and P. Shapira (1998). Summary of Manufacturing Extension Impact Studies, Georgia Tech Policy Project on Industrial Modernization. Atlanta, GA: Georgia Institute of Technology.

## Discussion of Philip Shapira's paper

**John Barber:** Is the Manufacturing Extension Program (MEP) a regional program or a national program administered locally? If a regional program, you would expect variation in services and this might serve as a fruitful basis for comparison.

**Phil Shapira:** It is a hybrid. A partnership at the state and local level that is partially funded by the federal government. The federal share went from 50 percent to 30 percent. There is lots of variation in centers and firms also vary by region. This is picked up in some studies where different services have yielded different results.

**Gretchen Jordan:** The GPRA report requires a strategy with verification and validation. Agencies must show the logic and the change toward outcomes over time. But the emphasis is not just on outcomes. Scherer used outcomes information and a logic model to demonstrate outcomes. If regions are successful you can compare to less successful regions and demonstrate causes.

**Phil Shapira:** High performance cases were examined in one NIST study. Very good results were studied by the Cosmos Corporation. The finding was that high performance interacts with strategic change in a company and a high degree of customization to needs. The studies recommended doing fewer and better in depth projects than the current [MEP] arrangement.

**Maryellen Kelley:** Outsiders' recommendations don't necessarily translate into recommendations for implementation. And the effects of certain design elements can affect choices of policymakers. The factors mediating outcomes could be taken seriously by program administrators.

**Phil Shapira:** Some recommendations get taken up while others do not. This is a hybrid partnership between the States and the Federal Government – with a decreasing proportion of federal funding. This complicates the take-up of recommendations and may actually may lead to some States wanting to pull out of the scheme when the federal funding decreases.

## **Evaluation of Regional Innovation Policies**

**Patries Boekholt**

Technopolis Group, Amsterdam, Netherlands

Email: patries.boekholt@technopolis-group.com

---

### **Introduction**

In 1993 the European Commission DG XVI (Regional Policy and Cohesion) launched a pilot initiative called Regional Technology Plans (RTP) which aimed to initiate the development of a Regional Strategy for Research, Technology and Development Policy. The projects in this initiative were to be undertaken in so called 'less favoured regions' which had an Objective 1 and 2 status. The initiative was made possible under Article 10 of the European Regional Development Fund (ERDF), which stated that part of the Structural Funds should be allocated to studies and pilot projects to promote innovation in regional development.

This paper addresses some of the findings that came out of the evaluation study of the RTP Action, carried out by Technopolis in 1997-1998, with the contribution of the University of Athens.<sup>13</sup> In addition the paper discusses the emerging role of science and technology policy in regional development policy.

The evaluation team has looked at what has happened in each of the seven Regional Technology Plans individually as well as the supporting framework of the European Commission. This was the main basis to assess the overall Pilot Action, which includes the role of the Commission in providing the connecting framework for the Action and its management of the pilot Action. According to the original plan the evaluation study would not have had much opportunity to consider the implementation phases of each of these regions. We have allowed for some more time in order to take on board the first results in terms of RTP project launched. However since each of the regions are doing this at a different pace, and in many cases implementation is still an ongoing, we were not able to analyse the post-RTP phase as systematically as the formal RTP phases.

### **The evaluation approach**

The Evaluation team was asked to address three issues in particular:

---

<sup>13</sup> Boekholt, Patries, Erik Arnold (Technopolis) and Lena Tsipouri (University of Athens), 1998, The Evaluation of the Pre-Pilot Actions under Article 10: Innovative Measures Regarding Regional Technology Plans, Technopolis, Brighton.

**The economic and institutional impact**, i.e. the contribution of the RTP projects to regional development efforts in the regions concerned. The key questions were:

- Were the RTPs conducive to the development of a richer regional innovation system?
- Can they help establishing a strategic planning culture?
- Has it helped the public sector's planning capacity in innovation policy, and integrated this in its regional development agenda?
- Did the RTP lead to a strategic Action Plan with real projects?
- Did the RTP contribute to a qualitative and quantitative increase of investment in the field of innovation in the medium and long term?
- Have the project leaders used the BC aid in line with their initial proposals?

**The innovative and demonstration character of the RTP projects**, i.e. did the RTPs contribute to a novel way of policy formulation in the regions? The key questions here were:

- Did the RTPs contribute to a multidisciplinary, demand-led and bottom-up approach to innovation promotion?
- Have the RTPs been an innovative approach in the regions or were similar actions in the making?
- To what extent can the RTP model be generalised for diverse regions?
- What was the benefit of the additional EU support measures and international networking to the participating regions?

**Lessons learnt and recommendations** to the Commission, in the light of improving efficiency in the management, organisation and development of the next generation RIS/RITTS projects, The ERDF Policy Action and understanding the relationships between innovation policy and economic development processes.

The key elements addressed in the case studies were:

- the social process of defining the framework for a regional strategy
- the events and studies that were initiated as part of the RTP project
- the project management in the regions
- the European Commission's role in the pilot action
- the involvement of the business community
- the outcome in terms of discussion platforms, (pilot) projects, programmes and use of Structural Funds.

## *6 Evaluation of Regionally-Based S&T Programs*

The evaluation team visited all seven regions and spoke face-to-face to more than 50 persons directly involved in each of the RTP projects. Particularly the project managers and promoters were a key source of information. In addition we spoke to members of the Steering Committee and representatives of organisations involved in the RTP process. We also held some interviews by telephone. The choice of interviewees represented a mix of people from different communities (government, private sector, intermediaries) in each region. Annex I provides an overview of the persons we have spoken to. In each region the project promoters were very helpful in providing us with written material and helping the team in setting up an interview schedule. Drafts of the case studies were sent to the regional project managers, who gave their comments and additional information.

The European Commission was also very helpful in providing us with the necessary information. The complete files for each of the projects were especially prepared for the evaluation team in order that they were systematically documented and easy accessible. Discussions with the Commission's staff provided more information on the Action Line's objectives, and the programme management.

One of the principal evaluation issues that the Commission had asked the team to address was the RTP's economic impact. From the start of the project the Evaluation team and the Commission agreed that it was too early, if not methodologically extremely difficult, to establish impact on the performance of the firms as a result from the RTP Action in their region. Nevertheless, to have a crude measure of the awareness of firms of the RTP, and their opinions on its usefulness, the evaluation team conducted a telephone survey of 150 firms in three of the seven regions. This is described in more detail in chapter 7 of this report.

### **The RTP pilot Action**

The actual RTP projects started officially in 1994 in seven European regions, and were completed in 1996. The first four regions to join the RTP initiative were:

- Wales
- Limburg (Netherlands),
- Lorraine (France) and
- Leipzig-Halle-DeSsaU (Germany).

A year later three more regions joined:

- Central Macedonia (Greece),
- Castilla Y Leon (Spain) and
- Abruzzo (Italy).

All regions have completed their RTP exercise, with the exception of Abruzzo who

have changed their RTP exercise into a RIS exercise restarting the project in 1996.

**The pilot action was a very novel approach since it intended to integrate a number of complex issues all within the structure of an RTP project:**

- it was at the same time a study and a social change process;
- it aimed to merge policies for regional development with policies for research, technology and innovation;
- it intended to create better linkages between three communities, i.e. policy, business and RTD (Research, Technology and Development) in order to evolve a demand oriented policy strategy;
- it counted on the involvement of SMEs, a target group which is difficult to engage in long term strategy discussions
- it expected results in the short term (pilot projects) as well in the long term (a strategic framework).

The Regional Technology Plan exercises had a duration of approximately 18 to 24 months in which each of the regions would conduct an analysis of the region's innovation capabilities and needs, while at the same time start a process of developing a common strategy for future innovation policy planning. Organisations with a role in Research, Technology and Development (RTD) policy in the selected regions took the responsibility for the project and operated it in a more or less independent manner. A regional project management unit would be responsible for its day-to-day operation. The Commission provided 50% of the funding of these projects, the other 50% was financed from regional or national resources. The European Commission played a 'mentor role' in the background, the regions themselves were responsible for running the RTP projects, provided that they would respect the general philosophy of the RTP model.

What the Commission offered was a policy planning model which included both an indication of the contents and a structure for the RTP policy process. In terms of contents the RTP prescribed a 'demand driven' analysis phase during which the 'real' innovation issues in industry were investigated as a basis for policy action. In terms of process the Commission propagated a 'consensus based' approach, where government agencies were to involve a large group of stakeholders to discuss strengths and weaknesses of regional innovation system, define priorities, and set out (pilot) projects. Many public-private partnerships were established as result of the RTP projects.

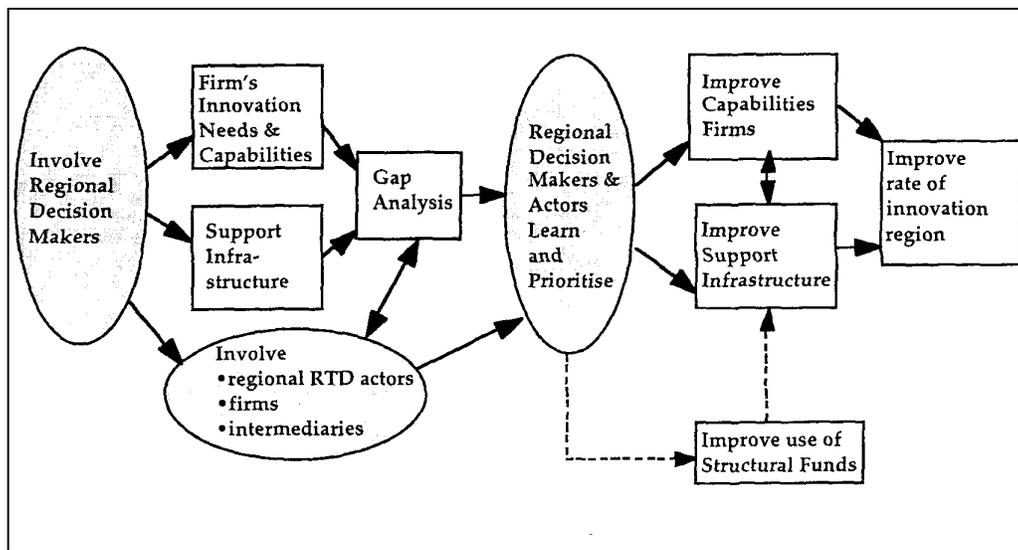
It was intended that the RTP would improve the planning capabilities in the regions, which would in turn lead to better use of the Structural Funds. European Commission officials who set up this initiative had perceived a lack of policy planning culture with many regional governments. Particularly in the area of science and technology, no experience had been developed, since this area had

## 6 Evaluation of Regionally-Based S&T Programs

traditionally been the domain of national policy makers. Particular concern related to the top down approach in regional technology policy initiatives either from centralist national authorities or inexperienced regional authorities. This had led to investments in RTD initiatives that had little or no effect on the performance of the regional economy.

The common structure of RTP projects is illustrated in Figure 1 which describes the whole process from start to implementation to the intended impact. In each of the regions the time and effort needed to complete this process was heavily underestimated, particularly arriving at a consensus based strategy and prioritisation of actions. The evaluation found that qualified project management, preferably by an influential regional actor was a prerequisite for a successful RTP. In those cases where the RTP process (almost) failed, the organisation who held the project management had little support from regional authorities and was not able to influence other actors in committing to the process.

**Figure 1** The logic of RTP projects



Nevertheless, the success of an RTP should not be judged solely on the basis of the outcome in terms of projects and budgets that it managed to generate. The success should also be judged to the extent to which the RTP has triggered new processes in innovation planning, compared to the situation prior to the process. In well organised regions such as Wales, Castilla Y Leon and Limburg with established organisations in the area of RTD policy, reaching a common consensus based strategy was less difficult than in a region such as Leipzig-Halle or Central-Macedonia where responsibilities for RTD policy strategies were not clearly defined.

## *6 Evaluation of Regionally-Based S&T Programs*

The RTP projects came at exactly the right time in terms of a change in attitude towards innovation support policies: the insight that RTD policy should also focus on the competitiveness and innovation capabilities of indigenous companies. This means not only addressing the large (R&D competent) companies or the new technology based firms, but also the SMEs and companies in the more traditional industrial sectors. In most RTPs much effort was made to have direct contact with many SMEs through interviews, company audits, seminars and so on. It must be noted that the European Commission contributed by insisting on this approach and constantly addressing this in their contacts with the regions. The result of taking into account the views of SMEs was a much wider approach to innovation issues, as were perceived by policy makers before.

The issues that came out of the ‘needs analyses’ covered a wide array of problems and opportunities, from technological bottlenecks and information needs, human resources (training and education), finance, management and marketing. The activities that came out of RTPs also covered this wide set of issues.

When RTP was conceived by the European Commission, innovation was still thought of largely as something connected to formal R&D and the traditional producers of knowledge in the state system: universities and research institutes. While there is brief reference to a broader notion of ‘innovation’ the idea that wealth-creating industrial innovation involves a rather wider range of activities than formal R&D has entered the RTP programme during the course of the initiative. Thus, the original RTP planning documents refer to “take-up of advanced technology” as a central issue for industry. The first pilot region Wales started straight away with emphasising that their RTP was about innovation and not just technology. One of the first things they would have changed in hindsight is the name of the RTP Action, giving the wrong impression that technology is at the heart of the initiative. So not only the regions went through this learning process, the European Commission services as well.

The main contribution of the RTP projects has been a renewed focus on innovation in indigenous companies as an important policy support area. For some regional governments innovation was a completely new territory. For other regions it meant a shift away from the high-technology route that many governments had taken before.

### **The RTP Evaluation: findings**

The RTP Evaluation study looked in depth at what happened during and as a result of the seven RTP regions that started in 1994. In one case the RTP had not been successful: in Abruzzo stakeholders could not agree on the objectives and activities that should take place in the RTP. The project never really took off and in the end the Commission did not allow this project to proceed in its original set up. With

## 6 Evaluation of Regionally-Based S&T Programs

new project leaders and organisations involved the project restarted in 1996 and is still ongoing as a Regional Innovation Strategy (P18) project.

In the other six RTP regions the project had

The outputs of the six completed RTPs to date are:

- The definition of **medium to long term innovation policy strategies**, which sets the priority lines and identifies projects which reflect these action lines. Each region has taken a different approach on strategy. A region such as Leipzig-Halle has concentrated on 5-6 concrete Initiatives (e.g. the launch of a Plastics Technology Network), to be implemented in the short term. In Wales on the other hand the choice has been to identify six Priorities Areas (e.g. Creating an Innovation Culture) in a more general way, asking many organisations to put forward proposals.
- The **implementation** of a significant share of the proposed projects, which involves the business community, the knowledge suppliers and intermediaries in the region. The RTP formed the trigger for many new initiatives and projects as well as improving existing initiatives.
- The re-definition of the **use of Structural Funds** in a majority of the RTP regions with a higher share for innovation oriented priority themes and projects, taking account of the priorities set in the RTP
- The continuation of innovation policy discussion **platforms** representing a broad group of regional partners. For example, the Steering Committee of Wales is still in operation and advises policy makers about the main direction of RTD policy. A representative from the business community has taken over the Chairman position and participation is extended to more regional stakeholders during and after the RTP process.

The overall conclusion of the RTP Evaluation is that Regional Technology Plans first of all have had an important **impact on the policy formulation process**, i.e. creating a policy planning culture where innovation and RTD are well embedded in the overall regional development strategies. In particular they:

- contributed significantly to initiating and establishing a strategic planning culture in the participating regions
- put innovation issues more prominent on policy agenda's
  - as policy issue for regional authorities
  - as theme in the Structural Funds
  - through new discussion platforms
- opened up public-public and public-private partnerships; We have seen several examples of (public) organisations and agencies who were competing with each other starting to co-ordinate their innovation support services as a result of the RTP process.

## 6 Evaluation of Regionally-Based S&T Programs

- broadened the scope from technology oriented to innovation oriented; The ‘demand oriented’ approach of the analysis phase brought up adapted policy insights regarding the bottlenecks for innovation, particularly with the smaller companies. It made clearer that innovation is not a matter of implementing new technologies or using research knowledge, but requires a mixed set of capabilities and resources.

Secondly, depending on the region, the **RTPs had a modest to very strong impact on the innovation oriented public expenditures** in projects, programmes and initiatives, of which some were funded with the support of European Structural Funds. In at least four of the seven RTP regions these public funds are supplemented through the leverage of private investments. In the other cases it is too early to assess the output in terms of expenditures.

One of the issues that the Commission asked the evaluation team to assess was the impact of the RTP on the business sector. Although all were well aware that measuring the impact of a policy actions in terms of increase of competitiveness is methodologically impossible, some indicators of an effect of the RTP on the business community could be distinguished.

The RTP project:

- increased the business input in the policy strategy formulation process through:
  - participation of entrepreneurs in Steering Committees
  - direct communications with firms in workshops, seminars, interviews, company audits, sector groups etcetera
  - a better understanding of the bottlenecks as a result of the analyses of the ‘demand’ side.
- a likely impact through the RTP’s output
  - indirectly through the improvement of innovation support services
  - directly through awareness of innovation issues and participation in (pilot) projects

The largest threat to the success of the RTPs was loosing the commitment of the actors that had been made enthusiastic about the change in policy culture and the plans for new initiative. This was due to the fact that the actual implementation of these plans and new working methods did not happen, or took a long period to become operational. Most of the regions did not have funding available for the priority actions and initiatives that came out of the RTP process. Regional policy makers had to convince other policy arenas to allocate Structural Funds to RTP related projects, or the timing for defining the Structural Funds themes and actions was out of tune with the RTP’s results. The absence of implemented pilot projects resulting from the RTP led to disappointment among those regional actors who have been actively involved in the RTP process.

## *6 Evaluation of Regionally-Based S&T Programs*

The Evaluation found that in three of the cases the influence of the RTP on Structural Funds use was limited or modest. Even in those RTP areas where the policy actors involved in the RTP, were able to significantly influence the formulation of the Single Programming Documents and the allocation of budgets, the time lapse between the RTP's Action Plan and the start of new projects in the Structural Funds was very long.

Despite these difficulties the evaluation of the RTP Initiative showed that the general policy planning model, first defined by the Commission in the early 1990s, and adapted over the years, can work in very different settings as long as the regions themselves have sufficient freedom to adapt contents and process to local conditions. The RTP Initiative is now continued under the name Regional Innovations Strategy (RIS) and Regional Innovation and Technology Transfer Strategy (RITTS) managed by DG XVI and DG XIII of the Commission. Currently more than 100 European regions were part of this policy planning process. The success of the approach has strengthened the view within the Commission that a shift is necessary in the use of Structural Funds: from 'bricks and mortar' to actions aimed at improving competitiveness of local industry.

### **An evaluation culture at regional level?**

Although the RTP Action, and its successor the PJS and RITTS projects had an impact on the policy planning processes in many regions, it did not have a great influence on creating an evaluation culture in with the policy authorities involved. Particularly in the later generations RITTS and RIS actions, the design and implementation of a monitor and evaluation system was one of the items that the European Commission had put forward as a crucial element of the policy planning exercise. In practice however this requirement is usually left to the very last of the already long exercise.

There are certain difficulties in evaluating regional innovation policies:

- Regional innovation programmes, particularly those funded with support of the Structural Funds are often fragmented in many small projects and pilots, rather than to a dedicated R&D programme which you would find on the national level. The use of public funds is much more dispersed. Evaluating the effectiveness and efficiency of these programmes is therefore more complex and involves a multitude of actors.
- Regional innovation programmes tend to have a broader range of objectives than R&D programmes and are more closely inter-linked with issues of competitiveness and entrepreneurship. Thus there is a wider set of factors influencing the outcome of innovation programmes, which increases the attribution problem.
- Regional policymakers cannot rely on a good set of indicators and data which

## *6 Evaluation of Regionally-Based S&T Programs*

allow them to benchmark and monitor the performance of their region and its S&T actors. In most countries R&D statistics are not available on the regional level, and at best they are national data sources which have been 'regionalised' after collecting the data.

Despite these difficulties the recent trend, partly inspired by the Commission's Communication Towards the European Research Area is to benchmark regional performance. This provokes similar methodological challenges as benchmarking on national policy level such as comparability of data, transferability of 'best practices' etcetera. This calls for further development of good robust RTD indicators at national and regional levels that manage to go beyond the classical RTD expenditure data.

---

### **Discussion of Patries Boekholt paper**

**Irwin Feller** (Pennsylvania State University, USA). I was involved in an earlier evaluation of the STRIDE project. It seems that you are not going to have regional innovation policy without regions or without regions with political and economic power. R&D planning is often a side show with no connection to the decisionmaker and resources. In the cases you describe, I am struck by the absence of universities in regional development. What about indigenous organizations doing research tied to local/regional needs? The absence of these sources is striking. This seems like a top down plan. With a plan it is assumed to happen. Where is the political and economic power in these endeavors? Where is the change in autonomy that will allow this to happen?

**Patries Boekholt**. Some things have changed since STRIDE. RTP says there is a role for regions in negotiation with national governments. Success depends on how the planning process works. Universities, for example, have been heavily involved.

**Arie Rip** (Twente University, The Netherlands). In the evaluation approach, the key recommendations are contraposed to earlier diagnoses. Why is the implementation missing in this approach?

**Patries Boekholt**. This is the mistake of not involving right actors from the start. Regional decision makers should be part of the project and not come in too late.

**Henry Etzkowitz** (State University of New York, USA). In Europe, some universities are regional organizers, including their offices of technology transfer. The staff [in those offices] play the role of network actor facilitators. The university often plays the role of regional government.

## *6 Evaluation of Regionally-Based S&T Programs*

**Patries Boekholt**, There are examples of countries where regional governments do not exist but the an active role of the university is shockingly low. But universities are becoming more aware of their expanding role. Yet, sometimes they have a different view of their mission than the EU presupposes them to have.

**John Barber** (Department of Trade and Industry, UK). Every region is different e.g. some are science-based some are not. There is a need for a wholistic plan that suits the needs of the region but that takes a 20 year view. Sometimes you have to trace secondary results. If there is a wholistic plan you need to develop criteria for evaluation. And in reality you need a whole series of evaluations because different things have to be evaluated.

**Patries Boekholt**. The active role of the universities is shockingly low and not linked to the economy around them.

**Henry Etzkowitz**. In which direction is it going?

**Patries Boekholt**. The universities are increasingly aware of their role, but there are differences between the countries e.g. universities do have different missions – education, research.

**Henry Etzkowitz**. In Sweden there is actually a third mission – innovation in the region.

## **Science and Technology Infrastructure Investments for Rebuilding Distressed Communities: A Case Study in the United States**

### **Peter D. Blair**

Executive Director

Sigma Xi, The Scientific Research Society

Research Triangle Park, North Carolina, USA

Email: pblair@nas.edu

---

### **Preface**

In the last several decades in the United States, many regions have experienced sweeping change in the structure, productivity, and character of their economies. The U.S. itself is in the midst of a long-term transformation from an economy dominated by manufacturing to one increasingly dominated by services and information. The transition to a new economic structure is happening at different rates and on varying scales across the country. However, many regions that have embraced these changes and aggressively initiated new enterprises to engage in that transformation in an organized way have fared best in terms of economic productivity and employment. Some examples include the great research parks such as the areas known as Silicon Valley in California, the so-called Route 128 corridor around Boston, Massachusetts, or the Research Triangle Park in North Carolina. These areas are typically major co-locations of major research universities, high technology industry facilities, and a supportive government to help facilitate collaboration and cooperation among various organizations by investing in common infrastructure and structuring tax, zoning, and other policies that support such activities. In this paper I chronicle a different kind of collaboration that focuses more on the workforce delivery of education institutions at all levels rather than their research productivity and delivery of primarily researchers and collaborative high-tech ventures.

The venture on which I report is located in Pioneer Valley in western Massachusetts, an area rich in American history, playing a key role in the American Revolution as well as in many developments since. The venture is the Springfield Technology Park and Enterprise Center which has been pivotal in the transformation of the economic revitalization of a deeply economically depressed region of the United States, which until only several years ago had not been able to participate in the remarkable national economic growth the nation had experienced in the 1990s.

### **The Springfield Technology Park and Enterprise Center**

Since the American Revolution, Springfield, Massachusetts and the surrounding Pioneer Valley has been among the U.S. geographic regions experiencing the deepest and most rapid economic transformations in the nation's history. In 1777 George Washington designated Springfield the fledgling democracy's first national arsenal for storing and protecting munitions. The town, overlooking the Connecticut River, was situated ideally for transportation access at the junction of a north-south river and a well-established east-west roadway. Ready access to raw materials and skilled labor added to the attractiveness of Springfield's location for the munitions industry so the manufacture of weapons began quite naturally in 1795. In 1797, the arsenal was designated by Congress as the Springfield Armory, which, for nearly a century, served as the nation's principal national armory for research, design, and arms manufacture.

Following the war of 1812 the design and production processes of weaponry evolved as a model for quality manufacturing. Visitors from around the world came to Springfield to learn about modern manufacturing techniques. The armory became the core of the surrounding region's economy, which when combined with the light manufacturing infrastructure supporting it, came to dominate the region's entire economic base. Throughout the 1800s Springfield was the core of arguably the first and one of the largest industrial centers in the nation. When Harpers Ferry was destroyed in the Civil War, the Springfield Armory became the only federal manufacturing facility of small arms until well into the 20<sup>th</sup> century. As the 20<sup>th</sup> century dawned and as two world wars and the growth of the military ensued, the regional armory buildings for manufacture of its mini- and mainframe computers of the early 1970s.

The new DEC facility assumed somewhat of a mantle of economic dominance in the Pioneer Valley, with dozens of supporting firms appearing in the surrounding region and vying to become suppliers to the new manufacturing activity. Again the economic dominance of the Armory endured. For example, every M16 rifle used in World War II was assembled in the Springfield Armory.

The availability of manufacturing skills and machine tools for munitions gradually spawned the manufacture of other products as well, such as typewriters, bicycles, and sewing machines, but the region remained principally dominated by the arms manufacturing business. As the U.S. military industrial complex grew in the post World War II period to its peak during the Cold War, and as the nature of modern weaponry evolved quickly during that period, the defense industry boomed and the Pioneer Valley enjoyed new growth in aircraft and other defense-related industrial activities.

The armory, which had seeded much of the region's defense-inspired growth, had evolved away from manufacturing, but the facility was not as well suited to the newer mission of research and design. The Pentagon's decision to phase out the Springfield Armory was announced in 1964 and it was eventually closed in 1968. The closure was an ominous early milestone in the beginning of a

dramatic regional economic downturn, as the region fell victim to international competition and successive waves of downsizing and manufacturing plant closures. By the 1980s the region had experienced a 40% decline in employment.<sup>14</sup> Finally, as federal defense spending was cut back substantially in the late 1980s and early 1990s, the Springfield region emerged in the early 1990s as one of the most economically distressed in the nation. Nonetheless, in the Armory facility, which had become the symbol of an economic past, lay the seeds of a dramatic economic rebirth.

### **Seeds of Rebirth**

When the 55-acre Springfield Armory was closed in 1968, the facility was divided gradually into three parcels. The first was a historical compound honoring the unique role of the facility in American history; the compound ultimately became a National Historic Site in 1974. The second parcel became the home for the new Springfield Technical Community College (STCC), the twelfth of the Commonwealth of Massachusetts' growing portfolio of currently fifteen community colleges, and to this date the only technical community college in the state. The third parcel became the site for a major manufacturing facility for Digital Equipment Corporation (DEC), which at the time was the world's second largest computer equipment manufacturer.

The Federal government, through the National Park Service, retained and maintains to this day the 20-acre historical section as a National Historic Site, which includes an Armory Museum. An entrepreneurial team comprised of Springfield Mayor Charles Ryan, local industrialist Joseph Deliso, State Assembly Representative Anthony Sciabelli, and the founding president of STCC, Edmond Garvey, engineered the college's takeover of a over 15 acres of the armory facility and the state of Massachusetts supported the renovation of the new STCC facility. Finally, DEC invested over 20 million dollars in renovating the remaining 15 acres of old region became dominated by and dependent upon a single major economic activity, this time the mini- and mainframe computer industry. As Moore's law<sup>15</sup> was proved time and time again in the computer industry in the mid 1970s and early 1980s, DEC's fortunes fell victim to the emergence of the personal computer and, despite the company's attempts to remain competitive, it began gradually scaling back the mini- and mainframe computer manufacturing activity at their Springfield facility, and eventually closed the facility in 1993.

---

<sup>14</sup> Bosworth, Brian, "Springfield Technical Community College (Massachusetts) Rebuilding a Tradition," Belmont, MA, ND.

<sup>15</sup> In 1965, Intel Corporation Chairman Gordon Moore delivered a speech in which he offered a memorable observation. In plotting the growth in computer memory chip performance, he noted that each new chip contained roughly twice as much capacity as its predecessor, and each chip was released within 18-24 months of the previous chip. With such a trend, he argued, computing power would rise exponentially over relatively brief periods of time, laying the way for personal and microcomputers. This observation became known as "Moore's Law."

## *6 Evaluation of Regionally-Based S&T Programs*

By 1993 the DEC facility was essentially abandoned, although several small spin-off ventures from the DEC presence became residual tenants. The facility overall was put up for sale and even though it was marketed as prime real estate in the heart Springfield, no buyer emerged in the community in the midst of a severe economic downturn precipitated in part by the facility's closure. Hence, as the 1990s opened, Springfield's economy was beset with among the highest unemployment in the nation, ranging between 7.6% and 11.8% between 1990 and 1995, and few prospects for turning the regional economy around. Indeed, as the U.S. economy began to emerge from a recession in 1991, the Springfield region found itself searching for tools to participate in and benefit from the longest and largest period of economic expansion in the nation's history.

The process of devising a set economic transformation tools in the Springfield region came about as a fascinating combination of vision, luck, perseverance, and clever politics, which translated the vision into one of the most remarkable coalitions of industry, academia, and government for engineering a regional economic transformation in the U.S. today. This case study reviews the basic ingredients of this transformation, chronicles how they came together, outlines what the combination produced, and describes the resulting current conditions in the Springfield region.

### **The Springfield Technology Park**

The key to Springfield's economic rebirth in the early 1990s was rooted in a radical economic transformation that would ride the wave of the new economy as the U.S. emerged from the 1991 recession. The transformation began with the vision of a local educational leader and built towards success with a coalition of academic, business, and government leaders in designing the Springfield Technology Park, which emerged as a key new engine of economic growth in the region.

#### ***The Vision: Turning economic disaster into opportunity***

In 1983 Andrew Sciabelli, a faculty member and nephew of one of the STCC coalition of founders, Anthony Sciabelli, became president of the struggling college in the midst of a struggling regional economy. Sciabelli, the younger, perhaps more quickly than most in his field, recognized the value of tuning the community college's curriculum very closely with the needs of local employers, even as the economy was experiencing sweeping change, not just in Springfield, but across the nation. STCC was among the first two-year colleges in the country to adopt economic development as part of its core mission.<sup>16</sup> Sciabelli's vision was translated into an entire philosophy for the college that to this day has made it a compelling example of how education coupled with industry can become a key

## *6 Evaluation of Regionally-Based S&T Programs*

catalyst in regional economic development, in a manner both well-suited to the region's economic assets but quite different from the path taken by the nation's great research parks, such as North Carolina's Research Triangle Park or the Route 128 corridor surrounding Boston.

In 1993, in what turned out to be a key component to Sciabelli's strategy, STCC sought to acquire the now vacated DEC facility and transform it into a technology park for "new economy" firms, particularly firms that could benefit from proximity to STCC. Sciabelli saw the vacant 15-acre DEC facility as an opportunity to leverage traditional college resources (skilled labor, workforce education, and skills upgrading), new programs (entrepreneurship and a highly focused technical curriculum), a geographic concentration of resources (with STCC and the Park), and a growing range of community outreach programs into new opportunities. There were many barriers standing in his way, however, and it would take successes on many fronts and at all levels of government to bring Sciabelli's vision to fruition.

### ***Luck: geography still matters***

As the role of advanced telecommunications networks in the U.S. economy has grown, locations with ready access to these networks have begun to carry a premium for many kinds of new economy firms. Springfield planners, largely through the regional planning activities undertaken by the Pioneer Valley Regional Planning Commission, discovered that the city is located at the nexus of many of the key optical data and telecommunications networks in the Northeast. Hence, a primary attraction of prospective tenants in the "Tech Park" would likely be easy and cheap access to fiber optic networks. Indeed, companies such as the Northeast Optical Network, Brooks Fiber, RCN, and others found Springfield to be an ideal location for network operations and the Tech Park provided a natural home for such operations, particularly if the Park managers were willing to help with the infrastructure investments.

### ***Formulating the plan and making the case***

The entire package of network access, cost effective space, skills training capabilities tuned to the needs of new firms, business amenities such as shared support resources and parking, a supportive community government, as well as other factors comprised a compelling case for developing the Springfield Technology Park. Sciabelli began to formulate the plan for implementing the Tech Park, but there were many other worthy causes as well before the Massachusetts legislature, where Sciabelli turned for help in finding financing for the development. The abandoned DEC facility, now vacant for five years, was available but needed a great deal of retrofit work to make it attractive to new firms.

Sciabelli and his project team made a convincing case of the benefits of the park to the Pioneer Regional Valley Planning Commission (PRVPC) and to all who

## *6 Evaluation of Regionally-Based S&T Programs*

would listen. In October of 1995, following a push from local legislators, the Massachusetts Legislature created the Springfield Technical Community College Assistance Corporation (STCCAC) to oversee the proposed project and eventually to administer the Tech Park. This laid the groundwork for subsequent State funding of the project. A team of eleven business and community leaders was appointed to the corporation's board, chaired by Sciabelli. STCCAC was designed to operate the Park separate from STCC, which is a state entity, but be essentially controlled by STCC. The advantages of this arrangement became more and more important as the project developed. For example, since STCCAC signed the leasing contracts for Park tenants, many innovative features could be incorporated in these leases to help develop necessary infrastructure that would not have been possible, or at least would have been very difficult, under traditional state leasing contracting procedures.

### ***The Pioneer Valley Regional Planning Commission's Plan for Progress***

Since the beginning of the economic downturn that reached its depths in the late 1980s, the PVPC had sensed that this downturn was different from previous ones. It was not just a cyclical or structural recession, it was a fundamental economic transformation stemming from the changing structure of the U.S. economy overall. The commission had been assembling an economic revitalization plan, which it released in 1994 as *The Plan for Progress*. The plan chronicled the fundamental shifts going on in the economy in the early 1990s, differentiating it from cyclical or structural recessions of the past and taking stock of the history of the region and its potential. The plan served as a roadmap for possible development. It included 21 different strategies, covering short-, middle- and long-term needs.

In June of 1996, following, Sciabelli's proposal, the commission endorsed the Tech Park concept, included it prominently in the region's economic plan, and endorsed STCC's acquisition of the former DEC facility. Another key feature of the planning process was the formation of teams of community leaders focusing on each strategy. Once a strategy has been enacted, the PVPC continues re-evaluating and re-engineering the strategies to address changing needs and shifting environments. For example, as the STCC incubator, The STCC Enterprise Center (discussed below), evolved as a key component in the PVPC plan, and as it has come to fruition, the plan has gradually evolved to include establishing a network of incubators throughout the region.

The PVPC plan identified the region's telecommunications infrastructure, i.e., specifically five points of presence for major network carriers, as a key asset early on in the course of its plan. Prior to recognition of it in the PVPC plan, this extraordinary asset had never been even identified, let alone analyzed for its value in spurring economic development.

***Clever politics: engaging all levels of government***

With the Tech Park plan continuing to take shape, in October of 1996, as business and other community leaders continued to endorse and lobby for the project's goals, the momentum led to a \$4.5 million grant from the Commonwealth of Massachusetts to purchase and begin to refurbish the former DEC facility. STCCAC assumed ownership for the property, which was purchased for \$3.8 million, and the balance of the grant was used to begin refurbishing space for new firms.

***The Springfield Enterprise Center***

The inertia of the development of the STCC Technology Park, its growing recognition as a key to future economic growth in the region, and the highly developed working relationship between the community, the STCC, and the Park tenants, laid the important groundwork for a high tech incubator, which was to become the capstone to the Park complex and to Sciabelli's vision of an educational and industry complex focusing on entrepreneurship.

Realizing Sciabelli's vision of an incubator began with a concept paper submitted to the local office of the U.S. Economic Development Administration (EDA). The concept paper outlined the key features of how the Springfield Enterprise Center (SEC) within the Technology Park and STCC complex would be structured, how it could attract telecommunications, data warehousing, call centers, and other new economy operations very competitively, and how the STCC could be an important source of skilled labor, retraining, and skills upgrading activities for resident firms.

The local EDA office was impressed with the features of the SEC concept and recommended it for consideration by the agency's regional office, which requested a full proposal including certification that the regional planning commission considered the project to be the region's highest priority. In March of 1997 STCC received a \$1 million grant from EDA's 302 Program. The next month the City of Springfield endorsed the project with capital funding loan guarantees and the Commonwealth of Massachusetts provided \$500,000 in capital funding. Following a local corporate capital campaign, the SEC facility renovation of one of the Armory buildings was completed in the fall of 1999 and opened the 39,000 square foot incubator for business in January of 2000. The SEC is dedicated exclusively to launching new high tech business initiatives and assisting recent start-up firms to achieve profitability. It currently houses six start-up firms with facilities for up to twenty firms with an office for pro bono business consulting services, network-ready space with fiber optic connections to all major networks, and very favorable rental rates and rent structures for new firms.

SEC director, Fred Andrews, notes that a key to the center's operation is the advisory board composed of local entrepreneurs and professional service providers who act as mentors for the start-ups. The board screens all potential incubator

## *6 Evaluation of Regionally-Based S&T Programs*

residents, reviews business plans, and assists in business development activities. The SEC offers all of the board's activities pro bono, and members are not permitted to be formally involved with companies until they graduate from the incubator. The board also includes representatives from the local Small Business Development Center and the Service Corps of Retired Executives (SCORE), which is housed in the SEC facility. The success of the capital campaign and the initial financing plan has enabled the SEC to operate debt-free from the beginning of its operations.

The SEC also is designed to work closely with and even co-locates with some of the STCC's educational programs such as the college's Entrepreneurial Institute, the Young Entrepreneurial Scholars program for high school students, and a wide range of continuing education programs. The Institute houses a business library and teleconferencing center, facilitates access to local banks and other capital sources, provides a service locating technically-skilled labor, and makes available a host of other services accessible by SEC tenants.

### **The Technology Park Today**

The evolution of the Springfield Technology Park is a remarkable success story and remains a key development feature of the Pioneer Valley region. The Park itself is now fully leased, housing a broad spectrum of firms (see Table 1), all of which fit into the pattern of taking advantage of the shared resources of the park, the STCC workforce training and education products, and the telecommunications infrastructure. Indeed, the Park is commonly referred to as a "telco hotel." The STCC has also established a routine process for adjusting its constantly changing curriculum to needs of park tenants, which, of course, keeps its course offerings very current and ensures that its graduates well positioned for employment in the park. Indeed, many STCC education and training programs are fully integrated into park tenant operations. Over 150 STCC graduates already work in the firms housed in the park, which comprise nearly a third of RCN's payroll and 15% of that of other companies.

What has emerged as a very important management feature of the park is that the facility is not managed directly by STCC, which is a state organization. Rather, as noted earlier, it is run by the STCCAC, which is independent but essentially controlled by the STCC. This organization permits much more flexibility in carrying out key aspects of running the park, especially negotiating leases, contracts, and other agreements, such as for expanding the park infrastructure. Hence, as the park has evolved, the ability to expand the features necessary to attract new tenants to the park have been much easier to acquire than if they were bound by, for example, state procurement procedures. Such features include high quality and load electric power supply (now including 8 back-up generators), heavy load bearing floors and tall ceilings to accommodate extensive conduit and equipment. The park also houses 6 telephone switches and 7 levels of fiber optics throughout the facility.

## 6 Evaluation of Regionally-Based S&T Programs

To date the resident tenants of the park have invested over \$225 million in build-outs and capital equipment. STCC uses creative leasing structures to finance capital build outs to avoid carrying debt, which is considered complicated or even precluded by government ownership. Hence, while the park is operating in the black, due to initial capital outlays being financed through the leases, it will take several years to develop the park’s full revenue potential.

According to the Park’s chief operating officer, Tom Holland, the park has a “three legged” plan for future operations: (1) To provide a major new revenue stream for the college; (2) to house key extension programs for the college, such as the Bell Atlantic sponsored fiber optics training program, and, (3) To locate a high-tech incubator to build the culture of entrepreneurship and other key community outreach features of the college. This last feature of the STCC and Technology Park complex is the key to leveraging the combination of assets to generate new value and new activity in the park and as opposed to being dependent solely upon firms relocating to the Springfield area.

### Integrated Educational Programs

The second part of Holland’s “three legged” plan, i.e., key educational extension programs, is a particularly important new innovative feature of the Tech Park complex. Sciabelli refers to the concept as an “enterprising college.” Its focus on tuning the educational activities to the needs of new businesses has many benefits and has paid off in many ways beyond making it attractive for students to come to the college and for graduates to find jobs. For example, STCC is partnering with the National Science Foundation to draft a curriculum for teaching telecommunications technologies across the U.S. STCC was also selected by Microsoft to train other colleges in how to teach information technologies.

Table 1
Springfield Technology Park Tenants*
1. <b>Brooks Fiber, a Division of MCI/Worldcom</b> (a telephone switching company)
2. <b>Choice One Communications</b> (provider of telecommunication services)
3. <b>CTC Communications Corp/fc.com</b> (a wide-area telecommunications provider and a provider of integrated telecommunications services)
4. <b>Equal Access Networks</b> (telecommunications infrastructure)
5. <b>Excitation L.L.C.</b> (manufacturer of lasers for medical applications)
6. <b>FutureWorks</b> (a one-stop career center serving jobseekers and employers in Hampden County)
7. <b>GlobalNaps, Inc.</b> (provider of telecommunications services)

6 Evaluation of Regionally-Based S&T Programs

8. <b>R. J. Greeley Company, LLC</b> (a full-service commercial and industrial real estate company)
9. <b>Group Four Transducers</b> (a manufacturer/distributor of force transducers for the weighing industry)
10. <b>Inlight Interactive</b> (interactive CDROM and health educational developer)
11. <b>NorthEast Optic Network, Inc. (NEON)</b> (facilities-based provider of technologically advanced, high-bandwidth, fiber optic transmission capacity for communications carriers on local loop, inter-city and interstate facilities)
12. <b>RCN (Residential Communications Network)</b> (an internet service provider)
13. <b>Springboard Technology</b> (computer repair company created by former Digital Equipment Corporation employees)
14. <b>Williams Communications</b> (a wide area telecommunications provider and a provider of integrated telecommunications services)
*STCC Technology Park, <a href="http://www.stcc.mass.edu/techpark/">http://www.stcc.mass.edu/techpark/</a> , March 31, 2000.

Following its successes in this area, STCC created the Northeast Center for Telecommunications Technology, which is a partnership aimed at developing education programs in telecommunications as well as textbooks and educational CD-ROMs. NSF awarded the Center a \$5 million grant to become the only NSF Center of Excellence located at a community college. Local business in the Tech Park and beyond are finding the center an invaluable resource. For example, Bell Atlantic has commissioned STCC to provide a substantial workforce training program in fiber optics in an \$8 million “next step” program. Indeed, businesses such as Cisco Systems, Nortel Networks, IBM, MCI WorldCom, Time Warner Corporation, Microsoft and many others have become “educational partners” in STCC’s educational and extension activities.

Many in the region and beyond consider the STCC complex to be almost revolutionary in its scope. For example, in an extremely unusual move the regional business journal for Western Massachusetts, *Business West* magazine, named Sciabelli “Top Entrepreneur for 1999.” The award recognized his pivotal role in revitalizing the greater Springfield business outlook. Former STCC Board of Trustees Chair Brian Corridan notes, “It’s easy to be an entrepreneur when times are good. It’s a lot harder when times aren’t so good, but even more necessary. Watching him through those difficult years is as indicative as watching what I call the ‘visionary Andy’ of today.”<sup>17</sup>

Business leaders have fully endorsed the Park concept. For example Telitcom President Geoffrey Little noted that the importance of the Park’s location at the crossroads of the nation’s telecommunications fiber optic networks “cannot be overstated.” He considers it a formidable resource leading to progressive

<sup>17</sup> “New School of Thought”, *Business West*, Vol. 16, No. 9, January 2000, p 7.

development in the park with call centers, new telecommunications carriers, and application service providers (ASPs).

## **Conclusions**

STCC is in many ways has provided the impetus for a new experiment in regional economic development. Indeed, even today, STCC remains the only community college in the U.S. to have established a technology park. The integration of a tech park, designed to capitalize on locational advantages for the telecommunications business, into the college's academic programs and the subsequent expansion of the park to include a high-tech incubator targeted in this same business segment has proved so far to be an overwhelming success. It remains to be seen how much new employment will be generated directly as a result of this strategy, but the impact on the economic vitality of the region has been substantial and the prospects continued growth seem promising.

In retrospect, the discovery of the locational advantage of Springfield for the telecommunications business (at the hub of key fiber optic networks) was crucial to attracting new business to the tech park. The addition of the tech park's strategy for accommodating facility "build outs" very flexibly for tenants, for providing many other kinds of infrastructure support (power, telephone switches, business services, etc.) and for coordinating with STCC to provide a very tailored source of new labor and expertise has been crucial to the sustainability of the park. The addition of the SEC shows great promise for making Springfield a center for telecommunications business innovation and entrepreneurship (as well as perhaps other business areas) that will have lasting impacts on the region's prospects for sustained economic growth in the fast changing U.S. economy.

Finally, the STCC Technology Park and Enterprise Center seems to provide an alternative model to a research park for using science and technology infrastructure to spur economic development – one supporting technical innovation but more focused on business innovation and provision of an infrastructure for promoting business development and tighter integration with academic programs not necessarily focused on research.

## ***Evaluation of Key Study Hypotheses for the STCC Case***

At the outset of the study of distressed communities, of which this case study is a part, the project team settled on ten major hypotheses that were to be evaluated for each case study. The STCC case confirms all these hypotheses and perhaps underscores the particular importance of vision and leadership as critical success factors (hypothesis 10). In the following, brief discussions of each of the ten hypotheses are included in the context of the STCC case.

1. *Enabling infrastructure (roads, water, sewer, etc.) is necessary but not sufficient to assure the effectiveness of technology infrastructure investment.*

## 6 Evaluation of Regionally-Based S&T Programs

The historical key infrastructure advantages of the Springfield location that led to the dominant role of the arsenal in the region's economic history were not sufficient in the new economy of the 1990s. New infrastructure, however, turned out to be crucial in the STCC case. The nexus of key fiber optic networks in the Northeast was crucial to attracting the early tenants to both the Technology Park and the Enterprise Center.

2. *Investments in technology infrastructure must be part of a larger local planning process to succeed.*

The Pioneer Valley Planning Commission process was crucial to the political success of the technology park concept at all levels of government: to receive the city and county governments' endorsements, to attract the initial seed funding for the park as well as support of the innovative management structure from the State legislature, and to establish credibility for EDA and NSF support for key programs housed in the park and for building the Enterprise Center.

3. *A certain density or critical mass is required for information and communications technology (ICT) to work, because of the costs of provision.*

The marginal additions to the existing nexus of fiber optic networks to accommodate various telecommunications tenants in the park were crucial and sizable relative to the size of the overall investment in the park, but the innovative management program allowed such investments to become part of the leasing arrangements for tenants. As noted in the case over \$225 million in build out investments have been made by the first round of tenants in the tech park.

4. *Technological innovations and advances are reducing the costs of communications and linkages.*

While this aspect was not explored in detail in this case, some evidence suggests that this is true in the STCC case. For example, advances in optical switching and other networking equipment is making new additions in the park less costly and with higher capacity. In the end, the innovations in these areas may actually be found to undermine the comparative advantage that Springfield has right now in the telecommunications business because as the network expands and the bandwidth grows across the network, the locational advantage of Springfield may become less important.

5. *Institutional mechanisms must be in place to help users deploy the hard infrastructure appropriately.*

The coordination among levels of government, and in particular the support of local agencies proved to be very important in the development of the technology park. For example, the Park managers' early efforts to work closely with local building inspectors established a culture of mutual trust that has proved invaluable

## *6 Evaluation of Regionally-Based S&T Programs*

in the major phase of growth in the park and major build out of infrastructure was required. Early awareness of special requirements related to the local implementation of fire code features and ADA compliance was especially important.

5. *Some hierarchical coordination or linkages among national, state and local governments are important for a community to be truly “connected.”*

As noted above, the Pioneer Valley Regional Planning Commission process was important to coordinating all levels of government in the early stages of the project. The informal political linkages, however, were just as important. The relationship of local state legislators with the key state legislative committees depended upon the support of local political officials. The local EDA office’s support was crucial to the ultimate regional office’s endorsement of funding proposals. Since the Springfield Arsenal is also a National Historic Site, other federal and state agencies involved in regulating historic preservation are involved in decisions about the facility as well. Indeed, the array of national, state, and local entities involved in the Technology Park is quite large and their interactions complex. Aggressive coordination of contacts among all these entities was crucial to the enviable pace of progress in the conception, construction, and operation of the Park.

6. *Linkages among government, education and industry are a critical success factor for ICT interventions. State policies and corporate actions are important determinants of local “success”.*

The case study notes the crucial role of organizations such as STCCAC, the PVPC’s key strategy working groups, SCORE, and the SEC Advisory Board were and are key to the Park’s activities. They provide important convening opportunities for collections of government, education, and business stakeholders. The state policies that permit and regulate the activities of such groups are important benchmarks for these activities. More importantly, however, the park originators view the combination of attitudes and commitment that these leaders bring to bear in participating in these groups as a key strength of their approach.

7. *Infrastructure providers must be a partner in assessment and planning, not just the spending, for the intervention to succeed.*

As noted above, build-out for new activities in the park are accomplished via a cooperative agreement executed through the lease agreement. Coordination of infrastructure improvements to make the park a competitive place to locate has proved to be important in getting the park fully leased. Just as important, however, is the general culture of encouraging and promoting aggressive infrastructure improvements, which is likely to be important for the long-term sustainability of the park.

## 6 Evaluation of Regionally-Based S&T Programs

8. *Workforce development is a critical complement to any infrastructure intervention.*

Workforce development was, indeed, crucial to the original design of the park. However, the nature of that development has been quite different from other case studies. Being the nation's only technology park connected to a community college rather than a major research university, the focus is on integrating college extension and other programs into the park's activities (especially the SEC), and on tuning the educational programs of the college very closely with the needs of the park tenants.

9. *Because of the long-term and expensive nature of ICT, vision and leadership are critical success factors.*

As noted above, this hypothesis is particularly underscored by the STCC case. STCC President Sciabelli, in particular, but many others as well made this happen against fairly substantial odds including doubt at all levels of government, skepticism on the part of regional business leaders, and when the educational mission of the college was struggling.

---

### Discussion of Peter Blair's paper

**John Barber** (Department of Trade and Industry, UK). We cannot use evaluative information in the policy process just through evaluation itself. The general public understanding of science and technology needs to increase. The PUSTAT database [a public understanding of S&T database] in the UK shows that children get excited about science and technology until they are teenagers and then they are lost. An increasing understanding of S&T is needed and this has to be realised in different dimensions: understanding about what scientists do as well as understanding about how policy is made. Policymakers need to help improve general understanding .

**Philip Shapira** (Georgia Institute of Technology). Yes, but shouldn't forget that elected officials are not so simplistic as often suggested. Many are sophisticated. They have street smarts and don't believe models that don't square with their experiential knowledge. Perhaps this is their own way of questioning the assumptions of those models. Also, it is important to remember that there are other audiences for evaluation besides elected officials, sponsors, program managers, and participants.

**Peter Blair.** I focused on legislators because they did not come up in the discussion thus far. It is important to think about how to tailor an evaluation to the audience or multiple audiences nonetheless.

## *6 Evaluation of Regionally-Based S&T Programs*

**Maryellen Kelley** (National Institute of Standards and Technology, USA).

Sometimes in programs like ATP, where program directors really think you have to target the evaluation to legislators, you still need to have a story—a convincing story based on something concrete. There must be a consistent effort to communicate even if you cannot do much about someone's ideology.

**David Guston** (Rutgers University, USA). Peter Blair emphasized the tie of the legislator to his/her constituents. They look for analysis that allows them to make arguments to their constituents.

**Stefan Kuhlmann** (Fraunhofer ISI, Germany). The way to communicate evaluation depends on the political system. E.g. we communicate the results of the evaluation with bureaucrats. Policy makers are often bureaucrats. And the elected officials don't have too much power over the top bureaucrats in the EU.

**Peter Blair.** But the elected politicians must discuss the results.

**Stefan Kuhlmann.** Politicians only discuss the general distribution of money.

**John Barber.** That's not true for the UK.

**Henry Etzkowitz** (State University of New York, USA). A two-step flow of influence in communication comes from respect. It is more influential. Maybe evaluation results need to come through a constituency. If you narrow the evaluation to too small a base, you don't have a constituency. Companies that graduate phase one of SBIR often turn to phase 2 and then later perhaps apply for an ATP award at the \$2 million level. The US program landscape is not integrated from above like the EU structure but from below by the actions of firms.

## 7. Evaluation in Multi-Goal Partnerships

Session Chair: Carsten Dreher (Fraunhofer ISI, Germany)

---

### Recognizing the Competing Values in Science and Technology Organizations: Implications for Evaluation

Gretchen B. Jordan<sup>†</sup> and L.D. Streit<sup>‡</sup>

<sup>†</sup>Sandia National Laboratories, Albuquerque, New Mexico and Livermore, California, USA; <sup>‡</sup>McNeil Technologies, Springfield, Virginia, USA  
Email: gbjorda@sandia.gov

---

#### Introduction

Since 1996, the U.S. Department of Energy's (DOE) Office of Basic Energy Sciences has sponsored a study to better understand the elements of research environments that contribute to scientists' and engineers' ability to perform excellent research and to develop new techniques to assess the Science and Technology (S&T) organization's effectiveness in providing these elements. This DOE study would also like to clarify what is the "best practice" in managing research and how to meaningfully compare the effectiveness of different organizations, given the different types of research they conduct and their differing circumstances.

In order to accomplish these goals, the authors found it necessary to propose a new way to describe and assess the effectiveness of S&T organizations, building on a concept called the Competing Values Framework developed by Rohrbaugh, Cameron and Quinn (1983, 1988, 1999). The authors first describe current motivations for assessing organizational effectiveness and existing models for doing so. Then they present the proposed framework that includes the Competing Values perspective on organizational culture and effectiveness within the context of the organization: its resources, products and purpose, and external environment. After describing the model of organizational effectiveness developed by Rohrbaugh, Cameron and Quinn, the authors adapt this model for S&T organizations. Attributes of the S&T organizational culture are defined, using as guidance work with more than 1500 scientists and engineers and an extensive literature review on what influences innovation, productivity, and "best practice." Finally, implications that the proposed framework has for evaluation planning, utilization, and data collection and for defining a balanced set of leading indicators are outlined.

## **Motivation for the Research**

### ***Desire to Improve and Report***

There is increased interest in assessment of the effectiveness of S&T organizations both because people want to know how to improve these organizations and because they need to report current effectiveness to senior managers and investors. For some, this is a matter of improving a situation that is already excellent or a result of the Quality movement (Endres, 1997; McLaughlin, 1995; Miller, 1995). Within the large U.S. public research laboratories, however, there is concern that the environment for research has been deteriorating. This worry has been voiced both by scientists within these organizations, as well as by the leaders whose responsibility it is to steward the nation's research infrastructure (U.S. House of Representatives, 1994).

Public concern that governments improve performance, clarify responsibilities and control, and realize cost savings has also increased the need to demonstrate the effectiveness of research organizations. This need is not unique to the United States. An Organization For Economic Cooperation and Development (OECD) study (1997) states that these concerns are noticeable in all OECD performance management frameworks, although to different degrees.

The call for more performance management has science managers concerned. Many feel the push toward measuring milestones that can be quantified on an annual basis creates an undesirable bias toward doing short term, less risky research. A recent study by the National Academies of Sciences Committee on Science, Engineering, and Public Policy (1999) concluded that while milestones are appropriate for applied research, agencies conducting basic research must set their expectations based on research quality, attainment of leadership in their field, and the development of human resources, and that these are three good indicators that the organization will provide long-term, valuable returns on the public or private research investment. Branscomb, in his 1999 article "The False Dichotomy: Scientific Creativity and Utility," suggests that a new conceptual model for public science is needed so that managers of public science can articulate the goals of public investment in research and measure its progress in a way that meet Congressional needs. Branscomb suggests that the new model should address both the motives for spending public money on research and the environment in which to perform the work.

### ***The Desire to Look at Differences in Goals, Organizational Design, and Circumstances***

Branscomb's call for a new model for discussing public science is echoed by many who seek ways to assess S&T organizations that recognize differences in goals, management environments, their circumstances, and expected outcomes. A 1993 report by the Federal Government of Canada suggests that methods for assessing

the socioeconomic impacts of government S&T will differ depending on the type of S&T (basic, applied and technology development), the purpose of the S&T being assessed, and whether or not the impact has already occurred. Organizational design also influences organizational outcomes. For example, Hull (1988) argues that efficient research performance is partly a function of the match between organizational design and the type of work performed. His research showed that organic, non-hierarchical systems are best for dynamic contexts, small organizations, and complex products.

Furthermore, there often appears to be a need to balance competing demands in managing S&T organizations. Udvardia (1990) suggests that “technological organizations need to engender environments that provide a delicate balance between giving the creative mind freedom to conduct its work while maintaining external constraints like goal setting and time-tables which are essential for the conduct of profitable business.” In other cases, a change in focus and culture is needed. The U.S. General Accounting Office (1996) suggests that one of the leadership practices that will reinforce results-oriented management is to redirect organizational culture from the traditional focus on inputs and activities to a new focus on defining missions and achieving results. Finally, there are differences in the resource bases of organizations and in their current and future circumstances that may influence organizational decisions, design and performance. A 1996 National Science and Technology Council Report suggested that “science agencies must devise assessment strategies ... designed to ... respond to surprises, pursue detours, and revise program agendas in response to new scientific information and technical opportunities essential to the future well-being of our people.”

### **Current Frameworks for Assessing the Effectiveness of S&T organizations**

Several authors have proposed frameworks for assessing S&T organizations that respond to the many requirements articulated above. In describing his approach, Szakonyi (1994) states that while “improving the effectiveness of R&D is the most important issue in R&D management ... there are still no methods that are widely accepted for measuring R&D effectiveness.” He suggests that there are major flaws in the last 30 years of effort, including lack of objectivity, credibility, and frame of reference. The audit model proposed by Chiesa, Coughlan, and Voss (1996) measures performance in seven areas of innovation and allows organizations not only to identify their strengths and weaknesses but also to determine methods of improving innovation processes and capacity. Kanter (1988) argues that recognizing the conditions that stimulate innovation first require understanding the factors involved in the innovation development process. Hurley (1997) concludes that even if an organization hired the right scientists, the highest level of discovery would occur only if they were put into a discovery-oriented environment. Crow and Bozeman (1998) suggest that their Environmental Context Taxonomy identifies laboratories of similar character and behavior, which then allows for a more accurate assessment of laboratory performance based on specific needs and goals of each S&T laboratory.

## *7 Evaluation in Multi-Goal Partnerships*

These models as well as Udvardia's (1990) Multiple Perspective Model and the Industrial Research Institute's Technology Value Pyramid are discussed briefly below before the authors propose a framework that builds on these and encompass four perspectives on effectiveness.

In Szakonyi's model (1994), how well an organization performs each of ten activities, such as selecting S&T, planning and managing projects, maintaining quality S&T processes, motivating technical people, coordinating S&T and marketing, is rated on a logical, objective six-point scale that ranges from "issue is not recognized" to "continuous improvement is underway." Benchmarking against the "average" department allows organizations to determine how well they are performing and guides their improvement.

Chiesa et. al. (1996) look at five dimensions of performance: resource availability and allocation, understanding competitor's strategies, understanding the technological environment, structural and cultural context, and strategic management to deal with entrepreneurial behavior. Central to their audit model are the interacting core processes of concept generation, product development, product innovation, and technology acquisition. These are fed by enabling processes of leadership, resources, and systems and tools.

Kanter's model (1988) looks at individual researchers, organizational structure, and the social and legal environment and suggests that it is most likely that innovations will develop in environments with "flexibility, quick action and intensive care, coalition formation, and connectedness." Some conditions are more important than others at different points in the innovation/development process.

Hurley's model (1997) suggests that an organization maximizes scientific discovery dependent upon individual characteristics such as scientific knowledge, personality characteristics, and organizational characteristics that include both resources and dynamics. Resources that foster discovery include money, equipment, libraries, competent technicians, and rewards. Organizational dynamics address psychological factors such as organizational stability, intellectual freedom, and a climate of enthusiasm, dedication, and encouragement.

Udvardia (1990) highlights creativity as the most critical element for the effective management of innovation. He presents his Multiple Perspective Model which includes three perspectives: the individual characteristics associated with creativity, the needed technical resources (material as well as human), and the organizational practices and managerial actions that aid or stifle creativity.

The Industrial Research Institute has developed the Technology Value Pyramid (TVP), a group of 50 metrics used to assess and predict S&T performance. As described by Tipping, Zeffren, and Fusfeld (1995), two of the five managerial factors that describe the innovation capability of the firm are the "Practice of R&D Processes to Support Innovation" which includes management practices, idea genera-

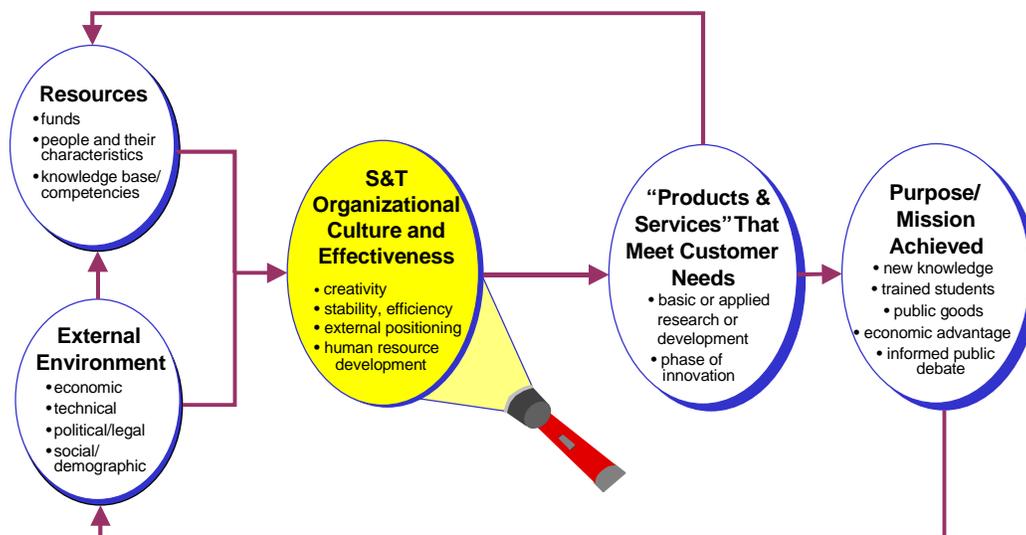
tion, and communication and the “Asset Value of Technology” which includes technology know how, people and proprietary assets.

Crow and Bozeman (1998) propose their Environmental Context Taxonomy classification system because the traditional way of classifying laboratories along the lines of university, industrial, and government overlooks the diversity in technical abilities, organization, economics, politics, and flexibility that S&T laboratories exhibit. They suggest that the level of government and economic influence that affects a laboratory has repercussions on the performance and capability of the organization, and is associated with research type and focus, resource quality, and laboratory structure.

### A New Framework for Assessing Effectiveness of S&T Organizations

All of the models described above link innovation, creativity or excellent S&T to organizational structure, culture, characteristics, or activities and processes. The models of Kanter (1988), Chiesa, et. al. (1996), Udwardia (1990), and Crow and Bozeman (1998) also include relationships with one or more aspects of the external environment. All the models, except for Szakonyi’s (1994) and Kanter (1988), explicitly include organizational resources. In Figure 1, the authors present a comprehensive framework that builds on the elements in these models to provide a complete and logical picture of the S&T organization, its “products,” and its internal and external circumstances.

**Figure 1. A Framework for Assessing Effectiveness of S&T Organizations**



## 7 Evaluation in Multi-Goal Partnerships

The important aspects of the proposed comprehensive framework are

- Organizational culture and effectiveness,
- The primary purpose or mission of the organization,
- The type of research or “product” of the organization,
- The level and quality of resources available to those doing the science, and
- The external environment (technology, political/regulatory, social/demographic, and economic).

The authors’ primary focus is the study of organizational culture and effectiveness (spotlighted in Figure 1) which they believe must be assessed within the larger context in which the organization operates. This context includes:

- *The primary purpose/mission of the organization*, here defined using categories developed by Laredo and Mustar (2000): new knowledge, trained students, public goods, economic advantage, and informed public debate.
- *The type of “products and services” of the S&T organization*, in this case the stage or type of research conducted. This component is meant to capture the results of the S&T activities as well as its uses. Although some would define the types of S&T more carefully to distinguish use-directed basic research from curiosity-driven research, for example, most define the types of S&T as basic, applied, and development. Kanter (1988) suggests it is important to distinguish the phases of innovation (idea generation, coalition building, idea realization, and transfer.)
- *The external environment’s influences* on the organization are typically broken into four categories: technical, economic, social and cultural, and political/regulatory. Each of these may be of varying influence and that influence may be stable or dynamic, and favorable or unfavorable.
- *The resources* currently available to an S&T organization include funding, staff and their characteristics, facilities and equipment, its knowledge base and core competencies. The characteristics of these resources define, in part, what an organization is capable of doing.

Within this framework, an effective organization is one whose culture, structure and management is optimal to turn its resources into outputs and accomplish the purpose of the organization’s effort, given the external environment in which it operates. The framework embraces a broad definition of effectiveness that includes creativity, productivity, efficiency, and employee morale and development. This is

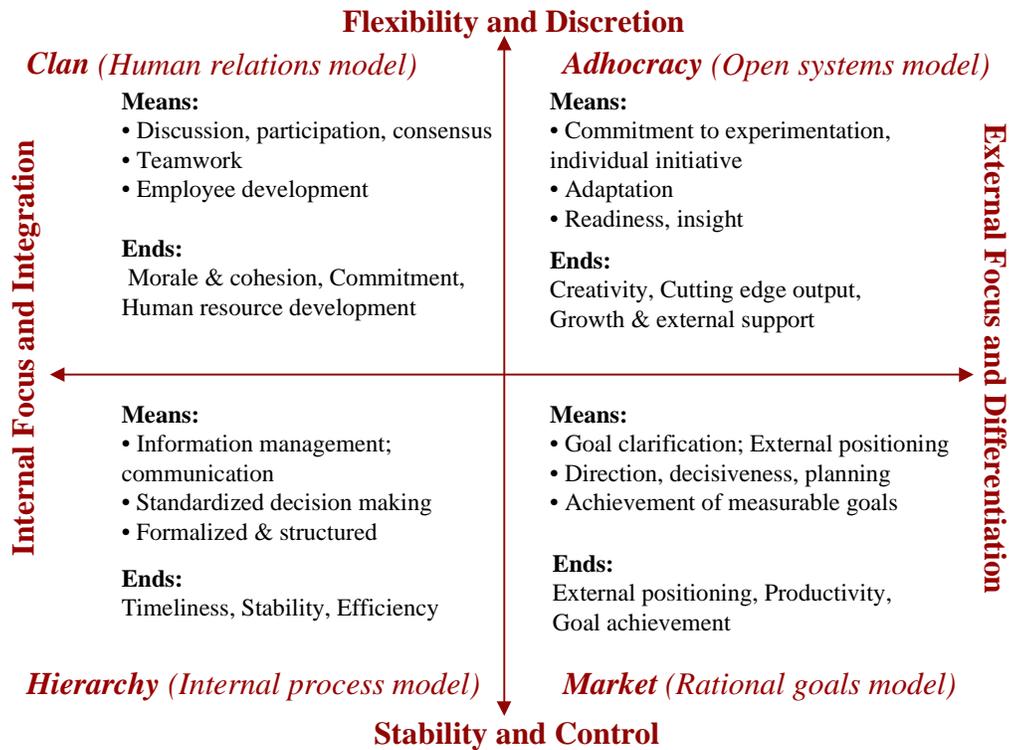
based on the Competing Values theory and is described in more detail in the next section.

### **The Competing Values theory of Organizational Effectiveness**

The DOE research environment study has chosen the Competing Values theory of Rohrbaugh, Cameron, and Quinn (1983, 1988, 1999) to describe the organizational culture and effectiveness of research organizations. The term “culture” is used broadly here to encompass all of the values, structure, focus, and behaviors of the organization. The authors strongly agree with the Cameron and Quinn (1999) argument that follows. Since organizational culture is a complex, interrelated, comprehensive, and ambiguous set of factors, it is impossible to include every relevant factor in diagnosing and assessing organizational culture. Therefore it is important to use an underlying framework, or theoretical foundation, that can narrow and focus the search for key cultural dimensions. The most appropriate framework should be based on empirical evidence, should capture accurately the reality being described (*i.e.*, be valid), and should be able to integrate and organize most of the dimensions that stakeholders value. It should also be congruent with well-known and well-accepted categorical schemes that organize the way people think, their values and assumptions about what makes a good organization, and the ways they process information. Cameron and Quinn (1999) have found that the Competing Values theory meets all of these requirements.

The Competing Values theory suggests three “value dimensions” that underlie conceptualizations of organizational effectiveness and therefore it can be used to organize the traditional and often conflicting models of effectiveness. The three value dimensions identified by Rohrbaugh, Quinn, and Cameron (1983, 1988, 1999) are *organizational structure*, *organizational focus* and the *means-ends continuum*. The dimension of *organizational structure* distinguishes between those values and activities that emphasize the organization’s flexibility and adaptability and those that stress control and stability. The dimension of *organizational focus* contrasts an emphasis on “internal and integrating issues,” such as the well being and employee development, with “external and differentiating issues,” such as the development and growth of the organization itself or its relations with entities outside itself. The *means-ends continuum* reflects emphasis on the objectives of the organization, such as productivity or human resource development, and the means by which it achieves these objectives, such as goal setting or enhancing morale. As illustrated in Figure 2, Rohrbaugh, Cameron, and Quinn (1983, 1988, 1999) use these value dimensions as axes with which to organize four of the most common theoretical models of organizational effectiveness: the *human relations model*, the *open systems model*, the *rational goal model*, and the *internal processes model*.

**Figure 2. The Rohrbaugh, Cameron, and Quinn Competing Values Theory of Organizational Effectiveness**



Each of the four models stresses the importance of different ends or criteria for creating an effective organization. The four criteria clusters define the core values on which judgments about organizations are made. For some, creativity and what it takes to foster creativity, is most important. For others, productivity, efficiency or morale and the means to achieve these ends may be most important. As Cameron and Quinn (1999) describe this,

What is notable about these four core values is that they represent opposite or competing assumptions. Each continuum highlights a core value that is opposite from the value on the other end of the continuum – that is, flexibility versus stability, internal versus external. The dimensions, therefore, produce quadrants that are also contradictory or competing on the diagonal. The upper left quadrant identifies values that emphasize an internal, organic focus, whereas the lower right quadrant identifies values that emphasize an external, control focus. Similarly, the upper right quadrant identifies values that emphasize an external, organic focus, whereas the lower left quadrant emphasizes internal, control values. The competing or opposite values in each quadrant give rise to the name for the model, the Competing Values Framework.

The concept of competing values is implicit in much of the current organizational development and management literature. Collins and Porras (1997), in their book

“Built to Last,” describe the successful habits of visionary companies, which they define as premier institutions in their industries, widely admired by their peers and having a long track record of making a significant impact. One of the twelve myths that were shattered during their research is the “Tyranny of the OR.”

[Visionary companies] reject having to make a choice between stability or progress; cult-like cultures OR individual autonomy; home-grown managers or fundamental change managers; conservative practices OR Big Hairy Audacious Goals; making money OR living according to values and purpose. Instead they embrace the ‘Genius of the AND’ - the paradoxical view that allows them to pursue both A AND B at the same time.

### **Defining the Competing Values Theory for S&T organizations**

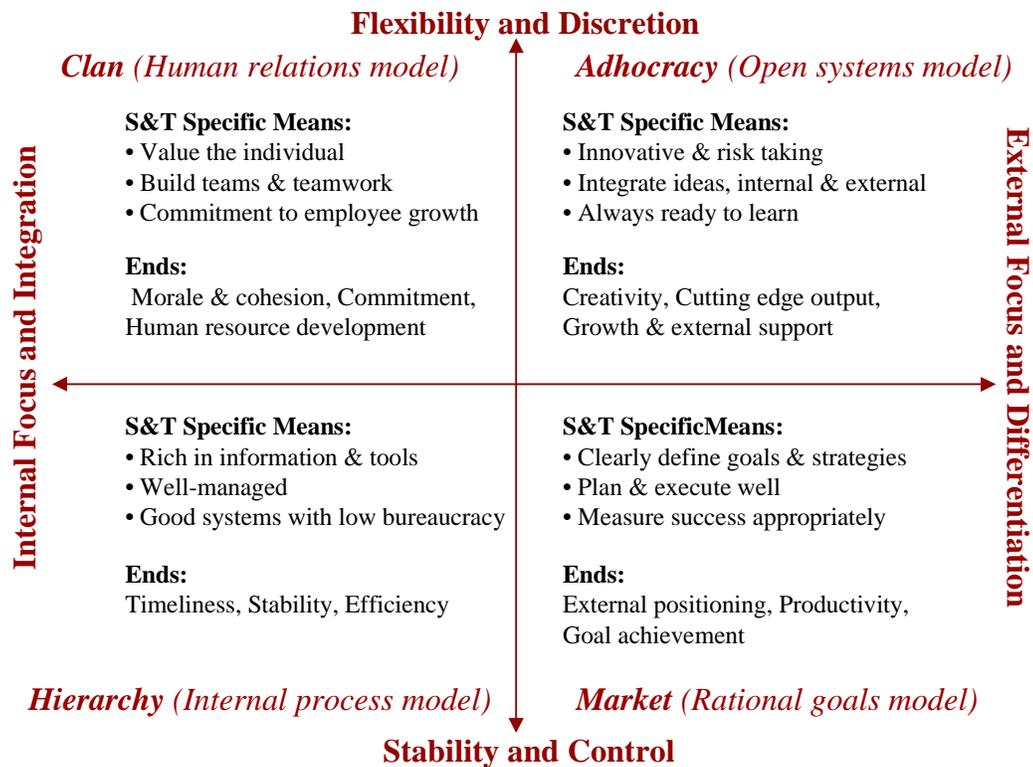
Altschuld and Zheng (1995) first proposed using the Competing Values theory to assess research organizations. They examined major assessment approaches and identified key issues in evaluating effectiveness for educational and social science research organizations whose outputs were intellectual rather than tangible products. They argued that research organizations need a framework because otherwise value judgments will be implicit rather than explicit. Since the Competing Values theory captures four different models of organizational effectiveness, it provides guidance for recognizing value biases. Using it as a framework may be more appropriate for research than any single model of effectiveness. For example, the goal-attainment approach is not by itself appropriate because research organizations have intangible goals. This lack of tangible goals and quantitatively measurable outcomes means that frameworks considering only efficiency and output measures reveal only part of the picture of effectiveness. The strategic constituency approach, while it is applicable because social references are important for publicly funded research organizations, creates opportunities for political manipulation which suggests it would not be wise to use it as the single model for S&T effectiveness.

Taking the lead from Altschuld and Zheng (1995), this DOE study set out to identify the attributes of a S&T organization’s culture from the perspectives of the four models represented in the Competing Values theory. This was accomplished through an extensive literature review and more than a dozen focus groups and questionnaires from more than 1500 scientists in DOE and industrial research laboratories. The study defined and organized the attributes that S&T managers and evaluators found essential to spur innovation and achievement of the desired performance. The twelve attributes that have been defined as the “means” that S&T organizations achieve the four criteria of effectiveness are shown in Figure 3. Just as Cameron and Quinn (1999) do not claim to have the only framework for assessing organizational effectiveness, this study does not claim to have determined the one right way of grouping and categorizing the attributes that are necessary in the S&T organizational environment. Rather, this scheme is proposed as one way to

organize thinking about the effectiveness of S&T organizations. It is based firmly in the S&T management literature and structured to take advantage of what is known about the competing values of organizational effectiveness.

The twelve attributes in the DOE study are further defined below, along with examples of supporting references from the literature. The authors whose work is noted below are: Bennis and Biederman (1997); Bland and Ruffin (1992); Brown and Eisenhardt (1995); Chiesa, Coughlan and Voss (1996); Ellis (1997); Endres (1997); Hauser and Zettelmeyer (1997); Hull (1988); Hurley (1997); Judge, Fryxell and Dooley (1997); Kanter (1988); Kumar, Persaud and Kumar (1996); Martin and Skea (1992); Menke (1997); Miller (1992, 1995); Montana (1992); National Research Council (1996); Pelz and Andrews (1976); Purdon (1996); Ransley and Rogers (1994); Roberts (1988); Rosenberg (1994); Szakonyi (1994); Tipping, Zeffren and Fusfeld (1995); Udwadia (1990); Van de Ven and Chu (1989).

**Figure 3. The Competing Values Theory Adapted for S&T Organizations**



***Attributes Influencing Morale, Commitment, Human Resource Development***

**Valuing Individuals** includes demonstrating respect for people, rewarding and recognizing merit, and offering competitive salaries and benefits. Kanter mentions that feeling valued and secure helps people be creative. Martin and Skea and Tipping et. al. conclude that performance is dependent upon morale. Bland and Ruffin, Ellis, Kanter and Van de Ven and Chu think that appropriate rewards are consistently related to productivity or innovation, and Judge et. al. suggest that intrinsic rewards are more important than monetary rewards.

**Building Teams and Teamwork** includes valuing cooperation and teamwork, emphasizing trust and integrity, and providing support to S&T projects. Chiesa et. al., Hurley, Judge et. al., Kumar et. al., Martin and Skea, Montana, and Udwardia mention teamwork as a prerequisite for effectiveness. Kumar et. al. says that cohesion in the project team and support from other teams affect project success. Udwardia argues that creativity is enhanced when trust and respect are promoted. The mix of skills within a team is also important. Hurley concludes that technical productivity and excellence require well-trained and highly developed technicians.

**Commitment to Employee Growth** includes hiring and retaining quality scientific and technical staff and providing opportunities for career development, as well as educational and professional development. Bland and Ruffin, Hurley, Judge et. al., Martin and Skea, the NRC study team, and Udwardia all cite quality of staff as important. In particular Bennis and Biederman and Judge et. al. suggest that recruitment should be careful to fit individuals within the group. Commitment to training and career paths is also important. One of ten best practices summarized by Ransley and Rogers is having career development in place and tied to strategy. Endres and Kanter mention employee development as important to innovation, with Kanter specifically mentioning jobs that are broadly defined.

***Attributes Influencing Creativity, Cutting-Edge Output, Growth and External Support***

**Innovative and Risk Taking** environments include a sense of challenge and enthusiasm, encouragement to pursue new ideas, and autonomy in scientific management. Bland and Ruffin and Udwardia note that intellectual challenges and positive group climate stimulates productivity and creativity. Creativity and effectiveness is also enhanced, according to Hurley, Kanter, Udwardia, and Van de Ven and Chu, in organizations that push change and risk taking. Hurley, Judge et. al., Kanter, Pelz and Andrews, Udwardia, and Van de Ven and Chu, believe that productivity and excellence are dependent on researchers having autonomy and the freedom to make decisions about their research.

**Integrates Ideas, Internally & Externally** requires organizations to be effective at the internal cross-fertilization of ideas, external collaborations and interactions, and at developing integrated, relevant project portfolios. Chiesa et. al., Endres, Hurley, Roberts, Montana and several others indicate that project success hinges on the ability of researchers to exchange and discuss ideas with colleagues in their organization as well as outside their organization. Hauser and Zettelmeyer, Menke, Ransley and Rogers, Roberts, Tipping et. al., and Udwardia expand idea integration to include an organization's ability to develop a portfolio that combines risks, needs, and goals with the ideas of decision-makers, program managers, researchers, and marketers.

**Always Ready to Learn** captures an organization's commitment to critical thinking, their ability to identify new projects and opportunities, and willingness to protect a researcher's time to think and explore. Pelz and Andrews, Tipping et. al., and Udwardia emphasize that effective organizations create environments where researchers feel free to learn, interact, disagree, and produce. According to Chiesa et. al., Hauser and Zettelmeyer, Kanter, Purdon, Ransley and Rogers, and Roberts, successful organizations must have the ability to consistently identify customer needs, emerging trends, and external opportunities. Kanter, Martin and Skea, Purdon, Udwardia, and Van de Ven and Chu reinforce the importance of having adequate time to conduct research, explore new approaches, and maintain mastery in their field.

*Attributes Influencing Timeliness, Stability, Efficiency*

**Rich in Information and Tools** includes good internal communication, strong research competencies and knowledge base, and good facilities and equipment. Bland and Ruffin, Chiesa et. al., Hull, Kanter, Purdon, Rosenberg, and Udwardia mention that consistent and direct communication between researchers, managers, and departments facilitates the production, circulation, and development of new ideas and technologies. Developing and taking advantage of expertise, diversity, and areas of core competency help organizations to maintain success, innovation, and "best practices" according to Bland and Ruffin, Endres, Kumar et. al., Menke, Purdon, and Ransley and Rogers. Research project and program success is enhanced by the quality of a laboratory's tools, equipment, and facilities, according to Chiesa et. al., Hurley, Kumar et. al., Martin and Skea, the NRC study team, and Udwardia.

**Well-Managed** refers to laboratory management that is decisive and informed, adds value to the research, and allocates resources well. Bland and Ruffin, Brown and Eisenhardt, Kumar et. al., Menke, Udwardia, and Van de Ven and Chu emphasize the importance of having managers with the technical knowledge, ability, and authority to make hard decisions about projects, employees, and resources. Bland and Ruffin, Hurley, Montana, Roberts, Szakonyi, and Udwardia mention the importance of committed managers with leadership and people skills. Kanter, Kumar et. al.,

## *7 Evaluation in Multi-Goal Partnerships*

Roberts, and Van de Ven and Chu state that the ability to allocate resources with forethought and strategy is an important management quality.

**Good Internal Systems** include laboratory services and laboratory systems and processes, such as financial accounting and purchasing, and minimal overhead rates and indirect burden. Hurley and Udwadia list the importance of library, computing, database, and communication services to research productivity. Chiesa et. al., Ellis, Menke, Roberts, Szakonyi, and Van de Ven and Chu expand this concept by introducing the importance of effective S&T processes and procedures ranging from project initiation and termination to human resources and hiring. Bland and Ruffin, Brown and Eisenhardt, Ellis, Judge et. al., Kumar et. al., Montana, and Tipping et. al. stress the importance of overall efficiency of organizational functions, low bureaucracy, low cost, and decentralization.

### *Attributes Influencing External Positioning, Productivity, Goal Achievement*

**Clearly Defines Goals & Strategies** includes an organization's ability to define a research vision and strategies for achieving it, maintain its commitment to fundamental research, and maintain strong relationships with its sponsors. Bland and Ruffin, Menke, Montana, Ransley and Rogers, and Tipping et. al. stress the importance of developing clear long- and short-term goals, a unifying and guiding organizational vision, and a well-communicated strategy for fulfilling goals and vision. Chiesa et. al., Hauser and Zettlemeyer, Kumar et. al., Menke, Miller, Montana, Purdon, Ransley and Rogers, and Roberts note that the organization's goals and strategies must be developed in response to sponsor need and input. To be successful, organizations must communicate with their sponsors and ensure that all projects address customer and end-user requirements and feedback.

**Plans and Executes Well** includes how an organization plans for and executes projects, whether they have sufficient, stable project funding, and if they invest in future capabilities. Brown and Eisenhardt, Chiesa et. al., Kumar et. al., Menke, Pelz and Andrews, Roberts, Szakonyi, Tipping et. al., and Udwadia list aspects of planning that enhance organizational performance that include having a formalized plan for activity integration, a clear definition of potential applications, and a strategy for choosing the right projects and focusing on realistic and relevant goals. Chiesa et. al. and Judge et. al. stress the importance of having sufficient funding that is stable and flexible, completing projects, and pursuing innovation. Essential to the success of long-term planning is the organization's investment in future capabilities. Chiesa et. al., Hull, Menke, Rosenberg, and Tipping et. al. state that future success depends on an organization's existing S&T processes as well as its ability to monitor and adapt to emerging technologies, industry change, and market fluctuation.

**Measures Success Appropriately** includes the criteria and methods the organization uses to evaluate both project and laboratory success, and the organization's reputation for excellence. Kanter, Miller, Montana, Ransley and Rogers, and Tip-

## *7 Evaluation in Multi-Goal Partnerships*

ping et. al. found that research projects are the most innovative and successful when they have milestones to reach, when they meet both research and business objectives, and when they are periodically and objectively reviewed. The NRC study team found that a reputation for excellence stemmed from a focus on continuous improvement, commitment to quality, and quality of research. Kanter and Martin and Skea note that an organization's reputation for excellence was a factor in driving future innovation and excellence.

### **Implications for Evaluation**

This new framework for assessing the effectiveness of S&T organizations recognizes that there are competing values implicit in the different perspectives of effectiveness, and that effectiveness must be assessed within the context of the organization's circumstances. These circumstances may be stable or changing and include the organization's resources, the stage or type of research, and its external technical, economic, social and legal/political influences. As discussed in more detail below, use of the proposed framework could improve evaluation planning, provide a balanced set of leading indicators for research performance, be useful for investigating differences for different types of research and environmental contexts, and guide the modification of existing data collection methods.

### ***Evaluation planning and utilization***

The better the evaluation planning, the more likely the evaluation is to provide information that users perceive as valuable and useful. Good evaluation planning starts with a clear definition of the purpose of the evaluation and its audience and a thorough picture of the components of the "program" or organization being evaluated. Whether the purpose of the evaluation is to provide managers with information on how to improve, or to provide senior managers and sponsors of the S&T with evidence that the S&T is well managed and meeting its objectives, an evaluation is more likely to be utilized and credible if it is based on a framework that includes various stakeholder perspectives.

The proposed framework provides guidance with respect to recognizing value biases and making them explicit. It provides a means for matching an organization's basic characteristics with evaluation strategies and facilitates the choice of effectiveness criteria such that assessment will be respected and accepted both internally and externally. The relationships between organizational structure and organizational effectiveness are recognized and can be investigated. Since the proposed framework also looks at the organization within its particular context, evaluations using this framework can examine relationships between organizational effectiveness and other measures of performance such as outputs and impact, or the stage or type of research.

***A Balanced Set of Leading Indicators***

Use of the proposed framework could also provide a balanced set of leading indicators for improvement and reporting. Leading indicators provide managers with early warning of opportunities for improvement as well as challenges that may detract from performing excellent S&T. The effectiveness of the S&T organization is a leading indicator for future S&T outcomes and provides a balance to the current emphasis on measuring those outcomes. Moreover, use of the Competing Values theory suggests that a balanced set of indicators for organizational effectiveness would include (1) creativity, (2) morale, (3) external positioning and productivity, and (4) efficiency of internal support structure and systems. For example, measures of creativity would include the extent of internal and external collaboration and the use of cross-functional teams. Other important leading indicators might be the alignment of organizational structure and culture with type of research and purpose as well as with resources and influences of the external environment.

The importance of using a balanced set of indicators should not be underestimated. Measurement always perturbs the system but will perturb it less if the set of measures or indicators are comprehensive enough to cover all aspects of the system and, where indicators push or pull in undesirable directions, offsetting indicators are included in the set.

***Investigating differences for different types of research and environmental context***

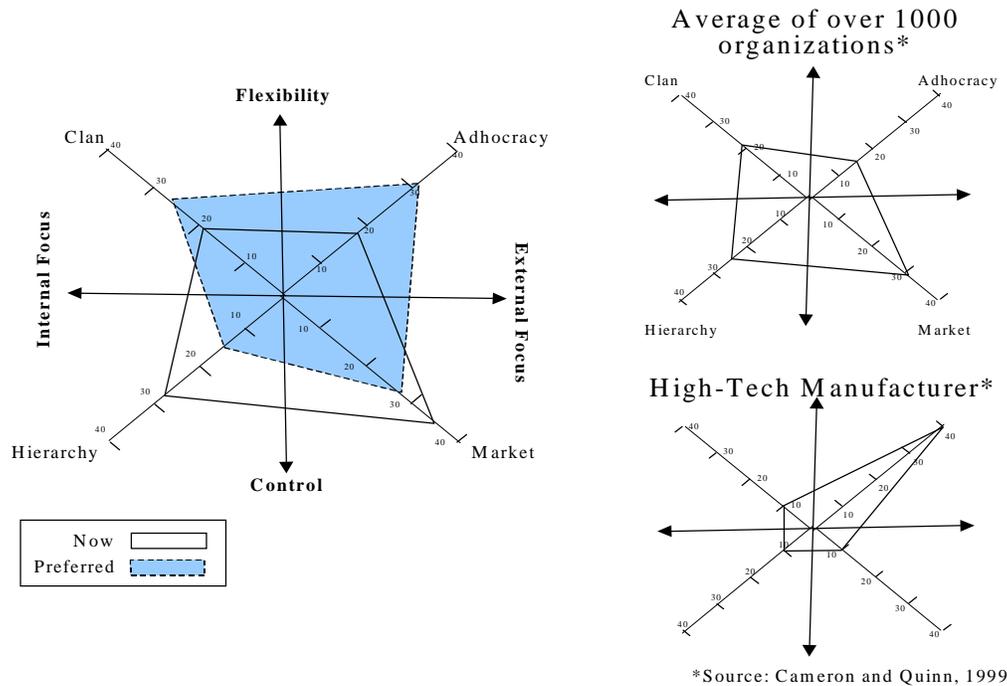
Use of the proposed framework is helpful for clarifying differences in organizational structure and focus, depending on the type or purpose of S&T or changing external circumstances. One of the tensions in the Competing Values theory is between control and flexibility. As Branscomb (1999), Udwardia (1990), and many others point out, more fundamental research requires a more flexible, non-hierarchical organization. The other tension is between internal focus and integration and external focus and differentiation. The trends toward increasing globalization and collaboration would indicate that organizations need to strike a balance between internal and external focus that is more on the side of external focus than it has been in the past.

Examples from experience with Cameron and Quinn's (1999) Organizational Culture Assessment Index (OCAI) demonstrate the potential usefulness of the framework in this area. A questionnaire is supplemented by qualitative methods to produce an overall organizational culture profile that assesses six dimensions of how an organization works and what it values. The OCAI identifies what the current organizational culture is like, as well as what the organization's preferred culture should be. For each of the six dimensions people suggest the weight that the organization gives or should give to values that represent each of the four quadrants of the Com-

## 7 Evaluation in Multi-Goal Partnerships

peting Values theory. An average for each quadrant is calculated from the six dimensions for both current and preferred and is plotted as shown in Figure 4.

**Figure 4. Examples of Organizational Culture Assessments**



The OCAI helps the organization diagnose its dominant orientation based on the four core culture types. The organization on the left in Figure 4 sees its culture as dominated by hierarchical structure and rules, as well as an orientation toward setting and achieving market or market-like goals. It would prefer to be more agile and flexible and more concerned about developing human resources. This might be in response to a dynamic and changing external environment that requires more cross-functional teaming, a different skill mix, and assurance to staff that they will be valued in the new environment. The figures on the right of Figure 4 are averages based on Cameron and Quinn's (1999) experience with more than 1000 organizations. As expected, the dominant culture profile of a high technology manufacturing organization is weighted toward flexibility and external focus, where effectiveness is viewed as creativity, cutting edge output, and external growth and support.

### ***Modifying Data Collection Instruments***

Another implication for evaluation is in the area of modifying data collection methods. Currently self-assessment, employee and customer evaluations, and reviews by peers, advisory committees and experts outside the laboratory, are the primary means used by U.S. federal laboratories to assess their workplaces and organizational effectiveness. Criteria for assessment differ and there is no common framework. Existing self-assessment and evaluation instruments fail to cover all of the perspectives of effectiveness in the Competing Values theory and thus are seen as incomplete or biased by those whose perspectives are not represented. In employee attitude surveys used by DOE laboratories, for example, the attributes identified in the literature and by DOE scientists as necessary for creativity and innovation, such as the freedom to pursue new ideas and the cross fertilization of ideas, are not usually investigated. Peer and expert review, while well regarded by scientists, typically provide information at the project level rather than the organizational level. Thus organizational assessments are rarely comprehensive or credible and are not comparable from one organization to the next.

It appears feasible and valuable to add a few standard questions to existing assessment instruments, including peer review, using anchored word scales to describe aspects of the S&T environment and the four views of effectiveness. This standard information could be compared across organizations, perhaps even aggregated from project level data and summarized. The relationships between this standard information and characteristics of the organization and its circumstances could then be investigated. Given increasing capabilities in “data mining,” this approach could increase the usefulness of current data collection methods.

### **Areas for further research**

Initial use of the framework in the DOE study indicates that, as Cameron and Quinn (1999) have found, the Competing Values theory is intuitively appealing to people and is useful in describing the tensions in managing S&T organizations. For example, managers of public research see that the proposed framework explains the tension between typical government hierarchical rules and the flexible environment needed for S&T. More research is suggested in several areas. First it is necessary to get more stakeholder input and refine the definition of the proposed framework through application to various types of S&T organizations in differing circumstances. More research is also needed to understand what is “good” or “best practice” for various circumstances, as well as how the framework and assessments can be used by managers to direct change and even to allocate resources. A related area for research, given the current interest in performance measurement and performance based management, is the relationship between the competing values in the S&T organizational culture and concepts such as the Balanced Scorecard and “Built to Last” strategies. The aim of these research questions would be to demonstrate

that use of the proposed framework provides credible, valid, and useful information to the managers of S&T organizations.

### **Summary and Conclusion**

There are two major requirements that motivate the interest in better methods for assessing the effectiveness of S&T organizations. First there is a desire to know how to improve S&T organizations and demonstrate effectiveness. Second there is the need to describe and recognize differences in effectiveness and management environments that are related to the type of research conducted, the resources available to the organization, and the influence of external circumstances. The framework for assessing S&T organizational effectiveness proposed by the authors builds on current models to include multiple perspectives of effectiveness, using the notions of Rohrbaugh, Cameron and Quinn's (1983, 1988, 1999) Competing Values theory. The proposed framework suggests that organizational effectiveness be assessed within the context of the organization's resources, external environment, and the type and purpose of the research.

The linkages described in the literature between specific attributes of S&T organizations that lead to effectiveness and outcomes begin to demonstrate that a framework can be defined that many would agree organizes and provides focus and credibility for assessing S&T organizational effectiveness. More research is necessary, but it appears possible that use of this framework could improve evaluation planning and utilization by bringing in the perspectives of multiple stakeholders. It could provide a balanced set of leading indicators for performance management efforts and facilitate investigating differences across S&T organizations. Valuable information could be gained by the addition of a few questions on organizational culture and circumstances to existing self-assessment, customer and employee surveys, and peer and expert review. If standardized, these questions would provide comparative data that would establish "best practices" depending on type of research and circumstances. Data mining would provide managers and policy makers with valuable information not currently available.

### **Acknowledgements**

The authors gratefully acknowledge the vision, support, and funding of Patricia Dehmer and Iran Thomas of the U.S. Department of Energy Office of Basic Energy Sciences, the insight and work of fellow project team member Steve Binkley, Sandia National Laboratories, and the assistance of the study champions in the participating research laboratories. Research has been performed under contract DE-AC04-94AL85000 with Sandia National Laboratories. Sandia is operated by Sandia Corporation, a subsidiary of Lockheed Martin Corporation. The opinions expressed are those of the authors, not the U.S. Department of Energy or Sandia National Laboratories.

## References

- Altschuld, J.W. and Zheng, H.Y. (1995) 'Assessing the effectiveness of research organizations: an examination of multiple approaches,' *Evaluation Review*, Vol. 19, 197-216.
- Bennis, W. and Biederman, P.W. (1997) '*Organizing genius: the secrets of creative collaboration.*' Reading, MA: Addison-Wesley.
- Bland, C.J. and Ruffin, M.T. (1992) 'Characteristics of a productive research environment,' *Academic Medicine*, Vol. 67, No. 6, 385-397.
- Branscomb, L.M. (1999) 'The False Dichotomy: Scientific Creativity and Utility,' *Issues in Science and Technology*, Vol. 16, Num. 1, 66-72.
- Brown, S.L. and Eisenhardt, K.M. (1995) 'Product development: past research, present findings, and future directions,' *Academy of Management Review*, Vol. 20, 343-378.
- Cameron, K.S. and Quinn, R.E. (1999) '*Diagnosing and changing organizational culture: based on the Competing Values Framework.*' Reading, MA: Addison-Wesley.
- Canada, (1993) "Methods for Assessing the Socioeconomic Impacts of Government S&T," Working Group on S&T Financial Management and Mechanisms, Federal Government of Canada.
- Chiesa, V., Coughlan, P. and Voss, C.A. (1996) 'Development of a technical innovation audit,' *Journal of Productive Innovation Management*, Vol. 13, 105-136.
- Collins, J.C. and Porras, J.I. (1994, 1997) *Built to Last: Successful Habits of Visionary Companies*, New York, Harper Collins.
- Committee on Science, Engineering and Public Policy (COSEPUP), National Academy of Sciences, National Academy of Engineering and the Institute of Medicine. (1999). *Evaluating Federal Research Programs: Research and the Government Performance and Results Act*. Washington, D.C.: National Academy Press.
- Crow, M. and Bozeman, B. (1998). *Limited By Design: R&D Laboratories in the U.S. National Innovation System*. New York: Columbia University Press.
- Denison, D.R. (1997) '*Corporate Culture and Organizational Effectiveness.*' Ann Arbor, MI: AVIAT.
- Ellis, L. (1997) '*Evaluation of R&D processes: effectiveness through measurements.*' Boston: Artech House.

## 7 Evaluation in Multi-Goal Partnerships

- Endres, A. (1997) *'Improving R&D performance the Juran way.'* New York: John Wiley & Sons.
- Hauser, J.R. (1996) 'Metrics to value R&D: an annotated bibliography,' Working paper for ICRMOT, March.
- Hauser, J.R. and Zettelmeyer, F. (1997) 'Metrics to evaluate R,D&E,' *Research • Technology Management*, Vol. 40, No. 4, 32-38.
- Hooijberg, R. and Petrock, F. (1993) 'On cultural change: using the competing values framework to help leaders to a transformational strategy,' *Human Resource Management*, Vol. 32, 29-51.
- Hull, F.M. (1988) 'Inventions from R&D: organizational designs for efficient research performance,' *Sociology*, Vol. 2, 393-415.
- Hurley, J. (1997) *'Organisation and Scientific Discovery.'* New York: John Wiley & Sons.
- Judge, W.Q., Fryxell, G.E., and Dooley, R.S. (1997) 'The new task of R&D management: creating goal-directed communities for innovation,' *California Management Review*, Vol. 39, No. 3, 72-85.
- Kanter, R.M. (1988) 'When a thousand flowers bloom: structural, collective, and social conditions for innovation in organizations,' *Research in Organizational Behavior*, Vol. 10, 169-211.
- Kumar, V., Persaud, A.N.S, and Kumar, U. (1996) 'To terminate or not an ongoing R&D project: a managerial dilemma,' *IEEE Transactions on Engineering Management*, Vol. 43, 273-284.
- Laredo, Philippe. and Mustar, Philippe. (2000). "Laboratory Activity Profiles: an Exploratory Approach." Forthcoming in *Scientometrics*, June.
- Martin, B.R. and Skea, J.E.F. (1992) 'Academic research performance indicators: an assessment of the possibilities,' Science Policy Research Unit, University of Sussex, April.
- McLaughlin, G.C. (1995) *'Total Quality in Research and Development.'* Delray Beach, FL: St. Lucie Press.
- Menke, M.M. (1997) 'Essentials of R&D strategic excellence,' *Research • Technology Management*, Vol. 40, No. 5, 42-47.
- Miller, R. (1992). "The Influence of Primary Task on R&D Laboratory Evaluation: A Comparative Bibliometric Analysis." *R&D Management*, Vol. 22, Iss. 1, pp. 3-20.

## 7 Evaluation in Multi-Goal Partnerships

- Miller, R. (1995) 'Applying quality practices to R&D,' *Research · Technology Management*, Vol. 38, No. 2, 47-54.
- Montana, A.J. (1992) 'Quality in R&D – if it isn't perfect, make it better,' *Research · Technology Management*, Vol. 35, No. 4, 38-41.
- National Research Council (1996) 'World-class research and development: characteristics of an army research, development, and engineering organization,' National Academy Press.
- National Science and Technology Council. (1996). *Assessing Fundamental Research*, Washington, D.C.
- Office of the Auditor General of Canada (1999), *Attributes of a Well Managed Research Organization*, Report of the Auditor General, November.  
<[http://www.oag-bvg.gc.ca/domino/reports.nsf/html/99menu\\_e.html](http://www.oag-bvg.gc.ca/domino/reports.nsf/html/99menu_e.html)>
- Organization For Economic Co-operation and Development (1997), *In Search of Results: Performance Management Practices*.
- Pelz, D.C. and Andrews, F.M. (1976) 'Scientists in organizations: productive climates for research and development,' Institute for Social Research, University of Michigan.
- Purdon, W.A.B. (1996) 'Increasing R&D effectiveness: researchers as business people,' *Research · Technology Management*, Vol. 39, No. 4, 48-56.
- Quinn, R.E. (1988) 'The Competing Values Model: redefining organizational effectiveness and change,' *Beyond rational management: mastering the paradoxes and competing demands of high performance*, San Francisco: Jossey-Bass.
- Quinn, R.E. and Cameron, K. (1983) 'Organizational life cycles and shifting criteria of effectiveness: some preliminary evidence,' *Management Science*, Vol. 29, 33-51.
- Quinn, R.E. and Rohrbaugh, J. (1983) 'A spatial model of effectiveness criteria: towards a competing values approach to organizational analysis,' *Management Science*, Vol. 29, 363-373.
- Ransley, D.L. and Rogers, J.L. (1994) 'A consensus on best R&D practices,' *Research · Technology Management*, Vol. 37, No. 2, 19-26.
- Roberts, E.B. (1988) 'Managing invention and innovation: what we've learned,' *Research · Technology Management*, Vol. 31, No. 1, 11-29.

## 7 Evaluation in Multi-Goal Partnerships

- Rohrbaugh, J. (1983) 'The competing values approach: innovation and effectiveness in the job service,' *Organizational Theory and Public Policy*, Eds. R.H. Hall & R.E. Quinn, Beverly Hills: Sage.
- Rosenberg, N. (1994) '*Exploring the Black Box.*' Cambridge, Great Britain: Cambridge University Press.
- Saleh, S.D. and Wang, C.K. (1993) 'The management of innovation: strategy, structure, and organizational climate,' *IEEE Transactions of Engineering Management*, Vol. 40, 14-21.
- Smart, J.C. and St. John, E.P. (1996) 'Organizational culture and effectiveness in higher education: a test of the "culture type" and "strong culture" hypothesis,' *Educational Evaluation and Policy Analysis*, Vol. 18, pp. 219-241.
- Szakonyi, R. (1994) 'Measuring R&D effectiveness I,' *Research • Technology Management*, Vol. 37, No. 2, 27-32.
- Szakonyi, R. (1994) 'Measuring R&D effectiveness II,' *Research • Technology Management*, Vol. 37, No. 3, 44-55.
- Tipping, J.W., Zeffren, E., and Fusfeld, A.R. (1995) 'Assessing the value of your technology,' *Research • Technology Management*, Vol. 38, No. 5, 22-39.
- Udwadia, F.E. (1990) 'Creativity and innovation in organizations: two models and managerial implications,' *Technological Forecasting and Social Change*, Vol. 38, 65-80.
- U.S. General Accounting Office (1996), "Executive Guide: Effectively Implementing the Government Performance and Results Act." GAO/GGD-96-118.
- U.S. House of Representatives, 103<sup>rd</sup> Congress, "Science in the National Interest", Hearing before the Subcommittee on Science of the Committee on Science, Space, and Technology, August 4, 1994, 4.Sci 2:103/149.
- Van de Ven, A.H. and Chu, Y. (1989) 'A psychometric assessment of the Minnesota Innovation Survey,' *Research on the Management of Innovations: The Minnesota Studies*, Van de Ven, Angle and Poole (Eds), New York: Harper and Row, 55-98.

## Discussion of Gretchen Jordan's paper

**Henry Etzkowitz** (State University of New York, USA). Why use four focus groups? What are the competing values that you found in the focus groups both at Sandia and Ford?

**Gretchen Jordan.** They didn't want to do the evaluation in labs if no one would actually use it. There was difficulty in having people participate. We used Ford because it was a willing participant. The focus groups always started out by having people define "excellent research" and found that they always had the same issues come up. If there were differences, it's because of differences in the funding mechanisms and environments. Sandia felt the heavy weight of bureaucracy. There are dynamic tensions in managing research.

**Henry Etzkowitz.** GM's transition hasn't happened yet but it is coming.

**Barry Bozeman** (Georgia Institute of Technology, USA). There is a contradiction built in in studies: you cannot have everything at once, this goes against the competing values models. If you go for one value, you lose something else: If we use the competing values framework, what do we do with the fact that government cannot cope with these trade offs between the values. What do you do with the pressure to pretend that a program or institution is good in everything?

**Gordon Jordan.** We need to help people understand that failure is common. It's the reason we all do benchmarking.

**John Barber** (Department of Trade and Industry, UK). There was a Paris meeting on the role of public sector research. How might we get some of your perspectives into these types of studies? Where would we find accounts of this type of research?

**Gretchen Jordan.** My paper has an extensive bibliography. I would like for the paper to be circulated but would also like feedback.

**Arie Rip** (Twente University, The Netherlands). There may be a missed opportunity because it didn't conceptualize "doing science." It's possible to do a double-version of what you are doing by combining with how scientists deal with the hypocrisy. Look at the work of Richard Whitley.

**Irwin Feller** (Pennsylvania State University, USA). I can finally understand why I have had a 10-year battle with university administrators - because I work in one type of organization and they work in another. This analysis will help in battles with them. It helps to understand the characteristics of good management.

**John Barber.** I suggest you also look at the hierarchy of structure. There are some real needs of science and technology organizations such as a library. They are defining positive aspects of structure rather than negative aspects.

## *7 Evaluation in Multi-Goal Partnerships*

**Iriwn Feller.** The more formal the system, the less flexible organizations are in their needs such as hiring, etc. An analysis may be too normative for real use.

**John Barber.** Use an organizational index to be able to see what should be done.

**Stefan Kuhlmann** (Fraunhofer ISI, Germany). The paper is very valuable but a central question exists. Within the matrix, how do you manage (in an evaluation) the effect that these aspects are competing? In all these contradictory areas, there is a lot of struggle between them so is it difficult to say which of these cells should be used? How do you operationalize these?

**Gretchen Jordan.** The management book has ways of balancing these conflicting demands. There is a fair amount of literature on balancing flexibility and autonomy and the focus that you give to scientists and engineers, so as to not have one person's values, etc driving the research focus. The general goal is to be flexible.

**Peter Blair** (Sigma Xi, USA). In a large organization such as Sandia, where you are in the matrix may actually move compared to where you are in the organization. You mentioned an index to describe this, but can you use a matrix to describe where you are in an organization?

**Gretchen Jordan.** We could try to apply an existing index of types of researchers. Sandia is particularly interested in spin-offs.

---

## **System Evaluation of the Promotion of Research Cooperation for Small and Medium-Sized Enterprises (SMEs) - the Concept of the Federal Ministry of Economics and Technology (BMWi)**

**Heike Belitz<sup>†</sup> and Hans-Peter Lorenzen<sup>‡</sup>**

<sup>†</sup>German Institute for Economic Research (DIW), Berlin, Germany

E-mail: hbelitz@diw.de

<sup>‡</sup>Federal Ministry of Economics and Technology (BMWi), Berlin, Germany

E-mail: hans-peter.lorenzen@bmwi.bund.de

---

### **Introduction**

In Germany, the potential to improve the implementation of research and development (R&D) results in industry is regarded as enormous. This is an important reason for system evaluations of the large research organisations which are basically publicly financed and additionally supported within the framework of project promotion programmes and for which the Federal Ministry of Education and Research (BMBF) is responsible. The results of the joint system evaluation of the German Research Foundation (*Deutsche Forschungsgemeinschaft*, DFG) and the Max Planck Society (*Max-Planck-Gesellschaft*, MPG) and of the system evaluation of the Fraunhofer Society (*Fraunhofer Gesellschaft*, FhG) were published in 1999. A system evaluation of the Helmholtz Foundation of Research Institutes (*Helmholtz-Gemeinschaft der Forschungseinrichtungen*, HGF, national research centres, *Großforschungseinrichtungen*) started in autumn 1999. An evaluation of the Institutes of the Science Foundation of the Leibniz Institutes (*Institute der Wissenschaftsgemeinschaft der Leibniz-Institute*, WGL, formerly "Institutes on the Blue List", "*Institute der Blauen Liste*") is being prepared.

Industry depends on the input of new know-how from research. Although the MPG and the DFG primarily carry out and promote basic research, their R&D results are increasingly implemented by industry. The participation of the MPG in a number of spin-offs in the field of biotechnology and the "ideas workshop" of the DFG are examples. In 1999, the FhG increased its revenues from contract research for industry to 37 % of its budget (excluding defence research) and thus serves as a model among the German non-profit R&D institutes.

As far as it is responsible, the BMWi promotes these efforts to improve the cooperation of various government supported R&D institutes with industry, insofar as they carry out industry-related R&D. The early involvement of industry in the selection of the R&D topics, the project design and the implementation is increasingly gaining significance.<sup>1</sup> While the BMBF within the framework of its

## *7 Evaluation in Multi-Goal Partnerships*

gaining significance.<sup>1</sup> While the BMBF within the framework of its special programmes promotes joint projects of companies and research institutes for the development of specific technologies or the application in selected fields, the BMWi uses the instrument of indirect project promotion in order to initiate or maintain in a different way SME-related research and development without such technology-related requirements. In this context, the question of the impact of these indirect support measures on the competitiveness of industry-related R&D institutes that work primarily for SMEs becomes more and more important also against the international background. State framework conditions and support measures must facilitate the activities of R&D institutes on the international market for R&D services and reward the successful implementation of R&D results in industry. Therefore it is necessary to review the whole range of indirect, traditional project support measures of the BMWi that aim at R&D cooperation.

In the past, evaluations of individual support measures were carried out, partly several times. The competent Ministries, the BMWi and the BMBF, together with the programme administration agencies<sup>2</sup>, examined whether the objectives of the relevant programmes were reached and commissioned external evaluations by independent institutions. The results of the controlling, the internal and external evaluations had an impact on the individual programmes. The assessment of the contributions of these programmes to the enhancement of the competitiveness of SMEs and of the R&D institutes that work primarily for them as well as to the ability of the relevant companies and research institutes to cooperate is therefore regarded as an essential improvement potential. Especially the non-profit research institutes may benefit, apart from the basic financing, from several support programmes and highly depend on these public funds. Thus the various programmes have an impact on them. The Federal Ministry of Economics and Technology therefore elaborated the Terms of Reference for the system evaluation of "industry-integrating research promotion", comprising all indirect support programmes that aim at cooperation and networking between SMEs and R&D institutes. In August 2000 the Ministry set up a commission of external experts who will carry out this evaluation.

Thus the initiated system evaluation of the industry-integrating research promotion of the BMWi belongs to the third layer of the shell model of evaluation procedures in the German research and technology policy according to Kuhlmann, namely institutions. The third shell comprises the new attempts - intensified in the nineties -

---

<sup>1</sup> Cf. the joint declaration of the Federation of German Industries (BDI) and the German scientific organisations of 7 April 2000 and the draft paper of the Scientific Council on the future development of the scientific system in Germany of 25 April 2000.

<sup>2</sup> These agencies are responsible for the conception, the organisation and the realisation of technological policy programmes and projects on behalf of the ministry.

## *7 Evaluation in Multi-Goal Partnerships*

to assess the performance of the industry-oriented, publicly supported research infrastructure.<sup>3</sup>

### **Factors that have an Impact on the System Evaluation Concept**

The BMWi's concept for the system evaluation of industry-integrating research promotion has been mainly influenced by the following factors:

- the related objective against the background of the definition of the new role of the BMWi in the technology policy after the change in government in 1998,
- the object: the programmes that were initially set up in two Federal Ministries and of which the BMWi is now in charge, whose objectives or target groups partly overlap and which compete for the scarce public funds and
- the experiences that have been made within the framework of other evaluations, in particular the system evaluation of the FhG.

### *Objective of the System Evaluation*

The objective of the evaluation is the creation of a transparent, consistent support system on the basis of the traditional BMWi support programmes which aim at co-operation and networking; this system should meet the future requirements of small and medium-sized businesses regarding modern industry-integrating promotion of research and development in an optimum way. A new support system is planned to be elaborated on the basis of the theoretical principles on state intervention to reduce the consequences of market failure in R&D, the empirical analyses of the success factors of R&D cooperation between SMEs and research institutes and the experiences with the efficiency of the R&D support instruments that have been used so far. This future support system should meet the following minimum requirements:

- take account of the needs of innovative SMEs,
- be easy to implement for SMEs and as transparent as possible,
- be flexible with regard to new requirements of industry (structural change, globalisation),
- make a substantial contribution to the creation of sustainable jobs,
- guarantee a minimum qualitative standard of the research projects that are to be supported,

---

<sup>3</sup> Cf. Kuhlmann, St.: *Moderation von Forschungs- und Technologiepolitik? Evaluationsverfahren als "reflexives" Medium*. In: Martinsen, R. and Simonis, G. (ed.): *Demokratie und Technik - (k)eine Wahlverwandtschaft?* Opladen 2000, pp. 305-334.

## *7 Evaluation in Multi-Goal Partnerships*

- prevent competitive distortions,
- treat projects in the eastern federal states in a preferential way,
- be complementary to the support measures of the BMBF, the federal states, the EU and the BMWi itself.

### *Object of the Evaluation*

The object of the system evaluation is the "industry-integrating" promotion of research and development for SMEs by the BMWi. This comprises all indirect support programmes of the BMWi that aim at cooperation and networking. These are basically:

- the promotion of R&D cooperation projects between companies and between companies and research institutes in Germany and abroad within the framework of the "PROgramme INNOvation Competence - PRO INNO",
- the project promotion within the framework of the R&D special programme "promotion of research, development and innovation in small and medium-sized enterprises in the eastern federal states",
- the "promotion of innovative networks - InnoNet",
- the Industrial Joint Research (*Industrielle Gemeinschaftsforschung*, IGF) within the framework of the Confederation of Industrial Research Associations (*Arbeitsgemeinschaft industrieller Forschungsvereinigungen*, AIF), including the inter-sectoral variant "future technologies for small and medium-sized enterprises - Zutech".

While the programme PRO INNO can be characterised as market-related element of research promotion by the BMWi because companies which themselves apply the research results participate in the supported R&D projects, the promotion of the Industrial Joint Research traditionally has been the pre-competition element of research promotion by the BMWi. The programme InnoNet supports larger pre-competition R&D network projects. The R&D project promotion in the eastern federal states covers the whole range of pre-competition and market-related projects. This programme does not directly aim at research cooperation but rather triggers it indirectly. The applying non-profit research institutes must make a contribution of their own which they cannot make unless they cooperate with enterprises.

Most programmes supported by the BMWi restrict the size of eligible businesses and their equity capital relations with large companies. The EU rules on research and development must be observed according to which the amount of the aid for R&D expenditures depends on the observation of specific criteria regarding the definition of SMEs. In addition, special rules favouring companies in the structurally weak eastern federal states and in the eastern part of Berlin exist; the BMWi

makes full use of these provisions within the framework of its support measures. Furthermore, the grants to the eligible R&D project expenditures of research institutes also differ depending on their classification pursuant to the tax legislation and the state basic financing. Table 1 gives an overview of the various types of research institutes.

**Table 1: Classification of Industry-Related Research Institutes in Germany**

	Industrial businesses and institutions			Public corporations and statutory bodies	
<b>Tax legislation</b>	Profit businesses	Non-profit businesses (tax reductions, can provide receipts for donations)			
<b>State basic financing</b>		No basic financing	Basic financing, but less than 50 %	Basic financing, more than 50 %	
<b>Examples</b>	R&D service providers	External industrial research institutes in the eastern federal states,  Sector-specific research institutes of the research associations	External industrial research institutes in the eastern federal states,  Fraunhofer Gesellschaft,  Sector-specific research institutes of the research associations	MPG,  R&D institutes of HGF and WGL	Universities, polytechnics,  Federal and regional R&D institutions

Until 1998 the BMBF had been in charge of the predecessor of the nation-wide programme to promote R&D cooperation projects, PRO INNO. The BMWi has traditionally been responsible for the promotion of pre-competition R&D projects in industry-related research institutes and universities - since the fifties organised by the SME-oriented, technology-related and sector-specific industrial research associations. Since German unification the BMWi has been in charge of the special promotion of R&D projects in industrial and non-profit external research institutes as well as in research-intensive and young SMEs in eastern Germany. The programme InnoNet to support larger network projects, i.e. cooperation networks consisting of at least two research institutes and four SMEs, is a relatively new promotion line of the BMWi. Except for the last, rather minor programme, all other industry-integrating programmes have similar scopes - in terms of annual federal funds and the number of supported, albeit differently structured R&D projects - (tables 2 and 3).

**Table 2: Support Funds for Industry-Integrating R&D Programmes 1998-2001**

Support programme	1998	1999	2000 (budget)	2001 (government draft)
	in DM million			
<b>PRO INNO and predecessor programme: cooperation projects</b>	210	212	189	202
<b>Industrial Joint Research (IGF)</b>	173	167	175	175
<b>Special programme for eastern Germany : project promotion</b>	169	175	180	170
<b>Innovative networks (InnoNet)</b>	0	0.4	9	13
<b>Total</b>	552	554	553	560

Source: BMWi

Apart from the four above-mentioned programmes that basically aim at the promotion of research projects for small and medium-sized enterprises, there are further support programmes of the BMWi that contribute to the establishment of research cooperations or are even an essential prerequisite for them. These programmes, which indirectly support research cooperation, therefore must be taken into account in the evaluation process. The personnel promotion programme in eastern Germany, for instance, in many companies in the eastern federal states secures the personnel basis for the absorption of external know-how and the ability to cooperate with research institutes and other businesses. Other R&D programmes of the BMWi also indirectly contribute to more cooperation. The ERP innovation programme, for instance, in the credit variant expands the financing scope of the borrowers who carry out R&D themselves or to a considerable extent award R&D contracts. The programme "Venture capital for small technology-based firms" (BTU) supports the establishment and growth of R&D start-up service providers which carry out research projects for SMEs. The programme FUTOUR promotes technology-oriented start-ups in the eastern federal states and thus supports the creation of new innovative SMEs. The assistance consists of a combination of consulting, a grant and dormant equity holding.

**Table 3: Scope and Limits of the Programmes of Industry-Integrating Research Promotion**

	<b>PRO INNO cooperation projects</b>	<b>IGF/ ZUTECH</b>	<b>InnoNet</b>	<b>Project Promotion in eastern Germany</b>
<b>Number of projects per year</b>	In the first project year 464 cooperation projects,  292 thereof projects between SMEs,  172 projects between SMEs and R&D institutes	About 1,200 projects currently,  400 new projects each year,  i.e. average duration about 2 - 3 years	18 joint projects in the first programme year	1,250 support projects,  more than 500 thereof new projects each year
<b>Limits</b>	A maximum of 2 projects per company.  A maximum of DM 250,000 per project for research institutes	Annual limits of support funds for research associations depending on their own funds	A maximum of DM  3 million per project	Percentage of supported R&D personnel: company: a maximum of 75 % non-profit organisations: a maximum of 50 % (in future 60 %)  (in future maximum grants per project)

This evaluation does not deal with support programmes of other institutions with similar objectives such as European research support programmes, research promotion carried out by the federal states, the support programmes of the BMBF within the framework of the special subject-related and technology-oriented programmes or the institutional support, especially support of the *Fraunhofer Gesellschaft*, as well as the activities of the state research institutes (the Federal Institute for Materials Research and Testing, BAM, and the Federal Institute of Physics and Metrology, PTB); in this study, they are regarded as exogenous factor whose impact, however, should also be taken into account.

#### *Key Questions Regarding the Terms of Reference*

The following key questions arise on the basis of the evaluation of individual programmes and the comparison of partly competing programmes in order to reach the above-mentioned objective of the system evaluation:

## *7 Evaluation in Multi-Goal Partnerships*

1. Is the public support of industry-integrating research justifiable in economic terms (e.g. due to market and system failures)? To what extent is the special promotion of R&D in eastern Germany justifiable after the year 2004? Which types of cooperation (contract research, cooperation between companies and research institutes, networks) should be initiated and facilitated through promotion? Which instruments (grants, venture capital, low-interest loans) are best suited to reach this objective? What position should the BMWi support hold in the three-layer support system: Europe - the state - the federal states?
2. Are there elements in the support systems of other countries from which important conclusions can be drawn for the industry-integrating research promotion in Germany?
3. How about the efficiency of the above-mentioned existing BMWi support programmes?
  - Are structural economic changes such as priority shifts between sectors and between the manufacturing industry and the services sector, new economy businesses, the demand for system solutions and internationalisation taken into account by the support programmes?
  - Are there any deficits regarding the institutional prerequisites of the relevant R&D institutes?
  - What is the impact of the evaluations of individual support programmes that have already been carried out or are being carried out?
4. What are the consequences of the experiences with the above-mentioned existing BMWi support programmes for the efficiency enhancement and the simplification and expansion of the system of industry-integrating R&D?
  - To what extent can the existing programmes be integrated as consistent modules in a BMWi support system of industry-integrating research and development?
  - How great should the public contribution to eligible projects in the R&D institutions be? Are there reasons for differing requirements regarding the contributions of companies involved in R&D cooperations and networks? Should the public contribution to networks with several enterprises and research institutes be greater than in the case of only two or three partners? To what extent should an institutional basic financing be taken into account?
  - Is it possible and useful to make a distinction between "pre-competition" and "market-related" as regards the R&D categories of the eligible projects or is it more appropriate to distinguish the projects on the basis of the number of in-

## *7 Evaluation in Multi-Goal Partnerships*

dustrial partners involved, their financing contributions and their intellectual property rights?

- How can the individual modules be simplified and improved with regard to access, application, selection procedures, flat rate determination of the supportable costs, cost estimation and accounting procedures?
- Are there support deficits as regards the cooperation of R&D institutions and SMEs which so far have not been covered by nation-wide programmes? To what extent can non-supported and supported loans and contributions of other programmes be integrated in a financing mix that suits the needs of industry-integrating research and development?

### *Experiences of the System Evaluation of the Fraunhofer Gesellschaft*

The approach pursuant to which evaluations focus on Terms of Reference is based on the above-mentioned examples of system evaluations, in particular the FhG evaluation. In this context, it was also necessary to analyse the future prospects not only in scientific terms, but also with regard to the market for R&D services and the underlying changes in the economic structure. As in the case of the FhG evaluation, the Terms of Reference were coordinated not only within the Ministry, but also with the directly concerned industrial associations of the R&D institutes (the Association of Innovative Enterprises, VIU, Dresden, and the Confederation of Industrial Research Associations, AiF, Cologne). In this process, the original ideas were further developed and differentiated instead of agreeing upon the "lowest common denominator". Thus an open discussion on the optimisation of the overall support system has been initiated to which all parties concerned agree.

### **Evaluation Procedure**

#### *Appointment of a Commission*

The BMWi has appointed a commission for the external evaluation whose members are to guarantee an independent assessment of the support programmes by experts. The selection of the commission members took account of experiences with relevant R&D programmes in Germany and abroad, the need for state support and the interests of the parties concerned (R&D institutes and companies). The commission comprising a total of eight members consists of two economists from Germany and abroad (one economist and one innovation researcher), two employers (two chief executives of innovative SMEs, one from western and one from eastern Germany), two scientists who are each appointed as experts for support policy by one of the two major competent organisations of the research institutions (VIU, AiF) and who are no members of these organisations, one business consultant with special experience in research promotion and one technical expert from a university who heads a jury in a competition-based support programme (InnoNet).

## *7 Evaluation in Multi-Goal Partnerships*

The BMWi Terms of Reference are the basis and framework of the commission's work. The commission, however, may modify or expand its tasks if it regards this as necessary in order to reach its objective. It will start its work in autumn and is planned to submit its final report in about one year's time after several meetings, possibly within the framework of working groups.

The BMWi makes available to the commission a secretariat for the retrieval of information and the collection of background data and for the organisation of the commission's work. In addition, funds are available for the award of small additional research contracts and the elaboration of reports of further experts.

### *Involvement of the Parties Concerned*

In the Ministry, four divisions are in charge of the R&D support programmes to be evaluated; various external programme administration agencies deal with the programmes on behalf of the Ministry (processing and examination of the applications, granting of the funds, control of the individual projects, support statistics, detailed analysis of support projects selected at random). The competent BMWi divisions, programme administration agencies and associations concerned have detailed knowledge of the relevant support programmes. Of course they all have their own interests in mind when they provide information about their support programmes in the public discussion and in the permanent internal discussion and assessment for the finetuning of the support by the Ministry. Three project groups for "better implementation of research results in innovative products, procedures and services" were set up prior to the system evaluation that has now begun; these groups consist of representatives of the Ministry, the research institutes, companies and their organisations (AiF, VIU). Two of these project groups dealt with the external industrial research institutes in the eastern federal states and with the pre-competition Industrial Joint Research, which is supported by AiF; they are also the main targets of BMWi support programmes which the system evaluation examines. These project groups have elaborated proposals on more problem- and application-oriented R&D in their institutions to meet the needs especially of SMEs; the implementation of these proposals has already started.<sup>4</sup> Consultations between the project groups on the relevant problems and possible solutions for the rapid implementation of their R&D results in industry has already resulted in closer cooperation between the various research institutes.

The existing external evaluation studies provide important information about the efficiency and handling of individual support programmes and their predecessors. Comparable information about the remaining programmes of industry-integrating promotion is, however, necessary for an analytical overall picture as the basis for

---

<sup>4</sup> Cf. "Ergebnispapier der Arbeitsgruppe Externe Industrieforschungseinrichtungen" of 30 May 2000 and "Die industrielle Gemeinschaftsforschung - Basis für das industriegetragene Innovationsnetzwerk des Mittelstandes" of 3 May 2000, available from the BMWi.

## *7 Evaluation in Multi-Goal Partnerships*

the system evaluation. Therefore the first task of the system evaluation is the establishment of criteria for the classification, comparison and finally evaluation of the programmes. For this reason, the competent divisions, programme administration agencies and interest groups concerned were already involved in the preparation phase and intensively contributed to the elaboration of the Terms of Reference. At this stage, they were able to raise questions regarding the functioning and the effects of the other complementary and partly substitutive programmes. In the course of this preparation process, the parties concerned already drew conclusions since the elaboration of key questions and detailed questions regarding the programmes triggered the process of agreement on the systematic classification and the comparison of the programmes that had so far been regarded in an isolated way. After the adoption of the Terms of Reference, the competent BMWi divisions made short statements which will be made available to the commission. In these statements, they describe the support programmes for which they are responsible from their perspective in the context of the overall system of industry-integrating research promotion.

There are of course also doubts about this attempt to classify the programmes and evaluate the system. They are based on the fear that within this concept an external commission may not be able to sufficiently acknowledge the special characteristics of individual programmes and their impact and that wrong conclusions may be drawn with regard to the continuation of important support instruments or new priorities. Since budgetary restrictions concerning the support policy are also likely in the future, some parties concerned regard this risk of the external evaluation as greater than the opportunity to optimise the whole support system. The commission's ability to work and the quality of its recommendations may be impaired unless all parties concerned are successfully involved in the system evaluation. The commission can prevent this by thoroughly discussing with all parties concerned and making transparent the criteria for comparison and thus the basis of the evaluation. The question remains open to what extent it will also call for the involvement of the interest groups in the evaluation process: This is possible in hearings of the competent BMWi divisions, the programme administration agencies and the relevant associations. Granting observer status to representatives of the associations concerned would be an expanded form of permanent or temporary involvement. The question of participation rights of parties concerned must be solved by the commission when it starts its work.

The commission will start its work on the system evaluation of the industry-integrating research promotion by the BMWi in October 2000. Thus only the concept and the preparatory steps could be described at this point. Whether and to what extent the commission consisting of experienced experts in cooperation with the Ministry, the programme administration agencies, the associations of the research institutes and further experts will succeed in concluding this new type of evaluation project cannot be said until the end of next year. The presentation of the concept at

this early stage is supposed to promote the discussion on the methodical approach of the system evaluation and to encourage constructive criticism.

---

## **Discussion of Heike Belitz and Hans-Peter Lorenzen's paper**

**Philip Shapira** (Georgia Institute of Technology). It is interesting to compare Belitz's presentation with Jordan's. Belitz spoke mostly about method but not about the process of evaluation management while Jordan greatly emphasized the evaluation management process. Is there a point of comparison between Germany and US? In Germany, the management of research institutes can be a very sensitive topic. If this is true, how does this affect how you manage the evaluation?

**Heike Belitz.** This is difficult to answer because the task of the Commission is to manage evaluation and decide on its implementation. Yet they use different methods and it is not easy to compare these evaluations and methods. You should ask the Commission in 1-2 months.

**Arie Rip** (Twente University, The Netherlands). They are curious characters in the Commission. They are trying to manage theory and true evaluation. You might want to try to manage the Commission so that you can get the best results from both sides.

**Heike Belitz.** There were discussions with the Commission on the topic of R&D cooperation and a proposal was discussed. Another important aspect is involvement of the parties concerned because this helps the acceptance of the results of the evaluation.

**David Guston** (Rutgers University, USA) The objectives of the evaluation "to secure a minimum quality of the proposal" did not explicitly assure the highest quality of the research evaluation.

**Heike Belitz.** The projects are very high quality and also very small projects.

**David Guston.** Do you have an articulation of the minimum quality or is it more intuitive?

**Heike Belitz.** There are competing initiatives within the Commission. If you want to do a real system analysis, should you work closely with other organizations within the Commission for a system analysis? Other initiatives have other goals. They will learn something from their smaller evaluations but they do cooperate with other departments within the Commission.

**Patries Boekholt** (Technopolis, The Netherlands). Since two years there is new policy-structure in Germany, and now the BMWi is sometimes competing with

## *7 Evaluation in Multi-Goal Partnerships*

BMBF. If you want to do a system analysis you need to involve the other ministry!  
Is this possible?

**Heike Belitz.** The next step is a co-operation in the framework of SME, since SME is within the range of BMWi.

## 8. Policy, Institutional and Portfolio Evaluations

Session Chair: Carsten Dreher (Fraunhofer ISI, Germany)

---

### Evaluating Scientific and Technical Human Capital: An Event History Approach<sup>1</sup>

Barry Bozeman<sup>†</sup> and Monica Gaughan<sup>‡</sup>

<sup>†</sup>Research Value Mapping Program, School of Public Policy, Georgia Institute of Technology, Atlanta, Georgia, USA

Email: barry.bozeman@pubpolicy.gatech.edu

<sup>‡</sup>Research Value Mapping Program and Department of Sociology, Oglethorpe University, Atlanta, Georgia, USA

---

#### Introduction

Despite considerable progress in approaches to evaluating science and technology policy, the great majority of studies remains focused on either peer review assessment or discrete products such as publications, patents or licenses. While these approaches continue to be useful we feel they do little justice to some of the more complex socially- and politically-embedded questions of science and technology policy. In particular, questions related to “capacity,” the ability of groups of scientists, engineers and the users of their work to grow and sustain and to make the most of the available talent reservoir. While there are several important reasons why capacity issues receive short shrift, one of these is that it requires a long-term view and longitudinal data, both rare in social research. Nevertheless, we feel it is vital to develop approaches to understanding capacity, particularly “scientific and technical human capital” (Bozeman, Gaughan and Dietz, forthcoming). The conundrum: how does one do this with the typically paltry data resources widely available?

---

<sup>1</sup> The authors gratefully acknowledge the support of the U.S. Department of Energy (DE-FG02-96ER45562) and the U.S. National Science Foundation (SBR 98-18229). The opinions expressed in the paper are the authors’ and do not necessarily reflect the views of the Department of Energy or the National Science Foundation. This work was performed under the *Research Value Mapping Program* [<http://rvm.pp.gatech.edu>] of the Georgia Institute of Technology. The authors wish to thank Jim Dietz, Jongwon Park, Juan Rogers, Jeff Bournes, Marie Chesnut, Jenna Dubois, Kathrin Jackson, Zahi Karam, Jungki Kim, Andy McNeil, Seth Sobel, Ryoung Song, and Larry Wilson for their assistance.

This paper provides a brief introduction to the scientific and technical human capital construct and an illustrative approach to doing research on capacity. Since our project is only at its beginning stages, our data are quite limited but nonetheless evoke the potential of our method.

In this paper, we develop a simple model for explaining scientists' progression rates to full professor. Rank is not, of course, an adequate surrogate for scientific and technical human capital, but the focus on career transitions does illustrate the general method we feel is useful. More sophisticated applications await further data, especially data on the award of public funding and individual participation in funded projects.

The methodological innovation of the paper is use of scientists and engineers curriculum vitae as a data source. This is a widely available data source, one obviously relevant, indeed virtually defining, scientific and technical human capital. It is also longitudinal and permits application of time-series modeling techniques such as event history analysis, the approach we employ here. Before describing our data, methods and results, we begin with a discussion of scientific and technical human capital (S&T human capital) and why it is a compelling focus for evaluative studies in science and technology policy.

## **Scientific and Technical Human Capital**

### ***A. Introduction to the Concept and Models***<sup>2</sup>

S&T human capital encompasses not only the individual human capital endowments normally included in labor economics models (Becker, 1962; Schultz, 1963), but also the sum total of researchers' tacit knowledge (Polanyi, 1967; 1969), craft knowledge, and know-how. but also the individual scientist's tacit knowledge (Polanyi, 1969; Senker, 1997), craft knowledge and know-how (Bidault and Fischer, 1994). S&T human capital further includes the social capital (Bourdieu, 1986; Coleman, 1988) and network ties scientist employ in the pursuit of knowledge, its validation and diffusion.

We argue that understanding the value of scientific and technical knowledge requires a view of the social context of scientific work (for a complementary argument see Audretsch and Stephan, 1999). Much of S&T human capital is embedded in social and professional networks, technological communities (Debackere and Rappa, 1994; Liyanage, 1995), or "knowledge value collectives" (Bozeman and Rogers, forthcoming; Rogers and Bozeman, forthcoming). These networks integrate and shape scientific work, providing knowledge of scientists' and engineers'

---

<sup>2</sup> This section draws extensively from Bozeman, Dietz and Gaughan, forthcoming and Bozeman and Rogers, forthcoming. For more detail on the S&T human capital models please see these papers, now available at the Research Value Mapping web site, <http://rvm.pp.gatech.edu>.

work activity, helping with job opportunities and job mobility, and providing indications about possible applications for scientific and technical work products. The value of knowledge and technology produced in formal and informal networks of scientists depends upon the conjoining of equipment, material resources (including funding), organizational and institutional arrangements for work and the unique S&T human capital embodied in individuals. At any level, from the individual scientist to the discipline, field or network, value is capacity- capacity to create new knowledge and technology. Thus, the key value question is “What factors enhance capacity, diminish it or simply change the reservoir of capabilities inherent in individuals and groups?”

S&T human capital is amenable to study at various levels of analysis. While the research reported in this paper focuses on the individual, the S&T human capital model works as well at the research group or even the scientific field level. It is also possible to consider S&T human capital at the level of “knowledge value collectives” (Bozeman and Rogers, forthcoming; Rogers and Bozeman, forthcoming), the aggregation of scientific knowledge producers and users if their knowledge and technology. Figure One depicts the flow of S&T human capital at the individual level, as influenced by project participation.

### ***B. S&T Human Capital and Research Evaluation***

In recent research on the social and economic effects of R&D projects supported by the US Department of Energy (Bozeman et al. 1997 and 1999), we came to the conclusion that most R&D assessment methodologies give insufficient attention to the socially-embedded nature of knowledge creation; transformation and use; and the dynamic, capacity-generating interchange between human and social capital. On this point, Zucker, Darby, and Brewer (1998) found, in studying the growth of the biotechnology industry, that the industry has grown up literally around so-called scientific superstars of the field (Zucker, Darby and Armstrong, 1998). It was investments in basic R&D—many of which were supported by the federal government—that led to start-up firms that clustered geographically around universities where these biotech superstars worked. The human capital capacity generated by government investments led to the economic wealth. But, in funding those projects, the government was not making *financial* investments, but *scientific capacity generating* ones. In our view, public R&D evaluation should center not on economic value or even improvements in state-of-the-art, but on the growth of capacity (Bozeman and Rogers 1999b).

With respect to S&T human capital, the primary task in public support of science and technology is to develop and nurture the ability of groups (whether networks, projects, or knowledge value collectives) to create new knowledge uses, not simply to develop discrete bits of knowledge or technology.

### **Use of CV Data for Research on S&T Human Capital**

The diverse and highly changing S&T human capital endowments, measured at multiple levels, not only make the study of scientists' and engineers' career trajectories more difficult than, say, standard labor models (e.g., less amenable to standard labor models) but potentially more rewarding. When a scientist or engineer changes jobs the implications are sometimes profound: the movement of knowledge value is in many instances a vital element of scientific discovery, technological innovation, and even economic development. Individual migration patterns of scientists and engineers can be likened more appropriately to the movement of the "web" of knowledge value that they possess—a web that continually takes on new shapes and patterns.

Scientists and engineers, especially academic ones, leave trails, trails marked in a somewhat standard fashion. The curriculum vita (CV) is one of the few widely available researchable records that remains underutilized. The utility of CV data for study of S&T human capital is striking. The CV provides not only a clear-cut indicator of movement from one work setting to the next but is, in a sense, a representation of certain aspects of knowledge value. The CV, unlike other data sources, often recounts the entire career of the scholar in some detail. Thus, it is not simply a list of credentials, but a historical document that evolves over time capturing changes in interests, jobs, and collaborations. Whether viewed as a historical record, a marketing tool, or a scientific resource, it is a potentially valuable datum for persons interested in career trajectories, research evaluation, or, more generally, science and technology studies. Not only is the CV nearly universal, it is in some respects standard, and it is relatively easily obtained (sometimes even from the public domain). Most important, the CV contains useful, concrete information on the timing, sequence, and duration of jobs, work products (e.g., articles, patents, papers), collaborative patterns, and scholarly lineage.

In addition to its value as a stand-alone source of data, a great advantage of the CV is that it can be used in conjunction with other sources of data. The availability of a wide array of citation data through the Science Citation Index is extremely valuable. These same databases also include information on the "power index" (i.e., the likelihood of citation) of journals. Similarly, the aggregate data provided by US National Science Foundation (NSF) databases, such as SESTAT, also serves as a potentially fruitful linkage. The problem, of course, is that the decision to use such benchmarks and cognate data requires making significant "up front" decisions on data collection strategies.

On the other hand, this approach is not without its limitations or problems, as we have discovered in the early phases of our research program (see Dietz, et al., forthcoming, for more details on the practicalities of research with CV's). In fact, several of the advantages of using the CV as a data source can also be viewed as disadvantages. First, because the information is self-reported, it is subject to being por-

trayed in a favorable light or even completely fabricated. But so too is any self-reported questionnaire or interview data. Second, the semi-structured format falls short of a purely standardized template, thus risking the elimination of valuable information or the inclusion of extraneous non-relevant data. Perhaps most significant, however, is the enormous work involved in coding the CV for subsequent data analysis. Not only is the coding time-consuming, but it is tedious and runs the risk of introducing error due to coder fatigue. In some cases it is possible to have as many as 2,000 variables for one CV.

Despite its limitations, the potential of the CV as a research tool is considerable. We have found that the costs of collecting, coding and entering CVs are higher, even higher than we expected, but we also expect those costs to be redeemed in terms of the quality of data available for longitudinal analysis of scientists' and engineers' career trajectories.

### **Modeling Transition to Full Professor**

In this initial study of S&T human capital our focus is quite limited. We examine the factors pertaining academic scientists' promotion to full professor. We do not claim that this is a vital issue for policy makers, though it is one that is of some concern beyond the individual's self-interest. Most institutions are concerned that academic rank progression exhibit "fairness" (by any of several definitions) and there is a concern with ensuring reasonable judgments and rewards. While most of us feel we have a good understanding of what is required for promotion- publish a great deal of high quality work, teach proficiently, work on committees and keep our colleagues distaste for us at least at reasonable levels- we probably know less about these processes than we think. Here are just a few questions about which there is limited evidence: What is the effect of having "alternative" careers, moving in and out of universities? What are the cohort effects governing career transitions? What are the gender and field dynamics? How much does amount published affect promotion and are their threshold effects? What are the impacts of grants and project participation on career trajectories and promotion? While we do not yet have the data to answer all these questions, we can get start on some of them, even with the limited data available for this study.

### **Data and Methods**

The data come from affiliates of eight NSF-funded Centers, Science Centers or Engineering Research Centers, we are studying as part of the Research Value Mapping Program's research on scientific and technical human capital. After having performed case studies of the Centers, we asked center affiliates to provide up to date curriculum vitae. About 55% complied, most transmitting them via email. It is important to emphasize that these data do not represent the population of natural and physical scientists; rather, they constitute a unique census of NSF-funded projects and centers, and of the scientists affiliated with these particular research enter-

prises. The curriculum vita data capitalize on the strengths of such a source: the name, field, and timing of degrees, career transitions, and publications and grants activity are readily coded in a standard format (for more information, see Dietz et al. forthcoming). At the same time, it is possible to consider individual scientists as they are embedded in their scientific research group.

In this analysis, we focus on how the organization of the early professional life course affects the transition rate to full professor. Therefore, we limit our analysis to those scientists who occupy, or who have occupied, an academic position. This results in a sample size of 189 academic scientists. Figure 2 shows the survival curve of the group, including scientists who do not experience the event (i.e. are censored). Censoring can happen when a person does not experience the event either because of a failure to be promoted, or because the person has not yet experienced the event, but may in the future. Clearly, both groups are of importance to the question of timing, and the analytic method we select allows us to include them in the multivariate analysis. On the horizontal axis are the years of risk, as calculated from last educational degree attainment. On the vertical axis is the proportion remaining in each time interval. At the beginning of the period (that is, in the first year following award of the last degree), none of the academics were promoted to full professor. Indeed, in the first 10 years of their careers, only 10% of scientists are promoted to full professor. The transition to full professorship is most rapid during the 10 to 15 year career period, after which the curve begins to flatten out.

This is a diverse and interesting group of scientists, and their CVs reveal heterogeneity particularly in the early career period. Descriptive statistics of the analytic sample are provided in Table 1. The scientists come from a wide range of Ph.D. cohorts: roughly 20 to 25% of the sample comes from each of the decades of the 1970s, 1980's, and 1990's. One-fifth of the sample is comprised of scientists awarded the Ph.D. in the 1950s or 1960's. Reflecting the unique nature of this sample, 13 scientists do not hold a Ph.D. The vast majority of the scientists--87%--are male; this may reflect the age of the sample. Forty-eight percent published prior to finishing the Ph.D, and 25% pursued academic post-doctoral studies. Although 66% ultimately pursued a traditional career trajectory (assistant to associate to full professor), 53% have occupied at least one nonacademic position during their careers. The scientists who were appointed as assistant professors (n=124) spent an average of 5 years in the academic rank. The associate professors (n=109) spent 6 years in the rank. Finally, the average full professor (n=97) has occupied his position for an average of 13 years. Overall career length is 21 years, with an annual publication rate of 2.3 articles. Each of these variables is tested to explain variation in the promotion rate.

## Results

The longitudinal nature of this data set lends itself particularly well to the use of event history analysis (a.k.a Cox proportional hazards, or survival analysis). In

such a regression analysis, one is interested in the timing of events. Here we analyze the rate of transition to the rank of full professor. The event, in other words, is the promotion to full professor, and the timing is conceptualized as the years elapsed since the final educational achievement. The survival curve in Figure 2 (described above) provides an excellent intuitive glimpse of our dependent variable. In fact, the dependent variable is the instantaneous rate of transition from one state to another known as the hazard rate, defined as:

$$H(t) = \lim [P(t \leq T < t + \Delta t \mid t \leq T) / \Delta t], \text{ where } T \text{ is the time of the promotion.}$$

A plot of the hazard rate is shown in Figure 3. The horizontal axis continues to represent the number of years since the last educational attainment. Note, however, that the vertical axis is the hazard function of the transition. What this denotes is the likelihood of making the transition to full professor in any time period. Turning again to the plot, note that the greatest rate of transition occurs in the 7 to 10 year period, followed by a fairly sustained rate of transition, and then a rapid decline in the promotion rate after 20 years.

We use partial likelihood estimation recommended by Cox (1972) and Allison (1984). This approach is fairly robust to violations of assumptions, although there is still the problem of nonindependence of observations due to clustering of scientists. This may affect standard errors and tests of significance, which should be interpreted with caution. In interpreting coefficients, we also examine direction and magnitude of effect. Results are reported in two ways. First, the log-odds coefficient resulting from the partial likelihood estimation function shows the change in the log-odds of the hazard given a unit shift in the independent variable, net of other covariates. Interpreting the log-odds coefficients is not intuitively appealing, so we also provide the odds of the hazard occurring. The odds are obtained by exponentiating the log-odds coefficient, and are interpreted in terms of the odds impact on the hazard, or promotion rate (Allison, 1984). Standard errors and probability values are provided for each coefficient, and overall fit is assessed using chi square tests.

Table 2 is key to the results as it shows the relationships among our models variables and between the predictor variables and the dependent variable, the hazard rates for promotion to full professor. The risk ratio, which has a base of one, can be interpreted as the variables likelihood of increasing or decreasing the “risk” of the event, promotion to full professor. If the ratio is 1.0, risk is neither increased or decreased when taking the variable into account. By way of further illustration, a risk ratio of 2.0 means that there is a two times greater risk, whereas a risk ratio of .50 means that there is, essentially, half the likelihood of the risk once we consider the effect of the variable.

*The Overall Model.* The model provides an excellent accounting of the risk of the event, transition to full professor. With a Chi Square of 30.00 and a p value of

.0001 we can consider, especially when taking into account the predictive value of the individual terms in the model, the risk is well accounted for by the overall model. The PhD cohort variables are introduced as a control for period effects, taking the 1950's Ph.D. cohort as the reference category. With the dummy defined as each cohort taken against the oldest cohort, the question becomes "what is the impact of membership in cohort  $t$  as compared to membership in the 1950's (and before) cohort, the one having had the longest time to reach full professor status?" None of the results is significant, though the 1990's cohort, as we would expect, has a stronger p value of .15. This implies that there are not radical cohort affect influencing the results.

*Gender.* According to the risk ratio, being male more than doubles the risk for the event, promotion to full professor. The results are complicated, however, by the fact that the percentage of women in the sample increases in each cohort, especially in the 1990's cohort where, of course, the risk for the event is greatly reduced. Modeling the interaction of incidence of women by cohort requires more data than we have at present.

*Publish before Ph.D.* Interesting, publishing before the Ph.D. is obtained has an effect opposite one's expectations. Those who have published before obtaining a Ph.D. are less likely to experience the risk event of promotion to full professor. We feel this is attributable to the fact that very recent Ph.D's (the 1990's cohort) are more likely to have published while graduate students but they simply have not had time to approach the peak for risk.

*Academic Post Doctoral Position.* Those who have had a post doctoral position are also at reduced risk for transition to full professor (the p value is .08). We think this finding inconclusive. It is perhaps owing to the fact that postdoctoral positions are much more common in the 1990's cohort, related, the sample includes several individuals who were occupying post doctoral positions at the time we gathered the data. Sorting out the full and perhaps complex effects of post doctoral positions can only be done once we have more complete data permitting us to consider field effects and to provide a more sophisticated and interactive model of cohort effects.

*Traditional Career Trajectory.* Our sample is unusual in that includes a substantial percentage of people (34%) who have not taken the traditional route of assistant-associate-full professor but for one reason or another, usually work in an industry setting, they have "skipped" certain stages. Our anecdotal knowledge from case studies tells us that the effects of non-traditional careers are quite diverse within this group. In many cases a non-traditional route has, essentially, reduced the likelihood of obtaining full professor status to zero. But in other cases the non-traditional track has had the effect of permitting individuals to reach full professor more quickly than is typical from the traditional route. The risk ratio for the traditional career trajectory variable shows that, overall, keeping to the traditional route nearly doubles the likelihood of exposure to the risk event of transitioning to full professor.

At the same time, the variable “any nonacademic position”- which simply indicates whether the individual has at any point since the Ph.D. had a non-academic position- is not significant in the model.

*Annual Publication Rate.* The annual publication rate, defined simply as the number of total publications divided by the number of years of one’s career, is significant and positive. What this tells us is that an increment of one publication per year, beyond the average rate, increases by 9% the likelihood of the risk event. Again, more sophisticated analysis is needed but awaits further data.

### **Conclusion**

We feel our approach- using curriculum vita data for longitudinal analysis with event history models- has considerable potential for analysis of the social context of science and, more particularly, to issues of evaluation. But in many respects the current paper is an illustration or “tease.” Many of the factors of greatest interest can be examined only after we have more complete data and from a more representative sample (or at least in comparison to a more representative sample). As policy evaluators, we are particularly interested in the impacts of public funding support on scientists’ careers and, in turn, the accumulation and diffusion of human capital.

Our next step will be to model the effects of public funding support, asking some of the following questions: What are the impacts of grants support? Does early career grants support have a greater impact? Does grants support affect job change or transition? Are their different effects according to the funding agency? Or there field effects of grants support? Does the length and amount of grant support play a major role in scientific and technical human capital accumulation? If we know the answers to these questions, as well as others related to such likely mitigating variables as gender, cohort and national origin, we feel that the allocation and management of publicly-funded programs and projects can be improved. If funding agents have information of the sort we hope to gather next, they should have a much better idea of the ways in which public funding of scientists can be used to maximally enhance scientific careers and the accumulation and diffusion of scientific and technical human capital.

**Table 1**

**SCIENTIFIC TRAJECTORIES  
Descriptive Statistics**

---

Eligibility: Scientists who have ever had a tenure-track academic position.

Total Sample Size: 182	#	%	Mean	Std Dev
Male (0=female; 1=male)	159	87		
Publish Before PhD	88	48		
Phd Cohort				
No PhD	13	7		
1950s	10	6		
1960s	30	17		
1970s	42	23		
1980s	52	29		
1990s	35	19		
Academic Post Doctoral Position	45	25		
Traditional Career Trajectory (Asst to Assoc to Full)	120	66		
Any Nonacademic Position	96	53		
Ever Full Professor (n=180)	94	48		
Years in Assistant Grade (n=124)			5	2.3
Years in Associate Grade (n=109)			6	5.8
Years in Full Grade (n=97)			13	9.5
Length of Career			21	11.7
Articles Per Year			2.3	2.7

**Table 2**

**SCIENTIFIC TRAJECTORIES**  
**Multivariate Event History Analysis**

Eligibility: Scientists who have ev had a tenure-track position  
Dependent Variable: Hazard/Risk/Transition to Full Professor

Total Sample Size: 180	Log Odds	SE	p value	Risk Ratio
Male	0.75	0.39	0.06	2.11
Publish Before PhD	-0.12	0.24	0.6	0.88
Academic Post Doctoral Position	-0.47	0.26	0.08	0.63
Traditional Career Trajectory	0.66	0.24	0.01	1.94
Any Nonacademic Position	0.13	0.22	0.56	1.14
Annual Publication Rate	0.09	0.03	0.01	1.09
PhD Cohort 60	-0.18	0.44	0.68	0.83
PhD Cohort 70	0.06	0.43	0.9	1.06
PhD Cohort 80	-0.32	0.46	0.49	0.73
PhD Cohort 90	-1.58	1.1	0.15	0.21
(Reference Category is Ph.D Cohort 50)				
No PhD	1.82	0.72	0.01	6.17
Model Chi Square	30.99			
d.f.	11			
p value	0.001			

## References

- Allison, Paul D. 1984. *Event History Analysis: Regression for Longitudinal Event Data*. Newbury Park, CA: Sage Publications.
- Becker, G. (1962), 'Investment in human capital: A theoretical analysis,' *Journal of Political Economics* 70, No.5, pp.S9-S49.
- Blossfeld, H, A. Hamerle and K. Mayer. 1989. *Event History Analysis*. Lawrence Erlbaum, Hillsdale, NJ.
- Bozeman, B. and Melkers, J. (1993), ' *Evaluating R&D impacts: Methods and practice*,' Boston, Kluwer Academic Publishers.
- Bozeman, B. and et al. (1999), ' *The Research Value Mapping Project: Qualitative--quantitative case studies of research projects funded by the Office of Basic Energy Sciences. Final report to the Department of Energy, Office of Basic Energy Sciences.*,' Atlanta, Georgia Institute of Technology.
- Bozeman, B., Roessner, J. D., Rogers, J., and Klein, H. (1997), ' *The R&D Value Mapping Project: Annual Report, 1997. Report to the Department of Energy, Office of Basic Energy Sciences.*,' Atlanta, Georgia Institute of Technology.
- Bozeman, B. and Rogers, J. (forthcoming), 'Strategic management of government-sponsored R&D portfolios: Project Outputs and "Scientific and Technical Human Capital" ,' *Environment and Planning C: Government and Policy*.
- Bozeman, B. and Rogers, J. (1999), 'Knowledge value collectives: A capacity-focused alternative for research evaluation,' *Submitted for publication*.
- Bourdieu, P. (1986), 'The forms of social capital,' In Richardson, J. G. ed., *Handbook of theory and research for the sociology of education* New York, Greenwood.
- Callon, M., Courtial, J. P., Crance, P., Laredo, P., Mauguin, P., Rabeharisoa, V., Rocher, Y. A., and Vinck, D. (1991), 'Tools for the evaluation of technological programmes: An account of work done at the Centre for the Sociology of Innovation,' *Technology Analysis & Strategic Management* 3, No.1, pp.3-41.
- Coleman, J.C. (1988), 'Social capital in the creation human capital,' *American Journal of Sociology* 94, pp.S95-S120.
- Coleman, J.C. (1990), ' *Foundations of social theory*,' Cambridge, The Belknap Press of Harvard University Press.
- Cox, D. R. 1972. Regression Models and Life Tables." *Journal of the Royal Statistical Society Series B* 34:187-202.

8 Policy, Institutional and Portfolio Evaluations

Dietz, J., I. Chompalov, B. Bozeman, E. Lane and J. Park (Forthcoming). "Using the Curriculum Vita to Study the Career Paths of Scientists and Engineers," *Scientometrics*.

Johnson, H. 1972. "Some Economic Aspects of Science," *Minerva*, 10 (1), 10-18.

Liyanage, S. (1995), 'Breeding innovation clusters through collaborative research networks,' *Technovation* 15, No.9, pp.553-567.

Polanyi, M. 1962. "The Republic of Science: Its Political and Economic Theory," *Minerva*, 1, 54-73.

Polanyi, M. (1967), *The tacit dimension*, London, Cox & Wyman.

Polanyi, M. (1969), *The logic of tacit inference. Knowing and being*, London, Routledge & Kegan Paul.

Rogers, J. and B. Bozeman. Forthcoming. "Knowledge Value Alliances: An Alternative to R&D Project Evaluation." *Science, Technology and Human Values*. .

Schultz, Theodore W. (1963), *The economic value of education*, New York, Columbia University Press.

Yamaguchi, Kazuo. 1991. *Event History Analysis*. Sage.

Zucker, L. G., Darby, M. R., and Brewer, M. G. (1998), 'Intellectual human capital and the birth of US biotechnology enterprises,' *The American Economic Review* 88, No.1, pp.290-306.

Zucker, L. G., Darby, M. R., and Armstrong, J. (1998), 'Geographically localized knowledge: Spillovers or markets?,' *Economic Inquiry* 36, No.1, pp.65-86.

## Discussion of Barry Bozeman and Monica Gaughan's paper

**Irwin Feller** (Pennsylvania State University, USA). This is a very valuable technique. On one of the first transparencies you talked about social capital. What do you mean by social capital and how does it appear empirically in the paper. There is at least one slight problem in tying the definition of social capital to grants—a selection bias. If a funding program award size is low, productive researchers won't apply to the program causing a selection bias.

**Barry Bozeman.** One of the reasons we are gathering the data from the centers and the comparison data is so that we can look at that caveat. The answer to the first question is that social capital is not a variable in this particular paper, although it is in our study (the Research Value Mapping (RVM) Program) in general. Ultimately we want to examine diffusion through networks.

**Dave Guston** (Rutgers University, USA). How are you going to operationalize tacit knowledge from curriculum vitae (CVs)?

**Barry Bozeman.** If it is ever possible to code tacit knowledge is an even broader question to consider. If this is possible, you can combine this method with the semi-structured interviews and questionnaires that we are conducting. We are not only interested in social capital but recognize that there is reciprocal relationship between social and human capital.

In addition, we are part of Ed Hackett's group working on collaborative research topics and will share data. You can also get at collaboration data from CVs and you can tie to citation data to available S&T workforce databases such as those maintained by the National Science Foundation.

**Henry Etzkowitz** (State University of New York, USA). On the next to last overhead you have statistics on males. If you have parallel data for females you will find strong gender effects.

**Barry Bozeman.** There are gender effects but right now they are washed out by their interaction with age cohorts. When we have more data gender effects will be addressed more directly. In fact, we are working on some papers now, which will focus on this.

**Luke Georghiou** (University of Manchester, UK). If we are talking about capital of any kind, capital depreciates. More recent and older publications have different effects.

**Barry Bozeman.** True. But event history analysis will take this into account.

**Erik Arnold** (Technopolis, UK). This reminds me of research on face identification processing and computer and mathematical modeling to look at how we recognize faces. In a similar way, CVs are a way of looking at how networks coagulate. Do you know of any exercises that have taken this approach in industry? What about national differences in CVs?

**Barry Bozeman.** We want to do this but people in industry have minimalist CVs which turn out to be a problem for the CV as a data collection approach. We do run into all kinds of nationality effects as well.

**Patries Boekolt** (Technopolis, The Netherlands). How can this approach be linked to program evaluation, e.g. the human potential programme of the EU? Specifically, how could this be used to look at the effect of a program for postdocs on their careers?

**Barry Bozeman.** The RVM approach is potentially useful for program evaluation, but we think it may be hard to find a client for it, mostly because of the small numbers of researchers affected by small programs. One of the things we found is that it is not the amount of the grants but the stability of funding that is most important in maintaining productive research groups. Those kinds of programmatic variables are simple but powerful and useful for very large programs.

**Gilbert Fayl** (European Commission). In connection with the European programs, we have a database containing 8-9,000 experts. However, there is a lot of variation in the database. We even have some Israelis in the database, but there is non-comparability of data because of non-standard CVs. I wonder if you have recommended a standard scheme or format for CVs?

**Barry Bozeman.** No, but we have observed that in general that shorter CVs may be becoming more common. It is unfortunate for us that we will lose information if that trend develops or continues. The internationalization of CVs also poses some standardization problems. Many of the researchers in our database were born in other countries.

## **The Assessment of Leibniz Institutes: The Relationship between External and Internal Evaluation**

Martina Röbbcke<sup>†</sup> and Dagmar Simon<sup>‡</sup>

Wissenschaftszentrum Berlin für Sozialforschung, Berlin, Germany

<sup>†</sup>Email roebbecke@medea.wz-berlin.de

<sup>‡</sup>Email dsimon@medea.wz-berlin.de

---

### **Introduction**

With reporting and assessment on the rise in society, the quality of the sciences and of their accountability, or public legitimation, has come under scrutiny. A key point of the discussion is performance and the limits of the scientific system's ability to manage itself. The fact that the evaluation of science has been externalized as a systematic and continuous process in the Federal Republic of Germany in recent years signals that expectations of and demands on science are rising among various social groups, especially political actors. It may also indicate an erosion of trust in the mechanisms that science has for managing itself. Proposals for responses to these trends can focus on science's appraisal and evaluation systems, which have been integral to the scientific system in the United States and European countries other than Germany.

### **Actors and Interests**

In the German discussion of evaluations and their goals, procedures, and methods in the sector of scientific and research, the various actors are readily identifiable. First, there are the policy-makers. They now tend away from creating basic conditions and input controls. Instead, they are moving toward output control. Evaluations are intended to "guarantee" that ever scarcer resources are distributed according to merit. Second, there are the intermediate scientific actors, such as the German Science Council. Its "seal of approval" vouching for the "quality" of scientific institutes prepares decision-making processes bearing on the future of the country's research institutes. Lastly, there are the members of the scientific community, at least a majority of whom expect evaluations to produce measures for quality development that will improve their ability to manage the system of science.

In the late 1980s the German debates about research evaluation and evaluation methods dealt primarily with introducing evaluations into the universities, a step that the universities hesitated to take. However, evaluations of major government-funded nonuniversity research institutes did not begin to draw attention until Germany's unification in the early 1990s, when the German Science Council reviewed those in the former German Democratic Republic and then in former West Germany.

To be sure, research evaluation has long been an important part of scientific life in these establishments, albeit to different degrees. These reviews are usually conducted by the scientific advisory boards of the research centers and are generally highly significant to the self-assurance and perspective of these institutes. What is new in the last few years is that all nonuniversity government-funded research institutes are expected to show proof of regular, systematic evaluation processes and that they all have had to submit to external evaluations, some of which had grave repercussions. A few institutes were even closed as a consequence.

Our contribution is about the institutes of the *Wissenschaftsgemeinschaft Gottfried Wilhelm Leibniz (WGL)*, which are major government-funded research centers outside the universities (so called "Blue List Institutes"). Historically, these institutes have been distinctly stamped by external evaluations, for such assessment was agreed upon long ago by their cosponsors, the federal and state governments. The German Science Council, which had been commissioned to appraise the institutes, concluded its evaluations in mid-2000 and presented its recommendations to all of them.

The results we present in this contribution were generated by a project supported by the *Stifterverband für die Deutsche Wissenschaft* and was conducted at the *Wissenschaftszentrum Berlin für Sozialforschung*. Its purpose was to identify objectives, procedures, and appropriate criteria for future evaluations of major nonuniversity research institutes funded by the government (Leibniz Institutes). Our study of such institutes was narrowed to a particular type with two striking characteristics. First, they deal with fields of research that necessitate treatment of both basic and applied research questions, usually in multidisciplinary research contexts. Second, the research tasks often also entail consulting and other services for reference groups both within and outside the scientific community. Since about the early 1990s this type of institute has gained attention in the discussion of science policy and the sociology of science. For with the rising importance of technical and scientific knowledge in international economic competition, the relation between technology and science (above all in the key technologies) and between basic and applied research has become a relevant topic.

At the same time, new investigations in the sociology of science have indicated changes in the ways knowledge is produced. For example, Gibbons et al. (1994) proceeded from a new type of knowledge production, "mode 2." It is characterized primarily by the great complexity of the issues it is used to address, which are not only generated within science but determined by the "problems of the real world" (Nowotny, 1997). One can therefore assume that this type of research institute has a certain amount of future viability.

The first aspect of interest to us in our study was the different objectives of evaluations. In principle, evaluations can have a broad spectrum of objectives, ranging from interventions of research policy and the creation of information for use in resource allocation to the generation of internal processes for improving quality. In

the case of major government-funded research institutes outside the universities, the objectives of evaluation are heavily influenced by precepts of science policy. The intention of the federal and state governments is to set priorities of science policy by having regular evaluations guide decisions on whether to continue funding the individual institutes. Goal conflicts may arise, however, for the intense pressure generated by an evaluation, the objective of which is to “increase the flexibility” of research institutes, is difficult to reconcile with attempts to ensure and improve quality that are based on a self-critical analysis of strengths and weaknesses.

Our second interest was the question of appropriate procedures for future evaluations. One of the basic insights from research on evaluation is that the different objectives of an evaluation have to be tied to different procedures. In addition, the selection and design of the procedures greatly affects whether the research institutes succeed at designing evaluations as a process of and at preventing excessive interference from outside the scientific system.

A third focus of our inquiry was the evaluation criteria for judging major government-funded nonuniversity research institutes. Whereas the search for suitable procedures can be based on models successfully developed in other settings, the identification of appropriate indicators is more difficult for a number of reasons. First, both the reliability and the validity of research indicators are disputed. Second, the collected data must be interpreted in an informed, independent manner, a step that entails new procedural problems. Third, convincing sets of indicators have to be created for very different institutes, for there are various dimensions to their responsibilities, such as differing subject matter, types of research, and areas of application. On the one hand, the institutes are oriented to relevant criteria generated within the scientific system and to recognition by the scientific community. Their behavior has to do with the specific reward system of science and to the effort to build a reputation. On the other hand, these institutes must also give due consideration to the expectations expressed in science policy, which has lately come to emphasize research on marketable products, procedures, and services. Evaluation of these research centers therefore calls for a diversified catalogue of criteria and questions.

Let us now turn to the objectives and procedures involved in evaluations of this group of nonuniversity institutes and highlight the issue of evaluation criteria.

### **Objectives and Procedures**

To ensure, monitor, and develop the research output and services of the major government-funded research institutes outside the universities, an integrated model of external evaluation and self-evaluation should concentrate on quality improvement, not exclusively on measures designed to ensure quality and facilitate quality control. However, it is not easy to say what an adequate definition of research quality is. Related questions are who defines and redefines quality and what instruments promote it.

Van Vught (1994), for example, stresses the “multidimensional” and “subjective” nature of quality, a term that occasionally has the character of a political concept exploited more and more by policy-makers in recent years. However, quality implies not only inquiring about the strengths of teaching and research but also uncovering weaknesses and discussing how to deal with them. The meaning of the term quality differs from one actor and constellation of interests to the next, particularly where the goals and consequences of evaluations are concerned. As stated by Felt (1999), we are talking about “quality as a moving target” (p. 13).

The discussion about the goals of evaluations in science and research revolves around accountability and improvement. The relationship between the two concepts is characterized in a variety of ways. In Dutch evaluation research, accountability and improvement are seen as opposite poles, as Scylla and Charybdis (Vroeijenstijn, 1995). Another point of view (Barz, Carstensen, and Reissert, 1997) ascribes to most national systems of evaluation a connection—albeit of a different kind—between public accountability and quality development oriented to the goals and concepts of the university.

In the discussion of instruments and methods for bettering the performance of universities, quality improvement is viewed and favored in many countries as a further development of quality assessment. In the universities the discussion of quality control and quality improvement goes beyond this viewpoint, touching on a variety of instruments such as the development of long-term strategies, organizational development, management development, and the reform of leadership structures. They are based on the idea that quality requires continual improvement that is expressed in other concepts of quality, too, such as certification.

We have taken up these ideas for nonuniversity research institutes as well, seeing for research institutes an exceptional opportunity in evaluations that not only rate the status quo but also aim to develop the institute and its research tasks. This approach in no way precludes decisions that ultimately lead to the termination of research areas or programs or even the closing of the institute itself. The subsequent evaluation should be concerned with checking whether and how the institute has acted on any recommendations. On the whole, the learning ability of the institute can thereby be enhanced as an essential element in the effort to increase performance.

We believe that the most likely way to improve research quality continuously is to integrate external evaluation and self-evaluation. The procedures of the Dutch universities and the Network of Northern German Universities (*Verbund Norddeutscher Universitäten*) in the Federal Republic of Germany provide the main models for our work. Two things are clear from the debate about quality. First, accountability is associated with external evaluations, and improvement is associated with self-evaluation (Altrichter & Schratz, 1992; van Vught, 1994; Westerheijden, Brennan, and Maassen, 1994). Second, questions of ensuring and improving quality arise directly from the existence and development of self-assessment systems, in-

cluding self-evaluation systems (e.g., Kells, 1992), which are analyzed as the prerequisite for an effective use of measures intended to guarantee quality.

As already stated, we set out from a model that integrates some tasks with external evaluations and others with internal evaluations. External evaluations have to do with the evaluation of research output and services of a research institute as measured against its assigned tasks and performance record. These evaluation also have to do with a check for the existence (meaning) and design (quality) of the institute's internal mechanisms for quality control and quality improvement (procedures for self-evaluation).

If external evaluation is coupled to a policy decision on whether to continue funding the institute, three tasks are performed by external evaluations: They conduct quality control, they ensure quality, and they improve quality. *Quality control* involves two reference systems. First, the institute's outputs must be judged according to its assigned tasks, its performance of those tasks, and its self-proclaimed objectives. Second, it must be judged whether the objectives are in keeping with the general demands that science policy makes on these nonuniversity research institutes (e.g., cooperation with the universities and the development of young researchers). *Ensuring quality* means checking the functionality and scope of the instruments and procedures of internal controls with a view to externally validating their quality from the outset. They are a crucial to judging institute performance. *Quality improvement* represents an attempt to identify shortcomings from an outside perspective, to find paths to solutions, and to make recommendations for the development of the institute.

Internal evaluations are intended to help ascertain the achievement level of the institute, to analyze strengths and weaknesses, and to elaborate perspectives for future development. Institutionalizing this kind of procedure is understood primarily as a collective learning process and only secondarily as an evaluation routine. It is to be conceived of as an ongoing, not a one-time, internal process of understanding and decision-making.

In terms of promoting the learning ability of institutions in order to improve and optimize their "products" and their structural and organizational conditions, evaluations can help counter the ever-present danger of routinization and institutional "paralysis" by initiating an action-oriented discussion process and thereby increasing the capacity for self-management. It is essential to bring the institute's personnel into this process and to forge a common basic understanding about objectives, tasks, and instruments. Experience in the universities has clearly shown that evaluation is accepted if it is about a process of becoming aware of shortcomings in order to improve quality and raise efficiency and if the initiative comes from within the institution itself. This experience should be analyzed and developed for research institutes outside the universities as well.

### Criteria

In this presentation we cannot delve into an extensive discussion of qualitative and quantitative criteria. Suffice it to say that judging the “products” (e.g., publications) of science and research by qualitative criteria has long been customary in the community of scholars and researchers. Quantitative criteria, such as number of publications or successful acquisition of third-party funding, has only recently become relevant. A prominent issue in the research institutes we studied is that the wide range of their tasks, objectives, and reference groups make it impossible for their output (products) to be judged solely by academic criteria. How, for example, does one adequately judge services such as advising political actors and providing data or certain technological products for industry? (We have formulated a few proposals in this regard.)

In discussing indicators, we distinguish between input (all external influences and internal requirements for the output production process), output (the results of the production process), and throughput (the processes and structures that facilitate the conversion of the input into the desired output). We would like to conclude by examining the third category of indicators, throughput.

A research institute’s performance is greatly affected by the organizational context of research, that is, by research operations. Decisions pertaining to structures and control are very important to scientific productivity. Every research institute has the task of developing structural or procedural solutions that ensure and promote its own ability to perform. These solutions include constant quality control of the output and a self-critical handling of the institute’s own structural, organizational, and control decisions. Quality control is one of the most important elements of throughput. It is the hinge between internal and external evaluation. Optimally, external evaluation is a meta-evaluation of internal evaluation.

The evaluation of major government-funded research institutes outside the university system clearly shows that there is no organizational model optimal for all research institutes and that the organizational structures should be appropriate to the special characteristics of the various institutes. However, there is little to guide investigation into what the appropriate organizational and control structures would be in each case. Organizational sociology and the sociology of science offer few, if any, studies on what it takes for successful research. Much work remains to be done in order to provide an empirical foundation for the pressing topics in the science policy debate at this time: “specific modes of work, the strategic capacity to act, and the procedures for ensuring quality” in science and research (see *Internationale Kommission*, 1999).

The research landscape as a whole is confronted by new expectations and challenges at various levels: (a) the formulation of new research topics that respond to the social and, especially, the economic need for innovation; (b) structural changes that can produce new forms of national and international cooperation and networking, particularly between universities and other institutions of higher learning;

(c) society's changed demands to know publicly how research funds are used; and  
(d) new ways to ensure quality by introducing internal assessment and management mechanisms.

We have tried to outline an evaluation procedure predicated on the learning ability of institutes and designed to make quality improvement a continuous process of research institutes outside the universities. This model is based on the interaction of external and internal evaluations. We have emphasized self-evaluations by institutes in order to strengthen their capacity to cope independently with the convoluted interests of external actors. To acquire that ability, these institutes need a self-critical approach and decision-oriented processes of self-reflection and self-assessment that are understood as an essential part of the research process rather than as a state of emergency.

### **References**

- Altrichter, H., Schratz, M. (Hg.), *Qualität von Universitäten. Evaluation: Impulse für Innovation?*, Innsbruck 1992
- Barz, A., Carstensen, D., Reissert, R., *Lehr- und Evaluationsberichte als Instrumente zur Qualitätssicherung, Bestandsaufnahme der aktuellen Praxis*, Gütersloh 1997
- Felt, U., *Evaluation im wissenschaftspolitischen Kontext*, in: Röbbbecke/Simon (Hg.) 1999, a.a.O., S. 11 ff.
- Gibbons, M., Limoges, C., Nowotny, H., Schwartzman, S., Scott, P., Trow, M., *The New Production of Knowledge. The Dynamics of Science and Research in Contemporary Societies*, London / New Dehli 1994
- Internationale Kommission zur Systemevaluation der Deutschen Forschungsgemeinschaft und der Max-Planck-Gesellschaft, *Forschungsförderung in Deutschland*, Hannover 1999
- Kells, H.R., *Self-Regulation in Higher Education. A Multi-National Perspective on Collaborative Systems of Quality Assurance and Control*, London 1992
- Nowotny, H. (1997), *Die Dynamik der Innovation. Über die Multiplizität des Neuen*, in: *Technik und Gesellschaft, Jahrbuch 9*, Frankfurt/Main / New York, S. 33 ff.
- Röbbbecke, M., Simon, D., *Zwischen Reputation und Markt. Ziele, Verfahren und Instrumente von (Selbst)Evaluationen außeruniversitärer, öffentlicher Forschungseinrichtungen*, WZB Discussion Paper P 99 002, Berlin 1999
- Röbbbecke, M., Simon, D. (Hg.), *Qualitätsförderung durch Evaluation? Ziele, Aufgaben und Verfahren von Forschungsbewertungen im Wandel*, WZB Discussion Paper P99 - 003, Berlin 1999

van Vught, F., The Humboldtian University under Pressure. New Forms of Quality Review in Western European Higher Education, in: Maassen / van Vught (Eds.), 1994, a.a.O., S. 185 ff.

Vroeijenstijn, A. I., Improvement and Accountability: Navigating between Scylla and Charybdis. Guide for External Quality Assessment in Higher Education, London/Bristol 1995

Westerheijden, D. F., Brennan, J., Maassen, P. A. M., Changing Contexts of Quality Assessment. Recent Trends in West European Higher Education, Utrecht 1994

---

## **Discussion of Martina Röbbcke and Dagmar Simon's paper**

**Erik Arnold** (Technopolis, UK). Your paper suggests that external evaluation should be a kind of a meta-evaluation of the internal evaluations of the institutes and their projects and programs. This assumes that we are dealing with commensurable things. I suspect the idea of meta-evaluation is a bit too limited and that external evaluations may want to do some different things from internal evaluations.

**Dagmar Simon**. At the institute we have a tradition of self evaluation, and external evaluation has the task to understand the different internal evaluations.

**Carsten Dreher** (Fraunhofer ISI, Germany). Self-evaluation suggests you are not willing to put strategic thinking into the hands of someone outside. What is your reaction?

**Martina Robbecke**. It is possible to set strategic aims internally and there is a pressure to do this. Institutes are financed publicly, they have to open up their strategic aims. This should be no problem. External evaluation controls the processes of internal evaluation.

## **Benchmarking University-Industry Relationships: A User-Centered Evaluation Approach**

**Louis G. Tornatzky**

Senior Fellow, Southern Technology Council and Senior Principal Associate,  
Battelle Memorial Institute, Laguna Beach, California, USA

Email: tornwildcat2@home.com

---

### **Introduction**

For over seven years the Southern Technology Council (STC) has been fielding a novel approach to evaluating science and technology programs. This has involved the translation of industrial “benchmarking” methods, practices and metaphors to what are essentially evaluation research topics. The work has been primarily focused on the domain of university-industry technology transfer, but has also been applied to issues of interstate personnel migration, technology business incubation, and the role of universities in regional economies.

The novelty of the approach does not lie in advances of methodological sophistication. Rather the strengths of benchmarking lie in: (1) its ability to capture the enthusiastic participation of the objects of study (in this case primarily research universities); (2) the power of its metaphors and presentation tactics to engage key institutional and political stakeholders; and (3) its successes in fostering institutional change and knowledge utilization.

The paper will discuss the following: (1) the organizational, economic and political context of this body of work; (2) the general approach of organizing and executing a benchmarking research effort; including issues of measurement and results presentation; (3) examples of STC benchmarking research projects; and (4) prospects for continuing and expanding this approach to evaluation practice.

### **The Context: Economic Development in the South and the Role of the STC**

For much of the past 30 years the South<sup>3</sup> has been involved in an ambitious process of transforming its economy from one that emphasized low wage industry, a mar-

---

<sup>3</sup> There are different viewpoints on which states constitute “The South” and they revolve around issues of economics, culture, and Civil War history. At the very least, the region includes the deep south states of Alabama, Arkansas, Louisiana, Mississippi, Georgia, South Carolina, plus the more northern but historically and culturally southern states of North Carolina, Tennessee, Virginia, and Kentucky. Northern Florida is perhaps more southern in orientation than is south Florida, with the latter’s heavy influx of Cuban immigrants and northern retirees. By the same token, Texas and Oklahoma are both western and southern, Missouri can be seen as both southern and part of the

ginally skilled workforce, and traditional incentive-laden industrial recruitment, to one that is increasingly prominent in the high value-added knowledge economy. This has been a tremendous challenge, and if one examines the human resources (MDC, 2000) and technological assets of the South, it is obvious that the majority of states in the region are still lagging on many metrics defining the so-called New Economy (Atkinson et al, 1999).

Interestingly however, these same states have approached the challenge of economic development from a much more self-consciously regional perspective. A prevailing myth in the region – perhaps true - is that the U.S. South has more regional organizations of all sorts than any other part of the country. Accurate or not, the region has adopted a more regional approach to technology-based economic development. In 1972 for example, under the leadership of then-Governor Jimmy Carter of Georgia and William Winter of Mississippi, and other southern visionaries, the Southern Growth Policies Board was established to function as a forum for forward-looking economic development policies in the region. The Board established a staff function in Research Triangle Park, NC, and ever since has been prominent for its analyses of and advocacy for a more thoughtful approach to economic development. The Board has also periodically convened a Commission on the Future of the South, which pulls together regional leaders for a yearlong process of updating and reformulating regional problems and potential solutions.

One outgrowth of the influential 1986 Commission report<sup>4</sup>, *Halfway Home and a Long Way to Go*, was the creation of the Southern Technology Council (STC). Chaired by a sitting governor on an annual rotation basis, the Council has governor-appointed members from each of 15 member states<sup>5</sup> plus participation from prominent technology-based corporation, and standing representation from the Southern Legislative Conference. The members of the Council tend to have science and technology backgrounds, but are primarily oriented toward the economic development potential thereof. Correspondingly, the organization's mission has been:

To strengthen the regional economy through the more effective development of technology. It fosters cooperative initiatives among regional science and technology organizations (industry, government, and education) and functions as a forum for information and recommendations about best practices, strategies, policies, and programs.

---

Midwest and West Virginia in an interesting mix of history and orientation. Nonetheless, all of these states are currently members of the Southern Growth Policies Board and the Southern Technology Council, the organizational locus for the research described here.

<sup>4</sup> Then-Governor Bill Clinton was one of the principal architects of this report and many of the themes therein were echoed in his subsequent national agenda.

<sup>5</sup> In addition to the states mentioned in footnote 1 above, the Commonwealth of Puerto Rico is currently an STC member and Maryland has been an STC member.

Much of the STC mission, approach, and substantive agenda was originally articulated in a 1988 report, *Turning to Technology*, which among other things emphasized more effective linkages between the region's universities and its industrial community. The historic interest in better university-industry partnerships becomes obvious if one examines state-level indices on industry and academic R&D performance (National Science Foundation, 1998). In several of the states of the South, because of the dearth of technology-intensive industry, the fraction of gross state product accounted for by academic R&D rivals that of its industrial counterpart. So, in addition to being one of the major expenditure items in states' annual budgets, the region's research universities are arguably one of their more significant public assets in building a different kind of an economy.

Before turning to some of the methods and approaches of benchmarking, we can summarize the perspectives and interests of the "customers" for the STC program of research. First of all, the members of the Council, and the constituents that they represent, are oriented toward practical results and somehow "doing better" as a result of their participation. They are not interested in solving academic questions or addressing theoretical issues. Second their states are fiercely competitive in terms of economic development, and they want to know how their state universities are doing relative to peers elsewhere in the region or the country, and if they are lagging, how they can improve their performance. Third, they want results conveyed in language that is accessible to them as intelligent lay persons, and which can be translated into policy actions such as program initiatives or legislation. If a governor or key legislator cannot quickly grasp an analytic finding, it is akin to the falling tree in a people-free forest.

### **An Overview of the STC Benchmarking Approach**

The benchmarking approach to evaluating science and technology programs that has been developed and implemented by the Southern Technology Council includes the following elements: (1) recruiting a group of organizations to participate in the benchmarking program; (2) articulating, with the participation of the benchmarking group, important domains of science and technology program activity; (3) developing methodologically-defensible, but stakeholder-understandable performance metrics; (4) gathering primary and/or secondary data on those metrics, and using the information to identify exemplary of "best-in-class" organizations; (5) identifying and qualitatively describing best practices used in exemplary organizations that might be emulated by others; and (6) packaging results so as to maximize the practical understanding of performance benchmarking data, as well as to encourage the adoption of best practices among organizations in the benchmarking group; and (7) fostering the development of public policies that might enable the adoption of those practices .

### **Recruiting a Benchmarking Group**

Benchmarking differs significantly from more typical approaches to evaluation research in that it involves “doing with” rather than “doing to.” It is also multi-organizational in context; benchmarking doesn’t work unless a large fraction of the peer group of organizations involved in a science and technology program domain are involved. Comparisons among this peer group – confidential and otherwise – are what provide the motivation and value for participants.

– There are several ways to develop a benchmarking group. Perhaps the most productive approach is to persuade a professional association or affiliation group to endorse the benchmarking project (in concept at least), and have them host a briefing and discussion among representatives of potential organizational participants. In the absence of a logical organizational sponsor, the benchmarking research team will be obliged to do intensive and extensive recruiting in order to get potential participants to a briefing.

– At such a meeting, the benchmarking research strategy is presented, there may be discussion of alternative approaches to measurement and metrics, and preliminary understandings are reached about data confidentiality and reporting strategy. The latter is particularly important; what will participants get as value-added product, and how will their confidentiality be protected (see below). The STC team has organized several of these sessions, typically accompanied by wine and cheese, snacks, or a free lunch. Since not all potential organizational participants will be in attendance, it is important to succinctly capture the gist of the discussions and decisions, and convey that in a memo to attendees and other potential participants in the benchmarking research. Eventually, each organization that should be involved will need to be contacted individually by phone and letter, and persuaded to participate. This mix of recruitment procedures in the STC benchmarking work has yielded 80-90% participation rates.

– It should also be understood that even when benchmarking analyses will be conducted using existing secondary data, and no operational cooperation is really needed from a group of organizations, one should behave as if primary data were being gathered. Since benchmarking always involves reports that draw comparisons among organizational entities, it is important to get guidance and input from a smaller group of organizations who can function as a proxy for the larger population of entities. Every benchmarking research project should have at least an advisory committee composed of stakeholder organizations.

### **Agreeing on the Foci of Benchmarking**

An extremely important initial task, which needs to be addressed jointly by the research team and the benchmarking group of organizations (or an advisory group proxy), is to define what areas of organizational activity should be benchmarked in

terms of performance. Part of this involves constructing a crude approximation of a causal model of the general area of organizational activity, particularly sorting out which behaviors or activities are considered outcomes or results, versus those which are contributory or intermediate thereto. For the university-industry benchmarking research reported here, the general focus has been the technology transfer function. However, within that domain, we have considered licenses, royalties and start-up companies formed as outcomes, and invention disclosures, patent applications and patents as inputs. This fairly straightforward conceptualization was developed as a result of several rounds of discussion with institutional technology transfer officials. Underneath these foci, of course, is another layer of more diverse and institution-specific behaviors we have considered practices and policies, and which we have approached using a more qualitative approach, described below.

### **Gathering and Analyzing Performance Benchmark Data**

A second key task of benchmarking research concerns collecting and analyzing performance data on the metrics agreed to by the benchmarking group of organizations. In most cases this will involve gathering primary data supplied by those organizations, facilitated by questionnaire or interview procedures conducted by the research team. The gathering of performance data across the benchmark group of organizations can have two objectives. One is to benchmark performance *per se*, in order to develop various reports that display peer group comparisons. For this purpose we have developed several practical guidelines of how individual measures and metrics should be constructed, based on our experience to date:

- *Use Fewer Measures.* Since the research team's data collection success is dependent upon cooperation with representatives from the benchmark group of organizations, one should not lard questionnaires with "nice-to-know" items. Limit the questions to those enthusiastically supported in discussions with participating organizations.
- *Avoid Multivariate Metrics.* Since the primary customers for benchmarking studies are educated laypersons in positions of institutional, government or industry leadership, they need to understand what a performance metric means and how it was constructed. Ratio metrics are intuitively understandable, and in fact essential if one is going to draw comparisons across institutions of different sizes. However, once one starts to use multivariate scaling or similar approaches, the practitioner and policy audiences have been lost, and one is simply creating an academic study for academics. Of course, however, if the purpose of the research is primarily hypothesis testing, then considerable statistical power can be gained from multivariate procedures. In this context, some recent studies are noteworthy (Siegel et al, 2000; Thursby and Kemp, 2000).

- *Don't Ask for Data Which Doesn't Exist.* For any important domain of organizational behavior that might be performance benchmarked, there typically exists a management information system, albeit rudimentary in scope. This is what informants will mine to respond to a benchmarking questionnaire. They will not have the time and inclination to create new data elements, for a study objective that may be seen as peripheral to their work. For example, in our ongoing study of university industry technology transfer we would have liked to conduct analyses on the industry sector and firm size distribution within university license portfolios. However, when we queried potential informants, we found that only a small fraction of schools had these data at hand.
- *Go for the Intuitive Rather than the Obscure.* Underlying any data collection is an explicit or implicit conceptual schema about the phenomena being studied. These conceptualizations can be simple and intuitive or complex and obscure. Rest your measurement and metric development strategy on the former. Measures and metrics need to “make sense” to the primary audience for benchmarking research, which is composed of very practical people. Our approach is paralleled by others (e.g., Zacks, 2000) who address similar audiences.
- *The Issue of Economic Geography.* As noted above, many of the clients for regional performance benchmarking data and reports are in intense competition with one another. Therefore we have been sensitive to developing performance metrics that are intuitive proxies for local and state-based economic impact. For example, in the area of university-industry technology transfer, we have consistently gathered data on the fraction of licenses involving in-state licensees, as well as the proportion of licenses involving start-ups (since the latter are typically local, at least in the early stages of their corporate development).

A second purpose of gathering performance benchmark data, however, is more instrumental. That is, to develop defensible metrics via which one can identify “best in class” institutions in an area of performance, primarily as a precursor to a more qualitative “best practice” analysis of a smaller sample of organizations. In these cases, it may be politically difficult or logistically impossible (see “Don't Ask for Data Which Doesn't Exist” rule above) to gather primary data in a performance domain, and a fallback strategy is to use ranking or rating procedures. For example, in a study of best practices in operating technology business incubators (Tornatzky et al, 1995a) the research team developed an aggregate of ratings from a panel of national experts, and then focused data collection on the sub-sample seen as exemplary. In another study (Tornatzky, 2000) a similar procedure was used to identify a small sub-sample of universities (top 10%) that were seen as exemplary in terms of business and economic development partnerships.

### **Benchmarking Best Practices.**

Benchmarking and defining organizational “best practices” complements performance benchmarking, and as discussed above generally follows from a performance benchmarking effort. Its purpose is to better understand the ingredients of “doing better” on one or more performance measures, and as such calls for different and more qualitative approaches to data collection and analysis. It focuses on the upper end (in terms of performance metrics) of the distribution of benchmarked organizations, and tries to understand how that upper quartile or top 10% of organizations differs from their peers. Analytically then, the benchmarking challenge turns to better understanding positive relationships between specific practices/policies and performance outcomes.

- Some of this may be quite quantitative, whereby variations in certain practices can be statistically related to variations in outcomes, such as via regression analysis. For example, the staffing ratio of technology transfer offices (in terms of FTE per \$X millions of research expenditures) seems to be related to outcomes such as the rate of patenting and licensing. Nominal practice variables can also be related to performance metrics using non-parametric approaches (e.g., the presence or absence of a written intellectual property policy, and how this relates to performance indices such as royalties). Both of these kinds of practice variations can be uncovered using structured questionnaires and checklists
  
- However, relying exclusively on a quantitative approach for documenting organizational best practices is a flawed approach. Organizational practice involves a rich mix of behaviors, roles, norms and incentives. Much of this is unique to a specific institution, and represents organizational innovation. This needs to be described in plain language that is accessible to the audience for the benchmarking process, and in forms that are actionable for practitioners. For example, demonstrating that a conflict of interest policy is essential for fostering university-industry partnerships represents one useful datum; including in a report some illustrative policy statements from two or three exemplary schools represents much richer lore. Based on customer feedback, the STC benchmarking studies on best practices have increasingly migrated toward rich, case study data collection and presentation. For the most part, this qualitative data collection has involved extensive phone interviews of key informants, as site visits become prohibitively expensive given the size of typical benchmarking groups, even when one wants to gather information primarily from a smaller group of “best-in-class” organizations. We have also found it useful to record phone interviews, as nuances often are missed when merely taking notes. In addition, it has proved useful to have two or more researchers on one end of the phone, with one moving through a list of open-ended questions, and the other functioning as a second set of ears and chiming in with prods, follow-up questions, requests for clarification, and so on. Given the goal of obtaining as much operational detail as possible, we often ask for internal reports, policy

documents, and other supplemental materials from the informant. In many cases, additional detail is provided by email exchanges over several weeks.

### **Packaging and Disseminating Results**

There are several distinct types of reports that can be generated from a program of benchmarking research. One general category focuses on performance benchmarking studies; the second encompasses best practice analyses. All are oriented toward practitioners and the policy community.

Within the category of performance benchmarking reports, the STC has produced two types of products, most of these in the area of technology transfer: (1) the summary region-wide report; and (2) the institution-specific customized report. Both types share several characteristics. In terms of language and format they are very much oriented toward a non-research audience. Social science jargon is eliminated with a vengeance, as is obscure usage and vocabulary. While the body of the report usually provides a broad conceptual discussion of methodology issues and approaches - such as sampling, instrumentation, and statistical approach - technical details thereof are always inserted as endnotes. There are no exhaustive reviews of the existing literature, although a few relevant citations are briefly discussed for intellectual context. There is always a section that deals with action or policy implications, and always a concise, readable executive summary. The core characteristic of performance benchmarking reports is to present useful comparisons about how your organizational entity (e.g., state, university) is doing relative to others, and these are presented in various ways. However, in virtually all of the performance benchmarking reports that have been produced, there has also been an effort to protect the confidentiality of participating organizations, although using somewhat different approaches.

For example, the summary reports that have been produced in the area of technology transfer are typically organized around the various categories of benchmark measures: inputs; outcomes; and economic development impacts. For each metric, we present data on the distribution of scores across the whole sample, with a graphic showing the quartile breakdowns. We also always identify by name those schools that are in the top 10% of the distribution on any performance metric, and give them some plaudits in the text. However, nowhere else in the report is the name of a specific institution attached to a specific score or quartile placement. This is known within the team as the “protect the guilty” rule. However, state policy leaders and other regional constituents can get a vivid picture about how the universities in each state rank relative to others. We include a summary table for each state, in which several metrics are displayed as quartiles, and in which the quartile placement of each institution in that state is indicated by an anonymous icon (e.g., a mortar board hat). So, an interested state legislator can see at a glance, how the two or three universities in his/her state stacked up against the regional competition. Moreover, following the “protect the guilty” rule we have re-

fused to divulge which bad hats are associated with which specific institutions. We have, however, disclosed the identity of “good hats” which no has seemed to mind.

The second type of product that we have produced is dramatically different in format and distribution. This is the institution-specific “report card”, which is in effect customized for every member of the benchmarking group. It too is organized around the small number of performance metrics. However, each school is provided data on how it ranks relative to the entire sample, as well as logical peer groups. For example, if a given institution is a publicly supported, Land Grant institution with a medical school, it will be provided with additional data on how it ranks, on each metric, relative to those groupings of institutions. Obviously, this type of a report demands that the benchmarking group be of appreciable size, so that these smaller sub-group comparisons do not run out of degrees of freedom. Another approach to generating comparison groups for individual institutions is to have them self-select their peers. Most institutional leaders and planners have some conception of the 6-10 universities that are most like them or that they aspire to emulate. This approach has been used to some degree in the STC benchmarking work, but more extensively by a benchmarking initiative that is being led by the KPMG Higher Education Practice and which is looking at research administration performance and practices. That effort has gotten to the point whereby participants are able to self-generate peer comparison group data by accessing a project website (KPMG, 2000).

The differing nature of these two types of products has demanded different approaches to dissemination. The summary regional reports have been actively and widely disseminated to various stakeholders in the STC states. Within institutions they go to governing boards, president or chancellor’s offices, chief research officer, chief academic office (e.g., provost), the technology transfer officer, and key deans. These are in turn are the only individuals who receive the institution-specific report card product. In addition, the summary regional reports are sent to relevant legislative committees in every STC state, to governors’ offices, to state economic development organizations, and to a number of the larger, technology-based companies in the region. There is also an aggressive effort to organize press coverage, through news releases as well as other informal contacts with the media. Every report has the same recognizable graphics and layout, so as to achieve some benefits from branding and product differentiation. Within the past seven years the STC has disseminated over 5,000 copies of technology transfer benchmarking reports throughout the South.

### **Fostering Policy Changes**

Practices, “best “ or otherwise, are often enabled or constrained by policies, and those policies can operate at institutional or state levels of origination and application. A significant theme of the STC benchmarking program has been to foster a more robust policy debate among stakeholders in the region, hopefully leading to

improvements in policies underlying university-industry relationships. Several tactics have been employed:

**The Decision-Maker Report.** Most of the benchmarking reports in the STC series have been focused on a particular issue (e.g., licensing to startups) or encompassing a particular time frame. We have concluded that decision-makers need a larger, summary perspective with clear suggestions for needed changes. As a result, we have produced a variety of products of this nature. In the technology transfer area, two have been produced. One, primarily targeted at university leadership (Tornatzky et al, 1999), reviewed the state of practice and policies and made several suggestions for intra-institutional reform. This was disseminated to over 600 institutional leaders. A second report, produced in partnership with the National Governors' Association, Center for Best Practices, covered much of the same ground, but emphasized policy and program options at the state level (Tornatzky, 2000). This effort was also bundled with several other commissioned papers dealing with issues of the New Economy, and accompanied by regional briefings and meetings around the country.

**The Thematic Conference.** One of the more important findings in the area of technology transfer performance benchmarking was the large fraction of university technologies that were being licensed to out-of-state and out-of-region companies. Among policymakers this led to greater awareness of the problem and of the need to explore ways to regionally anchor the value-added of technology transfer and commercialization (e.g., via greater facility in doing start-ups). Facilitating this process was a thematic conference on *Keeping it Home* that was co-hosted by STC and the University of Mississippi. A number of changes in the region were influenced by the conference, particularly upgrading of technology transfer functions and a greater attention to the economic geography of technology transfer.

**Technology Program Planning and Consultation.** In terms of moving the results of benchmarking research into changes in policies and practices, STC has the advantage of having both an analytic mission as well as a technical assistance mandate. As the results of the various benchmarking studies became disseminated and visible around the region, there were increasing requests for STC participation in state-level program planning and review projects, as well as institutional reviews within several research universities. These included efforts in Alabama, Kentucky, Louisiana, Mississippi, Oklahoma, Puerto Rico, South Carolina, Tennessee, West Virginia, and Virginia. The typical product for a state-level effort was a new or revised strategy for technology-based economic development, with the universities as a more robust partner therein. The institutional products tended to incorporate more operational recommendations regarding staffing, internal policies, culture, and rewards.

**Influencing Legislation.** An interesting story unfolded in Oklahoma over a two-year period as an outgrowth of benchmarking research. Somewhat chagrined by their universities' standing in early rounds of the STC technology transfer performance benchmarking, STC staff were asked by the state technology organization<sup>6</sup> to conduct, on a contract basis, a more detailed examination of performance, policies and practices. This led to another look at the performance numbers, as well as a detailed case analysis of each of the research institutions in the state. A report was developed and the findings were hashed over in a series of focus group sessions in the state. The universities gradually began to improve the staffing and support of their technology transfer activities, but it soon became apparent that there were some significant legal obstacles to building a fully enabled technology transfer function within the state-supported institutions. There were constitutional prohibitions against the universities and/or their faculty taking an ownership position in private sector activities that somehow derived from work supported by public funds. This effectively squelched equity participation by faculty and the university in start-ups based on university-developed technologies. This, in turn, prevented the university from actively participating in one of the growth areas of technology transfer and one that ironically tended to have favorable implications for economic geography. That is, it was becoming clear on the national scene that faculty start-ups tended to stay in the community or the state. Solving this problem ended up grabbing the attention of both the Oklahoma legislature and the governor. Eventually a constitutional initiative was put on the ballot in 1998 and approved by a comfortable margin. Of interest, other benchmarking data from STC on "brain drain" (see below) entered into the campaign as an argument for supporting the constitutional changes.

### **Benchmarking University-Industry Relationships: Illustrative Studies**

STC has been involved in a number of benchmarking studies, not all of which will be reviewed here. In the area of performance benchmarking of university-industry technology transfer, four waves of analysis have been conducted. (Waugaman et al, 1994; Tornatzky et al, 1995b; Tornatzky et al, 1997; and Waugaman, 2000). As part of this review we will present a composite of this group of projects.

Several practice studies have been conducted as well, mostly deriving from the series of performance benchmarking projects. These include: an analysis of best practices in using external patent counsel (Waugaman et al, 1994); a policy and practice analysis of technology business incubators, many of which are linked to research universities (Tornatzky et al, 1995a); best practices in working with start-up companies deriving from university technology (Tornatzky et al, 1995c); culture and rewards supporting faculty work with companies (Tornatzky and Bauman, 1997); and an analysis of universities which are exemplary in terms of industry and eco-

---

<sup>6</sup> The Oklahoma Center for the Advancement of Science and Technology (OCAST).

conomic development partnerships (Tornatzky, 2001). We will describe the last as an interesting illustration of a practice-oriented study

Finally, there is a line of benchmarking research that involves both performance and practice perspectives, and also looks at the 50 states as the units of analysis. This is the STC work on interstate migration of recent science and engineering graduates of universities (Tornatzky et al, 1998) and will be the third project described in more detail.

### **Benchmarking Technology Transfer Performance**

As indicated above, four separate studies have been conducted in this series. Several procedures have been consistently applied in all of the studies, and others have evolved as the program has expanded.

First of all, following our own guidelines above, we have been very careful to enlist the active participation of institutions in the South. The sample of institutions has grown from 25 in the initial study (Waugaman et al, 1994), to over 75 in the current round of data collection (Waugaman, 2000). Every study has had a small advisory group, drawn from the ranks of technology transfer managers and chief research officers. This has been supplemented with regional caucuses held in conjunction with annual meetings of the Association of University Technology Managers, in which we have presented our plans for that year's round of data collection, and gotten helpful and sometimes brutal feedback. As a precursor to the current study, we supplemented this face-to-face feedback with a region wide survey<sup>7</sup>, which asked respondents to confirm the usefulness of measures and metrics that had been used in prior waves of data collection, and to explore their reaction to potential additions. One result of this process has been the addition of a questionnaire item asking for the fraction of industry-sponsored research that is accounted for by state-based companies.

We have also focused on simple, intuitive measures and metrics and stayed with them over the years. These have included ratio indicators of patent applications, patents granted, licenses, royalties, and start-ups, all of which have been normalized in terms of the size of the institutions research expenditures (e.g., patents per \$10 million of research). A great advantage of this continuity of measurement, along with maintaining a high participation rate among institutions, has been our ability to report longitudinal trends on the measures. We have observed a trend among other researchers and commentators to rely more on ratio measures as well, which obviously facilitates comparisons among institutions of different size.

Third, given the nature of the STC constituency, we have also consistently collected data on state and regional impacts of licensing and technology transfer (actually a proxy thereof), including the fraction of licenses, start-ups, and industry-sponsored

---

<sup>7</sup> Of 250 survey questionnaires sent out, responses were received from 126 individuals.

research involving state-based companies. This has been of keen interest to both institutional and policy audiences, particularly when it became apparent early on that a large fraction (e.g., 80-90%, depending on the metric) of university technology from the South was being licensed to out-of-state and/or out-of-region licensees. Interestingly, STC is the only organization that has consistently gathered data on the economic geography of technology transfer, an issue that has been consistently ignored by national associations in the field as well as federal agencies.

Over the years, the series has gotten better and more professional in packaging results and reports. As noted elsewhere, a distinctive and recognizable format and associated graphics has been developed, which has served to grab the attention of stakeholders. The institution-specific report cards have been extremely popular at that level, particularly when the results indicate that the home state university is looking good indeed. In the more public, region-wide summary reports, the identification of “best-in-class” universities on each metric has been generally well received by both institutional and policy readers. We have typically been asked to produce additional copies of the reports for those institutions that are performing well on a number of indicators.

### **Best Practice Study of University Linkages to State Economies**

As indicated above, the STC has been involved in several practice benchmarking studies involving university linkages to the external business community. However, virtually all of these have been narrowly on a particular domain of programs, policies, and practice. Moreover, as our practice analyses accumulated, the team noticed that a small number of institutions turned up again and again as sources for novel practices and program models. This suggested that a useful analytic exercise might to characterize this small number of institutions, and describe their activities across a broad swath of industry partnerships and involvement in state economic development. In effect, develop case descriptions of those institutions that are doing most things “right” in terms of external partnerships.

Not surprisingly, in attempting to identify this hypothetical class of exemplary institutions, we found that there were no quantitative, objective metrics that one might use to develop a national ranking. As a result, a “reputation” approach was used instead. Fifty-five alleged experts in the areas of technology-based economic development, regional economics, and the process of innovation were contacted and asked to be judges. The forty individuals who ultimately agreed to participate were each given a list of 164 universities<sup>8</sup>, along with a two-page handout that briefly described 10 domains of activity that were of interest to the team. These included activities as disparate as manufacturing extension, technology transfer, explicit uni-

---

<sup>8</sup> The list included the top 150 institutions in terms of total research expenditures data as published by NSF, plus other universities from either STC or EPSCoR states that were typically “flagship” institutions locally, albeit modest in size by national standards.

versity mission statements regarding economic development, faculty rewards and culture, involvement in entrepreneurial development, and so on. The judges were asked to check off which institutions they felt were “actively and successfully participating in, or linked to, state and local economic development”. In turn, the research team simply compiled the votes, which resulted in a fairly crude approximation of a best practitioner national ranking.

The top 10% of institutions were identified as candidate “best practitioners.” The list of 16 schools includes: Georgia Tech, MIT, NC State, Penn State, Stanford, University of California-San Diego, University of Texas at Austin, Carnegie-Mellon, Purdue, University of Alabama at Birmingham, Wisconsin, Ohio State, Texas A&M, UNC-Chapel Hill, University of Utah, and Virginia Tech.

Data collection has involved a fairly tortuous process, which as of this writing is still not complete. A letter is sent to the president or chancellor of the institution, with copies to chief academic office, chief research officer, and manager of the technology transfer function. It briefly describes the research project, and asks the institution to identify a prime point of contact or liaison for data collection. Follow-up calls eventually end up with the liaison, who in turn is asked to identify upwards of a dozen “domain champions” corresponding to the activities mentioned above. Typically, the contact person is asked to identify the one or two people on campus who have the deepest knowledge about that particular area. In subsequent calls or emails to these domain champions, the project is again described, and the informants are asked to respond to a set of open-ended requests for information. Their response is typically some mix of mailed materials about their program, planning or evaluation reports, and emailed responses.

These responses are then aggregated into a draft chapter on the institution, which is then sent back to all of the informants with a number of queries and requests for additional information embedded in the text. Eventually, after a 3-4 month period, a draft chapter has been pulled together which meets everyone’s expectations regarding coverage and accuracy.

While this data collection is still underway, several illustrative findings might be shared, as follows:

- These highly “linked” institutions tend to have explicit mission language, at both institutional and unit levels, that encourages and champions relations with industry and contributing to state economic development;
- The institutions tend to grasp the economic geography issues that pertain to technology transfer, and tend to devote more program resources to fostering an entrepreneurial approach to commercializing technology (e.g., operating incubators, organizing local seed funds);

- Several of the institutions use a structure of branch campuses and/or out-reach offices to connect with local industry and economic development activities;
- Contrary to national trends on how federally-supported manufacturing extension programs are operated, these institutions have been in the industrial extension business for a long time, and intend to remain so;
- These institutions see external linkages as complementing traditional goals of research and teaching, rather than conflicting;
- Several institutions have highly elaborate systems to connect undergraduate students with local industry, such as via internships and/or co-op programs.

### **Benchmarking Interstate Migration of Recent Science and Engineering Graduates**

The impetus for this study (Tornatzky et al, 1998) came from the acute concern among members of the Southern Technology Council about problems of technology workforce shortages in the region, as well as an ongoing sensitivity to the economic geography of technology transfer. There was concern that too many young talented people from the South were leaving the region, and not necessarily being replaced by migrants from elsewhere. This led to a project, as well as a continuing area of research for STC. The project ended up becoming a mix of performance benchmarking at the state level (across the country), coupled to some early attempts to uncover policy and economic predictors, also at the state level.

When the project was launched in late-1997, the team's original data collection strategy was to contact alumni organizations in each of the major universities in the South, and try to extract information about the post-graduation locations of alumni. We quickly discovered that these kinds of data are not routinely maintained by alumni organizations, and moreover there are irreconcilable differences across institutions about how any of these types of data are gathered.

We quickly abandoned our plan of gathering primary data from institutions.

As a fallback strategy, the team explored various National Science Foundation databases, and struck upon their *National Survey of Recent College Graduates* (NSRCG). This survey is commissioned by NSF on a two to three year cycle, and is based on a national probability sample in excess of 25000 graduates. Information is gathered from graduates in several majors, organized broadly into computer and mathematical sciences, life sciences, physical sciences, social sciences and engineering. The survey itself gathers extensive information about current employment, subsequent schooling, personal background, and demographics, and questions are posed toward a "target week" one or two years after receiving their degrees. Of most relevance to the STC project, there is also data on the name and location of

their college, as well as where they are currently living and/or working. Access to the database was secured via a license from NSF to STC.

The data were first re-organized by the research team in terms of state-level performance benchmarking indices of student retention and net migration. These were computed for all of the states.<sup>9</sup> The performance benchmark metrics for retention were based on the fraction of graduates (with BS or MS degrees) who were working in the home state of the institution where they received their degree, one to two years hence<sup>10</sup>. The second metric – net migration – also added into the numerator of the computation those who came to state X from another state. In effect, the two types of performance metrics constituted either “brain drain” (losing graduates) or “brain suck” (keeping most of your own, but attracting talent from other states).

In the report itself, these complexities were simplified for the audience. Four two-color maps of the country were provided, each representing a different slice of the retention or net migration picture. Two shades of blue represented the top two quartiles of scores; two shades of red represented the bottom two quartiles of scores. Governors, legislators, and the media could see at a glance the status of a particular state. It was obvious to all that some states were net exporters of their best and brightest (e.g., Maine, Oklahoma, New Mexico, Iowa), while others were increasing their stake in the future economy by both retaining and attracting the same vital commodity (e.g., California, Texas). As one indicator of both the timelines of this benchmarking topic, as well as the understandability of our report, the authors have been interviewed nearly 40 times by regional and national media since the report was issued. As mentioned above in the discussion of the Oklahoma constitutional reform case, the data seems to be a powerful ingredient in policy discussions. People care about brain drain.

In addition to this fairly rudimentary state-level benchmarking analysis, the project also used multiple regression techniques to sort out a number of state-level variables that might account for the state differences in retention and migration performance. Those that had predictive power included several “new economy” proxies such as technology wage rates, as well as public university tuition levels. The latter was a negative predictor, such that lower tuition rates seem to attract graduates from elsewhere. This was interesting given recent discussions about quality of life as an important component of attracting high tech companies and talented people. Perhaps the most important predictor was whether or not an individual left the state at all after high school. The data indicate that once left, people are not likely to return home for graduate training and subsequent employment.

---

<sup>9</sup> Since the NSRCG was organized as a national probability sample, computation of state-level metrics created some problems. In a few states, the sample of observations was quite small, and for some analyses these states were discarded.

<sup>10</sup> The actual computation was based only on those former students who were working, as opposed to being involved in graduate training or unemployed.

A follow-on study is currently underway, supported by NSF and using a more recent version of the NSRCG database. It will be looking at the data from both individual and institutional levels of analysis. The former will consider issues of academic major, grades, gender, family status, community “connectedness” and socioeconomic background. The latter will consider whether different types of institutions (e.g., Land Grant, private) produce distinctive rates and patterns of retention. Ultimately, this analysis might lead to a more qualitative examination of institutional policies and practices contributing to in-state employment (e.g., co-op program, internships, state-focused career fairs). This will complement some of the findings discussed regarding our case studies of externally linked universities.

### **Prospects for Expanding the Use of a Benchmarking Approach**

Our experience is that a benchmarking approach to assessing program performance in science and technology programs, and potential best practices for adoption and improvement, has great appeal to users of and customers for evaluation research. There are a number of factors that might facilitate the use of this approach, as follows:

- When the science and technology program or activity being evaluated is relatively common among comparable organizations and institutions;
- When national organizations or federal agencies have routinely collected data which might easily be adapted to a benchmarking approach;
- When a regional organization (e.g., such as the Southern Technology Council) has popularized the use of benchmarking, and users are familiar with its methods and metaphors;
- When the organization that is the locus of the activity being evaluated can point to natural peers;
- When a major goal of the evaluation effort is to find ways (e.g., practices) to improve program performance;
- When there is a body of best practice analyses or equivalent studies, which might complement performance benchmarking;
- When the audience for the evaluation is heavily drawn from the private sector, in which benchmarking and continuous quality improvement are an integral part of culture and operations.
- Correspondingly there are some circumstances which would not be conducive to a benchmarking approach. They include:
- When the evaluation is of a truly unique program or activity;

## 8 Policy, Institutional and Portfolio Evaluations

- When there are no logical organizational peers, which might comprise a benchmarking group, ad hoc or otherwise;
- When the concepts and metaphors are counter-normative among the group of users and customers for the evaluation;
- When the primary goal of the evaluation effort is conduct a summative assessment.

Nonetheless, we believe that a benchmarking approach has great merit within the toolkit of the evaluation researcher. Much of the approach involves casting well-worn methods and procedures into a new frame of reference, using a somewhat different set of metaphors, and in particular focusing on the users of evaluation research and their need to induce productive change within the organization being evaluated.

### References

Atkinson, R., Court, R., and Ward, J. *The State New Economy Index*. Washington, D.C.: Progressive Policy Institute, 1999.

KPMG Consulting, Higher Education Practice. For an overview of the Sponsored Programs Benchmarking Project, go to the web site:  
[www.higheredbenchmarking.com](http://www.higheredbenchmarking.com).

MDC. *The State of the South 2000*, Chapel Hill, NC: MDC Inc., 2000.

National Science Foundation, Division of Science Resources Studies (SRS), *Science and Engineering State Profiles: 1998*. Available only in electronic version at <http://www.nsf.gov/sbe/srs/nsf99311/pdfstart.htm>

Siegel, D., Waldman, D., and A. Link, 2000, "Assessing the Impact of Organizational Practices on the Productivity of University Technology Transfer Offices: An Exploratory Study", NBER Working Paper #7256

Thursby, J. G. and S.Kemp, 2000, "Growth and Productive Efficiency of University Intellectual Property Licensing", Research Policy, in press.

Tornatzky, L. (Ed.) *Making the Future: Universities, their States and the New Economy*. Research Triangle Park, NC: Southern Growth Policies Board, 2001.

Tornatzky, L. *Building State Economies by Promoting University-Industry Technology Transfer*. Washington, D.C.: National Governors' Association, Center for Best Practices, 2000.

Tornatzky, L.G. and Bauman, J., 1997, *Outlaws or Heroes? Issues of Faculty Rewards, Organizational Culture, and University-Industry Technology Transfer*. Research Triangle Park, NC: Southern Growth Policies Board.

*8 Policy, Institutional and Portfolio Evaluations*

Tornatzky, L., Batts, Y., McCrea, N., Lewis, M. and L. Quittman, 1995a, *The Art and Craft of Technology Business Incubation*, Research Triangle Park, NC: Southern Growth Policies Board.

Tornatzky, L.G., Gray, D., Tarant, S.A., and J.E. Howe, 1998, *Where have all the Students Gone? Interstate Migration of Recent Science and Engineering Graduates*. Research Triangle Park, NC: Southern Growth Policies Board.

Tornatzky, L.G., Waugaman, P.G., and J. Bauman, 1997, *Benchmarking University-Industry Technology Transfer in the South: 1995-1996 Data*, Research Triangle Park, NC: Southern Growth Policies Board.

Tornatzky, L.G., Waugaman, P.G., and L.Casson, 1995b, *Benchmarking University-Industry Technology Transfer in the South: 1993-1994 Data*, Research Triangle Park, NC: Southern Growth Policies Board.

Tornatzky, L.G., Waugaman, P.G., Casson, L., Crowell, S., Spahr, C., and F. Wong, 1995c, *Benchmarking Best Practices For University-Industry Technology Transfer: Working with Start-up Companies*, Research Triangle Park, NC: Southern Growth Policies Board.

Tornatzky, L. G., Waugaman, P. G., and D. Gray, 1999, *Industry-University Technology Transfer: Models of Alternative Practice, Policy and Program*, Research Triangle Park, NC: Southern Growth Policies Board, 1999.

Waugaman, P.G., *Benchmarking University-Industry Technology Transfer: FY 1998 Data*, Research Triangle Park, NC: Southern Growth Policies Board, 2000.

Waugaman, P.G., Tornatzky, L.G. and B.S. Vickery, 1994, *Best Practices for University-Industry Technology Transfer: Working With External Patent Counsel*, Research Triangle Park, NC: Southern Growth Policies Board.

Zacks, R., 2000, "The TR University Research Scorecard", *Technology Review*, July/August, pp. 88-90.

---

## Quality Improvement and the Evolution of Academic Research Systems: Five Hypotheses on the Evaluation of University Research

**David F.J. Campbell**

Institute for Advanced Studies, Vienna, Austria

Email: David.Campbell@ihs.ac.at

---

### Introduction

Contemporary advanced societies commonly are described as “knowledge-based”, implying that knowledge, know-how, and expertise should be regarded as important factors that determine to a large extent economic performance and economic competitiveness (European Commission, 1997; IMD, 1996, 12). Crucial in such an understanding is the conviction that research or R&D (research and experimental development) should be regarded as core processes that are responsible for knowledge production. Furthermore, within the context of national (or supranational) research and innovation systems again academic research plays a pivotal role. The term academic research addresses university as well as university-related research. Referring to standardized OECD terminology, university research coincides with R&D that is performed by the higher education sector; and university-related research coincides with R&D being performed by the government and private non-profit sectors (thus “university-related” research can be regarded as a terminological equivalent to the German-speaking term of “ausseruniversitäre Forschung” — see BMBF, 1998, 14). Academic research clearly represents a sciences-based and sciences-induced activity, where a major emphasis is placed on basic research and on the combination of basic and applied research. Since academic institutions and universities, in particular, also conduct tertiary teaching and education, academic research is closely associated with the development and build-up of highly qualified human capital (OECD, 1998a).

Focusing on current R&D funding trends of the advanced OECD countries, the following conclusions can be drawn (Felderer & Campbell, 1994; OECD, 1998b, and OECD, 2000a): First, when R&D expenditure is calculated as a percentage share of GDP (Gross Domestic Product) — the so-called “national research quotas” —, then R&D expenditure only grew moderately during the second half of the 1990s (and even “declined” in the first half of the 1990s). Second, when R&D expenditure is calculated in constant \$ — in 1990 prices and purchasing power parities — per a population of 100,000, then in principle the same trend results: after a stagnation at the beginning of the 1990s, an increase in the second half of the 1990s (see Figure 3, OECD 1999b, and Campbell, 2000, 132). Third, when R&D expenditure of the universities (the higher education sector), for the OECD average, is also expressed as a percentage share of the GDP, then university research developed stable, with a slight increase (see Figure 1 and OECD, 2000a): would university R&D expendi-

ture be recalculated in constant \$, then probably even a real increase will show up, since there was a real GDP growth for most of the OECD economies during the 1990s (OECD, 1999a, 50).

Taking into account that the primary competence of university research focuses on basic research, often with a temporal mid-term and long-term perspective (OECD 1998c and 1999b), this implies the following conclusions: (a) The importance of academic (basic) research is viable for advanced societies. (b) Universities and university research continue to play an important role for the national (and supranational) research and innovation systems; thus, the concept of the knowledge-based societies is empirically substantially supported, also by referring to the pivotal role of the sciences within such processes (the universities primarily carry out sciences-based R&D). (c) In contrast to national and business R&D, the university R&D behaves “non-cyclically” and does not interfere that extensively with the economic growth cycles (see again Figure 1). Thus university research should be considered as an “element” of stability for the research and innovation systems (Campbell, 2000, 135).

The moderate growth of university and national R&D refers to the necessity of substantially increasing the output — and particularly the output quality — of research, while at the same time there is only a limited input increase. This implies that the “how” question of R&D funding expresses the same importance as the “how much” question of “quantitative” funding. This also implies that a greater attention should be devoted to the structures and functionality of research-performing institutions within the context of the national systems of research and innovation. Furthermore, a more emphasized “paralleling” of basic and applied research is also regarded as a key issue: within the (academic or business) institutions; and through an amplified linkage of academic and business institutions, on the other hand. For the policy orientation of governments a stressed allocation of public resources to research and university research should rise on the policy agenda: stimulating and supporting research through public funds may be qualified as a field of public policy activity, which should be regarded as pivotal for governments in addition to welfare policy or public economic policy. In that respect it should also be kept in mind that R&D expenditure, in constant \$ and per specific population units, was always higher in the U.S. than in Japan and the EU (see again Figure 3). This might also serve as one explanation for the success of the U.S. economy during the 1990s (called the “New Economy”) (Campbell, 1999, 369; Campbell, 2000, 135, 140-142).

Evaluations commonly are regarded as means and instruments that address such organizational issues of academia (Campbell & Felderer, 1997; 1999, 5-6; Kuhlmann, 1998). A core definition for the evaluation of university research could be as follows (Campbell, 1999, 369): *interpreting or judging the quality, efficiency, relevance, viability, and effectiveness of university (and university-related) research*. Evaluations help to optimize academic institutions; evaluations reinforce principles of accountability for academic institutions; and evaluations emphasize the application of explicit and rational criteria for decision-making and policy-

making. In summary, evaluations aim at improving national (and supranational) academic (university and university-related) research systems, so that they can deal “effectively” with new challenges that arise during the transformation process of industrialized countries towards advanced information societies.

In the following, five hypotheses are proposed for discussion. These hypotheses point at crucial issues in reference to the proper application of academic (university and university-related) research evaluation and the further development and improvement of evaluations.

**1. First Hypothesis on University Research Evaluation: peer review and indicator as the two basic methodic principles for evaluation and the conceptual multidimensionality or research quality**

In most OECD countries, university research is primarily public funded. In Europe, again, the basic public funding (called GUF, General University Funds) expresses an important role (see Figure 2 and OECD, 1999b). This GUF funding-component should fulfill the following functions: support basic research (also with a long-term perspective); enable researchers to perform “blue sky” and “curiosity-driven” research activities; hopefully supports the linkage between research and teaching (Campbell & Felderer, 1997, 56-57); is considered as an important “cultural element” for academic “intellectual freedom”. In contrast to GUF funding is the earmarked funded university research, that means university research, which is organized in the context of research projects or research programs. Earmarked funding aims at: “ex ante” quality controls; supports comparisons between planned and actual research outcome; might encourage aspects of “relevance”.

With regard to evaluation there are significant differences between those two funding modes: earmarked funded university research is always “ex ante” evaluated (during the application process), and often also “ex post” (and parallel) evaluated by the funders of that research. Public basic funded (GUF) university research, however, is mainly receptive for “ex post” evaluations. Thus it can be derived that, as a final consequence, GUF funded university research demands the implementation of *comprehensive institutional “ex post” evaluations*.

Concerning the methodic approaches for university research evaluation, the following two “binary” options emerge: *peer review*, which is judgment based on expert opinion; and *indicators*, which is judgment that is based on “quantitative” data or information. Each of those approaches expresses its immanent and specific strengths and weaknesses, which can be captured by the following standardized typology (see Scheme 1 and Campbell, 1999, 374):

1. *Peer review strength — complexity*: The information, that peer review or expert panels take into account, is extended and more comprehensive than indicator-based information. Also experts can conduct an analysis of a much higher complexity than indicator systems.

2. *Peer review weakness — subjectivity*: One main problem of peer review refers to the possibility that peer judgment may be biased by the specific composition of peer panels. Thus it is pivotal to find balancing criteria for the selection of peers.
3. *Indicator strength — objectivity*: The crucial demand, put on indicators, is that they reflect “data” or information that can be measured, that means counted. Thus an important criterion of “inter-subjective” validation is fulfilled.
4. *Indicator weakness — superficiality*: How do we know that the information, which is measured, also is the important or relevant information? In addition, there also might occur the problem of actually validating and confirming the “quantitative” data information — this is important for policy procedures.

In practical policy terms, peer review and indicators-systems mostly will be combined. Still, it can be observed that, for instance in the context of “ex post” evaluations, the peer review often represents the dominant approach, which is supplemented by indicator information. The UK (United Kingdom) and the Netherlands developed comprehensive institutional “ex post” evaluation systems of their university research, and those evaluation systems and evaluation outcomes are determined by peer review (HEFCs, 1995, 1999; VSNU, 1994, 1998). One explanation for this may be that we are still more willing to believe in expert judgment than in “pure” indicator systems.

With regard to the conceptual modeling of university research quality, in the UK only one quality dimension is applied, which simply is addressed as “quality”. In the Netherlands, however, four dimensions are taken into account: “quality”; “productivity”; “relevance”; and “viability” (see Scheme 2). Therefore, in the Netherlands an explicitly multi-dimensional quality approach is used. In the UK this multi-dimensionality is implicitly embedded. One reason, why there is only “one” British quality dimension, may be explained by the fact that earmarked research funding is much more dominant in the UK than in the Netherlands: while earmarked funding already refers to “relevance” criteria, the basic funding — of UK university research — should support more clearly “classical” quality concepts. Derived from a multi-dimensional quality modeling also additional dimensions could be added on a “higher”, “advanced” or “meta”-level: for example the *effectiveness*, when there is a desire to judge how “effective” university research is. In such an understanding the effectiveness would resemble an aggregated (or second-stage) “factor”, based on the other quality dimensions (see again Scheme 2).

## **2. Second Hypothesis on University Research Evaluation: combining evaluations of research with evaluations of teaching and education**

In the past, a well-defined division of competency existed that functionally separated the universities from the university-related (“*ausseruniversitäre*”) institutions; at least within the European or Continental European context. Universities per-

formed disciplinary-based scientific research, which was mainly basic research; and universities were responsible for conducting tertiary teaching and tertiary education. University-related institutions, in contrast, expressed only a minor competence in tertiary teaching; the research profile of university-related institutions, however, was more application-oriented and devoted to a practical and interdisciplinary problem-solving. During the next years we may witness a gradually increasing functional overlapping between university and university-related institutions. Universities must place a greater emphasis on interdisciplinarity and application-oriented research. University-related institutions, on the other hand, will engage more actively in tertiary teaching and education, implying that university-related organizations simultaneously will perform research and teaching. As a final consequence a growing number of academic institutions will do both, research and teaching, leading also to an intensification of cooperation between university and university-related institutions. Parallel to the functional overlapping between universities and university-related institutions one can also expect the institutional diversification process within the university and university-related sectors to progress (Campbell & Felderer, 1999, 22-23).

Once having accepted that evaluations are necessary, *the emerging dual performance mode of research and teaching of academic institutions creates a demand for combining the evaluation of research with evaluations of teaching (and education), since the achievements of academic institutions should be reflected in their functional complexity.* Concerning the evaluation of teaching (or of educational programs), it can be convincingly argued that the methodic “binary” dichotomy of peer review and indicators — which applies to research — also refers to teaching. However, because of complexity reasons the practical evaluation procedures for research and teaching (education) mostly are separate and distinct. Thus it represents a major challenge to design systematic interfaces and linkages between research and teaching (education) evaluations, for the purpose of comprehensively reflecting the performance of academic institutions. This also refers to the question of how to use collaboratively results of research and teaching (education) evaluations for a decision-making focus on structural reform and the allocation of resources.

### **3. Third Hypothesis on University Research Evaluation: reconciling the differences of cultural pluralism versus the functional needs of advanced societies**

Evaluation systems of academic research must be open to and receptive for an immanent learning process. One possibility for fostering such a learning process is carefully to compare evaluation systems of different countries, since this enables one to derive or to extract those international examples that successfully can be implemented within the domestic system. One question to be answered, of course, is how to interpret differences in evaluation systems across countries. Clearly, several answers are possible. One approach would be to stress cultural differences: evaluation systems of academic research are different so that they can reflect cultural differences. An alternative approach is to emphasize different functional needs of society, depending on its stages of development: consequently, evaluation systems are

different because societies — or specific sectors or subsystems of society — have progressed further or are still lagging behind. Within such a conceptualization differences in evaluation systems reflect maturity differences of academic research systems and also of academic research evaluation.

With regard to the implementation and application of academic research evaluation, *it appears necessary to refer to cultural pluralism as well as to the functional needs of advanced societies, and to reconcile both for the purpose of developing proper evaluation strategies.* There are different national cultures and also different national academic cultures, and they must be recognized by particular evaluation systems. On the other hand, one should be careful not to regress all differences to cultural differences and by this to exhaust culture as a “supervariable” that should explain everything, since this would lead to an undecidable “cultural relativism” (particularly in a multi-cultural and multi-lingual setting, as represented by Europe, this argument is important). Functional demands of advancing and of advanced societies should be taken seriously.

If a standardized comparative typology of European university systems is attempted, with regard to the application of institutional “ex post” evaluations of university research, then a *type A country cluster* may be defined that contains the UK and the Netherlands. Their evaluation systems of university research apply a “systemic and consistent approach”. Germany and Austria resemble a *type B country cluster*, using a “pluralized and situational approach” for their university evaluations. Finland and Switzerland may also be assigned to the B cluster — at least for the first half of the 1990s (see Scheme 3 and Campbell, 1998). When the differences between A and B are interpreted with regard to “functional needs of advanced societies”, this might imply a conversion from B to A. However, if a “cultural pluralism” thesis is assumed, then those differences in evaluation system might continue. Obviously, also a mutual conversion between A and B would be possible, resulting in a “mutual overlapping” (see Scheme 4).

#### **4. Fourth Hypothesis on University Research Evaluation: emphasizing the linkages between academic research and the societal environment**

Derived from the observation that the input of R&D funding and also of academic research funding increased only moderately during the 1990s, we already concluded that the “how” challenge of academic research gains in importance. A team of researchers, under the guidance of Michael Gibbons, attempted to define those principles that determine advanced knowledge production and, consequently, successful academic research. They arrive at a categorization of five key principles, which they summarize under the term “Mode 2” (Gibbons *et al.*, 1994): knowledge produced in the context of application; transdisciplinarity; heterogeneity and organizational diversity; social accountability and reflexivity; and quality control. Two core ideas of those principles are: first, that the sustained development of basic research demands the permanent reference of an application-oriented context. Second, research locations diffuse massively across society and many research locations emphasize the

concept of research networks. The shortening “time horizons” of (business) innovation cycles and of the applicability of information create the following paradoxical situation: on the one hand, there is a demand for rapidly channeling research results into applications (since time lags outdate the usability of information). On the other hand, research with a bias towards the application end can only improve, but not innovate, and it is primarily basic research that emphasizes the long-term perspectives that are so crucial for knowledge-based societies. This increasing importance of basic research clearly upgrades the significance of university and of academic research and also helps explaining, why university and university-related institutions express such a pivotal functionality for advanced societies.

Therefore, the crucial message seems to be *that linkages should be emphasized that connect academic (university and university-related) research with society or the “societal environment”*, when society is interpreted as the overall context that embodies the national academic research system. Three issues appear to be important (Campbell, 2000, 141). First, basic and applied research ought to be designed as parallel processes that are connected through interactive linkages (the old concept understood basic and applied research as two processes that were coupled through a sequential “first and then” relationship). Second, networks should be regarded as a structural type of research that will play an increasingly important role for the national research and innovation systems. Since networks connect different research organizations across sectors, that means university and university-related institutions with companies, academic research results can diffuse instantly and focus more precisely on potential user groups. As a by-product of such networking, multiple cross-fertilizations between basic and applied and between disciplinary and interdisciplinary (transdisciplinary) research are also supported. Of course, there is also the challenge of reconciling these network processes with the principles of a “fair competition”. Third, in the context of academic research evaluations the “relevance of research” should be treated as a crucial dimension. As already mentioned earlier, quite often four dimensions are considered as appropriate for evaluations: quality, efficiency, relevance, and viability. Effectiveness, a possible second-stage “meta-dimension”, can be modeled upon the other four dimensions (see again Scheme 2). *One possibility for operationalizing or indicating relevance is to assess the linkage patterns of academic research with its societal environment.* When “relevance” is regarded as a crucial dimension for evaluating university research, this can create the amplifying (and also intended) effect that university research communities deal more deliberately with relevance issues.

##### **5. Fifth Hypothesis on University Research Evaluation: reinforcing the evolution of academic research systems by applying comprehensive evaluations**

Academic research and academic research systems must change and demonstrate flexibility, so that they can adapt reactively and proactively to new conditions and demands that arise when societies develop and progress. The research intensity of advanced knowledge-based societies demands processes of a permanent remodeling of university and university-related institutions. We already mentioned the chal-

lenge of interactively linking basic and applied research in a “parallel” design. *Thus there operates an evolution of research systems and of academic research systems* that can be observed empirically and which also must be demanded normatively, if academic research focuses on sustaining its importance and competitiveness. One pivotal question, of course, is: How should the relationship of evolution and evaluation – within an academic context – be conceptualized? In that respect there are two crucial ideas. First, *the evolution of academic research systems is reinforced and supported by the comprehensive application of evaluations*. Evaluations should be regarded as a proper instrument for supporting the evolutionary adaptation processes of academic research systems, because evaluations generate a systematic feedback on the quality and efficiency — and also the relevance and viability (and effectiveness) — of academic research. Second, *evaluations are exposed to an evolution of evaluation*. Since academic research systems change, obviously also the evaluations themselves, that assess academic research, must change: evaluations, therefore, ought to demonstrate an openness for self-reflexive learning. Such a permanently ongoing “meta-evaluation” of evaluations may be conceptualized as an evolutionary process, emphasizing cross-referential interactions between the evolution of academic research systems and the evolution of evaluation systems of academic research. This still leaves the question unanswered, whether or not evaluations should be regarded as internal or external components or structural elements of academic research systems (both conceptualizations are possible). Therefore, we may conclude: *evaluation processes, that express a receptiveness for an evolution of evaluations, markedly support the evolution — and the evolutionary capabilities — of academic research systems*. This might point at a co-evolution of academic research and academic research evaluation.

### **Acknowledgment**

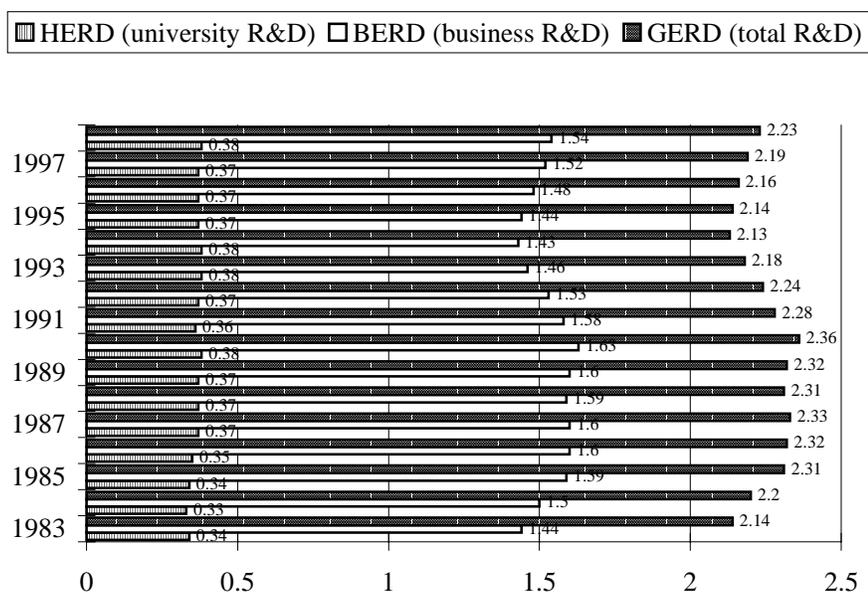
The idea of fostering the evolution of academic research systems by applying evaluations (“evolution through evaluation”) was brilliantly elaborated by Dr. Wilhelm Krull, Secretary General of the Volkswagen Foundation, in the context of a highly stimulating conversation in Berlin on the evening of June 7, 1999, during the conference “Evaluation of Science and Technology in the New Europe”.

### **References**

- BMBF* – Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie (1998). Faktenbericht 1998 zum Bundesbericht Forschung. BMBF: Bonn.
- Campbell*, David F.J. (1998). Evaluation von universitärer Forschung im europäischen Vergleich. *BUKO-Info* (4), 14-17 [bundeskonzferenz@buko.at].
- Campbell*, David F.J. (1999). Evaluation universitärer Forschung. Entwicklungstrends und neue Strategiemuster für wissenschaftsbasierte Gesellschaften. *SWS-Rundschau* 39 (4), 363-383 [swsrs@aon.at].

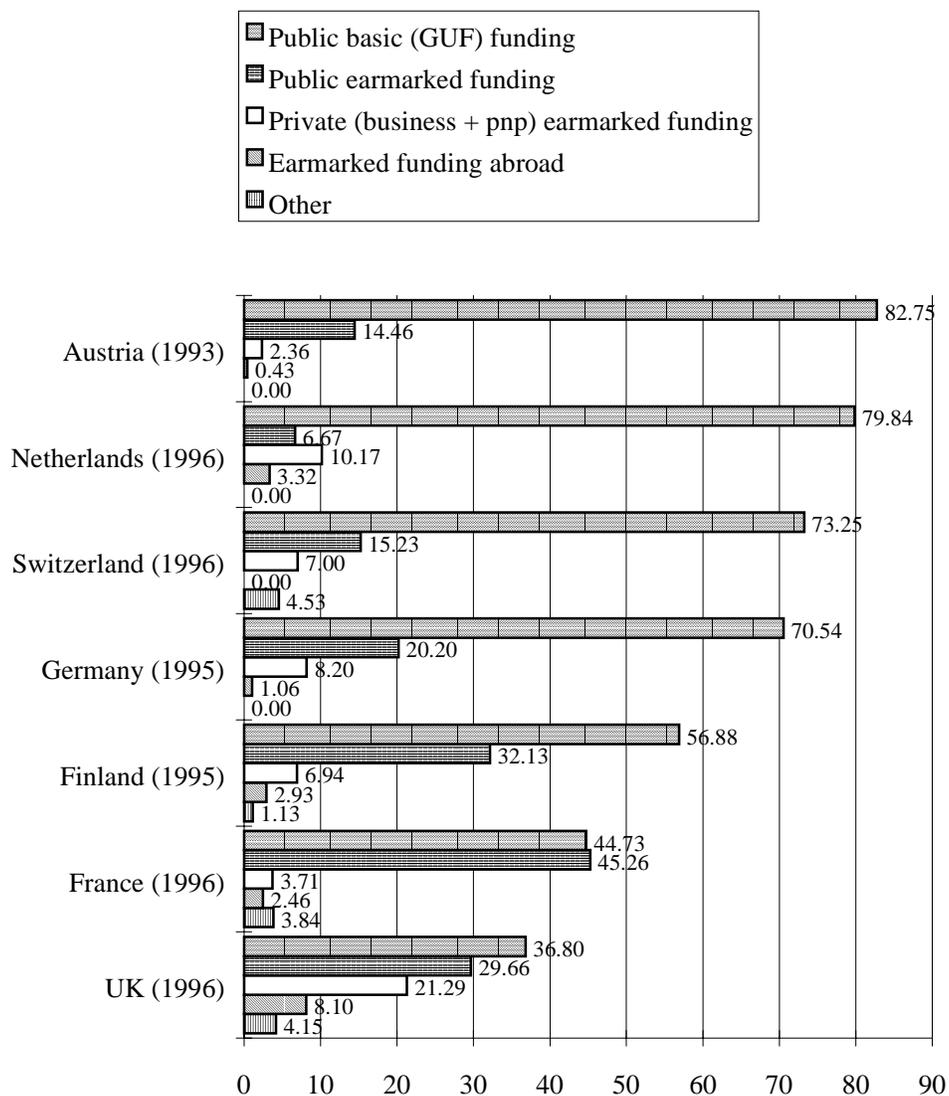
- Campbell*, David F.J. (2000). Forschungspolitische Trends in wissenschaftsbasierten Gesellschaften. Strategiemuster für entwickelte Wirtschaftssysteme. *Wirtschaftspolitische Blätter* 47 (2), 130-143 [<http://www.wk.or.at>].
- Campbell*, David F.J. / Bernhard *Felderer* (1997). Evaluating academic research in Germany. Patterns and policies. Vienna (Institute for Advanced Studies): Political Science Series No. 48 [[library@ihs.ac.at](mailto:library@ihs.ac.at)].
- Campbell*, David F.J. / Bernhard *Felderer* (1999). Empfehlungen zur Evaluation universitärer und außeruniversitärer Forschung in Österreich. Institute for Advanced Studies (IHS): Political Science Series No. 66 [[library@ihs.ac.at](mailto:library@ihs.ac.at)].
- European Commission* (1997). Second European report on science and technology indicators 1997. Brussels.
- Forschungs- und Technologiepolitik. Baden-Baden: *Felderer*, Bernhard / David F.J. *Campbell* (1994). Forschungsfinanzierung in Europa. Trends, Modelle, Empfehlungen für Österreich. Vienna: Manz.
- Gibbons*, Michael / Camille *Limoges* / Helga *Nowotny* / Simon *Schwartzman* / Peter *Scott* / Martin *Trow* (1994). The new production of knowledge. The dynamics of science and research in contemporary societies. London: SAGE Publications.
- HEFCs – Higher Education Funding Councils* (1995). 1996 Research Assessment Exercise: Guidance on Submissions. Bristol.
- HEFCs – Higher Education Funding Councils* (1999). 2001 Research Assessment Exercise: Guidance on Submissions. Bristol.
- IMD – Internationale Institute for Management Development* (1996). The World Competitiveness Yearbook 1996. IMD: Lausanne.
- Kuhlmann*, Stefan (1998). Politikmoderation. Evaluationsverfahren in der Nomos.
- OECD (1998a). Education at a glance. OECD indicators 1998. Paris.
- OECD (1998b). Main science and technology indicators. Paris.
- OECD (1998c). Science, Technology and Industry Outlook 1998. Paris.
- OECD (1999a). Historical Statistics 1960-1997. Paris.
- OECD (1999b). Basic Science and Technology Statistics. Paris.
- OECD (2000a). Main Science and Technology Indicators. Paris.
- VSNU – Vereniging van Samenwerkende Nederlandse Universiteiten* (1994). Protocol 1994. Utrecht.
- VSNU – Vereniging van Samenwerkende Nederlandse Universiteiten* (1998). Protocol 1998. Utrecht.

**Figure 1: The gross domestic expenditure on R&D for the total national R&D, the business enterprise sector R&D and the higher education sector R&D during the period 1983-1998. Country sample: average values for all OECD countries as a % of GDP.**



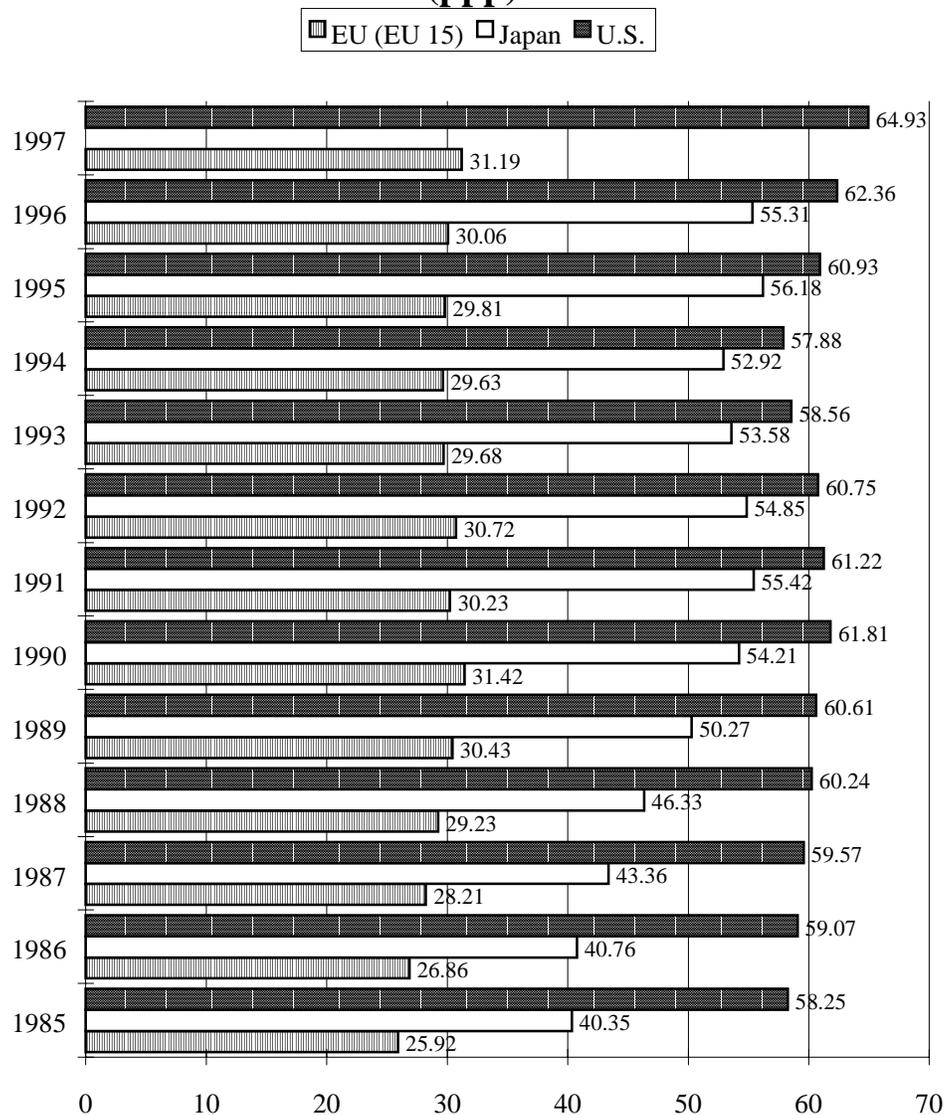
Source: OECD 2000 ("Main Science and Technology Indicators").

**Figure 2: The funding of university research in Europe through public basic funding and through earmarked funding. Mode of calculation: funding shares in % of the total funding.**



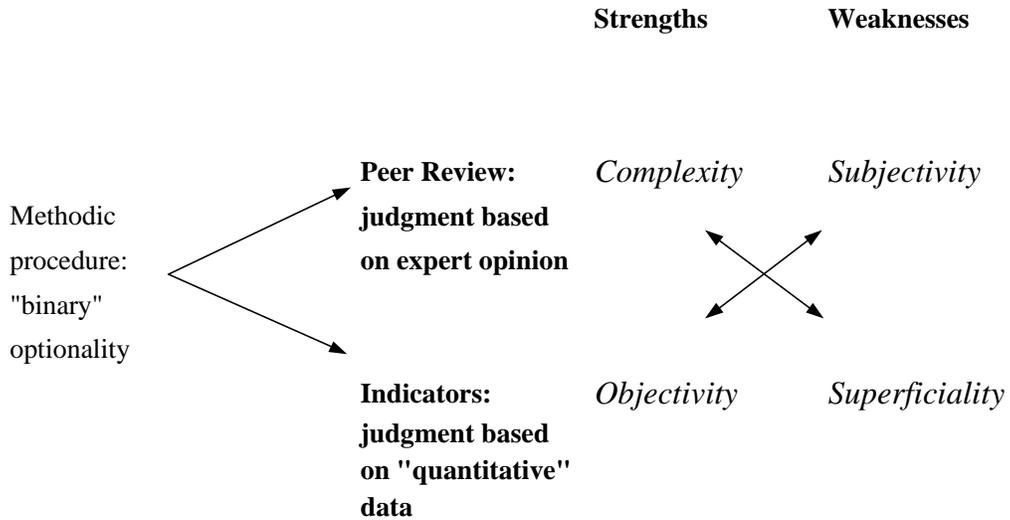
Source: Author's (David Campbell) own calculations based on OECD 1999 ("Basic Science and Technology Statistics").

**Figure 3: The gross domestic expenditure on R&D per a population of 100,000 (1985-1997). Currency unit: million constant \$ in 1990 prices and purchasing power parities (ppp).**



Source: Author's own calculations based on OECD 1999 ("Basic Science and Technology Statistics") and Campbell (2000).

**Scheme 1:** Standardized comparison of the strengths and weaknesses of "peer review" and indicators.

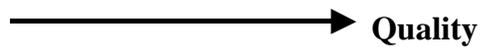


Source: Author's own conceptualization based on Campbell (1999) and Campbell/Felderer (1999).

Copyright by David F. J. Campbell, Vienna 2000

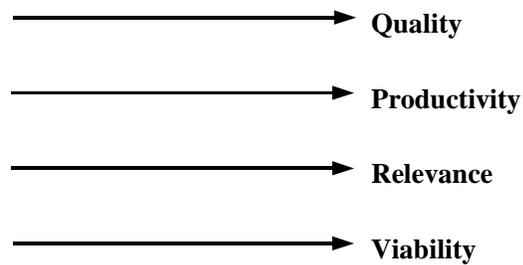
**Scheme 2: A typology of different conceptualizations of the dimensionality of university research quality.**

*The UK model of "ex post" evaluation of university research: one (comprehensive) dimension of quality.*



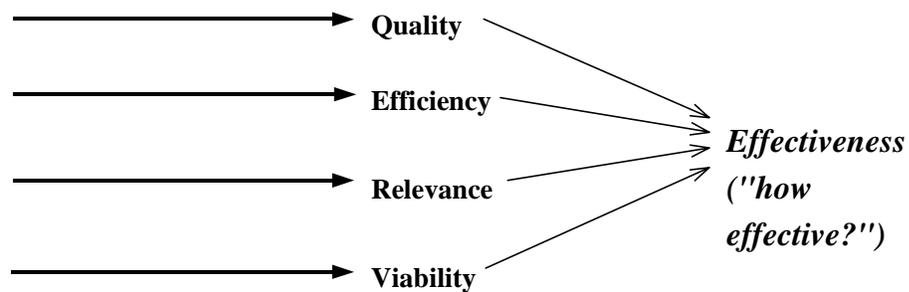
Source: HEFCs (1995, 1999).

*The Netherlands model of "ex post" evaluation of university research: four dimensions of quality.*



Source: VSNU (1994, 1998).

*The integration of the concept of "effectiveness" in a model of "ex post" evaluation of university research: four to five dimensions of quality.*



Source: Author's own conceptualization based on Campbell (1999).  
Copyright by David F. J. Campbell, Vienna 2000

### **Scheme 3:**

**A standardized comparative typology of national European higher education with regard to the application of "ex post" evaluations of university research.**

#### **Type A Nations (Countries)**

United Kingdom (UK)

Netherlands

*Systematic and comprehensive evaluations, at national level, with explicit references to grading scales: "systemic and consistent approach".*

#### **Type B Nations (Countries)**

Germany

Austria

[Finland?]

[Switzerland?]

*"Ad hoc" (meso-level) evaluations, disciplinary independent, and without references to explicit grading scales: "pluralized and situational approach".*

Source: Author's own conceptualization based on Campbell (1998).

Copyright by David F.J. Campbell, Vienna 2000

**Scheme 4:**

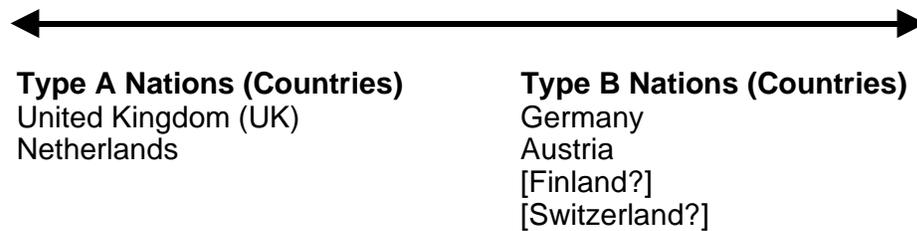
**A standardized comparative typology of national European higher education with regard to the application of "ex post" evaluations of university research.**

**Different possible scenarios:**

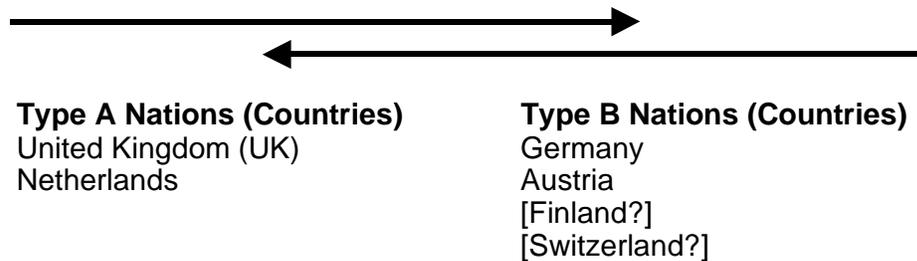
**Scenario One: Functional needs of advanced societies.**



**Scenario Two: Cultural pluralism.**



**Scenario Three: Mutual overlapping (and mutual influences).**



Source: Author's own conceptualization based on Campbell (1998).

Copyright by David F.J. Campbell, Vienna 2000

## Discussion of papers

**Henry Etzkowitz** (State University of New York, USA). Lou, is there a tension between the two dimensions of licensing and startups in your experience? In some states there is pressure for regional effects say in favor of licensing. In others, startup retention is related to availability of venture capital.

**Louis Tornatsky**. Yes, those institutions that do well in startups, often will run an incubator or are involved in a venture scheme that are wrapped into the broader technology policies.

**Arie Rip** (Twente University, The Netherlands). This question is addressed to both David and Lou. The association of Dutch universities is using benchmarking. There is more interest in the movement toward this exercise. Do you find there is a trend toward benchmarking for evaluation?

**Louis Tornatsky**. It is really more of a different way to approach the problem.

**Dave Guston** (Rustgers University, USA). Lou, is your ability to recognize and use simple measures tied to your ability to do follow up with the organizations you follow (i.e., in tech support)?

**Louis Tornatsky**. There are some things that are missing. For example, industrially sponsored work, and how much is located in the states that are reported on.

**Stefan Kuhlmann** (Fraunhofer ISI). Lou, is there a similar exercise (to the work of the Southern Technology Council) in the northern states.

**Louis Tornatsky**. No. In the (US) south there are more regional organizations than anywhere in the country.

## 9. Appropriate Methodological Matrices for S&T Evaluation

Chair: Peter Blair (Sigma Xi, USA)

---

### Real Options for Public Sector R&D Investments

**Nicholas S. Vornatas and Matt Lackey**

Science Technology and Public Policy Program

Georgia Washington University, Washington DC, USA

Email: vonortas@gwu.edu

---

#### Introduction

An increasing number of academics and financial managers have grown dissatisfied with the conventional methods for evaluating long-term investments such as the net present value (NPV) and return on investment (ROI). They join research administrators in both the private and the public sector who have emphatically argued that these methods do not capture the full value of an investment opportunity, often unduly penalizing investments with expected payoffs later in the future. Funding of strategic, long-term research and development (R&D) projects has thus been largely justified in the past on the basis of qualitative arguments and “gut feelings” of such administrators concerning the windows of opportunity such R&D opens in new technological areas.

An important underlying reason for this state of affairs is that capital budgeting and strategic planning have historically been treated as two distinct domains for resource allocation, the former dealing with measurable returns (profits/cash flows) of a project and neglecting more intangible strategic benefits associated with the project. The result has always been an alleged chronic deficit between the calculated value of strategic, long-term projects and their “true” value. The deficit is due to the oversight by the conventional NPV model of the inherent strategic value of the project and the flexibility associated with active management to alter the project's trajectory once undertaken. Several factors contribute to the strategic and flexibility deficit, including:

- Uncertainty of the outcome
- Timing of the investment
- Irreversibility of committed resources
- Inaccurate use of the discount rate

It is to this end, placing a value on the deficit, that economists have turned to the use of an option-pricing model. It is argued that the theory of options on financial securities can be applied to non-financial investments, resulting in “real options” that allow placing a valuation on the flexibility and strategic characteristics of long-term investments. That is to say, real options can formally address not only the measurable returns but also the intricacies of market and technological uncertainty, timing, irreversibility, and the discount factor associated with strategic, long-term investments.

The literature on real options has largely been confined to investment in the private sector. Following on the line of prior research (Vonortas and Hertzfeld, 1998), we argue that the approach can be extended to the public sector. We also sketch out some parallels with and enhancements of existing project appraisal techniques that, for most practical purposes, can arguably make them functionally equivalent to the real options approach to the ex ante appraisal of long-term, strategic R&D investments.

### **In Search of R&D Options**

It has been suggested that the decision to invest initially in an R&D project with a highly uncertain outcome is conditional on revisiting the decision sometime in the future. This is similar in its implications to buying a financial call option. A financial call option will permit (but not oblige) the owner to purchase stock at a specified price (exercise price) upon the expiration date of the option. An initial R&D investment will permit (but not oblige) the investor to commit to a particular technological area – that is, buy the entitlement of the future stream of profits – upon the predetermined date for revisiting the initial investment decision. The analogy between the R&D option and the stock option can be summarized as follows:

- The cost of the initial R&D project is analogous to the price of a financial call option.
- The cost of the follow-up investment needed to capitalize on the results of the initial R&D project is analogous to the exercise price of a financial call option.
- The stream of returns to this follow-up investment is analogous to the value of the underlying stock for a financial call option.

The downside risk of the initial R&D investment is that the invested resources will be lost if, for whatever reason, the follow-up investment is not made. This is analogous to the downside risk of a financial call option which, in case that the option is not exercised, will be the price of the option.

## *9 Appropriate Methodological Matrices*

Increased uncertainty decreases the value of an investment for risk averse investors. In contrast, and in combination to with the possibility of higher returns, increased uncertainty increases the value of an initial R&D project if it is considered an option to a potentially very valuable technology. This is analogous to the effect of uncertainty (volatility) on a financial call option.

A longer time framework decreases the present discounted value of an investment. In contrast, the value of an initial R&D project may well increase with time if considered an option to longer-term, high-opportunity investments. This is analogous to the positive effect of time on the value of a financial call option.

It has, thus, been argued that when an investor commits to an irreversible investment the investor essentially exercises his call option. In other words, the investor "... gives up the possibility of waiting for new information to arrive that might affect the desirability or timing of the expenditure; [the investor] cannot disinvest should market conditions change adversely." (Dixit and Pindyck, 1994, p. 6).

On the other hand, for most investments there exists some abandonment value in terms of a salvage value or the opportunity to simply shut down should the project become unprofitable. The abandonment option is parallel to a financial put option – where a price is paid for the opportunity to sell a security at a favorable price should the security decline in value. Thus in each investment in R&D there is an inherent value in the ability to stop investment or redirect resources to another project.

Real investment opportunities, then, usually involve multiple options whose individual values most often will interact and should be valued together (Trigeorgis, 1996). This feature should make them unsuitable for conventional discount cash flow methods that approach capitalized budgeting from a decentralized point-of-view, valuing projects on a stand-alone basis. Which is exactly why managers and policymakers, having long recognized that projects can have intangible strategic benefits that may lead to competitive advantage, have not trusted such methods in the past. The options approach is an attempt to capture these benefits in the valuation analysis. It is imperative that one accounts for all options in order to attain the best estimation of valuation.

The important implication is that the question of how to go about exploiting future opportunities reverts to a question of how to exercise the corresponding call options optimally. Academics and financial practitioners have studied this problem in stock option pricing theory where the value of a stock option has been formally expressed as a function of the underlying parameters. It is now argued that the basic principles arising from this work can be transferred to the arena of "real" (i.e., non-financial) investments. Specifically to our present interests, it is argued that the basic principles can be transferred to the problem of selection among highly uncertain, long-term R&D (and other kinds of) investments in the public sector.

## 9 Appropriate Methodological Matrices

One can start by differentiating between the various stages of a projected R&D program and evaluating the stages in sequence. Each stage provides information for the next thus creating an opportunity (option) for subsequent investment in a new technological area. For example, an early R&D investment by the public sector in an emerging technological area may be considered the mechanism for creating the option for (enabling) the private sector to undertake the follow-up investment required to innovate in that area.

By explicitly recognizing the “choice to invest” aspect of earlier-stage R&D projects, this mechanism greatly enhances the ability of decision-makers to justify long-term R&D investments made by the public sector. Moreover, by differentiating among the various stages in an R&D program, this mechanism allows the use of more appropriate discount rates that better reflect the differential risks of technologies in various stages of development.

### **Extended NPV and Real Options**

The criticism of conventional NPV, pointing at the difficulties of this project appraisal method to account for the “true” value of uncertain investment projects, is usually well taken. However, as analysts have delved into the details of the available alternative methods more recently, it has surfaced that NPV critics may be addressing a project appraisal approach that is no more, at least in theory. That is to say, NPV weaknesses have been projected to an extreme while the ability of the method to expand and incorporate many of the contemporary advancements have not been properly accounted for.

To begin with, one of the major attacks on NPV is the assumed discount rate. It is correctly pointed out that not all cash flows are subject to the same risk and therefore should not be discounted at the same rate. Detractors of NPV often assume that managers will use the same discount rate for all cash flows. While this is often the case, it is not out of necessity. One of the major benefits of NPV (as opposed to the Internal Rate of Return (IRR)) is that it is simply a summation of different time periods allowing the use of a different discount rate for each cash flow (net benefit):

$$NPV = CF_1/(1+r_1)^1 + CF_2/(1+r_2)^2 + CF_3/(1+r_3)^3 + \dots + CF_n/(1+r_n)^n$$

In addition, it is argued that the NPV approach does not really eliminate the case for project delay as it is frequently accused of doing. A positive NPV does not necessarily mean that a project should be best undertaken immediately; it may be even more valuable if undertaken in the future. Similarly, a project with a currently negative NPV might become a profitable opportunity if we wait a bit. Taking into account the option to delay is accomplished by evaluating the project at each alternative investment date (option) and choosing the one with the highest NPV (Brealey and Myers, 2000). Likewise, a project may have a negative NPV because of excess maintenance or capacity in later time periods, which can be abandoned

## 9 Appropriate Methodological Matrices

resulting in a possible altered positive NPV project. Each abandonment option scenario is evaluated, and the one with the highest NPV is chosen. By evaluating each alternative incrementally, the option to switch resources can be evaluated using the NPV method. While computationally complex, such analysis is theoretically correct.

Furthermore, while indeed small changes in chosen parameter values (such as discount rate) can have significant affects on NPV project valuation, the same argument can be made about the options approach. Similar methodologies are often used for both models – parameters are derived from a “twin” portfolio asset that is freely traded in the open market. For investments whose underlying asset is not traded – and, therefore, has no market price – the proposed solution is to assume a process and find a duplicate “twin” portfolio of traded assets to simulate the cash flows of the underlying asset (Brennan and Schwartz, 1985; Dixit and Pindyck, 1994; Trigeorgis, 1996).

This problem particularly afflicts long-term R&D investments as many such projects have no reasonable “twin” that trades in the open market, necessitating alternative methods.<sup>1</sup> One proposed practical alternative is to estimate the market potential of the product(s) deriving from the R&D project in question and use it as the underlying (Perlitz et al., 1999). This would seem to bring us full circle to the practices in the NPV. The practicality of the method is questionable (let alone its theoretical correctness) as it hinges on the analysts’ ability to foresee and place dollar values on future uses of early stage, strategic R&D.

The value of the option is inherently tied to its degree of associated risk which is approximated by the volatility of the underlying asset. For an R&D project risk can be divided into three categories: (a) technological risk, (b) market risk, and (c) risk due to exogenous events. Since it is unlikely that adequate historical risk data exists for the project, it is necessary to once again use a “twin” portfolio to derive this value. If a traded “twin” cannot be established – which is often the case with R&D projects as mentioned above – it will be necessary to choose a risk premium associated with the project from such models as the Capital Asset Pricing Model. This is exactly how the discount rate is determined in the NPV model: the discount rate is the opportunity cost of capital associated with the project and is made up of two parts, the risk free rate that accounts for the time value of money, and the risk premium which accounts for the riskiness of the project.

One way in which NPV can deal with the problem of appropriate parameter selection in the presence of risk is *sensitivity analysis*, whereby managers give optimistic and pessimistic estimates of the variables in addition to the expected value. The valuation model is then run using each variable's optimistic value and

---

<sup>1</sup> Which the debate over ex ante appraisal of uncertain investment projects has always had to grapple with.

## 9 Appropriate Methodological Matrices

then each pessimistic value one at a time. The expected project valuation is compared to pessimistic and optimistic valuations for the change in each variable. Using this approach will allow management to see which variables are more sensitive to miss-estimation and their subsequent effect on project valuation. Such variables may be worthy of more study (Brealey and Myers, 2000).

Sensitivity analysis also has limitations in that individual variables typically have influence on other variables and hence there is a covariance. To address this, analysts may use what is known as *scenario analysis*, which involves analyzing the change in different combinations of variables. For complex projects with many variables one needs to consider all possible combinations of variables and their outcomes. Which is what Monte Carlo *simulation* allows, in theory at least. Still, the analyst must specify interdependencies in variables and time periods, by no means an easy feat.

Decision-makers thus have several choices within the NPV model (sensitivity analysis, scenario analysis and simulation) to determine the probabilities associated with different outcomes. These probabilities can then be used in decision trees.

Decision tree analysis (DTA) is used to analyze sequential investment opportunities and help decision-makers identify strategies with the highest value. It does so by helping recognize the interdependencies between initial and subsequent decisions, given that the importance of recognizing the chain of options and possible outcomes to an investment cannot be underestimated. “[I]f subsequent investment decisions depend on those made today, then today’s decision may depend on what you plan to do tomorrow.” (Brealey and Myers, 2000, p. 275).

Using the rationale briefly outlined above, some experts support the view that the NPV method can be expanded and enriched by DTA, introducing various techniques to determine outcome probabilities, and by the additional consideration of flexibility and active management to approximate “real options” analysis (Trigeorgis, 1996). One could think of an expanded framework, for example, where the overall value of the project is the sum of the conventional, static NPV and the option premium, consisting of the flexibility value and the strategic value:

$$\text{NPV}^* = \text{NPV} + \Sigma (\text{value of flexibility options}) + \Sigma (\text{value of strategic options})$$

Rather than dismissed, then, conventional NVP can be part of an expanded framework that incorporates the notion of investment options, thus replicating the “real options” approach. The opportunity to invest can be more valuable than immediate investment due to the ability to defer investment until conditions become more favorable or to abandon completely if circumstances remain unfavorable.<sup>2</sup>

---

<sup>2</sup> It is important to keep in mind the differences between projects that offer decision-makers the ability to defer investment until receiving more information and those that require immediate

## 9 Appropriate Methodological Matrices

This expanded framework is argued to allow managers to combine traditional finance theory and strategic planning.

### Example

This section presents a simplistic example in order to demonstrate basic methodological differences and similarities between different project valuation models. The example builds progressively from conventional NPV to decision tree analysis and binomial options to the expanded evaluation method incorporating the full value of investment options. In this example, the basic observation for the reader should be as follows:

The conventional NPV simply requires from the analyst too much in terms of an accurate assessment (guess) of the future value of the underlying asset (technology);

Decision tree analysis and the binomial distribution method improve assessment by introducing more sophisticated techniques to arrive at a generally more accurate estimate of the future value of the underlying asset. These techniques take into account valuable information that the analyst can generate concerning the future value of the underlying asset. Still, the strategic value of the project and the value of flexibility associated with active project management remain outside the evaluation;

The expanded assessment framework combines all of the above, and enables the calculation of the “investment option premium”. It is the size of this premium that has underlined the arguments of the proponents of the “real options” approach to the appraisal of strategic, long-term R&D expenditures.

The values for investment and returns are chosen arbitrarily. While different values would alter the outcome, the inherent properties of each method would remain unchanged.

Suppose, then, a risky investment opportunity that requires an initial investment outlay  $I_0 = \$104\text{m}$ . This expenditure will produce an underlying asset in the form of a valuable perpetuity, that is, an annuity projected to infinity. Its present value is defined by:

$$PV = CF/r$$

where CF is the annual cash flow and r is the discount rate. Cash flow can be simply considered as the difference between the benefits from and the operating costs of the project. This equation is derived directly from the previously defined relationship:<sup>3</sup>

---

commitment. The latter do not involve a call option; they will have, instead, a put option to abandon. Put options must also be included in the evaluation of a project.

<sup>3</sup> Letting  $CF/(1+r) = a$  and  $1/(1+r) = x$ , and then substituting into  $PV = \sum (CF_1/(1+r_1)^1 + CF_2/(1+r_2)^2 + CF_3/(1+r_3)^3 + \dots)$ , we obtain  $PV = a(1 + x + x^2 + \dots)$ . Multiplying both sides by x, we get  $PVx =$

## 9 Appropriate Methodological Matrices

$$PV = \sum (CF_1/(1+r_1)^1 + CF_2/(1+r_2)^2 + CF_3/(1+r_3)^3 + \dots)$$

The value of the project value in four steps, using the methods of: (a) conventional, static NPV; (b) decision tree analysis; (c) binomial option; and (d) expanded NPV.

### A. Conventional NPV

The first step in this model is, of course, to estimate the annual net benefit and the discount rate, approximated by the cash flow and the opportunity cost of capital respectively. Let us assume a projected annual cash flow of \$18m. In addition, we assume that the “twin security” determines the discount rate at 20%. Then the value of the project is given by:

$$V_0 = CF/r = 18/.2 = \$90m$$

The NPV is simply given by:

$$NPV = V_0 - I_0 = 90 - 104 = -\$14m$$

The project should be foregone.

### B. Decision Tree Analysis (DTA)

The obvious weaknesses of the preceding approach is the assumption that events will turn out exactly as predicted and that the discount rate remains constant, both highly unlikely in the presence of significant uncertainties and long time periods. Improving the assessment of the potential future values of parameters and the overall estimation of the project value requires delving deeper into potential path trajectories.

Suppose now that we can perform a DTA whereby sensitivity analysis, or scenario analysis, or simulation is undertaken to obtain a more accurate forecast of the projected outcome. As a result, two possible scenaria are projected with equal probability ( $q = 0.5$ ):

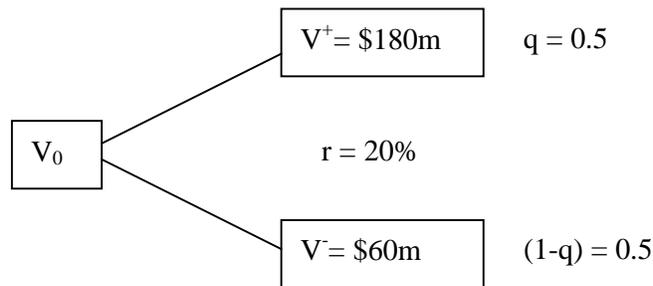
A favorable scenario that will produce a perpetuity of \$36m, which provides for a project value of  $V^+ = 36/0.2 = \$180m$ , and

An unfavorable scenario that will produce a perpetuity of \$12m, which provides for a project value of  $V^- = 12/0.2 = \$60m$ .

---

$a(x + x^2 + \dots)$ . Then,  $PV(1-x)=a$ . Substituting for  $a$  and  $x$ ,  $PV[1 - 1/(1+r)] = CF/(1+r)$ . Multiplying both sides by  $(1+r)$  and rearranging gives  $PV = CF/r$ .

## 9 Appropriate Methodological Matrices



The present value of the project would simply be the sum of the products of the outcomes and their respective probabilities divided by the opportunity cost of capital.

$$V_0 = [qV^+ + (1-q)V]/(1+r) = [0.5(180) + 0.5(60)] / (1+0.2) = \$100m,$$

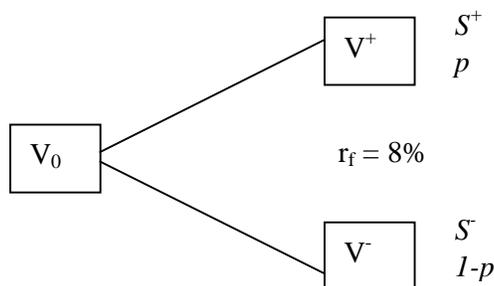
NPV is given by:

$$NPV = V_0 - I_0 = 100 - 104 = -\$4m.$$

The project should be forgone. In general, DTA is expected to account more accurately for future possibilities and to arrive at better project value estimates.

### C. Binomial Distribution

Now suppose that future values of the underlying asset (i.e., future values of the project) follow a multiplicative binomial distribution. Its derivation makes use of the finding of investor riskless attitude toward options pricing (e.g., Cox and Ross, 1976). If so, the value of the option can be made using a riskless interest rate. Since the value of the option depends only on the underlying asset and not investors' risk-aversion, it is possible to use certainty equivalent cash flows and assume a riskless rate for the model (Trigeorgis, 1996). A simple example of the model is as follows:



## 9 Appropriate Methodological Matrices

where the  $V$ 's are defined as before. A "twin security" is now defined with a value  $S^+$  in the case of an up move and  $S^-$  in the case of a down move. The pseudo probability (risk free probability) of an up move is  $p$ , where  $p = [(1+r_f)S^-] / (S^+ - S^-)$ . The probability of a down move is just  $1-p$ . The risk-free discount rate,  $r_f$ , is set at 8%.

The values of the up and down moves and of the corresponding probabilities must either be known or, similarly to DTA, must be derived from sensitivity, scenario, or simulation analysis. Assume that the Twin security offers a perpetuity of \$20m, allowing the same scenarios as before: an up move of value \$36m and a down move of value \$12m. Then:

$$p = [(1+0.08)20-12] / (36-12) = 0.4,$$

Using the same formula as before, the value of the investment would be the sum of the probability of an up-move times the value of the up-move plus the probability of a down move times the value of the down move, discounted at the risk free rate.

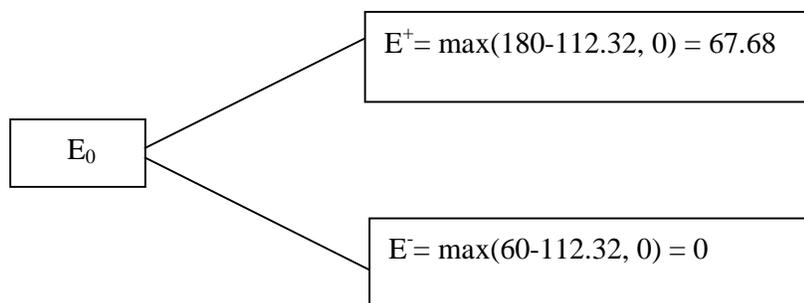
$$V_0 = [pV^+ + (1-p)V^-] / (1+r_f) = [0.4(180) + 0.6(60)] / 1.08 = \$100m$$

As expected, this is identical to the result from DTA analysis.

### **D. Options-Extended Framework**

Decision tree analysis and the binomial options method improve assessment by introducing more sophisticated techniques to arrive at a generally more accurate estimate of the future value of the underlying asset. However, these methods do not inherently account for the full value of the investment options reflecting the strategic value of the project and the value of flexibility associated with active project management remain outside the evaluation. All preceding methods do not capture the value of the option to pursue later actions if uncertainties decrease and, as a result, the project becomes more favorable. This needs some additional structure.

A simple way to determine the value of the option to defer investment is by substituting the equity value of the option payoff,  $E^+$  or  $E^-$ , in the previously used relationship as given by the figure below.



## 9 Appropriate Methodological Matrices

By deferring for a year, the initial intended outlay  $I_0 = \$104\text{m}$  increases by 8% to  $I_1 = \$112.32\text{m}$ . The two possibilities are again the favorable scenario with a project value of  $\$180\text{m}$  and the unfavorable scenario entailing a project value of  $\$60\text{m}$ . Therefore,

$$E^+ = \max(V^+ - I_1, 0) = \max(180 - 112.32, 0) = \$67.68\text{m},$$

$$\text{and } E^- = \max(V^- - I_1, 0) = \max(60 - 112.32, 0) = 0.$$

The asymmetric nature of an option, shown in this case, implies the following. Waiting for one period may either result in a payoff of  $\$67,68\text{m}$  if the outcome proves favorable and we proceed with the investment or may result in a payoff of  $\$0$  if the outcome is unfavorable and we forego the investment. Therefore, the total value of the investment becomes:

$$E_0 = [p(E^+) + (1-p)E^-]/(1-r) = [.4(67.68) + 0.6(0)]/1.08 = \$25.07\text{m}$$

It should be observed that the option premium for the flexibility to wait and defer investment pending a favorable development is given by the difference between the evaluation of the options-extended framework and the NPV form DTA or binomial:

$$\text{Option premium} = E_0 - \text{Passive NPV} = 25.07 - (-4) = \$29.07\text{m}.$$

### Sequential Investment.

Thus far we have assumed that the information acquired one year later as to the successfulness of the project is free. Such knowledge is highly unlikely to just appear, however, when dealing with specific R&D projects (or, if it does, someone else has a first-mover advantage to the technology). It could be argued, for example, that “exploratory” or “enabling” research will provide us the information as to whether or not this project will turn out favorably. Such research can be considered an option to the specific technology with price that should not exceed  $\$25.07\text{m}$ . Put differently, an initial R&D expense of  $\$25.07\text{m}$  would allow us to invest in a full-scale project one year later for a NPV exactly equal to 0. Any initial outlay of less than  $\$25.07\text{m}$  to achieve this same opportunity would result in a positive NPV project.

### **Concluding Remarks**

This paper briefly summarizes recent methodological improvements in assessing both event probabilities and the risks involved in uncertain investments. It argues that such methods have matured enough and can now be exploited to greatly enhance ex ante, formal assessments of strategic long-term R&D projects. An increasing number of R&D-intensive companies have already been making use of the modern “real options” evaluation concepts to improve their strategy formulation processes.

## *9 Appropriate Methodological Matrices*

It is high time that the public sector devotes adequate attention to the new project evaluation methods too. This is especially so given that some of the formal quantitative arguments in these methods seem to replicate formally the qualitative arguments of R&D administrators who have traditionally been apprehensive of the appropriateness of conventional methods to appraise strategic R&D investments. Consideration of the new techniques seems especially appropriate in view of the modern approach to project evaluation not only as a tool for ex post appraisal and control but increasingly as a tool of intelligence gathering and forward strategic planning (Kuhlmann, 1999; Kuhlmann et al., 1999).

### **Refereneces**

Brealey, Richard and Stewart Myers (2000) Principles of Corporate Finance. Irwin/McGraw-Hill.

Brennan, M. and E. Schwartz (1985) "Evaluating Natural Resource Investments", Journal of Business, 58(2): 135-157.

Cox, J.C. and Ross, S.A. (1976) "The Valuation of Options for Alternative Stochastic Processes", Journal of Financial Economics, 4(3): 145-166.

Dixit, Avinash K. and Robert S. Pindyck (1994) Investment Under Uncertainty. Princeton, NJ: Princeton: University Press.

Kuhlmann, Stefan (1999) "Distributed Intelligence: Combining Evaluation, Foresight and Technology Assessment", The IPTS Report, 40 (December).

Kuhlmann, Stefan, Patries Boekholt, Luke Georghiou, Ken Guy, Jean-Alain Heraud, Philippe Laredo, Tarmo Lemola, Denis Loveridge, Terttu Luukkonen, Wolfgang Polt, Arie Rip, Luis Sanz-Menendez, and Ruud Smits (1999) Improving Distributed Intelligence in Complex Innovation Systems, Final Report, Advanced Science & Technology Policy Planning Network, Targeted Socio-Economic Research Programme, European Commission, DG XII.

Perlitz, Manfred; Peske, Thorsten and Randolph Schrank (1999) "Real Options Valuation: The New Frontier in R&D Project Evaluation?" R&D Management, 29(3): 255-269.

Trigeorgis, Lenos (1996) Real Options: Managerial Flexibility And Strategy In Resource Allocation, Cambridge, MA: MIT Press.

Vonortas, Nicholas S. and Henry R. Hertzfeld (1998) “Research and Development Project Selection in the Public Sector”, Journal of Policy Analysis and Management, 17(4): 621-638.

---

## **Discussion of Nick Vonortas’ paper**

**Gretchen Jordan** (Sandia National Laboratories, USA). Are there other similar studies? I think the Industrial Research Institute has a programme on options. One should use their data.

**Luke Georghiou** (University of Manchester, UK). I agree that this is an excellent way to think about R&D, especially for industry. But calculation is difficult because of the volatility and difficulty in getting data.

**Irwin Feller** (Pennsylvania State University, USA). This paper is very useful, especially in the theoretical conception. It is interesting to use the option model to analyze the engagements of universities in start-ups with this kind of models. .

**John Barber** (Department of Trade and Industry, UK). In a more practical way, you must understand the choice of technology. Evaluation should look at whether there were enough winners, not simply precise analogs and calculations.

---

## **Assessing Programme Portfolios via Multi-Module Approaches**

**Ken Guy**

Wise Guys Ltd.

Shoreham-by-Sea, UK

Email: ken.guy@technopolis.co.uk

---

### **Issue**

This paper uses the example of a recent evaluation of a portfolio of eleven energy technology programmes in Finland to discuss how evaluators can address the needs of policymakers and programme administrators for the production of useful information via cost-effective evaluation approaches.

### **Relevance**

As innovation policy becomes more sophisticated and policymakers advocate the launch of ‘packages’ of innovation-related actions, the challenge for evaluators has been to consolidate the experience gained in project and programme evaluations to devise cost-effective ways of evaluating portfolios of programmes and, ultimately, complete policy ‘packages’. A multi-module approach involving combinations of tried and tested evaluation methodologies offers one way of producing useful insights and helpful policy suggestions.

### **Analysis**

Innovation policymakers have come to appreciate that there is much to learn from the evaluation of past activities. Over the past twenty years or so, evaluations have become established as important inputs to programme and policy formulation, with project evaluations feeding into the development of improved programme management strategies, and programme evaluations informing policy discussions at national and international levels.

Over this period, however, there have been many changes and developments in the theory and practice of evaluation. One of the main policy drivers for early evaluations was the need to measure the impact of past activities and, in particular, the added value of innovation-related actions such as publicly-funded research and development (R&D) programmes. In large part evidence of this nature was needed to legitimate past policies, but attempts to measure impacts via ‘summative’ approaches were beset by methodological difficulties and evaluation practice thus took a different course. In particular:

## 9 *Appropriate Methodological Matrices*

- The rationale for evaluations has shifted from a desire to legitimate past actions to the need to improve understanding and inform future actions
- Correspondingly, the issue focus of evaluations has broadened away from a narrow focus on quality, economy, efficiency and effectiveness, and towards a more all encompassing concern with additional issues such as the appropriateness of past actions and a concern with performance improvement and strategy development
- Approaches to evaluation have evolved away from a purist model of ‘objective neutrality’, characterised by independent evaluators producing evaluation outputs containing evidence and argument but no recommendations, to more ‘formative’ approaches in which evaluators involve all relevant stakeholders in learning exercises, providing advice and recommendations as well as independent analysis
- The context in which evaluations are conducted has become more turbulent. As our comprehension of the complexity of innovation systems has increased, collective certainty in the ‘correctness’ of particular policy prescriptions has decreased
- This has led to more flexible and experimental approaches to the construction of policy portfolios, and to even greater demands for well specified systems of monitoring, evaluation and benchmarking to aid analyses and feed into strategy development

All these changes have brought new challenges for both policymakers and evaluators. More than ever before, policymakers are having to think holistically about ‘packages’ of innovation-related actions, both in terms of their construction and their subsequent evaluation. The challenge for evaluators, too, has been to devise cost-effective ways of ‘stepping up the ladder’, first from individual project evaluations to programme evaluations, and then on to the evaluation of portfolios of related programmes and, ultimately, to the evaluation of complete ‘policy packages’.

In this paper we demonstrate how challenges of this nature can be met via the use of multi-module evaluation missions. Drawing on the experience of one such mission in Finland (Guy, *et al.*, 1999), we show how a range of evaluation methodologies in common use at project and programme levels can be combined in a cost-effective manner to feed timely assessments of the relevance, calibre and impact of a portfolio of eleven R&D programmes into policymaking processes.

### **The Need for a Portfolio Evaluation**

In the late 1980s, Finland’s Ministry of Trade and Industry (KTM) initiated a series of R&D programmes in the field of energy technology. Subsequently, in 1993, it launched a further suite of eleven Energy Technology Programmes scheduled to run over the period 1993-1998. Aimed at the development of efficient and

## *9 Appropriate Methodological Matrices*

environmentally sound energy technologies intended to be competitive in the international marketplace, the programmes sought to involve the research, industrial and public sectors in some FIM 1.2 billion (300 MEURO) of research and development activity. The eleven technology areas spanned the supply, distribution and use of energy in a number of sectors.

In early 1995, the Technology Development Centre of Finland (Tekes) assumed responsibility for the funding, management and administration of the programmes. As the final year of activities began, Tekes commissioned a team to conduct a major review of all eleven programmes over the course of 1998. This international team was assembled and led by the UK office of Technopolis Ltd. and comprised a total of 21 people drawn from the private and public sectors.

The broad aim of the exercise was to review the experience of the eleven technology R&D programmes and to make suggestions for the future. In particular, the intention was to cover a number of distinct levels. Most important were the **Programme** and **Portfolio** levels:

- At the individual **Programme** level, the review was to comment on the relevance, calibre and impact of programmes, concentrating in particular on the following:
  - **Relevance** - were programme and project level goals in line with Finnish interests and comparable agendas in other countries?
  - **Efficiency** - how well were the programmes implemented and managed?
  - **Quality** - how did the scientific and technological quality of the work compare with work conducted elsewhere?
  - **Effectiveness** - to what extent were the goals of the programmes met?
  - **Impact** - what has happened as a result of the programmes?
  - **Strategy** - what should Tekes do next?
- At the level of the **Portfolio** of energy technology programmes, the intention was to comment on the past and continued appropriateness of the programme mix, concentrating in particular on the technology areas covered and, to a lesser extent, the management and administrative structures in place to cover them

In addition, coverage was also expected to extend to three other levels:

- At an overarching **Policy** level, the aim was to suggest R&D directions and strategies that the Finnish Government might like to consider for the future
- At the **Project** level, the brief was only to comment on high points and low points
- At the level of **Participant**, commentary was requested on lessons for the future for academics, research institutes and industrialists

### **A Multi-Module Approach**

Evaluators typically recommend that 0.5% to 1% of a programme's budget should be spent on an evaluation. Few cost more than this, but evaluators are often asked to work with smaller budgets, especially when large programmes are involved and small percentages translate into large absolute values. Understandably, this phenomenon is even more acute when the task is to evaluate a portfolio of eleven programmes simultaneously. In the Finnish case, rather than construct an evaluation costing 0.5 to 1% of the portfolio's budget, the challenge was to conduct a timely and informative review for less than 0.1% over the last 12 months of the portfolio's lifetime.

Given the ambitious scope of the project and the limited resources and time available to conduct the review, the approach adopted was to construct an evaluation composed of a number of carefully streamlined modules, each of which could only provide a limited perspective on programme performance and impact when conducted in isolation, but all of which, when taken together, could provide a multi-faceted overview of utility to policymakers.

The review of the programmes was organised around the following five main work modules:

- The collection and preparation of **Background Material** on Finnish energy technology policy and the content, management and administration of the eleven energy technology programmes
- An analysis of **Questionnaires** distributed to programme participants. Over 900 questionnaires were sent to the leaders of all projects - both more 'academic' projects (Research Projects) and projects primarily carried out by industry (Industry Projects) – while another 500 plus questionnaires were distributed to a 50% sample of industrial organisations participating in the Research Projects
- An **Expert Panel** mission to Finland, employing the services of six energy and technology policy experts and four evaluation specialists to conduct interviews with programme directors and a sample of participating organisations. Here the task was to review individual programmes, presented and discussed under the broad headings of Relevance, Efficiency, Quality, Effectiveness, Impact and High/Low Points
- A **Peer Review** of selected publications by a team of twelve international scientists. A sample of over 60 publications produced by programme participants during the course of the programmes was reviewed for six of the 11 programmes. The aim here was to complement the mission of the Expert Panel by focusing in more detail on the quality of the work conducted in specific parts of the portfolio

## *9 Appropriate Methodological Matrices*

- A series of **Impact Case Studies** conducted by evaluation specialists. This was an exercise designed to explore the impacts of projects in a variety of company settings in order to understand the implications for future policy

Although the modules were designed to elicit evaluation data at policy, portfolio, programme, project and participant levels, the work culminated in a report focused on the **programme** and **portfolio** levels, in line with the evaluation customer's needs. After presenting the results of each individual module, the results of all modules across the portfolio as a whole were synthesised. Overall performance and achievements were then summarised and a number of prescriptive recommendations made concerning future activities.

### **Evaluation Results**

The headline statement emerging from the review team was that the portfolio of programmes constituted an appropriate course of action for Finland as a whole. Work was performed in an admirable fashion, and expected impacts were largely in line with those normally associated with publicly-supported ventures of this nature across the world. Naturally there was room for improvement, but overall achievements were impressive.

The relevance of the programmes to Finnish needs was in little doubt. The review team considered that the portfolio of energy technology activities supported by Tekes was appropriately focused on areas in which Finnish industry either already occupied a strong position globally or areas in which niche markets could be developed. The sheer scale of the publicly-funded energy R&D effort, the associated development of industrial capabilities and the export possibilities arising from the eleven programmes were particularly impressive.

The calibre of R&D programmes is a function of how well projects are performed and the quality of the resulting work and outputs. For the energy technology portfolio, the review team was convinced that projects were conducted in a satisfactory manner and were generally of a high quality. The review of publications for six of the programmes also confirmed that the work was relevant, well conducted and likely to lead to appreciable impacts, especially in areas related to energy supply and energy and the environment.

Concerning impacts, the review team was satisfied that the substantial and largely successful body of work conducted within the portfolio had had and would continue to have significant impacts on a number of fronts, including the development of new and improved products for export, energy savings and environmental improvements within existing industrial sectors and, perhaps most importantly, the creation of a highly networked cluster of companies and research teams capable of building on these successes in the future.

## *9 Appropriate Methodological Matrices*

Immediate outputs from the Research and Industry projects were in line with expectations for collaborative R&D programmes, with Research Projects geared towards new methods and tests and publications, whereas Industry Projects placed a greater emphasis on new products, pilots, prototypes and processes. These Industry Projects results were sometimes exploited within in-house business units, but in many cases they were being developed further within R&D units prior to subsequent exploitation and evaluated by partner and other organisations. Few participants had experienced significant commercial returns by the end of the programmes, though half expected these to materialise in due course.

For the remainder, however, the returns were expected to be less tangible, primarily because the type of projects in which they were involved were not expected to lead to large, short-term commercial returns. Like many other programmes of this nature, the energy technology portfolio supported relatively few projects designed to bring interesting but risky ideas to the market. Some were less adventurous than might have been expected, but others were more characteristic of collaborative R&D programmes elsewhere. Typically these comprise a project mix in which some projects help connect participants to the national research infrastructure, others allow increased access to problem-solving capabilities, and some enable participants to embark on exploratory projects they could not otherwise contemplate, or allow them to try out new technology concepts.

Despite the limited scale of immediate commercial returns, the majority of participants were satisfied that the benefits of involvement outweighed the costs. They were convinced that the projects in which they were involved had contributed to the realisation of many of the programme goals associated with the portfolio, especially the spread of new knowledge and technology within Finnish industry and the promotion of new collaborative links and networks within the Finnish energy cluster, all of which were expected to contribute to the realisation of economic and exploitation-oriented goals in the longer-term.

Although positive about many aspects of the programmes it reviewed, the review team also had some reservations about the portfolio as a whole. One problem concerned the existence of a project portfolio over-populated with short-term, low-risk R&D projects, primarily a consequence of the dominant role of industry in project selection, with few projects addressing the type of higher-level energy and environment-related issues which necessitate longer-term research agendas. Given the diversity of project types generally supported by collaborative R&D programmes, one would have expected a richer project mix.

The most worrying feature across the portfolio, however, was the low level of additionality associated with many projects and programmes, with little likelihood of the state leveraging appreciable private sector spend on R&D – a national priority at the time. Ideally this occurs when strategically important long-term work

## 9 Appropriate Methodological Matrices

is supported which would otherwise not be carried out, often because the risks and uncertainties are too great. When the state invests in this kind of work, the risk is shared and new avenues explored. If promising results emerge, the way is then open for firms to continue to invest their own resources in the more expensive shorter-term work that follows on the route to exploitation. In this way, a small amount of state investment can leverage a much larger R&D contribution from the private sector. For the Finnish energy technology programmes, the review team was concerned about the low levels of additionality associated with many of the projects. In particular, Table 1 shows that industrial participation in the longer term Research Projects was of low strategic importance to firms, with little prospect of state funding leveraging future spend.

To improve the efficiency, effectiveness and impact of future energy technology R&D initiatives, the review team recommended, *inter alia*, the following courses of action:

- Tekes was urged to ensure the involvement of a broad range of actors in the development of future programmes and in management and decision making processes, especially project selection. In particular, the inclusion of SME, research community and policy/regulatory interests would ensure that future programmes did not simply serve the short-term interests of a small number of large companies
- New programmes, it was suggested, should have a greater focus on the development of breakthrough technologies. Innovative ideas should come to the fore and there should be less ‘business as usual’ R&D funded by the public purse
- The review team called for comprehensive *ex ante* appraisals prior to the formulation of new programmes in order to ensure that eventual project mixes reflected the needs and competence of research actors

## Conclusions

The streamlined approach taken in some of the modules of a Portfolio evaluation almost inevitably means that some corners have to be cut, with corresponding losses in the ‘stand alone’ value of individual components. When the results of the separate modules are synthesised, however, the multi-module approach offers an extremely cost-effective way of combining some of the traditional ‘depth’ associated with the detailed evaluation of published outputs (peer review) and individual projects (impact case studies) with the ‘breadth’ of shallower but more wide-ranging reviews of complete programmes (expert panel overviews). Moreover, the approach allows results to be synthesised in such a way that comparisons can be made across programme portfolios (see Table 2 for a simplified synthesis of the Finnish energy portfolio evaluation results). Armed with comprehensive overviews of this nature, plus a fine-grain appreciation of achievements at lower levels, multi-module approaches lend themselves to

## 9 Appropriate Methodological Matrices

recommendations at project, programme, portfolio and policy levels. Moreover, they offer policymakers concerned with the development of complex innovation policy portfolios a much more useful set of insights than more highly focused attempts to measure impacts via elaborate ‘summative’ approaches.

### References

Guy, K. *et al.* Energy Technology Programmes 1993-1998: Evaluation Report, Tekes: Helsinki, ISBN 951-53-1408-9, 1999

### Acknowledgements

The work upon which this article is based was commissioned by Tekes, The Technology Development Centre of Finland

**Table 1. Additionality for Industrial Participants in Research Projects**

	<b>Pure Additionality</b> (would not have been carried out without the programmes)	<b>Process Additionality</b> (would have been carried out, but with reduced funding, objectives etc.)	
<b>High Strategic Importance</b>	18%	31%	49%
<b>Low Strategic Importance</b>	40%	11%	51%
	58%	42%	100%

**Table 2. Performance across the Whole Portfolio**

<b>Programme</b>	<b>Relevance</b>	<b>Calibre</b>	<b>Impact</b>	<b>Benefits</b>	<b>Additionality</b>	<b>Future</b>
<b>Combustion and Gasification</b>	Broad and High	High	High	Moderate	Moderate	Strong
<b>Bioenergy</b>	Moderately Broad and High	High	High	Moderate	Moderate	Strong
<b>Advanced Energy Systems and Technologies</b>	Broad and High	High in Parts	Moderate to High	High to Moderate	Moderate	Debatable
<b>Energy and the Environment in Transportation</b>	Narrow but High	High to Moderate	Moderate to High	Moderate	High	Moderate
<b>Energy and Environmental Technology</b>	Moderate to High	Moderate to High	High in Future	Low to Moderate	Moderate	Strong
<b>Fusion</b>	Moderate	Moderate	Low to Moderate	Moderate	Moderate	Debatable
<b>District Heating</b>	Moderate to Low	Moderate to Low	Low	Low	High	Weak
<b>Electricity Distribution Automation</b>	Moderate	Moderate	Low in Future	Moderate	Low	Debatable
<b>Sustainable Paper</b>	Narrow and Moderate	Moderate	Low	Moderate	Low	Debatable
<b>Energy Use in Buildings</b>	High	Low	Moderate to Low	Low	Moderate	Moderate
<b>Energy in Steel and Base Metal Production</b>	Low	Moderate	Low	Moderate	Low	Weak

## **Discussion of Ken Guy's paper**

**Maryellen Kelley** (National Institute of Standards and Technology, USA). How do you know what falls in the Devil's Island box? What are their characteristics?

**Ken Guy.** In Finland, there's a desire by some in industry to be a part of firms that are participating in these projects. There's an image associated with participation, partially as a result of the ways programs are selected in Finland. They are not a competitive process but instead based on people bringing their projects and trying to make them work together rather than choosing those that are of a high quality.

**Peter Blair** (Sigma Xi, USA). There's an issue of scale and marginal costs for advancement of technology. Marginal improvements are easier to come by. Does this get accounted for when trying to measure progress?

**Ken Guy.** In most instances they don't have the time or resources to do a full evaluation but instead have only pieces. It works out OK as the information generated is useful.

**Susan Cozzens** (Georgia Institute of Technology, USA). In trying to design an assessment system for the NSF, they had to think in terms of the full portfolio. But as soon as you talk about goals, program-level people start to take apart the goals or say that it's someone else's responsibility. Have you encountered these types of problems?

**Ken Guy.** Not really since they are relying on the expertise of people who put the program together. They were policy analysts. They weren't concerned with the performance evaluation since they were comparing programs. They were focused on a particular set of deliverables and asked experts to look at how the programs compared with each other.

**Susan Cozzens.** So the program level people didn't have to buy-in?

**Ken Guy.** No. Even though benchmarking is important and everyone wants it, we thought we'd get people bragging or would have gotten different picture.

**Arie Rip** (Twente University, The Netherlands). You have a behavioral theory of actors in your approach. In some ways there is a project fallacy: it is not the particular project bringing the impact, but the whole range of activities of the actors. Everything is part of the larger context, this has to be added.

**Ken Guy.** We did this but not just in terms of behavioral theory. We looked at other literature too. The pattern of behavior is not just theory but also based on what they know about how the institution operates.

---

## **Role of Appraisal Monitoring and Evaluation in Policy-Making**

### **John Barber**

Director of Technology, Economics, Statistics and Evaluation - Department of Trade and Industry (DTI), London, UK

Email: JM.Barber@ukgateway.net

---

### **Introduction**

This paper examines the role which appraisal, monitoring and evaluation of Government programmes for supporting Science, Technology and Innovation (ST&I) can play in making of policy in these areas. It is written from the point of view of someone who is involved in the policy process. What is said mainly relates to evaluation of support for Technology and Innovation – a paper that focussed mainly on funding of academic science would have a somewhat different emphasis.

Policy analysts within government, at least in the UK, may generate original ideas without being able to publish but at the same time tend to appropriate new ideas produced by academics without remembering exactly where they came from. Thus if any readers of this paper feel that the results of their research are being used without attribution this could be the case but it could also be true that we in DTI arrived at them independently. In any case the author is heavily indebted to a number of DTI colleagues both past and present and to those academics and consultants who have undertaken evaluations on our behalf or with whom we have frequent dealings. Any mistakes are of course the sole responsibility of the author.

Monitoring and evaluation of ST&I programmes may contribute to policy making in a variety of ways some of which are well established and understood and some which are more problematical. The paper uses the terms appraisal, monitoring and evaluation according to the following definitions:

Appraisal is the ex ante assessment of proposed new programmes to decide whether or not they should go ahead and if so in what form.

Monitoring is the collection of information on the progress of programmes during their lifetime.

Evaluation is the ex post assessment of appropriateness, efficiency and effectiveness of programme after they have ended or have been running for some time.

## *9 Appropriate Methodological Matrices*

Although each of the above is a discrete activity in its own right in order to be fully effective they should be carried out as an integrated process from the initial conception of the programme to the final verdict on its performance. In the UK Department of Trade and Industry this is done via the ROAME system. ROAME stands for Rationale, Objectives, Appraisal, Monitoring and Evaluation. A paper<sup>4</sup> describing the detailed operation of the ROAME system can be found at <http://www.dti.gov.uk/tese>. Note that in ROAME appraisal refers to the rules which govern the selection of individual projects within the programme in question.

### **Knowledge inputs to Science, Technology and Innovation Policy**

Appraisal, monitoring and evaluation are only one input into the body of knowledge, which supports the making of public policy towards Science, Technology and Innovation. Other knowledge inputs include:

- Academic Research into Science, Technology and Innovation as economic and social phenomena and into the operation of policy in these areas;
- Policy analysis by government officials, academics, OECD, consultants etc.
- Benchmarking of national S, T and I performance both in absolute terms and relative to other countries in order to identify the country's strengths and weaknesses in these areas;
- Foresight and similar forms of technology assessment which will identify the opportunities and threats which the country faces;
- Consultation and networking with industry, academia and other participants in the national innovation system.

These other knowledge inputs play a key role in defining the rationale and high level objectives of proposed new programmes and thus set the context and expectations against which the programmes will subsequently be evaluated. They will define the route whereby the direct effects of programmes will be expected to contribute to economic performance. The views which the various stakeholders hold about this 'route' will very much condition how they regard the results of monitoring and evaluation. For example those who are sceptical about the contribution of long term research to economic performance will be difficult to persuade of the value for money of government programmes which appear to stimulate such research. Such beliefs very much conditioned the attitude of many

---

<sup>4</sup> Evaluation of Support by the UK Department of Trade and Industry for the Exploitation of Science Technology and Innovation by UK Civil Industry

policy makers to the evaluation of the UK Alvey Programme undertaken by SPRU and PREST in the late 1980s.

### **Ex-Ante Appraisal**

Appraisal is a key part of the policy decision process but it also establishes the framework within which monitoring and evaluation will subsequently take place. From the latter point of view it has a number of important roles.

First it establishes a rationale why a programme (or policy) should go ahead. The use of financial and other resources to support S, T and I is only justified if it will achieve economic and/or social benefits which would not otherwise be obtained if the programme or policy did not go ahead. In other words without action to overcome some problem or realise some opportunity national economic or social benefits will be foregone. Government action, which did not realise additional welfare, would appear pointless and a waste of resources but the definition and measurement of additionality is a difficult issue, which will not be analysed here.

There are differences across countries as to what does or does not constitute a valid rationale for policy action in this area. In the case of the UK the ROAME system was developed out of a pre-existing framework<sup>2</sup> which justified technology and innovation policy in terms of market failure. Since then the appraisal and evaluation of a great many DTI programmes has shown that the scope of 'market failure' is much wider than the original framework envisaged and goes way beyond the analyses contained in the academic economics literature. Consideration of standard neo-classical market failures such as risk and uncertainty, public goods, externalities, imperfect information and indivisibilities have to be expanded to include, for example, bounded rationality, game theory, the influence of business culture, evolutionary economics and the functioning of large organisations. This would be a rich area for 'Mode 2' research, interdisciplinary, focusing on a practical policy problem but contributing to theoretical knowledge. Most economists, however, would wish to move on quickly perhaps noting, like the early cartographers, that 'Here be Dragons!'

The most important function of the rationale is to identify the problem which the programme is designated to address and high level objectives (GPP, Competitiveness, National Welfare etc) whose achievement will be enhanced by the solution of that problem. The programme must be well designed to achieve that

---

<sup>2</sup> For a description of this framework see 'Current Policy Practice and Problems from a UK Perspective' by John Barber and Geoff White published in *Economic Policy and Technological Performance* Eds. Partha Dasgupta and Paul Stoneman published by Cambridge University Press under the aegis of the Centre for Economic Policy Research, 1987

## *9 Appropriate Methodological Matrices*

solution and there must be a reasonable expectation that the benefits of the programme will outweigh its costs.

Having determined a clear rationale the ROAME system moves on to establish a hierarchy of objectives for the programme starting with the high level objectives referred to in the previous paragraph. These objectives should determine the criteria against which those running the programme will be judged and should be agreed with the latter. Evaluation is a social process and to be effective it requires the co-operation of those involved in the programme itself. This is much more likely to be forthcoming if they have agreed in advance the framework within which the results of their efforts will be assessed. In any case the policymaker will learn little from asking people to do one thing and then judging them against something completely different.

There is also a need to establish a set of operational objectives against which the progress of the programme can be monitored during its lifetime. The monitoring arrangements should be designed to aid programme management and to provide an early warning of problems both to the programme managers themselves and to the policymakers who control budget allocations. They should be designed to provide information of use to subsequent evaluation.

An effective system of ex-ante appraisal will not only underpin the subsequent evaluation of the programme but should also, in UK experience, significantly increase the chance that the programme will be found to have been successful. It will thus greatly help to improve the value for money of public expenditure in support of Science, Technology and Innovation.

### **The role of monitoring and ex post evaluation**

Monitoring and evaluation are designed to provide information on the progress, impacts, results, effectiveness and appropriateness of programmes and policies. The monitoring of the progress of programmes during their lifetime was described immediately above. It will typically be in terms of such operational objectives as the number of applications considered, the number of projects approved, their progress against their technological objectives etc.

The impacts of programmes can be thought of in terms of the activities or variables which it is designed to affect directly for example, the amount of extra R&D stimulated, the number of researchers trained, the number of articles published etc. The results of the programme can be thought of in terms of its effects on higher-level economic and social objectives such as additional sales or profits generated. These higher level objectives will be further along the path towards the high level objectives which were identified in conjunction with the formulation of the programme rationale. However tracking further along that path will require an

## *9 Appropriate Methodological Matrices*

economic model which shows, for example, how increasing the sales of innovative products by the widget industry will increase GDP. Ideally that model should be the same one as was used implicitly or explicitly to formulate the programme rationale. The question of the underlying model is picked up again below.

The effectiveness of the programme can be measured in several ways. One would be to relate its impacts to its costs, for example how much additional R&D per pound, dollar or euro of total programme costs. Alternatively one may try to compare estimates of the economic value of the programme's results with its costs. It will rarely be possible to compare the costs of the programme with the value of its effects on the high level objectives.

A programme can be said to be appropriate if its rationale and high level objectives can be shown to be valid and to have remained valid for the life of the programme and if it had a good strategic fit with that rationale and those objectives. A programme may prove inappropriate because the rationale was misconceived from the start, because circumstances change or because the programme was poorly designed or implemented. For example a programme designed to support a particular type of R&D may be rendered inappropriate because the only applications it received were to undertake R&D of a different type or in a different timescale.

Given the above monitoring and evaluation can serve the following purposes in policy making:

- enable those people managing programmes and implementing policies to make them more effective
- enable those who design new programmes and new policies to make them more appropriate and to increase their likely effectiveness
- enable those responsible for allocating budgets to do so in a way which increases the likely benefit to the economy and society
- demonstrate that money (or other resources) have been well spent and for the purposes envisaged by managers, peers, Ministers, Parliaments and the general public
- provide as a by-product information about the way the economy and society works.

Each of these is considered in turn below.

### **Programme Management**

As indicated above programmes need a clear set of operational objectives including measurable targets whose achievement can be measured during the lifetime of the programme or policy. Information on the achievement of these targets should be collated into a form which is readily absorbed and understood by managers and can be readily communicated to programme funders, evaluators and other stakeholders. Procedures should be put in place to allow mid-course redirection of programmes or policies if monitoring returns indicate that things are not going to plan. At one extreme this may involve piloting of programmes which contain novel features not implemented before.

### **Programme and Policy Design**

The outcome of programme monitoring and evaluation must be effectively communicated to those designing similar or related programmes. This feedback loop is extremely important and some users of ROAME in the UK emphasise this by referring to ROAMEF. In the case of DTI support for technology and innovation all evaluation reports are circulated to the relevant programme committee and the programme committee chairman is required to provide a written statement of resulting action.

In order to foster effective feedback evaluation reports should clearly state findings of wider policy interest and the conditions or assumptions on which those findings rest. Appraisal of new programmes should take explicit account of the findings of evaluations of earlier programmes.

Also programmes should be time limited and follow-on programmes should not be agreed without consideration of evidence on the effectiveness of the existing programme. This will often require evaluations that are carried out before programmes are finished. In some cases it may be appropriate to carry out an interim evaluation around the mid-point of the programme's lifetime.

Limited resources for evaluation should be concentrated on novel programmes to maximise opportunities for learning. Where programmes are particularly novel it may be appropriate to conduct a pilot before the full programme is launched.

### **Budget Allocation**

For a particular budget e.g. Innovation and Industrial Technology or Scientific Research the elimination of programmes yielding poor value for money and the addition of programmes promising excellent value for money will improve the cost-effectiveness of the whole budget. It can also be made more effective by an integrated approach to ex ante appraisal, monitoring and ex post evaluation. This

## *9 Appropriate Methodological Matrices*

not only reduces the number of poorly designed programmes which go into operation but will improve the quality of ex-post evaluation. Feedback of evaluation results into the design of new programmes is an important part of the process. Stipulating that no programme should run for more than a certain period of time without being evaluated will further improve the effectiveness of the process.

In a fast moving world this gradual bottom-up approach needs to be complemented by a more strategic top-down approach. A strategic approach should ideally involve two elements

- (i) A specification of the innovation or economic system whose performance policy makers are trying to improve;
- (ii) A set of indicators which enable one to compare how different programmes (existing or prospective) will influence the variables in this system.

Since we lack a fully specified dynamic long-term economic model, which properly incorporates the role of Science, Technology and Innovation (i) is impossible with any degree of rigour. Descriptive frameworks which incorporate many of the main elements can be produced but typically these cannot provide even moderately reliable quantitative estimates even of the direct impact of one variable on another still less compute the direct and indirect effects of changing one or more variables on the whole system. Similarly the production of a set of indicators which is comparable across a range of programmes is very difficult unless those programmes are similar in their effects. This becomes even more difficult if comparisons are to be made between different major budget areas.

Thus although it is possible, though not always easy, to produce quantitative estimates of the impact or even the results of individual programmes we lack the means to compute the impact on higher level objectives which would allow reliable comparisons between programmes with even moderately different aims. At best we can hope to compare programmes in terms of very rough estimates of value for money based on a separate assessment of each programme in terms of its own costs, impacts and results. Comparisons become mathematically more complex because in the case of innovation particularly, programmes and policies may be mutually reinforcing rather than substitutes for one another so that the effectiveness of one programme will depend on the operation of others.

Given the difficulty of producing reliable measures of value for money for individual programmes any reallocation of budgets based on such measures should be partial and gradual. This conclusion is reinforced by the fact that shutting down of existing programmes suddenly and rapid starting up of new programmes can waste resources.

## *9 Appropriate Methodological Matrices*

There are interactive resource allocation procedures based on subjective assessment of programme effectiveness by experts. These are time-consuming social processes and are not very suitable for the participation of Ministers and other very senior policy makers. The UK National Audit Office (NAO) undertook a comparison of a number of DTI Innovation Programmes using pre-existing evaluation results within a decision theory framework devised by Professor L D Phillips of the LSE. This produced some interesting results but demonstrated the difficulties of comparing programmes which have rather different objectives

As indicated in earlier paragraphs, policy making towards Science, Technology and Innovation should be based on a range of sources of information and not just on the results of monitoring and evaluation. That said it is often difficult for senior policy makers to find the time to absorb and understand all the relevant information and analysis.

### **Demonstration of Value for Money**

There is an increasing demand for quantitative indicators of programme and policy effectiveness for example in UK Government Departments Public Service Agreements (PSA) and the US Government Performance and Results Act (GPRA). These demands tend to emphasise the need for outcome rather than input or process measures. However it may be inferred from what is written above that indicators based on programme impacts are not too difficult to produce, those based on programme results are often feasible but subject to significant margins of errors, while those based on the effects on high level objectives are speculative to say the least. The lack of a fully specified dynamic model of how S,T&I impact on the economy combined with the fact that Government action may be only one factor in the complex processes which determine the outcomes which the programme or policy is designed to influence means that it is impossible to devise reliable indicators of how programmes affect economic performance or social wellbeing. It follows that reliable quantitative indicators will tend to be specific to quite narrowly defined types of programmes and that a set of indicators which covered the full range of government programmes which support Science, Technology and Innovation could be very large indeed and probably not of much practical use.

When faced with these arguments those who argue for quantitative indicators respond that they are only intended to be one of a number of inputs into policy decisions. However long experience, including that of the author, suggests that quantitative evidence tend to be given disproportionate weight in decisions relative to qualitative arguments. Thus policy decisions may be taken on the basis of indicators which can only capture part of the effects of the of programmes and that those programmes whose results are difficult to quantify because, for example, they tend to be long term in nature may suffer.

Also where policy is determined by automatic or semi-automatic feedback from quantitative indicators it will be subject to Goodhart's Law whereby the usefulness of the indicators used tends to degrade over time. For example if allocations of funds to university research institutes is made to depend on their publications record they will find ways of increasing publications without enhancing the underlying quality, quantity and dissemination of their research and with possible deleterious effects on other objectives.

When required to produce quantitative indicators of the quantity, quality and relevance of their research institutes, for example, will soon learn how to give those responsible for the funding of their activities exactly what they are asking for. However it is far from clear that, after an initial positive impact, the request for quantitative measures of performance this will do much to enhance what the nation gets for the resources which it has invested. At worst the use of quantitative indicators may degenerate into a pure public relations exercise.

**To provide information on the way the economy and society works**

Since the routes by which Science, Technology and Innovation affects Economic Growth and Social Wellbeing are not well understood government interventions can be regarded as experiments which will provide evidence on how the whole system works. Although there are some research institutes such as PREST at Manchester and ISI at Karlsruhe, which utilise the results of the evaluations they undertake in their theoretical research, one gets the impression that in Europe at least the results of evaluations are insufficiently used as input to academic research particularly by economists.

---

**Discussion of John Barber's paper**

**Erik Arnold** (Technopolis, UK). There is a problem of reports being put on a shelf. The question is how one takes the next steps so that results can be used. Embedding certain types of design can back this up. Use a group of 3-4 people to go through the design. Build in some capacity for learning within the organization. They are looking for the next steps to embed the results in institutional practice.

**John Barber.** The program committee has collective experience and part of the role of his staff is to interrogate the committee to make sure they hear the advice. It rarely becomes adversarial.

**David Campbell** (Institute for Advanced Studies, Austria). Using the psychology example, to what extent can these be put in a comparative matrix to try to create a bigger impact?

## *9 Appropriate Methodological Matrices*

**John Barber.** They try to look at programs and projects. Currently their data won't allow them to that but it would be useful.

**Philip Shapira** (Georgia Institute of Technology). Should evaluation always be focused towards decision-making? Could it be more critical? Or used to advance learning? Is there a role for this new type of evaluation and should public policy support it?

**John Barber.** There is the problem of using programs for research purposes in UK. It may be more difficult to have researchers look at programs for purpose of generating scientific knowledge. They would likely get turned down but may be able to find alternative funding sources. They would have to ask if OK by firms they are evaluating. There is a lot of academic research and public sector evaluations are very similar.

## 10. Roundtable: Lessons from Existing Networks

Session Chair: Peter Blair (Sigma Xi, USA)

---

Presenters: **Gretchen Jordan** (Sandia National Laboratory, USA)  
**Susan Cozzens** (Georgia Institute of Technology, USA)  
**Gilbert Fayl** (European Commission, DG Research)  
**Stefan Kuhlmann** (Fraunhofer ISI, Germany)

---

*In this roundtable discussion, US and European participants discussed the activities and experiences of networks that focus on the evaluation of science and technology policies.*

**Gretchen Jordan** (Sandia National Laboratory, USA). The American Evaluation Association (AEA) subgroup on Science and Technology evaluation recently reported on network building and the importance of organizing the gray literature. Also, the Association for Policy Analysis and Management (APAM) and the American Society for Public Administration (ASPA) have S&T evaluation sections. In addition, the American Economic Association and the 4S (Society for the Social Studies of Science) have science and technology evaluation tracks.

**Peter Blair** (Sigma Xi, USA). Is there a listserve of some type?

Gretchen Jordan. Evaltalk<sup>1</sup> is one list serve. The sections of many of the referenced groups have their own webpages.

**Susan Cozzens** (Georgia Institute of Technology, USA). The Federal Research Assessment Network (FRAN) reached its high point of activity in 1993-4. FRAN's purpose is for federal agency personnel to share experience in evaluation practice and to move forward the state the art. Conceptual models underlie these processes, but it is also a goal to deepen understanding of the practical problems faced in evaluation. We found that staff members are some of the best people to involve (rather than the decision makers). They will follow the scholarly literature. There is a mailing list for distribution of information.

FRAN arose because of a grant I had in 1992-93. I talked with people in the six major funding agencies including National Science Foundation, the National Institutes of Health, NASA, the Department of Agriculture, and the Defense Department. Before FRAN we found that they were not talking to each other at the time. My job was to facilitate the interaction of S&T evaluators in Washington. FRAN

---

<sup>1</sup> See: <http://www.eval.org/ListsLinks/ElectronicLists/evaltalk.html>

## *10 Roundtable: Lessons from Existing Networks*

met at different agencies (who host the meeting) to talk about what they are doing in evaluation. We maintained regular meetings for a year. But, the agencies were worried about talking about what they were doing in an open network because of the disclosure of future contracts. There was another problem: academics from outside Washington could not attend; but consultants from beltway could. The consultants then began to use this as an opportunity to monitor evaluation policies and actions of the agencies.

Currently, I have a small new grant from NSF to restart FRAN. We may be far enough along on the GPRA curve to talk about evaluation instead of defending against GPRA. As part of this grant, we are going to be setting up a Web page with links to indicators pages and to relevant pages inside and outside government. There will also be a self-subscribing listserv.

**Peter Blair.** How big was FRAN?

**Susan Cozzens.** Fifty or sixty people, mostly from inside the beltway.

**Gilbert Fayl** (European Commission).<sup>2</sup> European-level networking activities should be seen against the background of the Framework program of Research, technology and development (RTD). This is probably one of the largest programs...it is not a coherent program but a fragmented set of sub programs. We are operating in a political environment which limits possibilities and is governed by rules and regulations. We are not developing methods and we should not. We need to rely on the best experts in methods. However, we have to respect the bureaucratic rules.

Participation in networks has two fold rationale: To increase the credibility of the work and at the same time using the EU RTD frame to lend support to the member nation activities. The EU is the secondary player in this respect.

There are three networks of interest:

1. Commission inter-service group on monitoring and evaluation—to coordinate the activities across the programs and across the commission services. This group has made certain achievements. We have succeeded to get certain recognition of our work and carry out monitoring. The group meets seven or eight times per year and identifies key issues.
2. CREST evaluation sub-committee. This is an advisory body to the EU council and Commission regarding science policy issues. Members of CREST are senior people. The subgroup was created in 1995 to carryout discussions and to formulate advice. CREST meets 2-3 times per year. The next meeting is monitoring the entire framework program.

---

<sup>2</sup> See also copy of presentation overheads in Appendix 10-1.

### *10 Roundtable: Lessons from Existing Networks*

3. European RTD Evaluation Network. Is the child of different nets which were established back in the 1980s. Three years ago, we blew life into this network once again. The aim is to promote a dialogue inside Europe and to jointly organize meetings. In this network, there is one representative from each member state and we have invited representatives from the eastern states as well. It includes a total of 41 countries. This net is operating quite well and we organize the meetings in the member state which hosts the presidency. And they have a side meeting specifically on RTD evaluation issues. They make efforts to bring in the best experts. The issue should go beyond the subject of evaluation and push dialogue in the front end of knowledge and impact assessment.

For the internal commission network, we have a network of 30-35 people. For the CREST we have one or several members from the member states (15-25). The central and eastern countries will sit in. And finally we have the Eureka network (15-30).

From all perspectives, we recognize there is a need to promote the evolving improvement of the evaluation systems of European members. The stars are well known. The framework program is a test bed for trying out methods, given the constraints of the political environment. It is important to increase knowledge of what can be expected from evaluation to the political leaders and taxpayers. We recognize, there is an opportunity to have dialogue beyond Europe. Things are slowly moving in international ties outside the European scene. We have begun a dialogue with the NSF and this dialogue should move forward. There is also a dialogue with Japanese.

There is a need to increase dialogue beyond Europe. This transatlantic dialogue is an excellent start. One option is to sit down and think about how to maximize the usefulness of this group. There is being organized a second international conference on evaluation in Belgium next year.

**Peter Blair.** In terms of financing for these groups, you mentioned 50-50 split between the EU and the host country for these conferences?

**Gilbert Fayl.** The commission is paying 50 percent and the host country pays the other half.

**Irwin Feller** (Pennsylvania State University, USA). My question regards European regional technology programs. To what extent does the mandate for Framework link to the regional evaluation?

**Philip Shapira** (Georgia Institute of Technology, USA). Under the transatlantic dialogue between the European members and the NSF, what are the issues being discussed? Where are the opportunities for private and non-governmental initiatives?

*10 Roundtable: Lessons from Existing Networks*

**Gilbert Fayl.** They deal with both sides: government and the private sector. From Europe, the framework has a lot of industrial participation and interest. But there is no strict limit.

**Stefan Kuhlmann** (Fraunhofer ISI, Germany). From a European perspective there are many efforts to link expertise in the area of RTD policy evaluation and in the broader field of innovation research. One can differentiate two kinds of networks: thematic networks and professional networks

1. The “Six Countries Program” was created in 1975. When innovation policies came about, the idea was to bring together experts on the subject, this net played an important role. One institution per country that paid a fee to be a member. The focus was on conceptualisation and repeatedly on the evaluation of RTD programs. Ten countries were eventually included, including several Scandinavian countries, the UK, continental Europe, Hungary and Canada. This group meets twice per year.
2. Since the early 1990s many other European networks were created, all focusing in one way or the other on the conceptualisation and evaluation of research and innovation policies. One could mention e.g. the SPEAR network on socio-economic effects; the HCM network on R&D evaluation; the Advanced S&T Policy Planning Group (ASTPP); RTD evaluation net and various STRATA networks (funded by the Commission and aiming at Strategic Policy Analysis).
3. International evaluation research group INTEVAL—methodological competence in general not just in R&D evaluation.
4. Professional networks: European Evaluation Society (EES; the biennial meeting will be held this October). Individual countries often created their own version, e.g. the UK Switzerland and Germany (German Evaluation Society (DeGEval)).

This looks like an inflation of network initiatives. A critical review would release certain benefits and costs of networks. The benefits include:

- Exchange experience; learn; teach
- Joint efforts toward new or improved concepts or methods
- Joint research projects; international consortia
- Increased influence on actual policy planning
- Institution building and joint research groups, e.g. Technopolis

The costs are:

- Time, money, personal energy

*10 Roundtable: Lessons from Existing Networks*

- Need for minimal management capacities
- Drain of knowledge to competitors for contracting
- Aging boys networks, monopolies

What is the value of one more network originating from this workshop?

- Transatlantic exchange learning projects and the exchange of personnel
- Influence on policymaking
- Gathering and linking experts, academics, other researchers, and policymakers
- Series of workshops and joint publications
- If public funding is sought there needs to be some insurance of independence
- Facilitate joint projects.

**Gilbert Fayl.** To have an active network and ensure its existence, it must have a pragmatic grounding. To have a network that is too specific will lead to death of networks. They have limited possibility to influence policy formulation, and not being in synch with demand of pragmatic work is a danger. The AAAS meeting is useless to me because it is too undirected and large. How can we increase the usefulness for taxpayers of the network. If the dialogue is too diluted the bureaucrats can't use it.

**Peter Blair.** Different parties may have different objectives of participating in a network.

Ken Guy (WiseGuys, UK). There are lots of networks. There may be a variety of networks but we do lack a network that we will all be attracted to and will fulfill our needs. Specifically, there are three needs a network must address:

- School room
- Business incubator
- Playground

The current networks are playgrounds for dinosaurs and do not let in small fuzzy mammals. The SPEAR network was opened up to a large number of members. It was a good marketplace. Customers matched suppliers. You had opportunity as Technopolis did to grow. The age range was young.

**Irwin Feller.** What I didn't see in Stefan's list of reasons to have networks, was to build a unified Europe and to build cross-national collaboration. There are heavy opportunity costs to networks. If you want to build a transatlantic network, it has to

## *10 Roundtable: Lessons from Existing Networks*

be focused and narrow. What problems do policymakers have that are not being met by evaluators? This is an important challenge to the evaluation community. That is concrete and specific. The most productive meeting is focused meetings with tangible products that are valuable to policymakers and the research community.

**Arie Rip** (University of Twente, The Netherlands). In 1990, two other SPEAR networks were formed. There were a number of initiatives of overlapping groups. The lesson learned was that there was enough insight and experience to articulate and the people that were there were the ones needed to make it productive. There may have been a lost generation because it seems like new people did not join, but old ones stayed in. We don't need new networks not even transatlantic or Pacific ones, but we do need to rejuvenate existing ones.

**Maryellen Kelley** (National Institute of Standards and Technology, USA). I visited people around country who were doing research on innovation and I wasn't looking necessarily at policy community. Many were in business schools studying the management of innovation. The American Economics Association has had a meeting including young folks and at the Academy of Management. In the US, we need useful client-oriented evaluations, and this requires people who have client orientations and not strictly academic approaches. At the same time we [ATP] are looking at firms and need people who understand this. Knowledge of firms and innovation were not in policy schools or in economics schools, but in business schools. But, the downside of business school faculty is that they are not interested in policy and it was my mission to involve these people with mixed success. There is a big community growing in business schools. The trick is to attract them to policy issues. We are talking about public-private partnerships and that requires more than just policy knowledge.

**John Barber** (Department of Trade and Industry, UK). There is among the policymaking community a number of countries that want to make progress in evaluation: the Spanish, Italians, and the Japanese for example. There are quite a number of academic groups that need a new peer group to relate to. There is a point about the limits of networks. There are other audiences aside from management folks. The foresight people are an example. Where there is no network we should try to draw those people in.

**Peter Blair.** We could try experimenting with electronic and video conferencing capabilities.

Philip Shapira. In considering whether to initiate a new transatlantic S&T evaluation network, there are three points that should be considered.

1. What issues are not being answered? Suggestions include:
  - Distributive impacts of S&T policies,

## *10 Roundtable: Lessons from Existing Networks*

- Methods of socioeconomic evaluation and technical issues,
  - Institutional issues and the decisionmaker's role in peer review,
  - Regional innovation and top down/bottom up,
  - Assessment approaches for complex science and technology
2. One of the challenges is to get beyond the dog and pony stage to where researchers from Europe and the US can get together in collaborative project efforts. Is the network metaphor a good one for facilitating this kind of project activity? If yes, should we reinvigorate the existing networks or should we establish a new variation? Are there other metaphors like clusters and linkages that can be explored?
  3. A network is a loose thing, while Irwin Feller was talking about a tight project. Is it really possible to undertake networks and projects? If so, it will take funding. The objective of transatlantic cooperation also needs to be considered and it is not clear that this objective can be directly linked to policy decision-making. A transatlantic network might focus instead on the advancement of knowledge, or might be the testing ground of the links between analysis and policy recommendations.

**Nick Vonortas** (George Washington University, USA). The formal agreement between the US and Europe on scientific collaboration seems like a logical venue for transatlantic collaboration. However, the negative of Arie Rip's point about dinosaurs is that when you open the doors to new people they are a different crowd and will view innovation in a very different way. Evaluation is done more or less by technologists.

**Erik Arnold** (Technopolis, UK). It is important to separate networks for policy-making needs on the one hand, and the opportunistic networks on the other hand. Networks survive and prosper when they are based on mutual self-interest. I would like to see some kind of agreement to play with dinosaurs and small furry animals.

**Luke Georghiou** (University of Manchester, UK). We don't have many opportunities for in-depth representation in existing networks. Most of my colleagues only go to these networks occasionally. A large biennial conference could be better.

**Appendix 10-1. Overhead Presentation**

**LESSONS LEARNED FROM  
EXISTING EU EVALUATION NETWORKS**

**GILBERT FAYL  
EUROPEAN COMMISSION, DG  
RESEARCH**

**US-EUROPEAN WORKSHOP  
BAD HERRENALB, 13 SEPTEMBER 2000**

---

  
DG RTD Evaluation Unit

**EU RTD EVALUATION NETWORKS**

1. Within the Commission:  
*Commission Inter-Service Group*
2. Within Member States' Administration:  
*CREST Evaluation Sub-Committee*
3. With European Experts:  
*European RTD Evaluation Network*

The EC RTD Programme Evaluation Unit provides the Chair and the organisational support

---

  
DG RTD Evaluation Unit

**COMMISSION INTER-SERVICE GROUP ON  
MONITORING AND EVALUATION**

- \* Objectives: to co-ordinate within the Commission and establish good practices
- \* Achievements: harmonised programme evaluation approach
- \* Members: representatives of FP Programmes in addition to other relevant Commission Services
- \* Periodicity: 7-8 meetings a year

  
EU RTD Evaluation Unit

**CREST EVALUATION SUB-COMMITTEE**

- \* **Mandate:** to prepare CREST discussion/advice on EU evaluation
- \* **Achievements:** improved knowledge/acceptance of Commission evaluation approach and resulting reports
- \* **Members:** representatives of Member States' administration
- \* **Periodicity:** 2-3 meetings a year

---

  
EU RTD Evaluation Unit

**EUROPEAN RTD EVALUATION NETWORK**

- \* **Aim:** to enhance co-operation/dialogue on evaluation
- \* **Achievements:** - exchange/dissemination of good practices  
- organisation of conferences/workshops
- \* **Members:** selected European experts
- \* **Periodicity:** 2 meetings a year

---

\* **Future prospects:** *establish structured dialogue with experts from USA, Japan, etc.*

---

  
EU RTD Evaluation Unit

**LESSONS LEARNED**

- \* Transfer of good practices in RTD evaluation should not be about naming champions but rather reinforcing the ongoing systemic improvement and co-operating among relevant actors. The EU RTD Framework Programmes provide the largest 'test bed' for trying out methodologies.
- \* Evaluation is only a small, but important, component of the Governance system. Its promises and expectations on what and when it can deliver should reflect this. Therefore:
  - Dialogue and networking across interests and professional specialisation is necessary;
  - Training of young scientists/administration in RTD management, including evaluation, is of paramount importance.
- \* There is a need for structured dialogue between European and non-European experts. But do not proliferate initiatives/meetings.

## **11. Issues in Linking RIT Evaluation With Policy**

**Chair: David Guston (Rutgers University, USA)**

---

### **The Triple Helix of University-Industry-Government Relations: Implications for Policy and Evaluation**

**Henry Etzkowitz**

Science Policy Institute

State University of New York at Purchase, USA

Email: [Henryetzkowitz@earthlink.net](mailto:Henryetzkowitz@earthlink.net)

---

#### **Introduction**

The US has strongest industrial policy in the world, bar none. The US is acting from the bottom up, sideways, criss-cross as well as top down, although in an indirect and hidden manner. Europe has a long way to go to catch up with the US because it has only emphasized certain limited areas of intervention, primarily from the top down. Nevertheless, other forms of intervention such as bi-lateral initiatives are beginning to appear, especially in cross-border regions such as Oresund (Copenhagen/southern Sweden). University-industry relations are gaining strength in regions, such as Sienna, where government industry relations previously predominated.

In Europe, the US, Latin America and Asia, issues of knowledge and technology transfer have moved to the forefront of attention in economic, social and industrial policy. As the sources of future development increasingly derive from innovation, attention must be paid to non-traditional sources that have the potential to become the basis for construction of new business and social models as well as the renovation of old ones.

The National Systems of Innovation (NSI) approach is especially well suited to analysis of bounded phenomena, within nations or individual firms. Although other sources are taken into account, incremental innovation is viewed as primarily occurring within the firm, through various forms of learning (Lundvall, 1988). A different model of the sources of innovation is required to account for discontinuous as opposed to incremental innovation.

Innovation is increasingly likely to come from outside of the individual firm or even from another institutional sphere such as the university where the focus of attention is on original path breaking developments, whether in science or technology. It was not an accident that U.S. universities were favored over government and industrial laboratories as the site for path-breaking military R&D during the Second

World War. Moreover, it can be expected that discontinuous innovations, which originate in a company, are more likely to be utilized in a different environment where the blinders of current taken for granted practices or commitment to existing technologies and products are less likely to have effect.

As innovation moves outside of a single organization, lateral relationships across boundaries, rather than hierarchical bureaucratic structures, become more important. To both analyze these developments and guide their future development, a new model of the relationship among the institutional spheres and their internal transformation is needed. In this article, I outline a model that takes account of border crossing and the co-evolution between technological and institutional transformation as well as a regional research project to elucidate these processes.

### **The Triple Helix Model**

The "triple helix" is a spiral model of innovation that captures multiple reciprocal relationships at different points in the process of knowledge capitalization. The first dimension of the triple helix model is internal transformation in each of the helices, such as the development of lateral ties among companies through strategic alliances or an assumption of an economic development mission by universities.

The second is the influence of one helix upon another, for example, the role of the federal government in instituting an indirect industrial policy in the Bayh-Dole Act of 1980. When the rules of the game for the disposition of intellectual property produced from government sponsored research were changed; technology transfer activities spread to a much broader range of universities, resulting in the emergence of an academic technology transfer profession. The third dimension is the creation of a new overlay of trilateral networks and organizations from the interaction among the three helices, formed for the purpose of coming up with new ideas and formats for high-tech development.

The triple helix denotes the university-industry-government relationship as one of relatively equal, yet interdependent, institutional spheres which overlap and take the role of the other. There has been a movement from separate institutional spheres, which represent, at least in ideology, the US situation. There has also been a shift from the model of the state encompassing industry and academia, in its strongest form in the former Soviet Union but versions could also be found in Latin American and European countries.

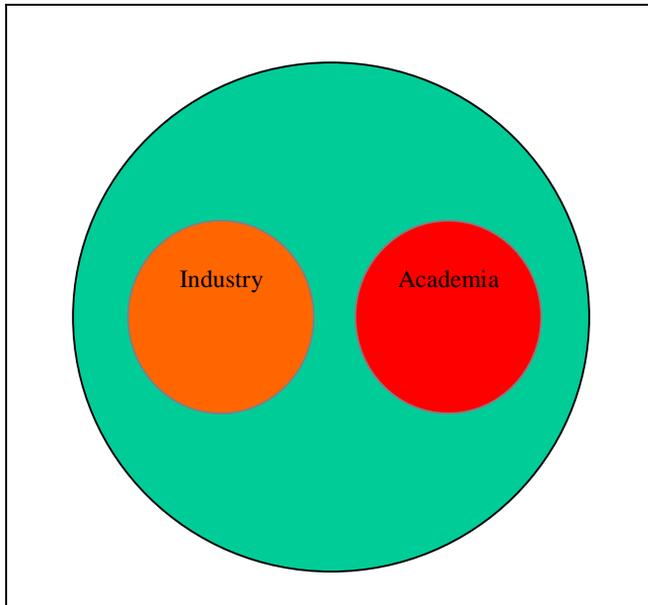
Bilateral relations between government and university, academia and industry and government and industry have expanded into triadic relationships among the spheres, especially at the regional level. Academic-industry-government relations are emerging from different institutional starting points in various parts of the world, but for the common purpose of stimulating knowledge-based economic development. Older economic development strategies, whether based primarily on the

industrial sector as in the U.S. or the governmental sector as in Latin America, are being supplemented, if not replaced, by knowledge-based economic development strategies, drawing upon resources from the three spheres.

A new institutional configuration to promote innovation, a “triple helix” of university, industry and government is emerging in which the university displaces the military as a leading actor. The dynamic of society has changed from one of strong boundaries between separate institutional spheres and organizations to a more flexible overlapping system, with each taking the role of the other. The university is a firm founder through incubator facilities; industry is an educator through company universities and government is a venture capitalist through the Small Business Innovation Research (SBIR) and other programs (Etzkowitz, Gulbrandsen and Levitt, 2000). Government has also encouraged collaborative R&D among firms, universities and national laboratories to address issues of national competitiveness (Wessner, 1999).

This is a different model of the relationship among the institutional spheres either than one in which the spheres are separate from each other and do not collaborate or one in which one sphere dominates the others. This picture, for example, depicts a model in which the state incorporates industry and the university (Figure I). This would represent the Former Soviet Union and some Latin American countries in a previous era, when state owned industries were predominant.

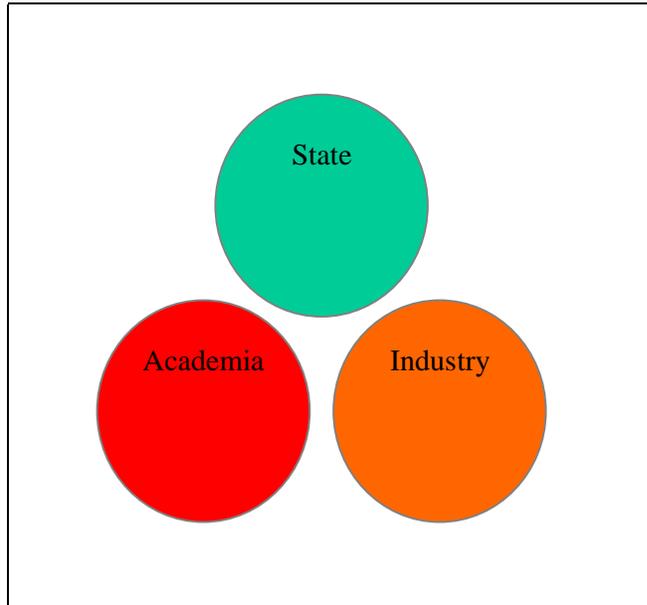
**Figure I**



*11 Issues in Linking RIT Evaluation With Policy*

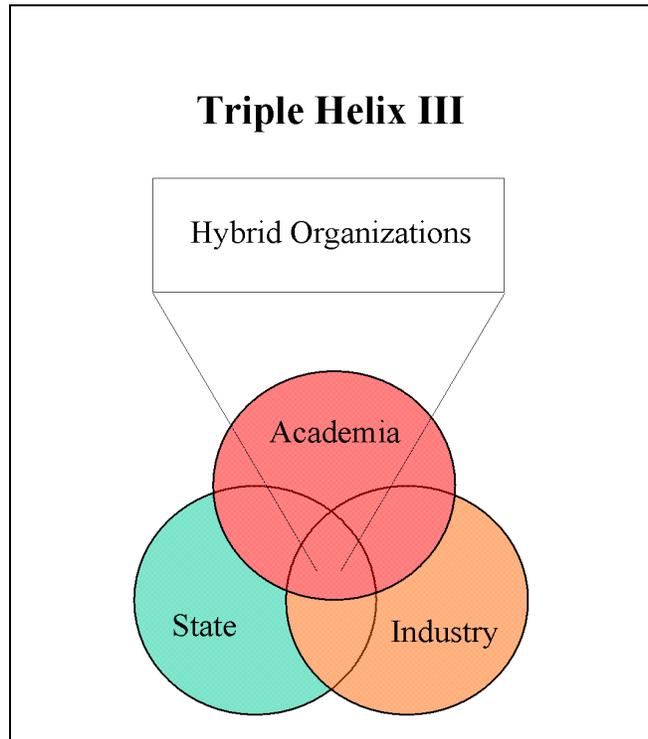
The model of overlapping spheres is also different from the model of institutional spheres as separate from each other, which, at least in theory is how the US is supposed to work (Figure II).

**Figure II**



From each of these previous models, whether it was the state dominating the other institutional spheres or the spheres separate from each other, we are moving to a model where the institutional spheres overlap and collaborate and cooperate with each other (Figure III).

**Figure III**



### **Normative Implications**

The triple helix model of innovation, with converging institutional spheres of academia, industry and government each taking the role of the other has been read in different ways in various parts of the world. In countries where the interface is well underway, whether occurring from the bottom up, through the interactions of individuals and organizations from different institutional spheres, or top down, encouraged by policy measures, the triple helix can be recognized as an empirical phenomenon. The US has been seen to exemplify the former and Europe the latter mode of triple helix development (Viale, and Campodall' Orto, 2000).

Both types of triple helix development may actually be under way in the US and Europe albeit at different rates and with varying emphases. Top down processes can be identified in the US. even through they are often hidden behind "bottom up" formats. Thus, Advanced Technology Program (ATP) program managers at the National Institute for Standards and Technology have been known to seek out technical leaders in industry to encourage them to initiate an "industry led" focus program. Nevertheless, as industry takes on the project as its own and draws academics as well, or vice versa, who can say where top down ends and bottom up begins. It may be more accurate to recognize both processes going on simultane-

ously and in tandem. Indeed, such a dual track for innovation promotion may be more productive than any single path.

Similarly in Sweden when young computer and business consultants join together to form an e-commerce firm, a new development is at hand in a society whose industry was led by a definable group of large firms for several decades. Certainly government supported entrepreneurship programs and incubator facilities are available to support these initiatives. Once again bottom up meets top down in a creative fashion, creating a broader context for innovation than would likely arise from either approach in isolation.

In other parts of the world, Latin America, for example, where industry and university have traditionally existed apart from each other, with academia as part of the governmental sphere, the triple helix is sometimes taken as a normative model. Some view it as a goal to strive for in bringing about change to enhance the prospects for innovation. Other observers see the coming of the triple helix as representing the downfall of the existing system of innovation, represented by government owned corporations sponsoring laboratories adjacent to university campuses.

Privatization of companies, it is believed, will reduce the resources available for R&D, including collaborations between the state-owned company laboratory and university researchers. On the other hand, many of these collaborations were not sufficiently market driven and resulted in innovations that lacked a context to be put to use, having been based upon a negotiation between two public laboratories, neither of which was closely enough tied to production and use (Mello, 1998).

This gap is not only a peculiarity of Latin American public research but has been noted in the large corporate laboratories in the U.S. that had been separated from production facilities and were operating as isolated entities, until quite recently. In the later case the reintegration of the laboratory into the firm and directing it more closely toward company goals has been occurring at IBM and GM, in recent years. Typically as corporate R&D facilities are moved closer to product development, longer term R&D is conducted in collaboration with other firms, university research groups and government laboratories.

### **Policy Implications: The dynamics of innovation spaces**

The level (multi-national, national and regional) is also to be taken into account. At the regional level, one can also look at this overlapping of institutional spheres as involving knowledge, consensus and innovation spaces, created at the intersection of the spheres. There is no necessary order to this sequence. A reversal of traditional orders of staged development is among the possible outcomes. Any one can be the basis for the development of the others but a fully developed triple helix will eventually comprise all three elements.

These spaces are created as a consequence of a change in values among promoters of regional economic development from a sole focus on “business climate” and subsidies to firms to creating the conditions for knowledge-based economic development. One indicator of this shift is the increased involvement of universities and other knowledge producing and disseminating institutions, such as Academies of Science, in regional development. The first step in a three-stage process of knowledge-based economic development is the creation of “knowledge spaces” or concentrations of related R&D activities in a local area. The existence of such “reticulated” agglomerations has been identified as a precursor to knowledge-based regional economic development (Casas, Gortari and Santos, 2000).

### **Knowledge Space**

The concept of knowledge space was developed by Dr. Rosalba Casas at UNAM as a way of conceptualizing some of the decentralization of research institutes from Mexico City to other regions of Mexico. This provided a base for the development of research projects and new technology related businesses in areas of the society which had not previously had this potential.

Some of this decentralization took place due to the earthquake. Other because it was decided it was not best to keep everything concentrated in one place but to move some of the technical resources to other parts of the society. Nevertheless, just as the existence of research universities by the 1920's and 30's in the US represented a potential for knowledge-based regional economic development, similarly these research institutes moved to other parts of Mexico still only represent a potential until they are put to use.

### **Consensus Space**

How are knowledge spaces transformed from potential to actual sources of economic and social development? The second stage is the creation of a “consensus space” a venue that brings together persons from different organizational backgrounds and perspectives for the purpose of generating new strategies and ideas. The concept of knowledge-based regional economic development is derived from activities of the New England Council, representing academic, business and political leaders. Based on the formation of firms from research at MIT in the 1920's, MIT President Karl Compton proposed to utilize the region's comparative advantage, its extensive academic base, to systematically create new firms from scientific research (Etzkowitz, In Press).

After reviewing the existing ideas for economic development, which were typically to reduce taxes or to attract branch plants, it was realized that these approaches would not work in New England because they were too far from raw materials and

distribution distances were too long. However, the special resource that the region had were its universities, such as Harvard and MIT in the Boston area, and examples of new firms that had been started from universities. There were only a few individuals who had started such firms so the idea was to establish an organizational support structure to promote high-tech firm formation.

In the 1930's, New England business and political leaders were open to new ideas, given the failure of traditional business models of regional development. Joint Venture Silicon Valley (JVSV), through its open regional "brainstorming sessions," played a similar creative role in Silicon Valley during the recession of the early 1990's (Etzkowitz, 1998). The New York Academy of Science has recently taken this role in the New York metropolitan region, drawing together a leadership group from different institutional spheres for a series of discussions (Raymond, 1996).

### **Innovation Space**

The third stage is the creation of an "innovation space" a new organizational mechanism that attempts to realize the goals articulated in the consensus space. From the analysis of the resources in a region and the creation of a consensus space bringing the different actors in a society together, a new innovation space was created, in this instance the venture capital firm to provide business advice, technical assistance and financing to start new firms.

In 1946 the first venture capital firm the American Research and Development Corporation (ARD) was founded which acted more or less as an incubator for these new firms in helping them with business and technical advice as well as financing. The "incubator" was not a formal entity at that time. Some of the firms were established in underutilized spaces on the MIT campus, in an informal adumbration of the later incubator concept.

Hybridization of organizational roles and functions, arising from the interaction that occurred in the consensus space is an expected development. The new institutional mechanism is typically a "hybrid organization," synthesizing elements of theory and practice from the different spheres. In the case of ARD, the elements were drawn from academia (MIT and the Harvard Business School), the financial industry (investment trusts and investment clubs) and government (changes in regulatory practices defining risky investments).

<b>Summary: conceptual framework for knowledge-based regional economic development</b>	
<b>Stage of development</b>	<b>Characteristics</b>
Creation of a <i>knowledge space</i>	Focus on “regional innovation environments” where different actors work to improve local conditions for innovation by concentrating related R&D activities and other relevant operations
Creation of a <i>consensus space</i>	Ideas and strategies are generated in a “triple helix” of multiple reciprocal relationships among institutional sectors (academic, public, private)
Creation of an <i>innovation space</i>	Attempts at realising the goals articulated in the previous phase; establishing and/or attracting public and private venture capital (combination of capital, technical knowledge and business knowledge) is central

### **Public Venture Capital**

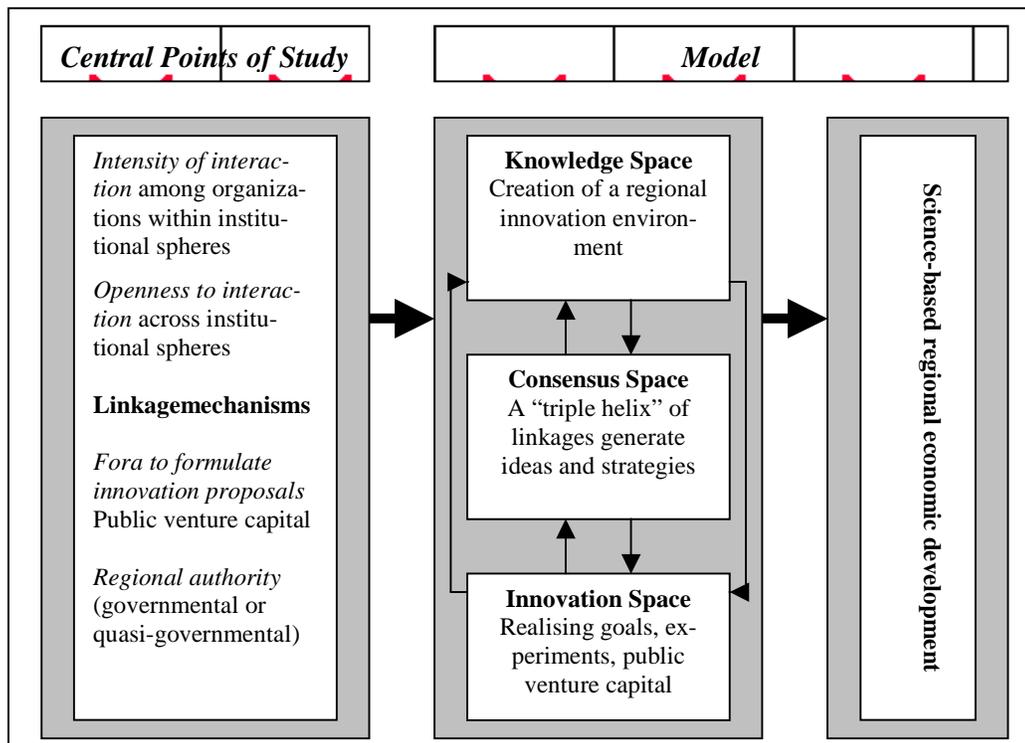
The invention of new ways to promote knowledge-based regional economic growth and the adaptation of old mechanisms to new circumstances is an on-going process. For example, the incubator facility has been creatively revised from an on-site support structure for new firms into a mechanism for linking proto-firms to available resources in a region. In the former case, a greenfield site had few available resources; in the latter a declining industrial region had a variety of financial and business instruments available but lacked the means to connect them to new ventures. Thus, an organizational entity with the same name can play quite different roles in different circumstances. Of course, the possibility also exists for a mismatch between innovation mechanisms and regional activities. It is the task of qualitative research, through in-depth interviews and focus groups to tease out the differences between such situations.

These integrating entities go beyond the activities of traditional boundary spanning mechanisms such as technology transfer offices that arrange interaction across delimited boundaries. Encouraging the establishment and extension of the activities of both these older (boundary spanning) and newer (integrative) linkage mechanisms have become part of the organizational strategy of regional groups that are established with the intention of intensifying the process of knowledge based economic development.

In recent decades, federal, state and local governments have created a variety of mechanisms to encourage knowledge-based economic development. These initia-

tives include the supply of bridging funds, grants and matching funds to support R&D and access to participation in joint projects with government laboratories. Public venture capital is a subset of “public investment,” a rationale for support of various government initiatives that enhance the health, education and welfare of the population.

These programs have in common the commitment of public funds to support the entrepreneurial development of technology in situations where private venture capital finds it too risky to venture. The gap between the creation of intellectual property and its translation into products and processes has been called “the valley of death.” The use of public resources to reduce risk in the development of new technology has long been accepted in the agricultural, military and health areas. In recent years, marked by controversy, public entrepreneurs have extended the role of government from the macro factors affecting innovation such as interest rates and money supply to the micro conditions of innovation.



**Implications for Evaluation**

The triple helix also has implications for evaluation method as well as for what is evaluated. Arie Rip’s slide of a falling ivory tower overlaid on an image of an intertwined “triple helix” in the opening talk of the Workshop exemplifies the changing context of evaluation. There is a shift in evaluation from an internal organizational focus to what is happening at the interface. This shift affects both what is evaluated and when evaluation takes place. It includes a shift from

## *11 Issues in Linking RIT Evaluation With Policy*

is evaluated and when evaluation takes place. It includes a shift from autonomous to interdependent institutional spheres, with the quantity and quality of these relationships seen as more significant. The other indicator of transformation came from evaluators themselves who noted the consequences of “overnetworking,” the burden of having to travel to too many meetings.

Evaluation needs to be focussed not only on what is happening within an organization in meeting its goals but in interaction with other organizations. This becomes especially clear in EU evaluation where the quality of the network and increasing interactions through the network for purposes of enhancing cohesion, breaking down national boundaries may be as important as R&D outcomes.

In addition to tension and conflict of interest, there is also convergence and confluence of interest. “Priorities have to be set no longer solely on intellectual grounds, but also with an eye to the resources available, the research agenda of the receiving units, and the ex ante assessment of the likelihood of success. (Leydesdorff, 2000). Evaluation becomes multi-valent as differing perspectives and success criteria must be taken into account. For example, the expected tradeoff between research quality and cohesion may become mutually reinforcing, with better results achieved on both scores.

Networks also play a role in “teaching and learning” including partners from developing regions is a way of raising their level through their participation with more experienced partners. These side-benefits of network participation need to be taken into account in evaluating networks. US evaluation of networks, such as those sponsored by the ATP, tends to be technocratic, focused on specific technological outcomes and spillovers to the virtual exclusion of human factors.

The interaction within a network may be as important as the product of the interaction. Beyond ex ante decisions about whether to proceed with a project, or ex post assessment of results, evaluation has increasingly moved to a “real time” mode of analyzing and benchmarking social processes as they take place and providing “feedback” for course corrections.

As evaluation attempts to capture social processes as well as inputs and outputs it draws upon social science analysis and techniques and more closely resembles other social research practices. The convergence of evaluation and general social science research is part of a broader movement toward the intersection of basic and applied research, a reflection in the social sciences in general, and innovation studies in particular, of a shift in social structure.

### **Regional Synergies**

Perhaps the most significant development in social structures in recent decades both Europe and the U.S. is the rise of innovative regions, based on various knowledge and technology bases. Traditionally, regions were based on natural geographical characteristics such as a watershed e.g the Mississippi Delta or a cultural area, with a common social characteristic such as slavery in the ante-bellum American South (Odum, 1936). More recently, Regions have been organized for functional purposes, whether it is economic development or flood control.

Typically, regions lack political boundaries although over time they may develop a quasi-political infrastructure. This can take place through informal councils based upon self-selection and co-optation such as the Pittsburgh High Technology Council and special districts such as the New York Port Authority based on compacts negotiated between adjacent political entities, in this case the states of New York and New Jersey.

The regional level offers an opportunity to get closer to the user perspective. If you take the viewpoint of the individual U.S. federal or European Union program it is difficult to see how various initiatives fit or do not fit into a coherent picture. The evaluator or policy analyst comes to this realization by taking the perspective of the user of various programs. In the U.S. some of the pieces include various state programs, which often fill gaps in the federal programs. These users on the regional level begin to fit the pieces together and see the broader picture.

### **Conclusion: The Triple Helix in Regional Development**

A trilateral series of relationships among industries, governments and universities is emerging in regions at different stages of development and with different inherited socio-economic systems and cultural values. As regions seek to create a self-reinforcing dynamic of knowledge-based economic development, the three institutional spheres are each undergoing an internal transformation, even as new relationships are established across institutional boundaries, creating hybrid organizations such as technology centers and virtual incubators.

The new networks within a region, established by means of concerted tripartite interactions, may allow the emergence or renewal of high-tech complexes and the creation and organization of new industrial sectors. Academic-industry-government cooperation requires new learning, communication, and service routines on the part of institutions that produce, diffuse, capitalize, and regulate processes of generation and application of useful knowledge. The paradigmatic institutions are the university, the firm, and the government, and the paradigmatic relationship is interactive concerted action embedded in projects, communication, and new kinds of shared values.

## *11 Issues in Linking RIT Evaluation With Policy*

A university-industry-government interaction at the regional level is not an entirely new phenomenon. The post-war "Route 128" high-tech conurbation can be traced to policy initiatives in the 1930's and even to the founding of MIT in the mid-nineteenth century for the purpose of infusing industry with new technology. What is new is the spread of technology policy to virtually all regions, irrespective of whether they are research or industrially intensive.

A normative injunction to attend to the commercial implications of research has arisen not only from the emergence of an entrepreneurial dynamic within academia but from government policies that changed the rules for disposition of intellectual property arising from federally funded research. Taking organizational forms such as technology transfer offices and the requirements of government granting programs for the support of research; the capitalization of knowledge changes the way that scientists view the results of their research.

Within specific regional contexts universities, governments and industry are learning to encourage economic redeployment through the development of loosely coupled reciprocal relationships and joint undertakings. For this to happen a local region must have some scientific and technological institutions and have produced or obtained access to other necessary kinds of innovation supporting instruments such as investment mechanisms and institutions to promote concerted action.

## **Appendix I: Outline for a Europe/U.S. Collaborative Research Project: Knowledge-Based Regional Economic and Social Development**

**Henry Etzkowitz and Magnus Gulbrandsen**

### **Introduction**

The following document is an outline draft proposal for an ongoing collaborative research project for the "Bad Herrenalb Group," to conduct a comparative analysis of regional levels in the US and Europe and compare the effect of the framework programs at the regional level to the state programs. In the US, it is difficult to arrange for the individual state programs to be evaluated in a comparative fashion. At the most the Southern Technology Council will examine the southern region but it will be very difficult to go beyond that.

Europe could learn from the state programs in the U.S. and there is much that the U.S. could learn from the framework programs. The premise of this study concept is that a combination of insider and outsider perspectives will lead to greater insight. Insiders are usually the object of study and outsiders the researchers. In this scheme, insiders and outsiders (European and US researchers), reverse and combine roles as they engage in collaborative studies of each other's regional innovation environments.

This study will use the "triple helix" model to analyze new linkages that transcend the traditional "contract" between science and society. The purpose of this research project is to elucidate some of the underlying dimensions on which future evaluations, policies and programs can be based. A comparative scheme is proposed to investigate the pre-conditions for success of the European Framework programs and State S&T programs. Both are oriented toward the regional level, although one derives from the multinational and the other from the sub-national levels. Both are valued at approximately US \$3.5 billion.

### **Theory**

The conceptual framework is based on three main and related nested ideas:

- (1) the "triple helix" which refers to the multiple reciprocal relationships among institutional sectors (public, private and academic) at different points in the knowledge capitalization process (knowledge, consensus and innovation spaces);
- (2) The concept of "regional innovation environment (RIE)", which consists of the set of political, industrial and academic institutions that, by design or unintended consequence, work to improve the local conditions for innovation (knowledge space);

- (3) The concepts of social capital and embeddedness which refer to the density of social relationships and trust in interpersonal relationships. We extend these concepts across institutional boundaries in inquiring into the conditions for production of social capital and trust across institutional spheres, allowing lateral rather than hierarchical coordination (consensus space)

**(1.0) Objectives:**

The purpose of this project is to produce useful knowledge about emerging forms of knowledge-based economic development in the United States and Europe and value changes that occur as institutional spheres interact more intensively. It is expected that significant differences can be identified in conditions, intentions, policies, mechanisms, processes and outcomes exist in regional innovation experiences.

*(4.1) General objectives:*

This study will:

1. elaborate the concepts of “knowledge” “consensus” and “innovation” spaces based upon data collected from a sample of regions with different conditions: high-tech; declining and excluded
2. explore the tensions, complementarities, interfaces and linkages between the academic, industrial, and governmental spheres in different kinds of innovation complexes.
3. analyze the intended and unintended effects of national policies and programs (public venture capital)) in the regional innovation environment.
4. identify and understand the conflicts of interests and tensions that the "triple helix" model implies at a regional level.
5. analyze bilateral interactions between academia and industry (e.g. technology transfer offices and firms) and government and academic (e.g. local economic development agencies and universities) as a precursor to trilateral interactions.
6. analyze tripartite concerted action (i.e. the formation of High-Tech Councils and other organizations) at the local level, as an impetus to knowledge-based economic development.

## *11 Issues in Linking RIT Evaluation With Policy*

### *(4.2) Specific Objectives:*

- To map the quality of the innovation environment of the regions both, in terms of the existing elements (economic, political, cultural, academic and organizational), and in terms of new formal mechanisms and institutions.
- To identify the factors and conditions that allow or limit multiple reciprocal linkages between academic institutions, industry and government in a regional context.
- To identify the actors and their backgrounds and initiatives in the innovation process i.e. identify “entrepreneurial gatekeepers” that span institutional boundaries.
- To identify the emergence of a new group of knowledge-based technologies and their related industrial sectors within each region.
- To identify and analyze informal and formal mechanisms of academy-industry and government concerted actions for innovation.

### **Research Questions**

Value conflicts in inter-organizational relations are typically generated either from the explicit statement of values to justify intended changes in policy or from the initiation of actions which are later realized to have important consequences for values. Studies of strategy formation in government and business have suggested the analytic utility of making a distinction between intended and realized strategies i.e. between organizational goals explicitly set forth and an underlying stream of actions which result in an implicit policy, which the organization often does not wish to recognize (Mintzberg & Waters 1985). How are value issues of university-industry-government relations defined by representatives from different sectors of the region? What are the themes used to legitimize the new regional ties? How are value conflicts emanating from university-industry-government interactions resolved?

What are the effects of the state programs in the U.S. and the Framework Programs in Europe on Regional Innovation Environments?

Are the US and European programs functionally equivalent, even though they derive from different levels, although both are above the regional level, albeit to significantly different degrees.

What are the differences and similarities in U.S./European experience in promoting innovation through key actions and public venture capital.

Does the US represent a “bottom up” form of the “triple helix” model and Europe a “top down” version or can both processes be identified in various formats.

### **(5.0) Main Hypotheses**

Is the existence of certain level of R&D activity a sufficient impetus to science-based regional development, the “island of innovation” hypothesis or must these activities be significantly related to each other in order to induce a “critical mass” of development activity, the “cluster” hypothesis?

Do these various regional experiences suggest the existence of alternative models of regional development or are they merely at different stages of a common process? For example, is there a single line of knowledge-based economic development based on the venture capital model or multiple cultural formats unique to each region (Saxenian, 1993).

In some areas, regional innovation systems may have to cope with traditions of isolation that still prevail in several economic, social and political spheres, that may inhibit the development of a regional innovation environment. Insular institutions and weak networks among institutional spheres are suggested as hypotheses to explain relatively low levels of high technology development in some research-intensive regions.

### **(6.0) Research Design**

Based on the ideas of knowledge, consensus and innovation spaces as the providing the basis for knowledge-based regional economic development, a sample will be drawn that will allow us to analyze the development of these “spaces” under contrasting conditions and at different phases of their development.

A sample will be constructed to include regions with different characteristics that are hypothesized to influence variants of the model such as research intensity. The objective will not be to conduct regional studies but rather to focus on the emergence of networks and organizations that facilitate science-based economic development. The analysis will concentrate on the implications of these linkages among academia, industry and government that formerly operated at arms length but are increasingly working together to promote innovation.

## **References**

- Braczyk, Hans Joachim et.al. eds. 1998. *Regional Innovation Systems*. (London: UCI Press)
- Casas, Rosalba, Rebecca de Gortari and Ma. Josefa Santos "The Building of Knowledge Spaces in Mexico: a regional approach to networking" *Research Policy* 29 225-241.
- Etzkowitz, Henry. In Press. *The Second Academic Revolution: MIT and the Origins of the Venture Capital Industry*. (London: Gordon and Breach)
- Etzkowitz, Henry and Magnus Gulbrandsen and Janet Levitt. 2000. *Public Venture Capital: Government Funding Sources for Technology Entrepreneurs* (New York: Harcourt)
- Etzkowitz, Henry. 1998. *From Zero-Sum to Value-Added Strategies: The Emergence of "Knowledge-Based Industrial Policy in the States of the United States"* *Policy Studies Journal*,
- Etzkowitz, Henry and Loet Leydesdorff. Eds. 1997. *Universities in the Global Knowledge Economy: A triple helix of university-industry-government relations* (London: Cassell)
- Leydesdorff, Loet. 2000. "The promotion of university-industry-government relations: a methodological contribution to its evaluation" *National Academy of Sciences Conference on "Policies to Promote Entrepreneurship in A Knowledge-Based Economy: Evaluating Best Practices form the U.S. and the U.K.*
- Lugar, Michael. 1994. "Science and technology in regional economic development: the role of policy in Europe, Japan and the United States. *Technology in Society*, 16: 9-33.
- Lundvall, Bengt-Ake. 1988. "Innovation as an interactive process: from user-producer interaction to the national system of innovation" In Giovanni Dosi et.al. eds *Technical Change and Economic Theory* (London: Pinter)
- Mello, Jose. 1998 "Management of Technological Diffusion: A Case Study in a Brazilian Public Technological Institute. Presented at the Seventh International Conference on the Management of Technology, Miami, Florida
- Mintzberg, H. & Waters, J. A.: *Of Strategies, Deliberate and Emergent*. *Strategic Management Journal*, Vol. 6, 1985, pp. 257-272.

Odum, Howard. 1936 Southern Regions of the United States (Chapel Hill: University of North Carolina Press)

Raymond, Susan ed. 1996. The Technology Link to Economic Development. New York: New York Academy of Sciences

Saxenian, AnaLee. 1994. Regional Advantage: Culture and Competition in Silicon Valley and Route 128 (Cambridge: Harvard University Press)

Viale, Riccardo, and Sergio Campodall'Orto. 2000. Neocorporatist or Evolutionary Triple Helix? Suggestions Coming from European Regions" Presented at the Third Triple Helix Conference, Rio de Janeiro

Wessner, Charles ed. 1999. The Advanced Technology Program: Challenges and Opportunities. (Washington D.C. National Academy Press)

---

## **Discussion of Henry Etzkowitz's paper**

**Arie Rip** (Twente University, The Netherlands) The triple helix presentation is different from Georghious's because he emphasizes the iron triangle more. What do you think of his analysis? Also, indirect S&T policy in the U.S. makes it difficult to evaluate it. You need to be a critical historian instead of a direct evaluator.

**Henry Etzkowitz.** Exactly. You can't do direct evaluation but have to look at all of the pieces to see where they fit at the state level because they have to fill in the gaps from the federal level. Comparative analysis of state and regional levels on an international focus is my proposal for future collaborative work among the U.S. and Europe.

**David Guston** (Rutgers University, USA). In the U.S., their covert micropolicy aims have additional goal of preventing brain drain from the federal labs that are not articulated in the letter of the law. In order to get at these goals you have to be a critical historian and to get at that you have to potentially expose political softness in the legislation.

**Henry Etzkowitz.** I don't think exposing these things is too much of a problem. It's important to show the effects - intended and unintended. The U.S. has the strongest industrial policy, while the Europeans do not. Europeans should study the U.S. policy focus.

**Louis Tornatzky.** For the comparative analysis question, STC studies do not produce robust data.

## *11 Issues in Linking RIT Evaluation With Policy*

**Henry Etzkowitz.** This is the opportunity because there are usually very spotty studies at the state level. We won't understand it unless it's done on a broader perspective. The collaborative potential is there to do these types of studies. There's the possibility of international collaboration because from a U.S perspective, it wouldn't be seen as competition from other states.

**Philip Shapira** (Georgia Institute of Technology, USA). These areas of policy comparison are not black and white. In Europe there are similar partnership developments. Also, the idea that states should intervene directly is a not a new idea. Thus, there is more of a similarity. But I do have a concern that the partnerships in the triple helix framework that the paper describes seem mostly to benefit traditionally advantaged groups.

**Henry Etzkowitz.** This is not simply university-industry, etc. but a combo of the three. We see an extension that they are moving into the more developing areas, such as you would see in Europe. HUD is now renovating buildings, etc. in Harlem; Some Internet companies are moving into Harlem and other Empowerment Zones. The next step is to see whether they are making connections to the universities in those areas.

## Evaluation as a Source of Strategic Intelligence

**Stefan Kuhlmann**

Fraunhofer Institute for Systems and Innovation Research, Karlsruhe, Germany

Email: sk@isi.fhg.de

---

### Introduction

*"Ursache und Wirkung. - Vor der Wirkung glaubt man an andere Ursachen als nach der Wirkung."*  
(Friedrich Nietzsche, Die fröhliche Wissenschaft, Aphorismus 217, 1882)

All innovation policies are of a complex nature, public efforts as well as corporate initiatives for innovation. They are difficult to pursue since innovation means change, and any change inevitably affects the interests of the many stakeholders involved; heterogeneous actors in public or private arenas will contest and negotiate: "Most technology policies ... represent a compromise among conflicting objectives" (Mowery 1994, 10). The main media for negotiation are power, money, and information. Various actors have different shares of these resources at their disposal. Here, one important source of policymaking authorities is the utilisation of "strategic intelligence"<sup>2</sup>. Whether located in the public or private sector, decision-makers concerned with technology choices, strategies and policies need a wide range of high quality intelligence inputs in order to make sustainable decisions.

Traditionally, policymakers and strategists dealing with technology and innovation have used a number of "intelligence" tools and techniques to provide themselves with the data they need to formulate appropriate policies and strategies. In the public sector, for example, innovation policy formulation has been improved in recent years via the use of *science and technology foresight*, *technology assessment* and *policy evaluation* exercises. All have yielded valuable information which has helped policymakers to make strategic technology choices and fine-tune courses of action.

In future, however, more will be needed. The complexity of the modern world, the crucial and simultaneously contentious role innovation plays within economies and social structures, the broad scope of intended and non-intended, direct or indirect potential impacts and outcomes of research and innovation policies in science, in-

---

<sup>1</sup> Cause and impact. - Before any impact you believe in other causes than after the impact.

<sup>2</sup> The ideas and concepts presented in this paper are based on a report produced by members of the Advanced Science and Technology Policy Planning Network (ASTPP), a thematic network set up as part of the Targeted Socio-Economic Research (TSER) Programme of the European Commission (see Kuhlmann et al. 1999).

dustry or polity, as well as the diversity of actors, interests and values represented in the related innovation policy arenas, make it imperative that intelligence tools are improved and access to the results of intelligence exercises carried out across the globe is enhanced.

In this paper the use of a number of intelligence tools in innovation policymaking is reviewed before going on to examine how they could be used in different combinations to enhance strategic intelligence inputs into the arenas of policymaking. Two illustrative examples of strategic intelligence use will be discussed. Critically, also the need for a system of "distributed intelligence" is examined which could provide public and corporate policymakers with access to strategic intelligence outputs produced in different locations for different reasons. Specifically, the design requirements of a "system architecture" for distributed intelligence are explored. Then the need for an effective European system of distributed intelligence is contemplated and illustrated by a fictional case, while the final part sums up the approach suggested in this paper.

### **Complex Innovation Systems and the Need for Improved Strategic Intelligence**

The likelihood of innovation in modern society's science, technology and industry has always been shaped by national, regional or sectoral "systems of innovation": innovation systems were discovered by the social scientists (first of all by economists<sup>3</sup>), as – with the increasing significance of international hi-tech markets – explanations for the differing degrees of competitiveness of economies, especially of their "technological performance" and their ability to innovate were sought. It was recognised that differing national, regional (e.g. Howells 1999) or sectoral (e.g. Kitschelt 1991) "innovation cultures", each rooted in historical origins, characteristic and unique industrial, scientific, state and politico-administrative institutions and inter-institutional networks, crucially affected the ability of economic actors and policymakers to produce and support successful innovations.

Innovation systems are encompassing, according to a meanwhile widely accepted understanding, the "biotopes" of all those institutions which are engaged in scientific research, the accumulation and diffusion of knowledge, which educate and train the working population, develop technology, produce innovative products and processes, and distribute them; to this belong the relevant regulative bodies (standards, norms, laws), as well as the state investments in appropriate infrastructures. Innovation systems extend over schools, universities, research institutions (education and science system), industrial enterprises (economic system), the politico-administrative and intermediary authorities (political system) as well as the formal

---

<sup>3</sup> See in particular Freeman 1987; Lundvall 1992; Nelson 1993; and Edquist 1997. Lundvall/Maskell (1999) provide a reconstruction of the genesis of the expression "national innovation systems". They all take as a theme, at least marginally, also the interface of markets and political systems (and, in particular, public policies by state governments) as a formative variable of innovation systems.

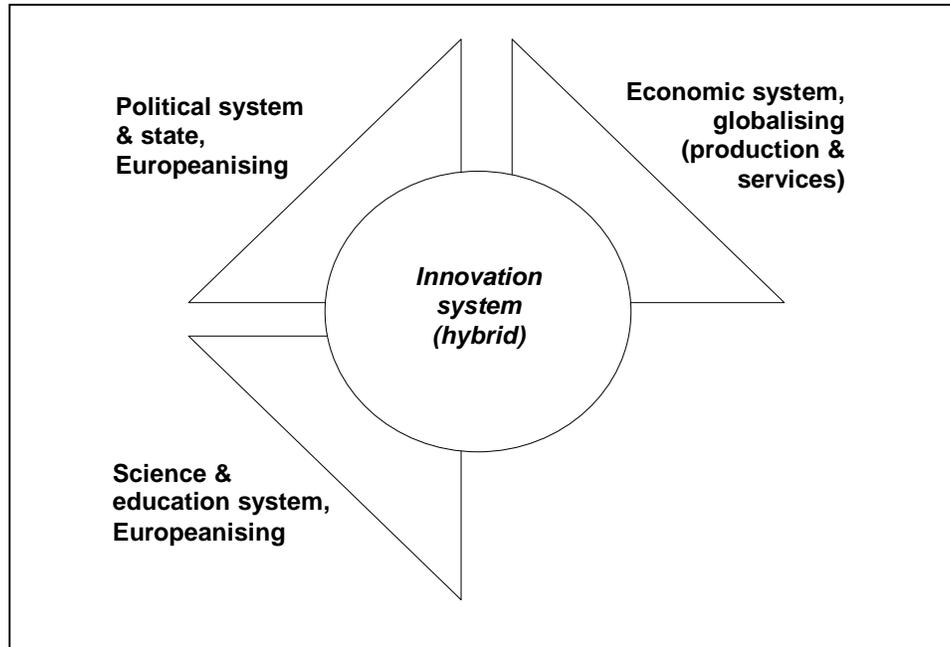
and informal networks of the actors of these institutions. As "hybrid systems"<sup>4</sup> (e.g. Kuhlmann 2001; see figure 1) they represent sections of society which carry far over into other societal areas, e.g. through education, or through entrepreneurial innovation activities and their socio-economic effects: innovation systems have a decisive influence on the modernisation processes of a society (OECD 1999a).

Each innovation system is different, just as one society is not the same as the others. Efficient innovation systems develop their special profiles and strengths only slowly, in the course of decades, or even centuries. Their governance is based on a co-evolutionary development of and stable exchange relationships among the institutions of science and technology, industry and the political system which also have been described as a Triple-Helix-Model of university-industry-government relations (Etzkowitz/Leydesdorff 2000). They make possible the formation of a characteristic, system-specific spectrum of different, unmistakable role definitions of the actors actively involved herein, come up with own negotiation arenas, and stabilise mutual expectations of behaviour. Finally, they produce particular, intermediary fora and bodies which facilitate the transactions of the actors. Innovation systems are "embedded" (Hollingsworth/Boyer 1997) in historically rooted, long-standing socio-economic structures.

---

<sup>4</sup> One might argue whether "innovation systems" should be considered as genuine (sub)systems in the sense of the sociological systems theory (e.g. Luhmann 1984). In the given context the author employs the notion "innovation system" simply as a *heuristic aide* facilitating the analysis of the embeddedness (Hollingsworth/Boyer, 1997) of innovation within the interplay of various societal subsystems.

Figure 1: Hybrid Innovation System



Public and private policymakers, both deeply rooted in the institutional settings of the innovation system, face a number of challenges, both now and in the future (see also Lundvall/Borrás 1998):

- (1) *The nature of technological innovation processes is changing.* The production of highly sophisticated products makes increased demands on the science base, necessitating inter- and trans-disciplinary research and the fusion of heterogeneous technological trajectories (Grupp 1992; Kodama 1995). New patterns of communication and interaction are emerging which researchers, innovators and policymakers have to recognise and comprehend. For example, if nanotechnology (miniaturisation) is to stimulate future innovation processes and new generations of technology as a new basic technology, an important precondition is transdisciplinary interaction with electronics, information technology, the science of materials, optics, biochemistry, biotechnology, medicine and micromechanics. The applications of nanotechnology accordingly encroach upon the fields of customised materials and biotechnical systems, even though they are envisaged as falling mainly into the area of electronics.
- (2) *The "soft side of innovation" is of growing importance* (den Hertog et al. 1997; Coombs 1999). Non-technical factors such as design, human resource man-

agement, business re-engineering, consumer behaviour and "man-machine interaction" are critical to the success of innovation processes. As a consequence, the learning ability of all actors in the innovation process is challenged and it becomes more appropriate to speak about a "learning economy" than a "knowledge-based economy" (Lundvall/Borrás 1998, 31).

- (3) These first two points are specific manifestations of what Gibbons et al. (1994) call the transition from *mode-1 science* to *mode-2 science*. Mode-1 refers to traditional science-driven modes of knowledge production. Mode-2 refers to knowledge production processes stimulated and influenced far more by demand, in which many actors other than scientists also have important and recognised roles to play.
- (4) The pressure on the science and technology systems and the innovation system to function more effectively is complemented by similar pressures to function more efficiently, largely driven by the *growing cost* of science and technology. This will require a much *better understanding of the research system itself* (Rip/van der Meulen 1997). In this respect, strategic intelligence (e.g. policy evaluations) can help sharpen insights into the internal dynamics of science and technology and their role in innovation systems.
- (5) European innovation policymakers have to co-ordinate or orchestrate their interventions with an *increasing range and number of actors* in mind (e.g. European authorities; various national government departments and regional agencies in an expanding number of member states; industrial enterprises and associations; trade unions and organised social movements etc.). Furthermore, the accession of new Eastern European member states will undoubtedly increase the importance of this aspect (Kuhlmann 2001).
- (6) Since the 1990s, industrial innovation processes *care less and less about national systems* and borders (see Reger et al. 1999; OECD 1999b). In particular big multinational companies developed from an "optimizing production machinery" to "globally learning corporations" (Meyer-Krahmer/Reger 1999). With the growing complexity of the knowledge required, isolated, individual actors are less and less in a position to master this adequately without external support. International innovation-oriented co-operation and the maintaining of corresponding networks belong meanwhile to the daily innovation routine – as horizontal co-operations within companies, as co-ordinated division of work between (even competing) companies, as well as within the framework of joint research and development projects between firms and public research institutions: The aim is the provision of "complementary assets" (Teece 1986) in the area of technological know-how, increasingly also in international networks (Meyer-Krahmer et al. 1998; Niosi 1999). Furthermore, innovation managers in large multinational corporations have to develop their strategies vis-à-vis

## 11 Issues in Linking RIT Evaluation With Policy

heterogeneous national innovation policy arenas with diverse actors, not at least a variety of non-governmental organisations.

Hence, policy-formulation in these circumstances is not straightforward. There is increasing pressure on policymakers and strategists to:

- acknowledge, comprehend and master the *increasing complexity* of innovation systems (more actors, more aspects, more levels etc.);
- help preside over the establishment of an *international division of labour in science and technology* acceptable to all actors involved;
- adapt to *changes in the focus of innovation policies* between international (growing), national (declining) and regional (growing) levels;
- *increase efficiency and effectiveness* in the governance of science and technology;
- *make difficult choices* in the allocation of scarce resources for the funding of science and technology;
- integrate "classical" innovation policy initiatives with *broader socio-economic targets*, such as reducing unemployment, fostering the social inclusion of less favoured societal groups and regions (as claimed in particular by the 5<sup>th</sup> Framework Programme of the European Commission), and reconciling innovation policy with a sustainable development of our natural environment as well as a careful use of natural resources (Kuhlmann/Meyer-Krahmer 2001).

Over the last two decades, considerable efforts have been made to improve inputs into the design of effective science, technology and innovation policies. In particular, formalised methodologies, based on the arsenal of social and economic sciences have been introduced and developed which attempt to analyse past behaviour (evaluation), review technological options for the future (foresight), and assess the implications of adopting particular options (technology assessment).

Achievements in these areas have been impressive. As a complement of evaluation, foresight and technology assessment, other intelligence tools such as comparative studies of the national, regional or sectoral "technological competitiveness", benchmarking methodologies etc. were developed and used. Policymakers at regional, national and international levels have all benefited from involvement in these processes and exploited their results in the formulation of new policies. Increasingly, however, it has become obvious to both policymakers and the analysts involved in the development and use of strategic intelligence tools that there is scope for improvement. In particular, there is a need to use such tools in more flexible and intelligent ways, combining them in individual exercises to satisfy the multiple needs of innovation policymakers.

There is a further need, however, to exploit potential synergies of the variety of strategic intelligence pursued at different places and levels across countries within what one could call a system of "*distributed intelligence*". Currently policymakers in different parts of the world independently call for localised strategic intelligence activities to be customised to their own particular needs. In this paper, however, it is argued that the results of many of these exercises may have a didactic value in other contexts. Furthermore, the competence which exists within the strategic intelligence community as a whole can also be exploited more broadly by policymakers in localised settings.

Before considering these issues further, however, more discussion is needed of the context within which infrastructural changes will be needed. In particular, a deeper discussion is needed of the complexity of modern-day innovation policymaking in multi-actor/multi-level arenas.

### **Innovation Policy in Multi-Actor/Multi-Level Arenas**

This section provides a set of assumptions concerning the functioning of advanced strategic intelligence in complex innovation systems. The analysis is based on a *twofold basic hypothesis*: innovation policymaking – using strategic intelligence – is pursued by political-administrative institutions and actors (e.g. ministries for science and technology), by research organisations and by R&D-based companies, seeking to learn in order to improve their institutional performance and the preconditions for institutional survival or even growth (*functional dimension*) and to contribute to socio-economic modernisation (*normative dimension*). In doing so science and technology and political actors find themselves confronted with

- given general issues of innovation policymaking (public and private policy), moulded by the emerging and constantly changing role of science and technology and innovation in economy and society (as sketched in section 2 above),
- given arenas and configurations with other actors in terms of resources (financial, knowledge), and of regulations and institutions (political, economic), partly determining and partly facilitating their actions.

*Assumption 1*: A linear model of policymaking as a consequential process<sup>5</sup> is no longer appropriate, at least not in the field of science, technology and innovation policies. Here, all typical steps are more or less interacting, thereby describing "loops"<sup>6</sup>: Ideally, *looping policy processes* provide "stopping points" where policy-shaping activities converge in such a way that effective acting is feasible. Entry

---

<sup>5</sup> Typical steps: formulation, agenda-setting, decisions, implementation, evaluation, formulation ...; see Brewer/de Leon (1983)

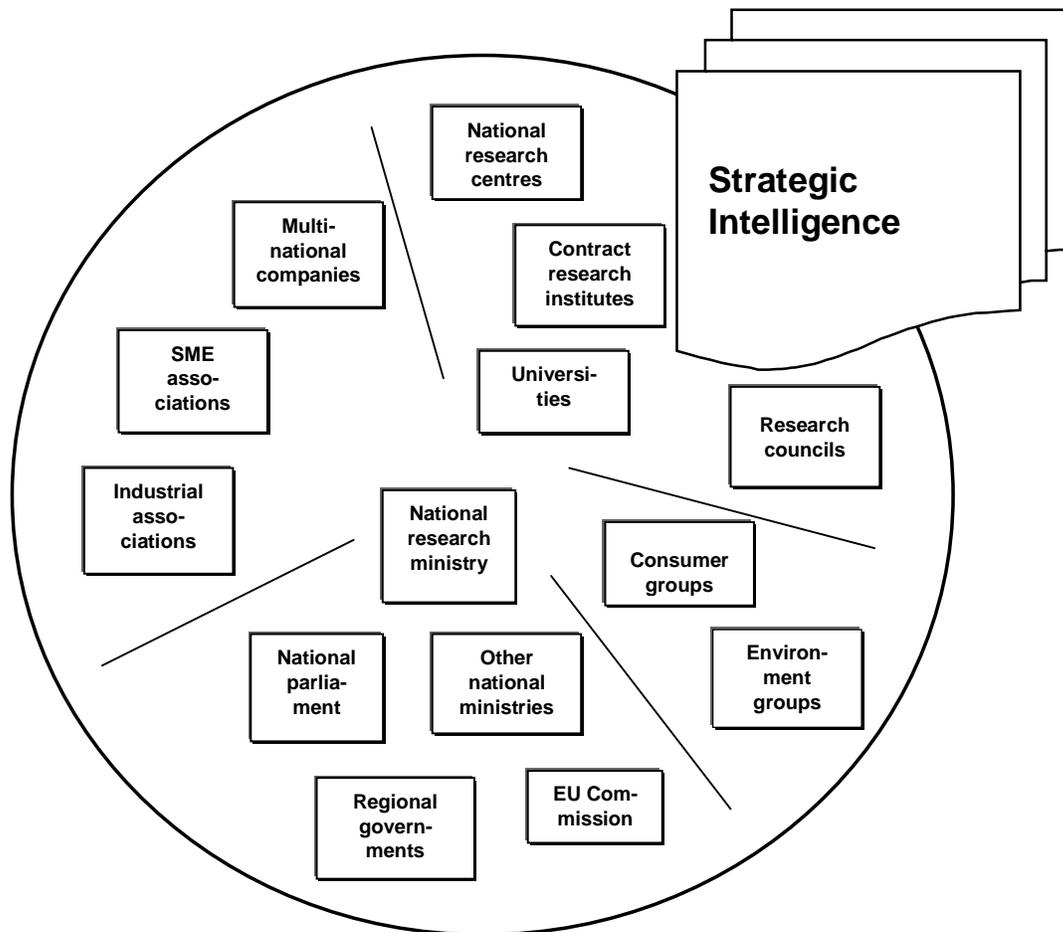
<sup>6</sup> Y. Dror (1968, 19<sub>1</sub> f) discussed already 30 years ago - though in the context of a rationalistic and quite sophisticated policymaking model – the many communication and feedback loops that connect all the phases and subphases of optimal policymaking with each other.

## *11 Issues in Linking RIT Evaluation With Policy*

options for strategic intelligence could be found at (1) such stopping points, (2) ad hoc opportunities, or (3) if an intelligence function has been institutionalised as a steady monitoring process. The emergence of strategic intelligence knowledge as a policy resource on the one hand and structural and institutional preconditions of using strategic intelligence activities on the other influence and transform each other. Often it is external pressure on policy actors and the related arenas that gives the impulse for the production and application of advanced strategic intelligence.

*Assumption 2:* Innovation policy is rather (and increasingly) a matter of networking between heterogeneous (organised) actors instead of top-down decision-making and implementation. Policy decisions frequently are negotiated in *multi-level/multi-actor arenas* and related actor networks (Marin/Mayntz 1991); given power structures and the shape of arenas, nevertheless, may vary considerably between member states (or regions) or corporations (see figure 2). "Successful" policymaking normally means compromising through *alignment* and "re-framing" of stakeholders' perspectives (Schön/Rein 1994; see also Kuhlmann 1998; de Laat 1996).

Figure 2: Actors in Innovation Policy Arenas and Strategic Intelligence



*Assumption 3:* Negotiating actors pursue different – partly contradicting – interests, represent *different stakeholders' perspectives*, construct different perceptions of "reality" (Callon 1992), refer to diverging institutional "frames". Different actors having different responsibilities (policymakers define programmes, allocate budgets; researchers define themes, purchase equipment; industry looks for competitive advantages ...) perceive different "stopping points".

*Assumption 4:* Contesting and negotiating actors use money, power and information as their main media. Various actors have different shares of these resources at their disposal. Strategic intelligence tools (policy evaluation; technology foresight; technology assessment) use in particular "information" and knowledge as negotiation media. Possible variables of arena configurations are the size of arena, related policy issues (e.g. distribution of resources; industrial; ethical/cultural), hampering or fostering institutional environments, the degree of self-organisation and of power of

## 11 Issues in Linking RIT Evaluation With Policy

actors, or established routines and traditions, and the stability of configurations. Within different "configurations" *strategic intelligence activities* may fulfil certain *functions* including

- analyses of changing innovation processes, the dynamics of changing research systems, changing functions of public policies;
- the identification of diverging "frames" of actors' perceptions;
- a more "objective" formulation of diverging perceptions of (even contentious) subjects, offering appropriate indicators and information-processing mechanisms;
- the organisation of mediation processes and "discourses" between contesting actors (or between representations of their views).

*Assumption 5*: Is there anything *specific* about strategic intelligence for *innovation policy*? Yes, research and innovation are open-ended activities, producing novelties, and related policymaking is rather "elusive" (Jasanoff 1997, xiv):

- There is a lingering uncertainty about the boundaries of research and innovation policies which could "with little imagination be stretched to encompass virtually every aspect of purposive state activity: health, education, welfare, defence, energy, environment".
- The relative autonomy of science and technological progress (Kuhn 1962) destabilizes the very meaning of policy which is conventionally defined as plan of action that is intended to determine people's behaviour.

Thus, the results of research evaluation and the efforts on foresight and technology assessment cannot produce automatically clear-cut alternatives for policymakers' decisions (for example, because of the life cycle of research issues, research groups and institutes). This is particularly – sometimes dramatically – visible vis-à-vis the consequences of scientific or technological breakthroughs.

How can the implications of these assumptions be taken up in innovation policy-making? Putting it abstractly, *alignment and consensus* production is a precondition for *successful policymaking*. Innovation policy decisions are taking place in multi-level/multi-actor arenas and related actor networks and thus no actor can easily make his/her own interests/objective/actions prior to those of the others. Consequently, foresight, evaluation and technology assessment are increasingly considered as tools to create alignment between actors, relying on the inputs of (competing) experts (cf. foresight triangle, in which foresight methods are positioned between the three poles of expertise, creativity and interaction).

Does the alignment process foster or prevent "*revolutionary*" decisions? Governments and others now and then try to force such decisions: examples are the Task

Forces of the European Union, or the creation of Technological Centres of Excellence ("Technologische Top Instituten") in the Netherlands in the 1990s. These were breakthrough policies that initialised new patterns, activities, aggregations within the innovation system, changing existing institutes or setting priorities and posteriorities. There is a *paradox*: the legitimisation of these breakthroughs partly results from evaluations of the performance of institutes or analyses of national systems of innovation. Strategic intelligence may, however, create also broader "roadmaps" orienting actors towards a more conscious decision-making exceeding "conservative" alignments.

To sum up, for many policy actors in search of alignment for decision-making, the *need* to access comparative contextual data and information produced via *strategic intelligence* exercises (own and of other actors) is becoming a necessity rather than a luxury.

Furthermore, the complex dynamics of innovation systems demand a pooling and sharing of such data in order to avoid expensive duplication of effort. The same complexity also means that there are only few tried and tested policy formulae that could be adopted and applied, with many policy initiatives constituting risky experiments unless guided by intelligence reports concerning the outcome of similar initiatives in other parts of the world. Current user needs for strategic intelligence may still have a strong local focus, but there is little doubt that the need for access to a *system of distributed intelligence* is increasing.

### **Innovation Policy Evaluation, Technology Foresight, Technology Assessment – Brief Overview**

Roughly, one can describe the basic concepts of technology foresight, science and technology policy evaluation, and technology assessment in the following way:

- "Technology foresight is the systematic attempt to look into the longer-term future of science, technology, the economy and society, with the aim of identifying the areas of strategic research and the emerging of generic technologies likely to yield the greatest economic and social benefits" (Martin 1995, 140).
- Practices of science and technology policy evaluation are wide-ranging, and their functions vary significantly (1) from the provision of legitimation for the distribution of public money and the demonstration of adequate and effective use of the funding by measuring the scientific/technological quality or the (potential) socio-economic impacts, via (2) targeting and "controlling" in the sense of improved management and "fine tuning" of science and technology policy programmes, to (3) an attempt to improve transparency in the rules of the game and the profusion of research funding and subsidies, and to enhance the information basis for shaping science and technology policies, in the sense of a government-led "mediation" between diverging and competing interests of various players within the science and technology system (see Kuhlmann 1997).

- Technology assessment, in very general terms, can be described as the anticipation of impacts and feedback in order to reduce the human and social costs of learning how to handle technology in society by trial and error. Behind this definition, a broad array of national traditions in technology assessment is hidden (see Schot/Rip 1997; Loveridge 1996).

In the following, a brief summary is presented of the experiences gained so far of the performance of foresight, evaluation and technology assessment in decision-making processes of science and technology policies.

### ***Science and Technology Foresight***

Science and technology foresight exercises are becoming increasingly attractive for governments, national research agencies and businesses in their efforts at coping with the increasing complexity of new technologies and decision environments, in an increased techno-economic competition world-wide (see Martin 1995; Cameron et al. 1996; Grupp 1998). Since the 1990s, quite a number of major foresight exercises have been launched in many European countries.

The majority of experts consider foresight essentially as a collective and consultative process, with the process itself being equally or even more important than the outcome. Foresight exercises are ways of obtaining opinions, conflicting or otherwise, about future developments, most of which are already established. Foresight in this sense is an essential contributor to the creation, either collectively or individually, of models of the future. Such models are important because they are capable of creating synthesis, they are disruptive and interfere with current modes of thought, thus forming and shifting values.

Foresight is different from prognosis or prediction. Implicitly, it means taking an active role in shaping the future. As a possible result our prognosis of today may be falsified in the future because of a new orientation resulting from foresight. Elder attempts at a "planning" of the future by developing heuristic models (in the sense of futurology) were based on the assumption that the future is pre-defined as a linear continuation of present trends (Helmer 1966; Flechtheim 1968; Steinmüller 1995; Linstone 1999). Albeit these approaches largely failed due to the in-built simplification of the actual dynamics of social, economic and technological developments, some studies nevertheless evoked a vivid discussion about the future (e.g. Forrester 1971; Meadows et al. 1972).

In reality, future developments underlie reciprocal influences which cannot be assessed exhaustively in advance, thus not predicted. There is, nevertheless, a need to "monitor the future prospectively": the accelerating changes that individuals as well as societies have to adapt to socially and psychologically, make it necessary to anticipate these changes before they become reality (Helmer 1967). A *new understanding of foresight* gaining acceptance in the 1990s (starting with Irvine/Martin 1984) made clear that a targeted shaping of future developments is strictly limited

and that the potential impacts of decisions can only partially be estimated. Hence, the new approaches to foresight are striving for relatively "realistic" objectives (Cuhls 1998). In the context of policy-making, the most important intentions are

- to find out new demand and new possibilities as well as new ideas,
- to identify a choice of opportunities, to set priorities and to assess potential impacts and chances,
- to discuss desirable and undesirable futures,
- to prospect the potential impacts of current research and technology policy,
- to focus selectively on economic, technological, social and ecological areas as well as to start monitoring and detailed research in these fields.

A popular foresight approach is represented by the Delphi method originally developed in the USA in the 1960s (Gordon/Helmer 1964; Helmer 1983; Cuhls 1998): Delphi belongs to the subjective and intuitive methods of foresight. Issues are assessed, on which only unsure and incomplete knowledge exists. Delphi is based on a structured survey of expert groups and makes use of the implicit knowledge of participants. Hence, Delphi is both quantitative and qualitative. It includes explorative-predictive as well as normative elements (Irvine/Martin 1984). There is not a single method, but different variations in the application which all agree that Delphi implies an expert survey in two or more rounds. Starting from the second round, a feedback is given about the results of previous rounds: the same experts assess the same matters once more - influenced by the opinions of the other experts. Delphi facilitates a relatively strongly structured group communication process, revealing conflicting as well as consensus areas. Delphi-based foresight exercises, therefore, were used repeatedly and increasingly in the context of policymaking (Grupp 1998), building on their capacity to facilitate an alignment of actors' expectations through interactions (Sanz/Cabello 2000).

Results generated through Delphi processes are welcomed by many policymakers and strategists since they offer semi-quantitative data – which, nevertheless, like the older, naive future-planning exercises, can be misunderstood and misused as "facts" about the future. At the same time, with explicit professional methods of foresight, a broad variety of stakeholders can be involved: scientists, managers, consultancy firms, social organisations, etc. In this respect, more *distributed intelligence* can be enforced. Through their participation, all these various actors get information, do their own intelligence building and feed back their perceptions (and values) into the system. Large explicit procedures are costly, but they improve the quality of the decision process also in another sense: allowing the reaction of various categories of "experts", they add dimensions of technology assessment and evaluation to the "pure" foresight exercise.

It is well known that sudden science and technology breakthroughs often have not been foreseen by the majority of main-stream experts but were anticipated by a few unorthodox thinkers. This is a classical problem of foresight and other methods of "prospection": how to detect feeble signals or the minority views that could be revealed as the very precursors of the future? The *paradoxical nature of foresight* tools is that they aim at two conflicting goals: *building consensus* and preserving *variety of visions*. The specific problem with administrative procedures is their natural bias towards selecting majority views (risk-taking is not the philosophy of administration - private or public). Distributed intelligence means here: combining views of heterogeneous relative weights.

### ***Innovation Policy Evaluation***

In most European countries as well as in North America, an "evaluation culture" in science, technology and innovation policies has evolved since the 1980s (Papaconstantinou/Polt 1997), including the *ex post* evaluation of research programmes and other policy initiatives, the evaluation of R&D centres and universities, and the evaluation of R&D funding agencies. In all these areas of application, evaluation plays a different role in decision-making, varying from case to case, including the following *functions* (cf. Kuhlmann 1998, 84-113):

- evaluation may provide legitimisation for the allocation of public money to R&D,
- evaluation may enhance an adequate and effective use of funding by measuring the scientific/technological quality or the (potential) socio-economic impact,
- evaluation may improve programme management and "fine tune" science and technology policy programmes,
- evaluation may release new ideas or legitimate already circulating ideas about changes in R&D centres and funding agencies, thus enhancing the fulfilling of their missions,
- evaluation may be an attempt to improve transparency of the rules of the game of science and technology funding decisions, and
- enhance the information basis for science and technology policies, in the sense of a government-led "mediation" between diverging and competing interests of various players within the science and technology system.

European countries differ in the extent to which they apply science and technology policy evaluation (Georghiou 1995). Some countries have longer traditions of evaluation cultures, others are relative newcomers in this field. In countries such as Greece, EU research programmes and their evaluation procedures have stimulated science and technology policy evaluation exercises and helped train national experts in evaluation. Also, the OECD has played an important role in diffusing models of

evaluations (OECD 1997). Scandinavian countries have exchanged models and ideas on evaluation as early as the 1980s in their mutual science and technology policy collaboration.

The European and North American "evaluation culture", meanwhile, has a broad range of *conceptual and methodological experiences* at its disposal. Methods of various types have been developed and utilised to determine attained or attainable effects; the most important are<sup>7</sup>: peer reviews, before/after comparisons, control or comparison group approaches, a variety of quantitative and qualitative analyses etc.. These approaches can be applied individually or in combination with various data and indicators (financial expenditure on research and development, patents, economic, social, technical indicators, publications, citations, etc.), data collection methods (existing statistics, questionnaires, interviews, case studies, panels, etc.), data analysis methods (econometric models, cost/benefit analyses, other statistical methods, technometrics, bibliometrics, peer reviews etc.; see e.g. Grupp/Schmoch/Kuntze 1995). All the procedures have different strengths and weaknesses, which makes using a combination of methods advisable.

At the present stage of evaluation research, and despite all the (necessary) efforts made to objectify the methods and the resulting indicators, one must *warn against* considering *quantitative indicators alone* to be adequate for evaluation purposes. The understandable desire for a tool-box of indicators which can be used in a standardised fashion is not realisable vis-à-vis our limited knowledge of the dynamics of innovation processes: a measurable research performance and related output do not automatically produce socio-economically effective innovations.

Experience proves that any evaluation is faced with *limitations*, some related to the methods development, others to budgetary and time restrictions (Airaghi et al. 1999). Any impact assessment is hampered by the fact that it takes many years for impact to be seen; however, those commissioning the evaluation seldom wish to wait for years to find out about impact. Further, those involved in the processes might have difficulties in remembering the events concerned if consulted much later. Also, attributing effects to the initiatives to be evaluated is a basic difficulty faced by all evaluation exercises. Further, indirect effects are not sufficiently taken into account because of the difficulty in measuring them. Lastly, socio-economic effects and contribution to societal needs are difficult and laborious to evaluate.

There have been many changes and developments in the theory and practice of evaluation over the past decade or so. In particular, in countries where evaluation has taken root fairly early, following *trends* can be observed:

---

<sup>7</sup> See e.g. Meyer-Krahmer/Montigny 1989; Bozeman/Melkers 1993; Callon/ Laredo / Mustar 1997; Shapira/Youtie 1998; Georghiou/Roessner 2000.

## *11 Issues in Linking RIT Evaluation With Policy*

- The major rationale for evaluations has shifted and evolved from a desire to *legitimate past actions* and demonstrate accountability, to the need to improve understanding and *inform future actions*.
- Correspondingly, the issue focus of evaluations has broadened away from a narrow focus on quality, economy, efficiency and effectiveness, and towards a more *all-encompassing concern* with additional issues, such as the appropriateness of past actions and a concern with performance improvement and strategy development.
- Approaches to evaluation have evolved away from a purist model of "objective neutrality", characterised by independent evaluators producing evaluation outputs containing evidence and argument, but no recommendations; to more formative approaches in which evaluators act as process consultants in learning exercises *involving all relevant stakeholders*, providing advice and recommendations as well as independent analysis.
- This has led to more *flexible and experimental approaches* to the construction of policy portfolios, and to even greater demands for well specified systems of monitoring, evaluation and benchmarking to aid analyses and feedback into strategy development.

Many evaluations thus reflect an increasing concern with the link between evaluation and strategy, with an eclectic mix of methodologies used within the context of individual exercises to satisfy the demands for understanding and advice. Increasing attention is also being paid within many institutional settings to the way in which evaluations can inform strategy.

### ***Technology Assessment***

Technology assessment, with its twin components of anticipation (of effects and impacts) and evaluation and feedback into decision-making, is done in various ways, depending on the key actors and the arenas (see e.g. Rip/Misa/Schot 1995; Smits et al. 1995; Loveridge 1996; Sundermann et al. 1999). Three strands, each with its own style, can be distinguished:

- Technology assessment in firms and in technological institutes, oriented towards mapping future technological developments and their value to the firm or institute, and used as an input in strategy development. "*Picking the winners*" (or "*avoiding the losers*") used to be the overriding orientation. This strand has developed relatively independently of "public domain" technology assessment, but links are emerging because of the need of firms to take possible societal impacts and public acceptance into account; biotechnology is the main example at the moment.

- Technology assessment for policy development and political decision-making about projects or programmes with a strong technological component (e.g. the electronic superhighway or modern agriculture) or important technologies (like genetic modification). One can call this "*public service*" *technology assessment*, and consider the U.S. Office of Technology Assessment (OTA) as the embodiment of this type of technology assessment. OTA has, during its lifetime, developed a robust approach to technology assessment studies, which can still be followed profitably. Other technology assessment bodies serving national parliaments and/or national governments were modelled on the OTA example, but have to attend to their specific situation and tend to include participatory technology assessment methods in addition to expert- and stakeholder-based approaches.
- *Agenda-building technology assessment* is the most recent strand. While it is particularly visible and more or less institutionalised in some European countries (Denmark, the Netherlands), participatory methods like consensus conferences are taken up all over the world. *De facto* agenda-building technology assessment has a longer history; for example, controversies over new projects or new technologies (and the studies and documents produced in the course of the controversy) induce learning (about potential impacts) and articulation (of the value of the technology). Agenda-building technology assessment merges into informed consultation processes to reach agreement on the value of new technology, as happens for instance through *Sozialpartnerschaft* in Austria.

Technology assessment is much more an advisory than a scientific research and policy-analytical activity. Increasingly, the *advisory activity includes participation*, and thus becomes joint agenda-building. One can compare this shift with the recognition, in foresight and evaluation exercises, of the importance and effects of the process as such, rather than just the data collection and analysis.

The strengths and weaknesses of technology assessment cannot be identified unambiguously, because of the variety in the contexts of use, and thus in goals and style. It is clear that *there is renewed interest in technology assessment*, and that this has to do with the increased possibilities of combining private-domain and public-domain technology assessment, and with the role of technology assessment in broader priority setting, technology road-mapping, and articulation of views about new technology.

It has been noted that *technology foresight methods might be used for technology assessment, and vice versa*. There may well be such opportunities, for example the Delphi method. The German study, *Technology at the Threshold of the 21st Century* (Grupp 1993) is a foresight study, but indicates the relevance of extending foresight methods to technology assessment. The experts involved in technology foresight studies are assumed to have some feeling for the effects and impacts of new technology, even if this is often limited to the promise of new technology. In other

words, an informal technology assessment competence is required, which could profit from exposure to foresight methods and experience. Foresight and technology assessment can jointly contribute to (distributed) intelligence about future developments and their value. A difference in style and context will remain: Foresight aims to open up spaces for thinking about new possibilities, technology assessment is oriented to selecting or at least modifying and modulating developments. The link with decisions and strategies implies that there will be more and more broadly based controversy than with foresight, which often remains limited to communities of experts.

### ***Towards an Integrated Use of Tools for Strategic Intelligence***

A brief survey of existing practices and experiences with the integrated use of the three intelligence tools for innovation policymaking foresight, evaluation, technology assessment in various European countries and the EU Commission led to the following conclusions (cf. Kuhlmann et al. 1999, 45-58):

(1) Though quite *some examples of integration* between the three bodies of experiences can be found in several countries, there is no systematic effort, either by policymakers, or by the research practitioners, to combine the strategic intelligence coming from the three different traditions. The *synergy* that could be gained by using a combination of methodologies, issues, processes and so on, is *not exploited* in the most effective manner yet.

(2) *Industry* has an *older tradition* of combining approaches when defining strategies to assess uncertain (technological) developments with potentially wide impacts, both commercial and societal.

(3) Present empirical and well-documented examples from *cross-border learning* show that it is valuable to learn even from different institutional settings, to avoid repeating mistakes and to pick up good practice experience more quickly.

(4) There is *no "blue-print"* of how the tools of evaluation, foresight and technology assessment can be best combined. The configuration should be considered from case to case, depending on the objectives and scope of the policy decision-making process in question. Integration seems to be useful for those cases where a combination of information looking back in time, looking at current strengths and weaknesses, looking at a wide set of stakeholders and at future developments can improve the insights needed to choose between strategic options. This also asks for further exploration of the limits of integration to avoid unnecessary "heavy weight" exercises.

In general, one could state that *the greater the potential socio-economic impact* of technology and innovation, *the stronger* the case is for using the full array of available techniques for *strategic intelligence*.

## Improved Strategic Intelligence for Innovation Policy – Principles and Examples

In the preceding sections it has been demonstrated that a growing need exists for strategic intelligence to underpin policymaking in the area of science, technology and innovation. Also, it has been demonstrated that it is not necessary to start from scratch when attempting to meet these needs. In the past a whole array of instruments have been developed to provide strategic intelligence. Among the best known are the three strategic intelligence tools discussed in detail in this report: foresight, policy evaluation, and technology assessment. The use of these tools, however, could be improved considerably, as could access to the results of related exercises.

Basically, there are two parallel and complementary routes which can be taken to improve the quality, efficiency and efficacy of strategic intelligence.

The *first route* (dealt with in the present section) aims at *improvements* in the use and deployment of *existing instruments and tools*. A great deal could be gained via the further development of these instruments and via their use in interesting *new combinations*, either with each other, e.g. combined evaluation and foresight exercises feeding into strategy development, or, alternatively, via comparison of the results of the parallel use of the same instruments at different levels (e.g. national vs. international) or in different places (national vs. national).

That there is potential for further developing these instruments is perhaps demonstrated by the extent of developments to date. Foresight and technology assessment, for example, have changed considerably over the last three decades, with forecasting (prediction) being supplanted by foresight (scenario construction), and technology assessment metamorphosing from an "early warning system" into a policy instrument capable not only of identifying possible positive and negative effects, but also capable of helping actors in innovation processes to develop insights into the conditions necessary for the successful production of socially desirable goods and services (see e.g. Smits et al. 1995). As a relatively new trend one can observe a *shift from solely analytical to more process-oriented instruments* (IT-supported group decision rooms, consensus development conferences, and platform and scenario workshops), a shift which takes into account the growing complexity of innovation systems and the need for assistance in strategy development to go beyond the provision of empirical data on the development of new technologies.

Starting from the above sketched availability of integrated tools and of new process-oriented approaches – and in order to justify the direction taken in this article – one could stipulate a number of *general principles* of strategic intelligence for complex innovation systems:

- (1) *Principle of participation*: strategic intelligence realises the multiplicity of actors' and stakeholders' values and interests involved in innovation policymaking. Foresight, evaluation or technology assessment exercises take care of the

diversity of perspectives of actors and make an attempt to give them a voice (multiple perspective approach). Strategic intelligence avoids maintaining one unequivocal "truth" about a given innovation policy theme.

- (2) *Principle of "objectivisation"*: strategic intelligence "injects objectivised" information into the policy arena, i.e. the results of policy/strategy evaluations, foresight exercises or technology assessment, and also of analyses of changing innovation processes, of the dynamics of changing research systems and changing functions of public policies. Thus, strategic intelligence facilitates a more "objective" formulation of diverging perceptions by offering appropriate indicators, analyses and information-processing mechanisms.
- (3) *Principle of mediation and alignment*: strategic intelligence facilitates debates and "discourses" between contesting actors in related policy arenas, thus mediating and "moderating", supported by "objectivised" information to be "digested" by the struggling parties. Mutual learning about the perspectives of competing actors and their interest backgrounds can ease an alignment of views.
- (4) *Principle of decision support*: strategic intelligence requires forums for negotiation and the preparation of policy decisions. The outcome of participatory, objectivised and mediated alignment processes will facilitate political decisions – not least as a response to the political quest for democracy vis-à-vis technological choices –, and effectuate the successful subsequent implementation.

In order to illustrate these principles, two examples of strategic intelligence for innovation policy will be discussed in the following.

***Using Foresight (Delphi) Results for the Evaluation of a Research Organisation: The Case of the Fraunhofer-Gesellschaft***

Using technology foresight results in order to evaluate a research institution enables evaluators to get an overview of the fit between perceived future developments in science and technology world-wide and the performance portfolio of a given publicly (co-) funded research organisation. By constructing an adequate index the results of e.g. a Delphi study may be compared with the research activities and/or the staff competencies of a given sample of research units.

The following example provides some evidence of the applicability of this approach. In 1996, the German Chancellor and the Prime Minister of the federal "Länder" decided to evaluate all major research institutions which are jointly financed by the Federation and the Länder (i.e. the Fraunhofer-Gesellschaft; the Max-Planck-Gesellschaft; the Deutsche Forschungsgemeinschaft; the G.W. Leibniz-Gesellschaft; the Helmholtz-Gesellschaft). The strategic aim of the envisaged "system evaluations" of these organisations was not a detailed analysis of the research performance of their units, but the *assessment of the actual functioning* of these organisations in the context of the German "research landscape" as a part of the

innovation system. International evaluation panels were formed in order to conduct these evaluations.

The Fraunhofer-Gesellschaft (FhG) is a semi-public contract research organisation consisting of 49 quite autonomous institutes, primarily active in the field of applied technological research (see Trischler/vom Bruch 1999). Among the most important issues of the FhG evaluation were questions like: Which technology-related markets promise the largest growth (world-wide and nationally)? Is FhG sufficiently represented in these markets? Does the technological portfolio of FhG fit with related technological developments world-wide?

The international panel in charge of the evaluation decided to employ – inter alia – the results of the German "Delphi '98" Study (Cuhls/Blind et al. 1998) as a benchmark for FhG's research and technology competencies. The report offered some 1,000 "visions" of "problem solutions" based on future scientific or technological achievements: in a Delphi process conducted on behalf of the German Ministry for Research (BMBF) these visions had been checked by some 1,000 experts from science, industry, and societal organisations. For each vision the "Delphi '98" Study presented information about its feasibility, the time horizon of its realisation, and also an assessment of the frame conditions fostering or hampering the realisation of a vision (e.g. the performance of the related public research infrastructure).

For the purpose of the FhG benchmarking, a "RETIED Index" was constructed, consisting of three Delphi criteria which were considered to be important for FhG, i.e. showing a future demand for R&D activities of the Fraunhofer institutes:

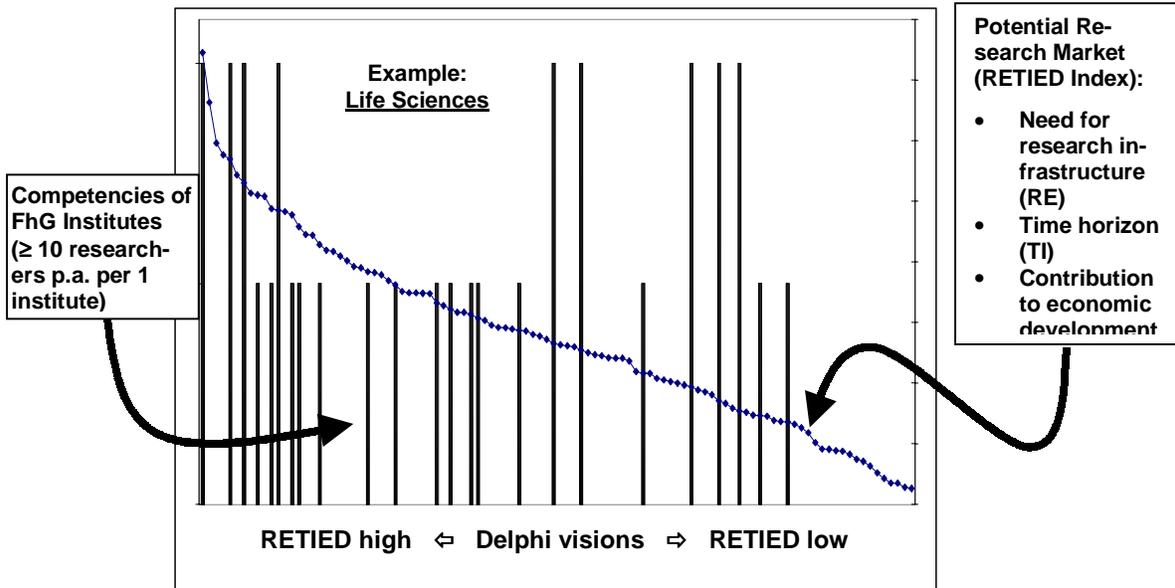
- (1) necessity of an improvement of the research infrastructure (RE),
- (2) time horizon of the realisation of a technological innovation (TI),
- (3) contribution of an innovation to the economic development (ED).

Within each thematic sub-field (e.g. information and communication technologies, life sciences, environment and nature, mobility), the Delphi visions were sorted according to this index (see figure 3, right hand).

As a next step the competencies of the Fraunhofer Society were assigned to the sorted visions: an internal group of Fraunhofer experts rated the competencies of FhG along various performance indicators (e.g. significant research competencies and personnel in at least one or two institutes) (see figure 3, left hand). Hereby a set of figures of "important visions" of future developments in science and technology was gained on the one hand and FhG-related competencies on the other. The matching of the two heterogeneous but inter-related strands of information revealed in an informative manner strengths and weaknesses of FhG's competencies vis-à-vis potential future research markets. The evaluation panel received these figures as a crucial input to the overall assessment of the adequacy of the given FhG portfolio.

With respect to the general principles for strategic intelligence presented above, this case of using foresight results as a means of evaluation may be assessed in the following way:

Figure 3: Combining Foresight Results with Evaluation Purposes - Example: System Evaluation of the Fraunhofer Society (FhG)



- (1) *Principle of participation*: the use of Delphi data – based on assessments of 1,000 experts – introduced an unusually broad representation of views of future research needs, coming from science, industry, and society experts outside FhG. Thus, the scope of views and expertise represented by the relatively small international evaluation panel was opened up considerably.
- (2) *Principle of "objectivisation"*: FhG had to present and defend its research portfolio vis-à-vis the evaluation panel, hoping for a positive assessment. The "injection" of non-partisan Delphi data into the arena worked as a relatively "objective" benchmark of required future FhG capabilities.
- (3) *Principle of mediation and alignment*: the matching of Delphi priorities and FhG competencies revealed strengths and weaknesses of the FhG portfolio, e.g. providing evidence of a weak position in life sciences (see figure 3). The FhG management, nevertheless, got the opportunity of commenting on each obvious (mis)match: there may have been good reasons why FhG should not invest too heavily into a certain field of technology (e.g. because of already existing strong competitors). Related discussions eased an alignment within the panel and with the FhG management.

- (4) *Principle of decision support*: The results of the FhG "system evaluation" – based inter alia on the Dephi/FhG portfolio matching – facilitated the preparation of political decisions and their subsequent implementation: the FhG management could utilize the matching figures as a means to achieve "objectivised" decisions on the prioritisation of research strategies, not least with respect to single institutes.

### ***Benchmarking of Innovation Clusters: Nano-technology Competence Centres***

The German Federal Ministry for Education and Research (BMBF) has been promoting "virtual competence centres" since 1998 in the area of nano-technology, classified as a future key technology, in which "the economy, science and private capital from all over Germany is brought together, in order to open up these areas and develop new products up to market maturity. The USA, Japan and China are already investing heavily in nano-technology". The new competence networks should in the future also be able to play in the international "concert of the major nano-players"<sup>8</sup>.

The innovation-political model of the BMBF "Nano-technology Competence Centres" (NCC) initiative is the *innovation cluster approach* meanwhile well-known and trendy in many countries (OECD 1999c), modelled after successful innovation clusters like Silicon Valley in the area of ICT, or the Cambridge region in the field of biotechnology. The new promotional measure provides financial *support for the networking activities* – in particular a joint office facilitating co-operation and exchange amongst the members, e.g. running an Internet platform, offering seminars, training activities; standardisation efforts etc. – of the 50-90 institutional members of each of the six selected newly established NCC, each busy in a different area of nano-technology development and application. The membership of an NCC includes university institutes, non-university research institutes, small and medium-sized enterprises, large companies, banks (venture capital), educational institutions (vocational training), local community authorities (supporting industrial start-ups), etc. The research institutes participating in the six NCCs receive the networking support on top of considerable amounts of public research money already channelled through BMBF's nano-technology research programme.

With the start of the NCC initiative, a *monitoring evaluation process* has been launched (conducted by the Fraunhofer Institute for Systems and Innovation Research, ISI): BMBF policymakers have been concerned about the risk of serious free-rider effects originating from the simple and quite likely interest of the participating research institutes in getting easier access to BMBF research funds, overriding BMBF's structural political aim of effectively linking and embedding their re-

---

<sup>8</sup> Press release of the Federal Ministry for Education and Research of 12.08.1998. Many of the institutions participating in the competence centres have themselves long been intensively involved in international science and technology networks, the national focus of the BMBF promotional measure seems almost arbitrary by contrast.

search into the envisaged market-oriented innovation clusters. Therefore, the evaluation exercise was designed to help the NCCs as a "critical friend" with their attempts at developing an effective innovation network relationship amongst NCC members.

One important element of the "critical friendship" was the provision of appropriate success criteria for target attainment of the NCCs, embedded as they are in (greatly) differing technological-industrial framework conditions. The problematic restructuring of East Germany's economy had demonstrated that if there is no industrial core, then even good promotional concepts for industrial research will fail or at least require considerably longer for successes to emerge.

Among other factors, the different conditions for success of the NCCs can be determined by estimating their respective technology areas according to the *maturity in the technology cycle* and by taking the *differing competitive conditions of the industrial branches* into consideration, in which the R&D results have to be transformed into marketable products:

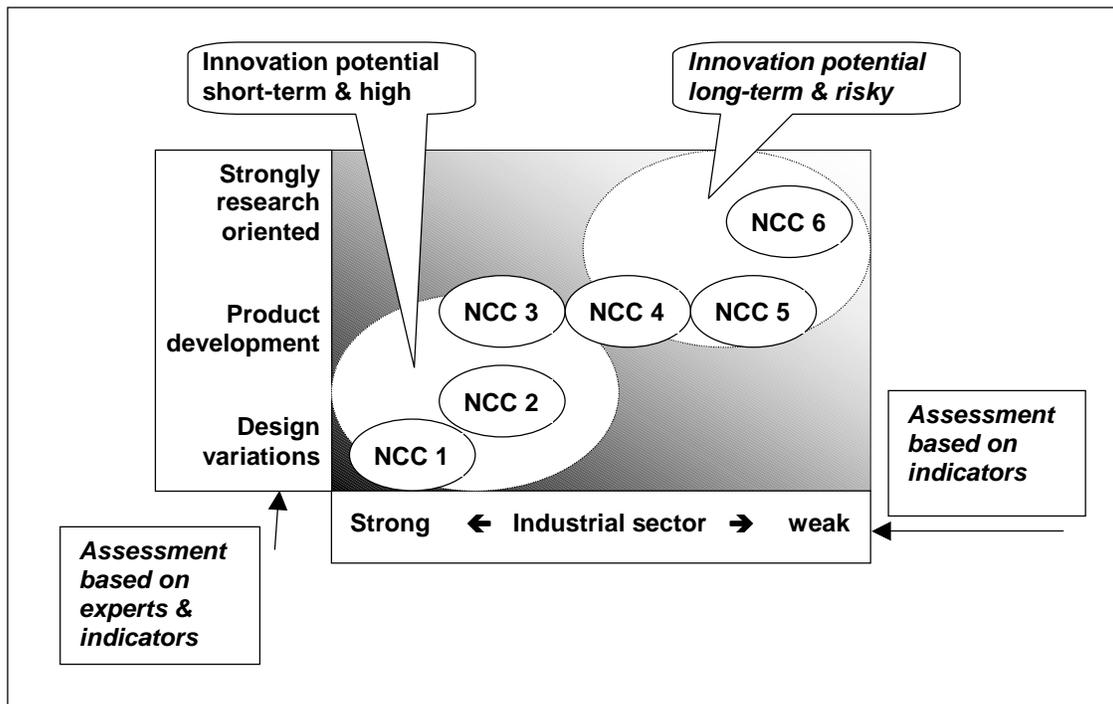
- Commercial successes are easier to achieve with close-to-market or application-ready technology developments, while basic research projects on the other hand will have greater difficulties in achieving successes. The NCCs however do not have a completely free hand in their prioritising between these two poles, depending on the maturity of their field of competence.
- Successes are more probable if the promoted NCCs are embedded in a strong and receptive industrial background, a criterion more likely fulfilled e.g. in chemicals, where German industry has traditionally played a leading role in the world market, than in microelectronics where in the past the German position in the world market has been that of a follower, or can now be characterised as defensive or catching up.

A well-founded basis for the assessment of technological maturity vs. longer-term innovation potential of the NCCs is provided by a thorough expert document "Innovation Push from the Nano Cosmos" (VDI-TZ 1998), a mixture of technology foresight and technology assessment. Here a broad spectrum of R&D projects was judged according to the degree of innovation and the anticipated time horizon up to market maturity. Since this database was no longer sufficiently up-to-date, the monitoring evaluation had also to rely on the additional expertise of the NCCs themselves. In estimating the competitiveness of the industrial environment in which the NCCs are embedded, the evaluation also referred to the annual report on Germany's technological performance (BMBF 1999) which is based – inter alia – on thorough empirical studies of the international competitiveness of those industrial branches which could apply the NCCs' R&D results.

The resulting benchmarking of the technology orientation and the industrial background – as depicted in figure 4 – reflects an "objectivised" perception (from the

viewpoint of the monitoring evaluation team). Figure 4 illustrates how the six NCCs' innovation potential can be located at different positions along the time axis. At a workshop aimed at motivating the centres to start strategic reflections on their respective technical-industrial affiliations, this benchmarking exercise was presented to representatives of the NCCs. The results were intended to guide the strategy development of the NCCs, and also to inform the further evaluation of NCCs' success or failure. Not surprisingly, at first the benchmarking received a lot of criticism from the NCCs: by revealing *significant differences of the innovation potential*, the exercise urged the NCC representatives to comment on a certain lack of compliance of their previous official statements (as laid down e.g. in BMBF funding applications) with the actual time horizons of their activities.

Figure 4: Benchmarking of Competence Centres with Respect to their Innovation Potential



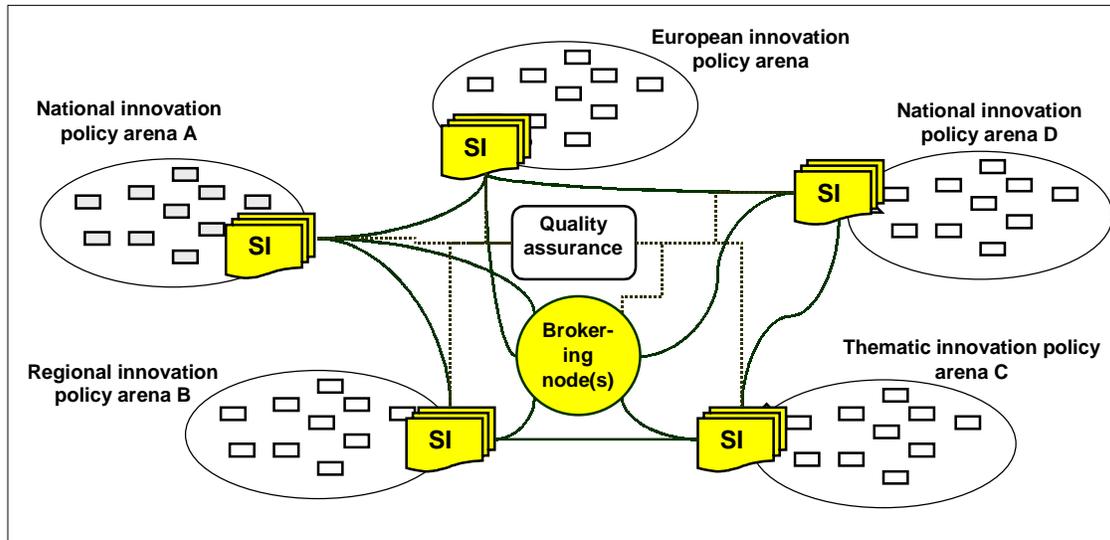
Eventually, this mobilisation of heterogeneous sources of strategic intelligence helped to introduce a *tone of "objectivity"* in the discussions with BMBF. With respect to the general principles for strategic intelligence presented above, this case may be assessed in the following way:

- (1) *Principle of participation*: the benchmarking exercise followed a multiple perspective approach by gathering information both from the NCCs and from a variety of other relatively reliable information sources.
- (2) *Principle of "objectivisation"*: the benchmarking exercise made an attempt to inject "objectivised" information into the BMBF/NCC arena. The actual innovation political problem was not the fact that the differing short resp. long-term innovation potentials of the NCCs were to be judged as better or worse – rather, an "objective" perspective on the actual and feasible innovation potential of a given NCC was a crucial premise of the design of realistic and tailor-made development strategies for and by each of the NCCs.
- (3) *Principle of mediation and alignment*: the benchmarking results provoked heated debates amongst the contesting actors, i.e. within and between the NCCs, and vis-à-vis the BMBF and the evaluation team. After a couple of months of discussion, a "moderating" effect could be noted, i.e. the actors had "digested" the information (including the evaluators, who had learned to refine their methodology). Mutual learning about the perspectives of competing actors had been eased.
- (4) *Principle of decision support*: eventually, the benchmarking exercise was one step of several within an ongoing learning process of BMBF and of the NCCs on how to "artificially" create and grow innovation clusters successfully: there is an obvious need to develop sustainable innovation strategies for and by the centres themselves, a precondition of which is sufficient room to manoeuvre for the NCCs. In early 2001 the BMBF launched a related guideline for the NCCs.

### **General Requirements for Distributed Intelligence**

The examples discussed in the previous section demonstrated that the application of strategic intelligence – in particular of its first three principles: participation; objectivisation; mediation and alignment – can be further effectuated if strategic information is gathered simultaneously from several independent and heterogeneous sources. Therefore, the *second route* to improved strategic intelligence leads us to the concept of *distributed intelligence*. This concept starts from the observation that policymakers and other actors involved in innovation processes only use or have access to a small share of the strategic intelligence of potential relevance to their needs, or to the tools and resources necessary to provide relevant strategic information. Such assets, nevertheless, exist within a wide variety of institutional settings and at many organisational levels, though scattered across the globe. As a consequence, they are difficult to find, access and use. Hence, rectifying this situation will require major efforts to develop interfaces enhancing the transparency and accessibility of already existing information, and to convince potential users of the need to adopt a broader perspective in their search for relevant intelligence expertise and outputs.

Figure 5: Architecture of Distributed Strategic Intelligence



Consequently, an architecture and infrastructures of distributed intelligence (see figure 5) must allow access, and create inter-operability across locations and types of intelligence, including a distribution of responsibilities with horizontal as well as vertical connections, in a non-hierarchical manner. Such an architecture of distributed strategic intelligence would, at least, limit the public cost and strengthen the "robustness" of intelligence exercises. Robustness, nevertheless, presupposes also provisions for quality assurance, boosting the trust in distributed intelligence based debates and decision-making. Five *general requirements* of infrastructures for distributed intelligence can be stipulated:

- (1) *Networking requirement*: the architecture of "infrastructures" for distributed intelligence should neither be designed as one monolithic block nor as a top-down system – rather the opposite: ideally the design allows for *multiple vertical and horizontal links* amongst and *across the existing* regional, national, sectoral, and transnational *infrastructures* and facilities of the related innovation systems and policy arenas.
- (2) *Active node requirement*: in order to guarantee a sustainable performance of distributed intelligence and to avoid hierarchical top-down control, the architecture would have to offer active brokering "nodes" (or "hubs") for managing and maintaining the infrastructure. *Three types of active nodes* can be distinguished:
  - (a) The first type provides *enabling facilities*, e.g. a "foresight bank" for policy-makers, research institutions, non-governmental organisations, or enterprises (see e.g. the fictional case described in box 1). The objective is to render results arrived at in one place directly accessible in another, without requiring direct

contacts between actors in both places. There is no need to contact the promoter e.g. of the UK or Japanese science and technology Delphi or the experts they have mobilized, to get hold of the results arrived at concerning a certain technology. The word "facility" is important here, since it tells not only about the need for developing and maintaining/updating a bank, it also underlines the importance of "harmonisation" efforts easing the absorption of "foreign" results and thus facilitating their circulation: "compatibility" issues do not only raise technical "inter-operability" problems, they also raise issues about reliability, i.e. processes through which results are arrived at. (b) The second type of node offers a "directory" allowing direct connections between relevant actors. Often policymakers and strategists need not only information about certain strategic intelligence exercises, e.g. "technology assessments" or "consensus conferences" which took place on the relevant subject and the compromises arrived at (as in the "Biomat case" sketched in box 1), but wish to grasp the reasons why such actions arose, what argumentation was developed by stakeholders, and through what process "consensus" or "dissensus" was arrived at, i.e. information they can get only through immediate interaction with the involved actors. Hence, this rather classical directory would work as an intermediary, facilitating direct connections between relevant actors. (c) A third type of node offers a "register" allowing free access to all strategic intelligence exercises undertaken under public auspices. The variety of policy arenas and of problems addressed require information not only dedicated to a given "instrument", but centred on processes and on the related combinations of instruments. The issue here is less the notion of best practice, but a topological notion: the universe of situations is so vast that the idea of "best practices" does not seem appropriate; instead it would be replaced by the concept of a register of practices with *indications how and why certain procedures and practices functioned*. Over time, the register would include more and more strategic exercises, requiring some structuring, so that reference to the register will make new exercises more relevant and/or less costly, hence facilitating *collective learning processes*.

(3) *Transparent access requirement:* clear rules concerning the access to the infrastructure of distributed intelligence have to be defined, spanning from *public domain* information areas to *restricted services*, accessible only for certain types of actors or after charging a fee.

(4) *Public support requirement:* in order to guarantee a high degree of independence the distributed intelligence infrastructure is in need of a regular and reliable support by public funding sources. This applies in particular to the basic services provided by the "brokering nodes"; adequate resources will make them robust. It does not, however, prevent the node providers from additionally selling market-driven information services, thus extending their financial base.

(5) *Quality assurance requirement:* the notion of "quality assurance" relates directly to issues of trust: how can actors in policy arenas trust in all the "intermediaries" mobilised in the course of the preparation or conduct of policy-making? Three major avenues of quality assurance can be followed: (a) bottom-up processes of institutionalisation amongst the providers of strategic intelligence may

**Box 1: SME considering a technological move – a fiction**

It is the year 2005 and the Managing Director of Biomat<sup>®</sup> has a problem. Her spin-off firm builds replacement human organs using metagenic technology, but after two years of success Biomat<sup>®</sup> is at a crossroads. Should it stick to metagenics, which is costly and prone to production problems, or should it use the latest ultragenic approaches - still unproven but likely to yield great cost reductions? She switched on her videophone and traced the local head of RIB, the Regional Innovation Bureau of ENDBITS, the European Network of Distributed Bureaux of Intelligence for Technology Strategies. RIB helped her to prepare a videonote on technology options. It ran a standard search on the European Foresight Bank, an electronic tool which logged all of the world's foresight outputs and used AI algorithms to cluster the results and build scenarios. Recent expert assessments all looked good for ultragenics, but RIB advised her not to rely solely upon foresight results – however positive. Social and regulatory problems were also possible, and the Bureau had heard of some problems in Austria. RIB used the Technology Assessment directory to identify the main Austrian experts in the field and confirmed that ultragenics had been subject to ethical challenges from a local religious foundation. RIB then called for more information and scanned the recordings of the Consensus Conference. Relief! The objections were based on a misunderstanding of the procedures for ultragenics (which unlike earlier approaches did not depend upon foetal cells) and the citizens' jury had come out in favour of the technology. Biomat<sup>®</sup> was ready to launch its ultragenics research programme, but was worried about the cost. RIB pointed out that all three of the European Research Framework Programme agencies offered support, but noted that recent evaluations praised the Prague office for its fast turnaround of proposals and claims. "Thanks RIB" said the Biomat<sup>®</sup> MD. "Life without ENDBITS just wouldn't be the same."

(L. Georghiou in: Kuhlmann et al. 1999)

play a crucial role, in particular *professional associations* (like e.g. the American Evaluation Association, the European Evaluation Society, and the growing number of national evaluation associations that have been established since the 1990s). Also, scientific and *expert journals* are indispensable means of maintaining and improving the professional level of services. Furthermore, education and training in the area of strategic intelligence for innovation policy have to be extended and improved, in particular on graduate and postgraduate levels of

*university teaching* (see e.g. the "science and technology policy programs" and the like offered meanwhile by quite a number of American universities). (b) A second means of quality assurance is the establishment of *accreditation mechanisms* for providers of strategic intelligence, based on a self-organising and vivid "scene" of experts. (c) A third and basic source of quality assurance would have to be guaranteed through a reliable support with *repeated and "fresh" strategic intelligence exercises* (e.g. evaluation, foresight, technology assessment) and new combinations of actors, levels, and methods initiated and funded by innovation policymakers across arenas and innovation systems.

### **Enhancing Distributed Intelligence for Innovation Policymaking on the European Level**

Presently, the concept of distributed strategic intelligence is gaining in importance in particular on the European scale. One can trace, on top of the national and regional efforts and in parallel with Europe's economic and political integration, the emergence of an architecture and infrastructures of a *European innovation policy-making system* (see e.g. Peterson/Sharp 1998; Guzzetti 1995). It has been established not only in order to run the European Commission's "Framework Programmes for Research and Technological Development" (FPs) but also – according to Article 130 of the Maastricht Treaty – aiming at a better co-ordination of genuine European, national and regional and policy efforts (Caracostas/Muldur 1998, 127ff), i.e. at transnational governance structures. Here, pressing questions arise about the inter-relationship between emerging transnational political institutions and the actual policy development within national innovation systems, not least vis-à-vis internationalising markets for technology-related products and producers. The European Commission's ongoing efforts at compiling and preparing the information basis for the implementation of the "*European Research Area*" (European Commission 2000) provide vivid evidence of the urgent need for an appropriately adapted *infrastructure of European distributed strategic intelligence*. Presently, public agencies, data base providers and policy analysts across Europe are delivering bits and pieces of knowledge and information to the EU Commission's DG Research in order to sketch benchmarks of national research and innovation policies, of indicators for the identification of "centres of excellence", etc. If there were more reliable linkages and robust "brokerage nodes" between strategic intelligence systems, the synergy effects could be significantly further advanced.

Still though, the production and the use of strategic intelligence in Europe are spreading across a diverse "landscape" of research institutes, consulting firms, and government agencies, which have emerged over decades in various national, political, economic, and cultural environments, thus reflecting different governance structures, only loosely inter-connected. So far, just a few facilities, like the Institute for Prospective Policy Studies (IPTS) and the European Science & Technology Ob-

servatory (ESTO), are attempting to work as "brokerage nodes" between the various strategic intelligence providers and users across Europe.

There is much to be gained from customised exercises involving combinations and permutations of foresight, evaluation, technology assessment, for all can be seen as decision support tools which look back to see what we can learn from the past (evaluation), consider potential options for the future (foresight), and examine the implications of particular lines of action (technology assessment). There is also much to be gained from their systematic use and exploitation in multiple settings, with outputs accessible and distributed across actors, levels, environments etc. (see the fictional case in box 2), for instance in terms of

- matching of various national research and funding priorities
- distribution of responsibilities between regional, national, or European public or private actors
- public acceptability of science and technology and an increase in the transparency of related policy actions (e.g. as the Internet-based "Futur-Prozess", recently launched in Germany extending the Foresight experiences of the 1990s; see [www.futur.de](http://www.futur.de)).

### **In Essence ...**

To sum up briefly, in this paper we have argued for a new approach which we have called a system of distributed intelligence. In particular, we have suggested the development of tools which can be used in different combinations to enhance strategic intelligence inputs into policymaking, and access to, and exploitation of, strategic intelligence in different locations for different reasons. Initiating and exploiting these intelligence tools in a systematic fashion across innovation systems will demand new architectures, institutions, configurations and their inter-linkages.

This paper started quoting Friedrich Nietzsche: "Before any impact you believe in other causes than after the impact". If we manage to develop and implement a new strategic intelligence infrastructure, then research and innovation policies could become more realistic, efficient, more relevant, and more democratic. Four *basic principles for effective strategic intelligence* were figured out in this paper:

- (1) *Principle of participation*: Foresight, evaluation or technology assessment exercises take care of the diversity of perspectives of actors avoiding maintaining one unequivocal "truth" about a given innovation policy theme.
- (2) *Principle of "objectivisation"*: strategic intelligence facilitates a more "objective" formulation of diverging perceptions by offering appropriate indicators, analyses and information-processing mechanisms.

**Box 2: The Emergence of a European Technology Programme – A Fiction**

September 2007: The newly established "Joint Office for Socio-technological Programmes" between the European Commission and the European Parliament is under multiple pressures. The Office is the result of new procedures adopted for the Sixth Research Framework Programme of the European Union (FP6). No clear pattern for putting a problem on the political agenda has yet emerged while candidate actions multiply to get hold of the billions of euros that are still pending! It is true that major technology foresight exercises are held every five years (if only to please the Japanese), and these produce listings of promising technologies. The 2006, UK-led Foresight Exercise, as well as the parallel Japanese Delphi Foresight, have identified new solid-state technologies which allow much higher conversion rates of solar to electrical energy. This creates new possibilities for centralised solar energy power plants. The Office is particularly keen to pursue this lead, and perhaps establish a development and demonstration programme, one reason being the disarray of the nuclear power programme, even in France. Quite a number of countries have passed laws against recycling nuclear waste.

Alternative ways of providing electrical power should be developed, and be part of the portfolio of public programmes. The Office prepares itself well by inviting C<sup>2</sup>, the Consultancy Consortium, to prepare a background report on the new solid-state technology for solar power and its societal impacts. The Consultancy Consortium, established in 2005, led by the respected consultancy firm John D. Big, pools the dedicated technology assessment and foresight studies of its members (which include the consultancy arms of some major research universities). The data remain confidential, and the Consortium charges a fee for delivering analyses based on them. An important feature of the Consortium is that they recognise a civic duty to deliver such analyses with the public interest in mind. This rule allowed the universities to come in, and shifted the role of the Consortium from that of a self-interested actor to a node in the network.

The action of the Office coincided with the publication, by the Association of Mediterranean Regions, of a study of solar energy options and critical issues, which ended up commending "centralised thermal solar" as the most promising solution. As environmental groups criticising the required concentration of mirrors for transforming the last untouched landscapes of the Mediterranean zone were quick to point out, at least one of the champions for these plans was a member of the Joint Office. Whether the coincidence was indeed a case of lobbying was not clear, but the suggestion added force to their general argument that major public investments should be postponed until the new technology has proven its efficacy, its reliability and demonstrated cost-effectiveness.

While the Office awaited the report of the Consultancy Consortium, it realised that it needed further, independent inputs to overcome a possible stalemate between proponents and opponents. It was decided to set up a series of "research programming" version of consensus conferences. The conferences were exciting events for the participants, not least because of new information and communication technologies: whizz-kids from Big Heart Company introduced comic strip balloons offering URL linkages to key words as well as to experts speaking out, which allowed all participants to contextualise what was being introduced.

The programme finally proposed to the European Council and European Parliament is novel in two ways. One, it is not a finished programme, but linked to ongoing activities of actual and prospective participants. Two, as a programme, it is implemented in two stages: firstly, there is a framework for articulation and implementation of programme goals. And secondly, the implementation of the programme has been delegated to the Association of Mediterranean Regions which has been very active in organising consensus conferences in all regions.

*Source: Ph. Laredo in: Kuhlmann et al. 1999*

*Principle of mediation and alignment:* strategic intelligence facilitates mutual learning about the perspectives of competing actors and their interest backgrounds can ease an alignment of views.

- (4) *Principle of decision support:* the outcome of strategic intelligence processes will facilitate political decisions and effectuate the successful subsequent implementation.

Thereby, no single "correct" or "best" configuration of tools, procedures, institutions and structures can be used in all contexts and situations. So far, the focus has been on national level policy configurations, but we can see that regions and supra-national organisations or even "thematic" organisation have become more important as policy arenas. Moreover, there is a growing need for new configurations which link up private and public actors and promote their interaction. By private actors we do not mean only companies, but also representatives of many other stakeholders (professional associations, consumer organisations, environmental organisations etc.).

The application of strategic intelligence can be further effectuated if information is gathered simultaneously from several independent and heterogeneous sources. Therefore, a *second route* to improved strategic intelligence leads us to the concept of *distributed intelligence*. This concept starts from the observation that policymakers and other actors involved in innovation processes only use or have access to a small share of the strategic intelligence of potential relevance to their needs, or to the tools and resources necessary to provide relevant strategic information. Such assets, nevertheless, exist within a wide variety of institutional settings and at many organisational levels, though scattered across the globe. As a consequence, they are difficult to find, access and use.

In distributed intelligence, a decentral architecture of information sources will be unfold – spanning across innovation systems and related policy arenas – working as brokering nodes which guide and enable the supply of strategic intelligence. Five *general requirements* of such infrastructures can be stipulated:

- (1) *Networking requirement:* distributed intelligence will not be designed as a top-down system – rather the opposite: ideally the design allows for multiple vertical and horizontal links across the existing sources of strategic intelligence.
- (2) *Active node requirement:* three types of active nodes can be distinguished: (a) The first type provides enabling facilities, e.g. a "foresight bank". (b) The second type delivers a "directory" allowing direct connections between relevant actors. (c) A third type offers a "register" allowing free access to all strategic intelligence exercises undertaken under public auspices, hence facilitating collective learning processes.
- (3) *Transparent access requirement:* clear rules concerning the access to the infrastructure of distributed intelligence are needed.

- (4) *Public support requirement*: distributed intelligence infrastructure is in need of a regular and reliable support by public funding sources.
- (5) *Quality assurance requirement*: three major avenues of quality assurance can be followed: (a) professional associations; expert journals; university teaching; (b) accreditation mechanisms for providers of strategic intelligence, based on a self-organising "scene" of experts; (c) a reliable support with repeated and "fresh" strategic intelligence exercises and new combinations of actors, levels, and methods initiated by innovation policymakers across arenas and innovation systems.

## References

- Airaghi, A. /Busch, N. /Georghiou, L. /Kuhlmann, S. /Ledoux, J.M /van Raan, A. /Viana Baptista, J. (1999): Options and Limits for Assessing the Socio-Economic Impact of European RTD Programmes. Report to the European Commission
- BMBF (1999): Berichterstattung zur technologischen Leistungsfähigkeit Deutschlands. Bonn.
- Bozeman, B. / Melkers, J. (eds.) (1993): Evaluating R&D Impacts: Methods and Practice, Boston et al (Kluwer Academic Publishers)
- Brewer, G.D./de Leon, P. (1983): The Foundation of Policy Analysis, Homewood, Ill., 17-21
- Callon M. / P. Laredo / Mustar, P. (1997): The Strategic Management of Research and Technology, Paris (Economica International)
- Callon, M. (1992): The Dynamics of Techno-Economic Networks. In: Coombs, R. / Saviotti, P. / Walsh, V. (eds.): Technological Change and Company Strategies: Economic and sociological perspectives, London et al. (Academic Press Limited), 72-102
- Cameron, H. / Loveridge, D. / Cabrera, J. / Castanier, L. / Presmanes, B. / Vasquez, L. / van der Meulen, B. (1996): Technology Foresight - Perspectives for European and International Co-operation, Manchester (PREST: Mimeo)
- Caracostas, P. / Muldur, U. (1998): Society, the Endless Frontier – A European Vision of Research and Innovation Policies for the 21st Century, Luxembourg/Brussels (Office for Official Publications of the European Communities) (EUR 17655)
- Coombs, R. (1999): Innovation in the Services. Overcoming the Services-Manufacturing Divide. Antwerpen/Apeldoorn (Maklu) (Nijmegen Lectures on Innovation Management, 3)
- Cuhls, K. (1998): Technikvorausschau in Japan. Ein Rückblick auf 30 Jahre Delphi-Expertenbefragungen, Physica Verlag, Heidelberg
- Cuhls, K. / Blind, K. et al. (1998): DELPHI '98 Umfrage. Studie zur globalen Entwicklung von Wissenschaft und Technik. Edited by the Fraunhofer Insti-

- tute for Systems and Innovation Research on behalf of the Federal Ministry for Education and Research (BMBF). Karlsruhe (ISI)
- Den Hertog, P. / Bilderbeek, R. / Maltha, S. (1997): Intangibles. The soft side of innovation. In: *Futures*, Vol. 29, No.1, 33-45
- Dror, Y. (1968): *Public Policymaking Reexamined*, San Francisco/CA (Chandler)
- Edquist, Ch. (ed.) (1997): *Systems of Innovation. Technologies, Institutions and Organisations*, London/Washington (Pinter)
- Etzkowitz, Henry / Leydesdorff, Loet (2000): The dynamics of innovation: from National Systems and "Mode-2" to a Triple Helix of university-industry-government relations. In: *Research Policy* 29, 109-123
- European Commission (2000): *Towards a European research area, Communication from the Commission to the Council, the European Parliament, the Economic and Social Committee and the Committee of the Regions*, Brussels, 18 January 2000, COM (2000) 6
- Flechtheim, O. K. (1968): *Futurologie - Möglichkeiten und Grenzen*, Frankfurt/M./Berlin
- Forrester, J. W. (1971): *Der teuflische Regelkreis. Das Globalmodell der Menschheitskrise*, Stuttgart
- Freeman, C. (1987): *Technology Policy and Economic Performance: Lessons from Japan*, London (Pinter)
- Georghiou, L. (1995): Research evaluation in European National Science and Technology Systems. In: *Research evaluation*, Vol. 5, No. 1, 3-10
- Georghiou, L. / Roessner, D. (2000): Evaluating technology programs: tools and methods, *Research Policy* (29) 4-5, 657-678
- Gibbons, M. / Limoges, C. / Nowotny, H. / Schwartzman, S. / Scott, P. / Trow, M. (1994): *The new production of knowledge. The dynamics of science and research in contemporary societies*, London et al. (Sage)
- Gordon, T. J. / Helmer, O. (1964): *Report on a Long-Range Forecasting Study*, Rand Corporation, Santa Monica/ California
- Grupp, H. (1998) (ed.): *Technological Forecasting & Social Change; Special Issue on Recent National Foresight Activities*, Vol. 57
- Grupp, H. (ed.) (1992): *Dynamics of Science-Based Innovation*, Berlin et al. (Springer)
- Grupp, H. (ed.) (1993): *Technologie am Beginn des 21. Jahrhunderts*, Heidelberg (Physica/Springer)
- Grupp, H. / Kuntze, U. / Schmoch, U. (1995): New Technology Indicators for the Evaluation of Research and Development Programmes. In: Becher, G. / Kuhlmann, S. (eds.): *Evaluation of Technology Policy Programmes in Germany*, Boston et al. (Kluwer Academic Publishers), 243-284
- Guzzetti, L., 1995, *A Brief History of European Research Policy*. Brussels/Luxembourg (Office for Official Publications of the European Communities)
- Helmer, O. (1966): *Social Technology*, Basic Books, New York/ London

- Helmer, O. (1967): *Analysis of the Future: The Delphi method*, Rand Corporation, Santa Monica
- Helmer, O. (1983): *Looking Forward. A Guide to Futures Research*, Beverly Hills/London/ New Delhi
- Hollingsworth, R. / Boyer, R. (eds.), 1997, *Contemporary Capitalism. The Embeddedness of Institutions*, Cambridge University Press
- Howells, J. (1999): *Regional systems of innovation?* In: Archibugi, D. / Howells, J. / Michie, J. (eds.), *Innovation Policy in a Global Economy*, Cambridge (Cambridge University Press), 67-93
- Irvine, J. / Martin, B. R. (1984): *Foresight in Science, Picking the Winners*, London/Dover
- Jasanoff, Sh. (1997): *Introduction*. In: Jasanoff, Sh. (ed.): *Comparative Science and Technology. The International Library of Comparative Public Policy*, vol. 5, Cheltenham (Edward Elgar), xiii-xxiii
- Kitschelt, H. (1991): *Industrial Governance Structures, Innovation Strategies, and the Case of Japan: Sectoral or Cross-National Comparative Analysis?* In: *International Organization*, 45 (4), Autumn, 453-493
- Kodama, F. (1995): *Emerging Patterns of Innovation. Sources of Japan's Technological Edge*, Boston (Harvard Business School Press).
- Kuhlmann, S. (1997): *Evaluation as a Medium of Science & Technology Policy: Recent Developments in Germany and Beyond*. In: OECD (ed.): *Policy Evaluation in Innovation and Technology, Towards Best Practices*. Paris, 443-460
- Kuhlmann, S. (1998): *Politikmoderation. Evaluationsverfahren in der Forschungs- und Technologiepolitik*, Baden-Baden (Nomos)
- Kuhlmann, S. (2001): *Governance of Innovation Policy in Europe – Three Scenarios*. In: *Research Policy, Special Issue "Innovation Policy in Europe and the US: New Policies in New Institutions"*, edited by Hans K. Klein, Stefan Kuhlmann, and Philip Shapira (forthcoming)
- Kuhlmann, S. / Boekholt, P. / Georghiou, L. / Guy, K. / Héraud, J.-A. / Laredo. Ph. / Lemola, T. / Loveridge, D. / Luukkonen, T. / Polt, W. / Rip, A. / Sanz-Menendez, L. / Smits, R. (1999): *Improving Distributed Intelligence in Complex Innovation Systems*. Brussels/Luxembourg (Office for Official Publications of the European Communities) (see also <http://www.isi.fhg.de/ti/Final.pdf>)
- Kuhlmann, S. / Meyer-Krahmer, F. (2001): *Internationalisation of Innovation, Interdependence, and Innovation Policy for Sustainable Development*. In: Sweeney, G.P.: *Innovation, economic progress and quality of life*, Cheltenham (Edward Elgar), -86-110
- Kuhn, Th. (1962): *The Structure of Scientific Revolutions*, Chicago (University of Chicago Press)
- Laat, B. de (1996): *Scripts for the future. Technology foresight, strategic evaluation and socio-technical networks: the confrontation of script-based scenarios*, Amsterdam (PhD thesis)

*11 Issues in Linking RIT Evaluation With Policy*

- Linstone, H. (1999): *Technological Forecasting and Social Change*, Vol. 62 (1999) p. 1-8
- Loveridge, D. (guest ed.) (1996): *International Journal of Technology Management; Special Publication on Technology Assessment*, Vol. 11, Nos 5/6
- Luhmann, N. (1984): *Soziale Systeme*, Frankfurt (Suhrkamp)
- Lundvall, B.-Å. (ed.) (1992): *National Systems of Innovation: Towards a Theory of Innovation and Interactive Learning*, London (Pinter)
- Lundvall, B.-Å. / Borrás, S. (1998): *The globalising learning economy: Implications for innovation policy*, Luxembourg (Office for Official Publications of the European Communities) (Targeted Socio-Economic Research).
- Lundvall, B.-Å. / Maskell, P. (1999): *Nation states and economic development – from national systems of production to national systems of knowledge creation and learning*, in: Clark, G.L. / Feldmann, M.P. Gertler, M.S. (eds.), *Handbook of Economic Geography*, Oxford (Oxford University Press)
- Marin, B. / Mayntz, R. (1991): *Policy Networks. Empirical Evidence and Theoretical Considerations*. Frankfurt/Main and Boulder/Colorado (Campus; Westview Press)
- Martin, B. (1995): *Foresight in Science and Technology*. In: *Technology Analysis & Strategic Management* vol. 7, no. 2, 140
- Meadows, D. L.; Meadows, D. H. / Zahn, E. / Milling, P. (1972): *The Limits To Growth*, Universe Books, New York
- Meyer-Krahmer, F. / Montigny, P. (1989): *Evaluations of innovation programmes in selected European countries*. In: *Research Policy*, 18, vol. 6, 313-331
- Meyer-Krahmer, F. / Reger, G. (1999): *New perspectives on the innovation strategies of multinational enterprises: lessons for technology policy in Europe*. In: *Research Policy* (28) 7, 751-776
- Meyer-Krahmer, F./ Reger, G./ Archibugi, D./ Durand, Th./ Molero, J./ Pavitt, K./ Soete, L./ Sölvell, Ö. (1998): *Internationalisation of Research and Technology: Trends, Issues, and Implications for Science and Technology Policies in Europe*. Brussels/ Luxembourg (ETAN Working Paper, prepared for the European Commission)
- Mowery, D. C. (1994): *Science and Technology Policy in Interdependent Economies*, Boston/Dordrecht/London (Kluwer Academic Publishers)
- Nelson, R. R. (Hg.) (1993): *National Innovation Systems: A Comparative Analysis*, Oxford/New York (Oxford University Press)
- Niosi, J. (Hg.) (1999): *The Internationalization of Industrial R&D*. In: *Research Policy (Special Issue)*, vol. 28, nos. 2-3, 107-336
- OECD (1997) (ed.): *Policy Evaluation in Innovation and Technology. Towards Best Practices*, Paris (OECD)
- OECD (1999a): *Managing National Innovation Systems*, Paris (OECD)
- OECD (1999b): *Globalisation of Industrial R&D: Policy Issues*, Paris (OECD)
- OECD (1999c): *Boosting Innovation: The Cluster Approach*, Paris (OECD)

- Papaconstantinou, G. / Polt, W. (1997): Policy Evaluation in Innovation and Technology: An Overview. In: OECD (ed.): Policy Evaluation in Innovation and Technology. Towards Best Practices, Paris (OECD)
- Peterson, J. / Sharp, M., 1998, Technology Policy in the European Union, New York (St. Martin's Press)
- Reger, G. /Beise, M. /Belitz, H. (eds.) (1999): Internationalisierung technologischer Kompetenzen: Trends und Wirkungen in den Schlüsseltechnologien Pharmazie, Halbleiter und Telekommunikation. Heidelberg (Physica/Springer)
- Rip, A. / van der Meulen, B. (1997): The post modern research system, in: Barré, R. / Gibbons, M. / Maddox, J. / Martin, B. / Papon, P. (eds.): Science in Tomorrow's Europe, Paris (Economica International), 51-67.
- Rip, A./ Misa, Th.J. / Schot, J. (eds.) (1995): Managing Technology in Society. The Approach of Constructive Technology Assessment, London/New York (Pinter)
- Sanz, L. / Cabello, C. (2000): Expectations and Learning as Principles of Shaping the Future. In: N. Brown, B. Rappert and A. Webster (Eds): Contested Futures: A Sociology of Prospective Science and Technology. Aldershot: Ashgate
- Schön, D. / Rein, M. (1994): Frame Reflection. Toward the Resolution of Intractable Policy Controversies, New York (BasicBooks).
- Schot, J. / Rip, A./ (1997): The Past and the Future of Constructive Technology Assessment. In: Technological Forecasting and Social Change 54 (1996), 251-268
- Shapira, P. /Youtie, Y. (1998): Evaluation Industrial Modernization: Methods, Results, and Insights from the Georgia Manufacturing Extension Alliance. In: The Journal of Technology Transfer, vol. 23, No. 1, 17-27
- Smits R. / Leyten, J. / den Hertog, P. (1995): Technology Assessment and technology policy in Europe: new concepts, new goals, new infrastructures. In: Policy Sciences (28), 272-299
- Steinmüller, K. (1995): Beiträge zu Grundfragen der Zukunftsforschung, Werkstattbericht des Sekretariats für Zukunftsforschung 2/95, Gelsenkirchen
- Sundermann, K. / Bröchler, S. / Simonis, G. (eds.): Handbuch Technikfolgenabschätzung, Berlin 1999 (Sigma)
- Teece, D. (1986): Profiting from Technological Innovation: Implications for Integration, Collaboration, Licensing and Public Policy, in: Research Policy 15 (6), 285-305
- Trischler, H. / vom Bruch, R. (1999): Forschung für den Markt - Geschichte der Fraunhofer-Gesellschaft, München (Beck)
- VDI-Technologiezentrum (1998): Innovationsschub aus dem Nanokosmos. Technologieanalyse. Düsseldorf.

## **Discussion of Stefan Kuhlmann's paper**

**Arie Rip** (Twente University, The Netherlands) Referring to the way you showed pictures about intelligence, it would be mutually interesting to see examples of what is happening already in the distributed intelligence system.

**Stefan Kuhlmann.** Our thinking about this conception is normative and not a clear picture of the reality. There are problems of having access to examples since they can only provide examples of cases with interested stakeholders. There are also examples that were organized as very public events and the information was put on the Web and disseminated in other media as well. The work of the Foresight Group in the EU is a good example. A group of 50 civil servants developed scientifically based visions of Europe in 20-30 years and made the study publicly available which makes it good for debating the future of Europe.

**Louis Tornatzky** (Southern Technology Council, USA). Etzkowitz has raised the issue of a large collaborative project to develop a higher level of strategic intelligence. Can we do a regional U.S. collaborative program by doing meta-analysis of existing evaluations and collect some data based on university-industry linkages? Can we harvest what we've done already since a lot of work already performed? It would be good to focus on a narrower domain and focus on strategic intelligence.

**John Barber** (Department of Trade and Industry, UK). How would "strategic intelligence" be operationalized? If you look at the various actors, defining the concept that all can agree on would be difficult. Those not used to benchmarking would really be very interested. How can we do this so that people can relate to information/knowledge flow? Does this mean you have to improve linkages and existing dialogue? Otherwise you'd have to repackage it for the various groups involved.

---

## **Commentary**

**Irwin Feller**

Pennsylvania State University, USA

---

The discussions in these sessions have focused around three main concepts: 1) decision making; 2) evaluation; and 3) comparative U.S. and European policy. With regard to decision-making, I have observed that evaluation research is used in one or more of the following ways:

- shelved
- irrelevant because there's no sponsor
- used and cited because rationalized an existing study
- accepted but it's irrelevant because of the changing political setting
- explicitly rejected (because not politically accepted)
- enlightenment: evaluation enters the debate either as offensive or defensive technique
- instrumental value and clearly led to policy decision (with a clear lineage between finding and policy; however, not supposed to accept evaluation as leading to clear policy as that is dangerous)

It appears that European colleagues are more interested in outcomes and learning. I think this is a close to the policymaking environment. Americans are much more focused on technique, methodology and truth seeking. This seems to be reflected in the themes of the presentations.

What are the types of decisions that the policy maker is dealing with? There are two types: 1) To maintain a program - what's the level of program? Should it exist? This looks at incremental change and can be interpreted as types of learning. 2) Truth-seeking level: Should we have radical policy change? Should we get rid of ATP? Etc. There's a relationship between the type of policy analysis and the type of question that you are asking. It relates in part to decisions that you'd like to make. Evaluators work within the constraint of a sponsor. This may give us new findings yet we're working at a fairly constrained level with only incremental change. Only at the truth-seeking level can we really find new methods. Another focus should be on enlightenment to find new policies. We need to see whether we're asking right question.

**Luke Georghiou** (University of Manchester, UK). There's an unstated assumption here that the sponsor is in the executive branch or an agency. But if you shift from a program-level sponsor to elected officials, what is the impact on how the evaluation is done? We should not simply try to reach the program level manager but also try

## *11 Issues in Linking RIT Evaluation With Policy*

to get a broader conception of what the program means. This means moving from incremental to instrumental. Evaluation is part of a political game. No agency is going to accept findings out of an executive agency. Legislators have developed their own evaluations but we didn't hear anything about this from our European counterparts. What are the European political structures that are similar to the U.S. framework?

**Ken Guy** (Technopolis, UK). Twenty years ago I met Chris Freeman for the first time. He talked about evaluation as a learning process and adopted an incremental approach. It was decided this was a 25-year process that we're still moving toward enlightenment. In terms of parliamentary differences, Parliament as an oversight role has less power in Europe than in the balance of power as reflected in the US. There's not a sample bias in the people here but a bias in the paper topics selected. One strong difference is the emphasis on econometric analysis in the U.S. as Europeans would be much less accepting of these types of analysis. It could be seen as charlatantry. The U.S. approach may not be seen as being rigorous enough.

**David Guston** (Rutgers University, USA). It looks like you were going towards truth seeking and the widest possible use by the stakeholders.

**Irwin Feller**. In trying to shape the terms of the debate, we must realize that changes happen over time. Many of these ideas would have been laughed at 20 years ago.

**Arie Rip** (University of Twente, The Netherlands). Where is the equilibrium in your second slide?

**Irwin Feller**. I haven't really thought it through. It's a just-in-time analysis. I am simply trying to make the point that you shouldn't be too constrained about the questions that you ask.

**Arie Rip**. Your punctuated equilibrium model could be related to just-in-time analysis. You would get stability eventually.

**Susan Cozzens** (Georgia Institute of Technology, USA). I am struck by how you have to learn these things over time if you are in policy analysis. That's a difference between policy research and policy analysis. The federal client for policy analysis was wiped out during the Reagan years and hasn't been rebuilt. In Europe the growth in this area is encouraging because you don't see this type of regular evaluation in the U.S. by the federal clients. In part, what you see in the quantification emphasis in the U.S. is a reflection that they don't have to report to federal agencies such as that which occurs in the European environment.

## *11 Issues in Linking RIT Evaluation With Policy*

**Stefan Kuhlmann** (Fraunhofer ISI, Germany). I have doubts whether this is really a clear reflection of what is happening in Europe. One colleague would have presented a European study of quantification of socio-economics but couldn't make this meeting. He is focusing on policy analysis which is always a mixture of quantitative/qualitative type information but the presentation is usually less quantitative because of who they have to report the data to. Europeans don't have checks and balances such as that in the US. Decisions being made by bureaucrats and not politicians is an important difference.

**Barry Bozeman** (Georgia Institute of Technology, USA). One character of my career is that I have worked along the spectrum. One quarter of my work is in program evaluation; one quarter in broader policy analysis not focused on a particular client; and one half of work is very academic that no one really reads and doesn't have much direct influence. Thus, studies will be used only if they fit a particular goal. Yet I never feels studies don't have an influence because somehow these things trickle out. The way they make a difference is because someone will ask what I think about a topic. That's the way utilization will work. I have had influence this way but not through program evaluation directly.

**Irwin Feller**. It's about being involved in structuring issues. It's about ad hoc consulting.

**David Guston**. We need to add one more item in the aspects of learning—the things the evaluators learn about their role in the process.

## 12. New Opportunities for Transatlantic Collaboration

Session Chair: David Guston (Rutgers University, USA)

---

*In this final session of the workshop, US and European participants discussed the theme of new opportunities for transatlantic collaboration in the evaluation science and technology policies.*

**Philip Shapira** (Georgia Institute of Technology). There are at several areas for follow-up collaboration. Some of these are more immediate. Others are longer-term.

1. ***Proceedings and publications.*** The Fraunhofer ISI and Georgia Tech will produce workshop proceedings that will contain all the papers presented at the workshop. These proceedings can be made available electronically, on the Web. We should also consider whether we wish to pursue an edited book collection or a special journal issue (or two).
2. ***Issue a communiqué.*** We could issue a brief announcement that discusses the workshop and any next steps. It would cover what went on here, what was said, and what we would wish to communicate with others, especially in the policy community. It would signpost the existence of the proceedings and future collaboration plans. Can we agree on a few major points that are valuable to communicate? If so, we can write up, circulate by email, revise, then distribute
3. ***Develop thematic issues for collaboration.*** I would like to re-emphasize the idea of formulating themes of shared interest on each side of the Atlantic, then seeing if it is possible to develop collaborative projects. Here are some of the possible thematic ideas for collaboration that were raised at our workshop:
  - Methodological innovation in the evaluation of advanced technology programs
  - Distributional impacts of S&T policies
  - Methods for review of social and economic impacts of S&T policies (as well as technical competence)
  - Comparative analysis of state and regional innovation policies and programs
  - Assessment and evaluation of science performance in complex organizations and labs
  - Methods for ex-ante evaluation of programs

## 12 Roundtable: New Opportunities for Transatlantic Cooperation

- Public policy and S&T human capital development
  - Meta evaluations of a targeted domain, e.g. incubators, university-industry linkages
4. **Discuss the form of future collaboration options.** There are a variety of options that frame how we could pursue further collaboration and some of the thematic ideas I have just described. These include:
- Continue at existing level of US-EU interaction – use contacts and ideas gained here through individual follow-up. This is the “don’t do much” option. Advantage: Not much additional cost or effort required. Disadvantage: Will mean that we miss exciting new opportunities for collaboration
  - Re-invigorate special interest groups in other networks (AEA, EEA). Comments: Lack of agenda control. Easy-to-do. Limited results, beyond exchange.
  - Develop a new international S&T special interest group. Emphasize meeting one year in US, alternate year in EU. Or, we could develop a series of special workshops and conferences. This could involve, say, a multi-year program, e.g. one year in Atlanta, next year in Manchester, next year Penn State?, Then Karlsruhe? Each organizer seeks own funds and sponsorship. Themes will evolve. Allows greater number of participants.
  - Internationalization of selected projects – engage Americans in some EU projects and evaluations and vice-versa. Allows “real” projects – but funding will need to be obtained.

**Maryellen Kelley** (National Institute of Standards and Technology, USA). It would be useful if you plan something in the future in the United States. There is an issue of legitimacy in evaluation. We have problems getting funding and trusting competency in evaluation. There is talk about budget problems. It could be possible to get together a group of government folks who would be interested. Please think about the user community.

**Arie Rip** (Twente University, The Netherlands). One item is missing: the development of young people recruitment and on the job training. One or two month stays for young people would be helpful. You can broaden that notion that in attracting evaluators and enhancing evaluators where some of the new methods and competences are trained and they can therefore be robust.

**Luke Georghiou** (University of Manchester, UK). Manchester would be pleased to organize one of the conferences. The multi-level system context is increasingly important and the evaluation of capacity is an emerging theme and is more in tune with current science and innovation policies. There is some tradeoff of this with legitimation and we have to note that in the communiqué. Performance indicators need to be complements to other evaluation methods and must be in context of

*12 Roundtable: New Opportunities for Transatlantic Cooperation*

proper models. With regard to benchmarking and international comparison, evaluators' techniques and studies could lend credibility, technique and rigor and insight into that process or set of indicators.

**Louis Tornatsky** (Southern Technology Council, USA). The anarchic element rebels against creating new organizations before doing the collaborative work. We should vote with minds and bodies and people; lets put together concept papers of a page and circulate and see if others are interested. The Southern Technology Council and SSTI [State Science and Technology Institute] are meeting in the next couple of weeks and might indicate interest.

**Nick Vornatas** (Georgia Washington University, USA). GWU [Georgia Washington University] would be happy to host an event. But organizers should coordinate ideas about themes, so that events are not alike all the time.

**Irwin Feller** (Pennsylvania State University, USA). Here is a costless target: The AEA [American Evaluation Association] meetings.

**David Guston** (Rutgers University, USA). Another possible organization, CSPO [Center for the Science, Policy and Outcomes] could be interested in collaboration on the issue of non-economic impacts of research and development.

**Maryellen Kelley**. Whenever there is a new Congress, there is an educational workshop. One useful thing would be to see if there could be a session on the issues of evaluation and policy in S&T. The message: we need to legitimate certain types of evaluation to our Congress and Government officials. We have a hard time in justifying spending on evaluation.

**Ken Guy** (Technopolis, UK). I'd like to see in the circulation briefs the types of diagrams that Irwin put up that clarified the differences in the us European evaluation goals, experience, culture etc.

**Luke Georghiou**. The EU will host a conference on S&T evaluation next year. It will be a large conference and the intention is to try to take messages to policy makers. There will be opening one-day plenary involving policymakers and then a second day, which would be a scientific paper session. The conference will be held in the presidency country, which will be Belgium.

**Louis Tornatsky**. For a follow-up conference, I would rather not have it involve only the choir talking to itself. I would rather see policymakers involved.

**Dave Guston**. There are opportunities to work with a Gordon Science and Technology policy conference, and the next will be in 2002.

*12 Roundtable: New Opportunities for Transatlantic Cooperation*

**Irwin Feller.** Any future conference must have many more policymakers and serve to promote dialogue between policymakers and researchers in the US and Europe. The goal and strategy would be to pick a topic that is salient to policymakers in both continents. What do you evaluators have to say that is useful? We show our wares and get their feedback. Focused questions with set answers are required for the sessions.

**Arie Rip.** This would go against the other function of meeting, which is the opportunity for people to present their research and there be an academic discussions. Maybe there are two parts of meeting: interaction with politicians and stakeholders, and then academic discussion.

**Susan Cozzens.** Sounds like we have an emerging conference. Georgia Tech could be a sponsor of the meeting even if in Washington. We can wait until after the next election to see if it is state level or national that is most important.

**Louis Tornatksy.** We should structure any meeting to influence the people making decisions about programs.

**David Campbell** (Institute of Advanced Studies, Austria). The question is whether there is an interest or desire to support transatlantic networks and if so then there should be some funding schemes and if so there is argument for why evaluators are emphasizing the transatlantic connections. It makes it easier to justify transatlantic nets if you have transatlantic projects.

**Henry Etzkovitz** (State University of New York, USA). We found from triple helix that policymakers were using the meeting as a venue to meet and make deals. The conference served as an opportunity to seek a confluence of interests. Introduce the networking model of research collaboration into the US. It is counter to the II model in the US. If we can introduce that mode, it will be successful. The politics are difficult to work out.

## Workshop Agenda

---

### U.S.-European Workshop on Learning from Science and Technology Policy Evaluation

Evangelische Akademie Baden, Bad Herrenalb, Germany.  
September 11-14, 2000

---

#### **Monday, 11 September 2000**

- 11:45            Registration  
12:30            Lunch (Evangelische Akademie)
- 14:00 – 15:30    **Overview of Workshop and Aims**  
Participant introductions and introduction to the workshop:
- **Stefan Kuhlmann** (Fraunhofer ISI, Karlsruhe, Germany and Workshop Co-Organizer)
  - **Philip Shapira** (Georgia Institute of Technology, Atlanta, GA, USA and Workshop Co-Organizer)
- 15:30 – 15:45    Break
- Introduction: Emerging Paradigms for Evaluating Research, Innovation, and Technology Policies in the U.S. and Europe**  
(Chair: Philip Shapira)
- **Irwin Feller** (Penn State University, University Park, PA, USA). *The Academic Policy Analyst as Reporter: The Who, What,...and How of Evaluating Science and Technology Programs*
  - **Arie Rip** (Twente University, Netherlands). *Societal Challenges for R&D Evaluation*
- Frameworks for Evaluating Science and Technology Policy**
- Luke Georghiou (PREST, University of Manchester, UK), *Evaluation of Research and Innovation Policy in Europe – New Policies, New Frameworks?*
  - Susan Cozzens (Georgia Institute of Technology, Atlanta, GA, USA). *Frameworks for Evaluating Science and Technology Policies in the United States.*
- Open discussion**
- 18:30            Dinner (Evangelische Akademie)  
20:00+          Informal get-together at Akademie bar

## Workshop Agenda

### **Tuesday, 12 September**

- 09:00 – 10:30 **Evaluation of Public Research, Innovation and Technology (RIT) Policies in the EU: Roundtable on Trends and Issues**
- **Liam O’Sullivan** (European Commission). *European RIT Evaluation Practices*
  - **Christian Uhlhorn** (Federal Ministry for Education & Research, Germany). *Evaluation of National Policies*
  - **Joseph Cullen** (Tavistock Institute, London, UK). *Understanding the cultural logic of EU science and technology Programmes*
  - Commentary (Nicholas Vonortas) and open discussion
- 10:30 – 11:00 Break
- 11:00 – 12:30 **Public Research, Innovation and Technology (RIT) Policies in the US: Federal Policy Developments and Evaluative Approaches**
- **Maryellen Kelley** (National Institute of Standards and Technology). *The Effectiveness and Additionality of the U.S. Advanced Technology Program*
  - **David Guston** (Rutgers University, New Jersey, USA). *The Expanding Role of Peer Review Processes in the United States*
  - **James Dietz** (Georgia Institute of Technology). *The Assessment of Federal Lab-Industry Interactions*
  - Commentary: Arie Rip and open discussion
- 12:30 – 14:00 Lunch
- 14:00 – 15:30 **Evaluation of Large-Scale Programs in Europe**
- **Laurent Bach** (Université L. Pasteur, Strasbourg, France). *Evaluation of the BRITE/EURAM Programme*
  - **Terttu Luukkonen** (VTT Technology Group, Finland), *Evaluation of the EU’s 4<sup>th</sup> Framework Programme in Finland.*
  - **Susanne Bühner**, Fraunhofer ISI, Karlsruhe, Germany.) *Evaluation of Nanotechnology Competence Centres*
- Commentary: Louis Tornatsky and open discussion
- 15:30 Break

## Workshop Agenda

- 16:00-17:30      **Evaluation of Regionally-Based Science and Technology Programs**
- **Philip Shapira** (Georgia Institute of Technology, Atlanta, USA). *Evaluation of Industrial Extension and Technology Deployment Programs in the United States*
  - **Patries Boekholt** (Technopolis, Amsterdam, The Netherlands). *Evaluation of Regional RIT Strategies in Europe*
  - **Peter Blair** (Sigma Xi, The Scientific Research Society, Research Triangle Park, NC, USA). *Science and Technology Infrastructure Investments for Rebuilding Distressed Communities: A Case Study in the United States*
  - Commentary (EU participant TBA) and open discussion
- 17:30              **Roundtable: Challenges and Trends in Program Evaluation**
- **Ken Guy** (Wise Guys, UK)
  - Commentary: Barry Bozeman and open discussion
- 18:30              Dinner at Mönchs Post Hotel – Klosterschänke

## Wednesday, 13 September

- 09:00 – 10:30    **Evaluation in multi-goal partnerships**
- **Erik Arnold** (Technopolis, Brighton, UK). *Can We Evaluate at a Systems Level? A Production Engineering Approach*
  - **Heike Belitz** (German Institute for Economic Research, DIW, Berlin), *System Evaluation of Publicly Supported Industrial Collaborative Research*
  - **Gretchen Jordan** (Sandia National Laboratory, Washington, DC, USA), *Recognizing and Assessing Competing Values in S&T Organizations and Implications for Evaluation*
  - Discussion
- 10:30 – 11:00    Break
- 11:00 - 12:30    **Policy, institutional, and portfolio evaluations**
- **Barry Bozeman** (Georgia Institute of Technology, Atlanta, GA, USA). *Evaluating Scientific and Technical Human Capital: An Event History Approach*
  - **David Campbell** (Institute for Advanced Studies, Vienna, Austria), *Evaluation of University Research in UK, Netherlands, Austria and Germany*
  - **Louis G. Tornatsky** (Southern Technology Council, Research Triangle Park, NC, USA). *University-Industry Benchmarking*.
  - **Martina Röbbcke, Dagmar Simon** (WZB, Germany), *Assessment of the Evaluation of Leibniz-Institutes*
  - Discussion

## Workshop Agenda

- 12:30 – 14:00 Lunch
- 14:00 – 15:30 **Appropriate methodological matrices for S/T evaluation**
- **Nicholas Vonortas** (George Washington University, Washington, DC., USA), *Real Options Approach for Evaluating Public Sector R&D Investments*
  - **Ken Guy** (WiseGuys, UK). *Assessing Programme Portfolios via Multi-Module Approaches*
  - **John Barber** (Department of Trade and Industry, UK), *Making Evaluation Useful for Decision Making*
- 15:30 Break
- 16:00 – 17:30 **Lessons from existing networks– Roundtable**
- Gretchen Jordan (AEA)
  - Susan Cozzens (Federal Evaluators Network)
  - Gilbert Fayl (EU RTD Evaluation Network)
  - Stefan Kuhlmann (German Evaluation Society, and INTEVAL)
  - Open discussion
- 18:30 Dinner (Evangelische Akademie)

## **Thursday, September 14**

- 09:00 – 10:30 **Issues in linking RIT evaluation with policy** (Chair: David Guston)
- **Henry Etzkowitz** (State University of New York, Purchase, NY, USA). *The Triple Helix of University-Industry-Government: Dynamics of Innovation Spaces and Implications for Policy and Evaluation*
  - **Stefan Kuhlmann** (Fraunhofer ISI, Karlsruhe, Germany), *Evaluation as a Source of “Strategic Intelligence”*
  - Commentary (Irwin Feller) and Discussion
- 10:30 Break
- 11:00 - 12:15 **New Opportunities for Transatlantic Collaboration – Roundtable**
- Philip Shapira (Georgia Institute of Technology)
  - Gilbert Fayl (EU Commission)
  - Nicholas Vonortas (Georgia Washington University)
  - Arie Rip (Twente University)
  - Luke Georgiou (PREST, Manchester University, UK)
  - Discussion
- 12:15 **Next steps and close of workshop**  
Stefan Kuhlmann  
Phil Shapira
- 12:30 – 14:00 Lunch

## Workshop Participants

---

### **U.S.-European Workshop on Learning from Science and Technology Policy Evaluation**

Evangelische Akademie Baden, Bad Herrenalb, Germany.  
September 11-14, 2000

---

**Erik Arnold**, Director, Technopolis Ltd., Innovation Policy Research Associates, Brighton, United Kingdom.

**Laurent Bach**, Professor, Université Louis Pasteur, BETA, Strasbourg, France.

**John Barber**, Director of TESE, Department of Trade and Industry, London, United Kingdom.

**Markus Beiner**, Volkswagen Foundation, Department of Social Sciences, Hannover, Germany.

**Heike Belitz**, Federal Ministry for Economic Affairs and German Institute for Economic Research (DIW), Berlin, Germany.

**Peter Blair**, Executive Director, Sigma Xi, The Scientific Research Society, Research Triangle Park, North Carolina, USA.

**Patries Boekholt**, Director, Technopolis B.V., Innovation Policy Research, Associates, Amsterdam, The Netherlands.

**Barry Bozeman**, Professor of Public Policy and Director of the Research Value Mapping Program, School of Public Policy, Georgia Institute of Technology, Atlanta, Georgia, USA.

**Susanne Bühner**, Fraunhofer Institute for Systems and Innovation Research, Karlsruhe, Germany.

**David F. J. Campbell**, Institute for Advanced Studies (IHS), Vienna, Austria.

**Susan Cozzens**, Professor of Public Policy and Chair, School of Public Policy, Georgia Institute of Technology, Atlanta, Georgia, USA.

*Workshop Participants*

**Joseph Cullen**, Dean for Scientific Affairs, The Tavistock Institute, London United Kingdom.

**James Dietz**, Senior Research Associate, School of Public Policy, Georgia Institute of Technology, Atlanta, Georgia, USA.

**Carsten Dreher**, Head of Innovation in Production Department, Fraunhofer Institute for Systems and Innovation Research, Karlsruhe, Germany.

**Jakob Edler**, Fraunhofer Institute for Systems and Innovation Research, Karlsruhe, Germany

**Henry Etzkowitz**, Professor and Director of the Science Policy Institute, State University of New York at Purchase, New York, USA.

**Gilbert Fayl**, Head of Evaluation Unit, European Commission, Directorate General Research, Bruxelles, Belgium.

**Irwin Feller**, Professor of Economics and Director, Institute for Policy Research and Evaluation, Pennsylvania State University, University Park, Pennsylvania, USA.

**Luke Georghiou**, Professor and Director of PREST, The Victoria University of Manchester, Manchester, United Kingdom.

**David Guston**, Associate Professor, Department of Public Policy, Rutgers University, New Brunswick, New Jersey, USA.

**Ken Guy**, Director, WiseGuys Ltd., Shoreham-by-Sea, United Kingdom.

**Sibylle Hinze**, Fraunhofer Institute for Systems and Innovation Research, Karlsruhe, Germany.

**Gretchen Jordan**, Principal Member of Technical Staff, Sandia National Laboratories, Washington, DC, USA.

**Maryellen Kelly**, Senior Economist, Economic Assessment Office, Advanced Technology Program, National Institute of Standards and Technology, Gaithersburg, Maryland, USA.

**Stefan Kuhlmann**, Head of Technology Analysis and Innovation Strategies Department, Fraunhofer Institute for Systems and Innovation Research, Karlsruhe, Germany.

*Workshop Participants*

**Terttu Luukkonen**, Chief Research Scientist, VTT Group for Technology Studies, Espoo, Finland.

**Dara V. O’Neil**, Research Associate, Georgia Tech Research Institute, Atlanta, Georgia, USA.

**Liam O’Sullivan**, European Commission, Directorate General Research, Evaluation Unit, Bruxelles, Belgium.

**Arie Rip**, Professor and Director of Netherlands Graduate School of Science, Technology, and Modern Culture (WTMC). Department of Philosophy of Science and Technology, University of Twente, Enschede, The Netherlands.

**Martina Röbbbecke**, Social Science Research Center Berlin (WZB), Berlin, Germany.

**Philip Shapira**, Professor of Public Policy, School of Public Policy, Georgia Institute of Technology, Atlanta, Georgia, USA.

**Dagmar Simon**, Social Science Research Center Berlin (WZB), Berlin, Germany.

**Louis G. Tornatzky**, Southern Technology Council, Research Triangle Park, North Carolina and Silverado, California, USA.

**Christian Uhlhorn**, Head of Strategy Division, Federal Ministry for Education and Research (BMBF), Bonn, Germany.

**Nicholas Vonortas**, Center for International Science and Technology Policy and Associate Professor, Department of Economics, George Washington University, Washington, DC, USA.

**Stefan Wörner**, Fraunhofer Institute for Systems and Innovation Research, Karlsruhe, Germany.