

## Newsletter 5/97 Plattform Technologieevaluierung

### Vorwort

*"La réalité, c'est tout ce que nous éprouvons et savons rendre avec force et avec amour".  
Élie Faure*

Wie kann man die Wirklichkeit erfassen und dadurch mitgestalten? In unserem Bereich, der Forschungs- und Technologiepolitik, verbirgt sich die Wirklichkeit - sprich das Gedachte, das Erforschte, das Getane - hinter einer Vielzahl unterschiedlicher Erscheinungsformen, hinter unmittelbaren und mittelbaren Wirkungen, aber auch hinter einer Bandbreite unterschiedlicher Wahrnehmungen der Realität. Evaluierungen beleuchten die Realität, helfen aber in einem gewissen Ausmaß auch mit, diese zu schaffen. Zu einem bewußten Methodeneinsatz und zu einer transparenten Vorgangsweise muß freilich auch eine Leidenschaft für die Sache treten. Das stellt eines der Motive für die Arbeit der Plattform Technologieevaluierung dar. Das einleitende Zitat aus der Feder eines Kunsthistorikers bietet eine gute Analogie: Das Vergangene und uns gleichsam in der Zweidimensionalität Überlieferte gewinnt Kraft und Ausdruck durch die methodische und engagierte Beschäftigung mit Autoren, Werk und Hintergrund. Durch diese intensive Arbeit an der Vergangenheit entstehen Orientierungshilfen für die Zukunft.

In der vorliegenden Ausgabe des Newsletters beschäftigen wir uns mit den Ergebnissen einer großen FTE-Evaluierungskonferenz, die die OECD im Juni dieses Jahres in Paris veranstaltet hat. Neben einem Überblick über die zahlreichen inhaltlichen und methodischen Beiträge gehen die Autoren Wolfgang Polt, Gernot Hutschenreiter und Michael Stampfer auf einige Spezialfragen ein, etwa auf die Evaluierung indirekter Förderungen und auf die Ausgestaltung von Institutionen im Zusammenhang mit Evaluierungen.

Eine im September d.J. in Wien abgehaltene Veranstaltung der Plattform brachte Einblicke in die skandinavische Evaluierungspraxis; als Vortragende konnte die finnische Expertin Terttu Luukkonen (VTT) gewonnen werden. Ein Beitrag von ihr wird im nächsten Newsletter erscheinen.

### Summary

In this issue, three contributions deal with the outcomes of an OECD Conference on RTD Evaluations, held in Paris in June 1997. All three articles are in English (which is not always the case - we regret !). Special attention is given to institutional questions, also to evaluations of RTD tax credit systems.

Oktober 1997; Michael Stampfer, Oliver Fritz, Gernot Hutschenreiter

# **Evaluation of technology and innovation policies - in search of best practices<sup>1</sup>**

*Wolfgang Polt*

The Directorate for Science, Technology and Industry of the OECD organised on 26 and 27 June an international Conference on Policy Evaluation Practices in Innovation and Technology, the aim of which was to bring together policy practitioners from OECD Member countries with researchers working on evaluations in order to present and compare the experience in policy evaluation methods and practices. The conference<sup>2</sup> dealt with different approaches to evaluation in different fields of innovation and technology policy (financial support to industrial R&D; large technology programs; technology diffusion policies), with different evaluation practices to be observed in different countries and addressed issues relating to the techniques used in evaluation and to the institutions involved.

The following paper summarises some of the main results of this conference and attempts to outline what could be distilled as general lessons for 'good practice' evaluations.

## ***The changing context of evaluation***

Evaluation of technology and innovation policies has recently seen some renewed interest both from the side of the practitioners as well as from a methodological angle - for a variety of reasons. The most obvious is probably the sustained budget stringency that many countries are facing. Against this background, many government activities are coming under scrutiny, science and technology policies not being exempt from this general trend.

The basic questions raised are: Is the activity in question (or to be evaluated) a task for government in the first place or should it be left to the private sector to secure the intended outcome? If there is a case for government intervention, what is the best instrument to achieve a given task? If carried out, was it cost-effective? And what were the (wider) socio-economic effects, which are in many cases the *raison d'être* of a government programme?

As if evaluations wouldn't have already had difficulties to answer these questions, the development of both innovation theory and policy in the past 20 years have complicated matters further. They have increasingly departed from the 'linear model' of technological change and have argued for and adopted more complex approaches, addressing such 'soft aspects' as 'information and awareness', 'organisational change', 'human capital', 'collaboration', 'innovation and learning capabilities of firms' etc. as central to the successful innovation process and hence for policy (see Figure 1).

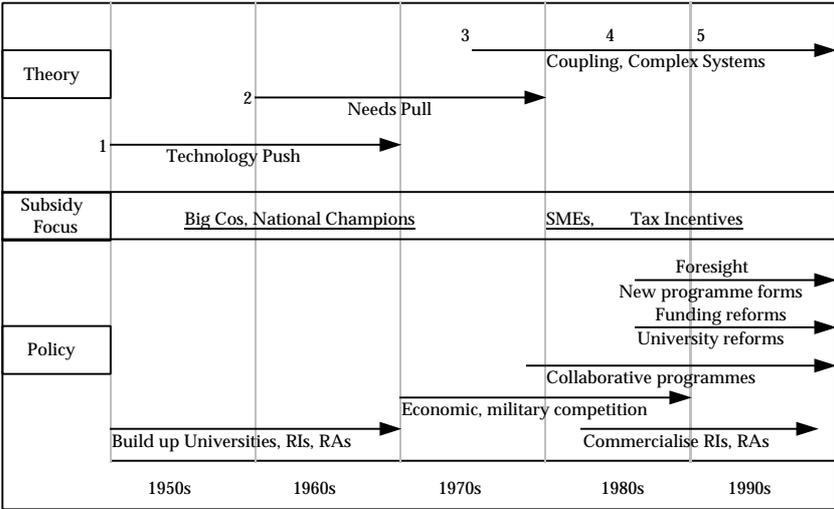
While the measure of external effects and societal benefits was already tricky, tracing back effects of policies on these 'softer aspects' is even more so. Together these developments and challenges ask for a further development of evaluation methodologies and tools and for an improved conduct of evaluations.

---

<sup>1</sup> This article is based to a large extent on the introductory chapter of the proceedings of the conference: George Papaconstantinou / Wolfgang Polt: Policy evaluation in innovation and technology: towards best practices - An Introduction. (forthcoming)

<sup>2</sup> The program of the conference is provided at the end of the paper

**Figure 1. Post-war shifts in theory, subsidy and RTD policy**



from: Arnold/Guy<sup>3</sup>

**Scope and coverage of evaluations**

Broadly speaking, evaluation refers to “a process which seeks to determine as systematically and objectively as possible the relevance, efficiency and effect of an activity in terms of its objectives” (United Nations definition), including the analysis of the implementation and administrative management of such activities. The scope and methods of evaluation differ according to the questions to be addressed and the character of the policy measure. Thus, they can be retrospective (“ex-post”), current or prospective (“ex-ante”) evaluations, producing information that can be used in the assessment of past policies, the monitoring of on-going initiatives or the forward planning of innovation and technology policies.

Evaluations can also have different purposes. In effect, part of the difference in opinions expressed in some of the contributions at the conference can be traced to different perspectives on what evaluations aim to achieve. Policy-makers and economic analysts stressed the role of evaluation in examining the justification of a programme, analysing its economic impacts, and thus providing information to guide resource allocation as well as more strategic decision processes which involve selection of instruments, or the thrust and direction of technology policies in general. Alternatively, many professional evaluators as well as policy-makers involved in hands-on running of programmes stressed the role of evaluation in improving the conduct, quality, responsiveness and effectiveness of a programme, thus raising its leverage effect. Fig. 2 provides an example for different perceptions regarding the Georgia Manufacturing Extension Alliance [GMEA] in the United States.

<sup>3</sup> all references given refer to the contributions at the conference. For the Title see the Annex.

**Table 1 Perception of usefulness of different evaluation approaches**

<b>Method</b>	<b>Program Justification</b>	<b>Program Management and Improvement</b>
<b>Management Information System</b>	State: 3 Federal: 2	Overall: 4-5 Analysis of MIS information is useful for direct program management
<b>Client Valuation Surveys; Customer Follow-Ups</b>	State: 5 Federal: 3	Overall: 3-4 Customer reports of poor service: 5
<b>Program Impact Analysis</b>	State: 1 Federal: 1	Overall: 4-5 But managers can find it hard to act on results
<b>Cost-Benefit Analysis</b>	State: 3 Federal: 3 In theory, CBA can be used to allocate resources between different policies; in practice, this is rarely done.	Overall: 1
<b>Longitudinal Controlled Surveys</b>	State: 3-4 Federal: 4-5	Overall: 2-3 Customer reports of anticipated needs are valued by managers
<b>Case Studies</b>	State: 3 Federal: 3	Overall: 2-4
<b>External Reviews</b>	State: 3 Federal: 4	Overall: 3-5

**Note:** Ranking (schematic): 5 = extremely important; 3 = somewhat important; 1 = not important. Ranking weights are schematic, based on experience.

Source: Shapira/Youtie

Clearly, these objectives are complementary; achieving them however often calls for different evaluation tools and institutions carrying out the evaluations.

Different types of policies also ask for different evaluation methods. Within the broad area of innovation and technology diffusion policies, the assessment of large-scale mission-oriented programmes or of fiscal incentives for industrial R&D are more amenable to the use of cost-benefit analysis. Newer programmes like those fostering pre-competitive R&D collaboration ask partly for other methods, not yet fully developed. The evaluation of diffusion-oriented programmes needs much more elaborate micro-level econometric analysis, and hence high-quality data and databases, than others. Yet other programmes focusing on the ‘soft side’ of the innovation process (like awareness, information and consulting programmes) demand the use of intensive case studies and user-surveys. Thus, the development and application of evaluation methods reflects also the stage of development of technology and innovation policy in general.

## ***Developments in evaluation methods and practices***

Evaluation methods and practices have developed alongside the evolution of technology and innovation policy and the understanding of the innovation process. Starting from the predominant model in the post-war period, the focus was first on the assessment of the quality of scientific research, with peer review and bibliometric techniques (impact analysis and citation counts) the main methods. Programme evaluation developed later, following the proliferation of government programmes to support industrial innovation, often through fostering collaborative research. These demanded more elaborate techniques both for the assessment of direct and indirect socio-economic effects and for the assessment of the conduct of the programmes. Econometric techniques, cost-benefit analysis and surveys were increasingly applied.

With the recognition of the complex, systemic nature of innovation processes, and with technology policy instruments covering a growing range of activities, evaluation techniques have had not only to develop quantitative techniques further (e.g. micro-level analysis for the assessment of diffusion oriented programmes) but had also increasingly to look at the “soft side” of innovation (e.g. in trying to capture networking, learning effects, etc.), where further work is still needed to identify their economic impacts. The proliferation and widening coverage of policy initiatives has led evaluation to adopt increasingly a portfolio approach, rather than focusing on individual projects; to a greater use of performance indicators; and towards a convergence between the activities of ex-post evaluation and continuous monitoring. This multi-faceted approach has also been made necessary by the multiplicity of actors involved in technology and innovation policies. Each of these actors (policy-makers deciding on a programme, programme managers designing and conducting it, firms participating etc.) has the need for different types of information, hence the need for a combination of methods shedding light on the basic rationale, the economic impact, the administrative conduct and the customer satisfaction derived from the activity.

## ***Evaluation criteria and tools***

A crucial issue in evaluation is the criteria to be used for judging programmes and policies. The basic rationale for government initiatives to stimulate technological development in the first place is the recognition that there is a difference between the expected private rate of return and the social rate of return, with the private rate being too low to induce firms to engage in innovative activities that would be beneficial from a societal standpoint. This “market failure” rationale suggests that while “additionality” and the existence of positive private returns to firms as a result of government programmes are preconditions for success, for policy to be fully justified the net social benefits of a government programme must be positive. A number of participants in the conference in effect suggested that the focus on additionality (the changes in behaviour and performance that would not have occurred without the programme) as a criterion for success is simply a reflection of the difficulty to accurately measure spillovers or externalities and thus the net social benefits of programmes.

The accumulated experience of three decades of technology policies, together with recent advances in innovation theory have shown the limits of a simple “market failure” rationale to policy. The preponderance of “government failure” has forced evaluators to be more careful in accounting for costs of programmes as well as for benefits, including those costs that are associated with the distortions to economic incentives that policy initiatives can bring about. On the other hand, the realisation that the benefits of individual programmes or policies can often be understood only in the context of their impact on a complex innovation system has given rise to the notion of “systemic failure” as a basis for policy. In terms of evaluation, this has translated into the more elaborate principle of “behavioural additionality”, which is intended to capture the many ways in which participation in a programme can change the innovative behaviour of a firm. As a number of conference participants noted, it has also

forced evaluators to recognise that identifying social benefits in diffusion policies involves a dynamic analysis that looks at the development of new capabilities, and of learning.

A large part of the conference papers and discussions was devoted to the critical presentation of different methodological tools for evaluation: cost-benefit techniques, econometric methods, case studies, in-depth surveys and peer reviews (for an overview see Table 2).

**Table 2. An overview of evaluation methods**

Methods	Qualitative & Semi-qualitative		Quantitative	
<i>ex-ante</i>	Peer Reviews  Questionnaires  Interviews	Techn. Forecast. Methods  Scenario Method  Cross-impact Matrices  Morphological Analysis	Cost-Benefit Analysis  Technometrics  Optional Pricing	Scoring Methods  Matrix Approaches  --analysis matrices, --decision-making matrices  --multicriteria analysis, --relevance trees  Systemic Approaches  --system analysis, --dynamic modelling
<i>ex-post</i>	Case Study  Histogramical Tracing of Accomplishments  Critical Scientific Events		Quantitative Indicators  Bibliometrics  S&T Indicators  Patent Data  Econometric Models	Financial Methods  Ratio Methods  Risk Profiles  Programming Models  Portfolio Models

from Piric/Reeve

There were clear differences between participants in the faith that they put to conclusions based on quantitative as opposed to qualitative techniques. Nevertheless, whatever the type of policy being evaluated (financial support to industrial R&D, large technology programmes, diffusion-oriented policies), it was equally clear that a combination of different approaches (quantitative and qualitative) is needed in order to cover all aspects of the evaluation process; different approaches are complementary, not mutually exclusive (for an overview of the advantages and disadvantages of various methods see Table 3). In effect the distinction is not that clear-cut: it often turns out that quantitative techniques produce mainly qualitative information. Thus, to increase the credibility of evaluation results, a number of alternative methods should be used to consolidate the foundations of policy recommendations.

Some of the most interesting recent developments in evaluation methodology concern the use of econometric techniques based on longitudinal micro-level data, where the impacts of programmes are examined by comparing the performance characteristics of firms that are clients of government initiatives (such as extension services) with those of non-client firms. The quality of results based on this approach is however conditional on the extent to which researchers can control for firm characteristics other than programme participation. Furthermore, this technique is only the first step in a full cost-benefit analysis: at its best it establishes the private benefits conferred to firms as a result of the programme; justification of a programme needs to account for social benefits against the total costs.

Many participants felt that evaluation schemes should be constructed around social cost-benefit frameworks which estimate the impact induced by the policy measure, its spillover benefits, as well as costs such as the marginal excess burden of taxation and compliance costs. Nevertheless, they warned against the spurious precision that cost-benefit calculations can give, and suggested that such schemes should combine qualitative and quantitative indicators of the impact of the policy measure and of the private and estimated social benefits from the induced change in behaviour. Ideally, they should be combined with the qualitative information from user-surveys, in-depth case studies and interviews to

produce the variety of information needed by the different users of evaluations. Single-approach evaluations might in effect be downright misleading, and putting too much emphasis on single quantitative estimations, while useful as a measure of cross-checking, might miss the essential qualitative effects of new initiatives. Furthermore, it is clear that quantitative techniques have to be developed further, especially with regard to the challenge of capturing the economic impacts of the “soft factors” of innovation (impact of programmes on learning, co-operative and innovative behaviour).

**Table 3. Synthesis of evaluation methods – relevance and drawbacks**

Method	Relevance	Drawbacks	Field of application	Analytical level
1. Peer review	- Screening of projects and research orientations	- Subjectivity of experts - Partial forecasts - Lack of independence of experts	- Selection, ongoing, and impact - Technological forecasting.	Micro
2. Matrix approaches				
- Analysis matrices	- Rich information	- Difficult to collect the required information	- Impact	Meso
- Decision making	- Rationalise and simplify choices	- Subjectivity - Lack of flexibility	- Selection - Ongoing	Micro
- Multicriteria analysis	- Profiles projects and R&D planning	- Constitution of a group of experts - Subjectivity in the choice of weightings	- Selection	Micro; Integrated
- Relevance trees	- Provides lots of information	- Subjectivity in the allocation of quantitative values	- Selection	Integrated
3. Systemic approaches				
- Systemic analysis	- Can be used to implement an evaluation - R&D strategies	- Not really suitable for evaluating as such	- Selection	Integrated
- Dynamic modelling	- Includes social, historical and ecological structures - Takes feedback phenomena into account	- Very difficult to implement	- Impact	Integrated
4. Financial methods				
- Cost-benefit	- Measures marketable outputs and commercial resources	- Difficult to collect the information	- Selection - Impact	Micro
- Ratios methods	- Simple instruments	- Some factors cannot be financially assessed		
- Risk profiles		- Results sensitive to arbitrary choices		
- Portfolio models		- Purely financial aspects		
5. Technological forecasting methods				
- Scenario method	- Allows the causality chain to be reversed	- Subjectivity	- Selection - Forecasting	Integrated
- Cross-impact	- Takes social transformations into account			
6. Quantitative indicators				
- S&T indicators	- Easy measurement	- Purely descriptive	- Selection	Integrated
- Bibliometrics	- Builds up fundamental scientific output indicators	- Partially descriptive indicators	- Impact - Selection	Micro Meso
- Technometrics	- Measures technology	- Mainly descriptive indicators	- Selection	Micro

	characteristics		- Impact	
- Econometrics	- Measures the full range of socio-economic impacts	- Theoretical and methodological background	- Impact	Integrated

from Capron/van Pottelsberghe

The need for an approach combining quantitative with qualitative information is also underscored by the fact that programme management also necessitates looking into the process and performance of different policy instruments. Given the high variance of returns in different technology projects, detailed case-studies are important to see what works and what does not. But whatever method is used, the importance of having a “counterfactual” in policy evaluation exercises was stressed: evaluation requires comparing the absence of policies with the impacts in the presence of policies. Furthermore, a number of participants stressed the fact that much of the evaluation work to date ignores the competitive environment within which client plants and firms operate, and in which the services provided by governments are supposed to improve performance; yet it is important to have an understanding of this environment in order to optimally design, provide and evaluate program services and to ensure that programme objectives are not at odds with those of the client firms.

### ***Institutions and practices - different 'national systems of evaluation'***

Techniques aside, evaluation is very much a social process, as it involves interaction of individuals, organisational beliefs, practices, and routines. The institutional set-up within which programmes and policies are evaluated in effect determines the nature, quality, relevance and effectiveness of evaluation practices. Many papers and discussions in the conference treated this issue, and addressed the question of whether there is such a thing as an “optimal” institutional set-up that is transferable across countries. On this question of country specificities and general principles in institutional arrangements, it was felt that while practical arrangements are country specific, basic principles/challenges in evaluation are not. It was suggested that the precise institutional framework for evaluation is of less importance than its functionality. What was necessary from a practical point of view was to reconcile evaluation designs with the varying needs of programme sponsors, service providers, and customers; to incorporate methods that can support programme learning and improvement as well as address issues of programme justification; to reconcile desired information needs with resource and information availability constraints; and to ensure that evaluation takes place on a programmed and properly resourced basis, guarantees the independence of the evaluators, and provides a mechanism for feedback of results into policy-making.

In general, evaluation practices are far from uniform across OECD countries; as the different contributions have shown, in the mid-90s programme and policy evaluation is characterised by different degrees of maturity as well as by a large measure of variety in terms of the tools used, the institutions involved, and of the place of evaluation in policy-making in general. Despite such differences, there are some common lessons that can be drawn. A basic lesson is that evaluations must build on the strengths and variety of the national systems of innovation in order to develop systematic evaluation practices embedded in the policy-making process; there is no optimal institutional design for evaluations that is transferable across countries. Nevertheless, there are some general conclusions to be drawn from the comparison of country experiences (see also Box 1: Best practice bullet points).

First, experiences from several countries show that evaluations are best designed together with the program or policy to be evaluated. Only such an early preparation would secure the collection and provision of the data needed and the common acceptance of the evaluation procedures and criteria among the institutions involved.

Secondly, as the results of evaluations are often taken up only in a “localised form” (that is, they are only implemented if the institution evaluated could implement the recommendations on its own), there is a need to secure the take up on a higher level of policy making. Thus, it seems necessary to put a formal

obligation on those responsible for policy making to react to the results of evaluations or to expose the results of evaluations to a public discussion, which would also result in a higher awareness of policy makers. A presumption in favour of publication of evaluation reports was thought to be very important in this respect in many countries; once in the public domain it is more difficult to ignore results of evaluations. Some countries have gone far in implementing such disclosure and feedback arrangements, but many still lack this form of feed-back mechanism.

Third, evaluations ought to be more “user-oriented”, that is to address the informational needs of the respective “clients” (policy-makers, firms, program administrators on various levels); hence they should encompass an appropriate mix of methods to produce this different types of information. This would also improve the take-up of their results. There are some promising examples of this approach, notably in countries where there is a strong budgetary pressure on programmes and institutions (e.g. Australia, the UK and the USA).

Finally, in almost all countries evaluations have so far been used mainly for incremental changes (i.e. for the improvement of programmes), but hardly to guide more fundamental shifts and re-orientations in technology and innovation policy. To empower evaluation for such a task, one would have to embed it into a wider system of information gathering and preparation, linking it to technology foresight and technology assessment exercises. In addition, the role of the evaluator would than have to change from being predominantly a “referee” to a “moderator” of this information gathering process and a “coach” for the strategic policy decision making process.

This raises the question of how far evaluation can go. Although welcoming an increased scope of evaluation techniques to produce a greater variety of information more reliably, policy makers during the conference warned against stretching evaluation too far. In their perspective, evaluation can help to guide informed choices, but not to substitute for a political decision-making process. While evaluations are increasingly used in the policy process (especially when results are easily interpretable by decision makers), they are not and could not be decisive, partly because policy involves trade-offs and values, and partly because often evaluations are often not good enough, due to uncertainty in the impact of many programmes. In effect, many policy decisions are based on intuition and first principles, and often evaluation is just used to justify certain decisions after the fact. Nevertheless, it was clear that there is a need for more and better evaluations, and especially for evaluations that go beyond individual programmes and compare the impact of different spending initiatives, examining in other words the appropriateness and efficiency of using different policy tools to achieve a given objective.

#### ***Best practice - bullet points***

- Create ‘evaluation culture’ (awareness, positive attitude, knowledge pool and expertise to conduct evaluations)
- Formulate guidelines and establish consensus on approach and measures (‘metrics’/criteria) to secure acceptance of results
- Prepare evaluations carefully, design together with the programme to be evaluated (to ensure timely provision of data)
- Identify and address different informational needs of the ‘clients’ of evaluations (technology policy makers, programme managers, administrative staff etc)
- Establish multi-disciplinary evaluation teams in order to be able to judge on scientific, economic, managerial, political dimensions of the programme/institution/activity.
- Secure independence of evaluators, avoid ‘vested interests’ to dominate the process. A possible way to avoid such unwarranted influence is probably to formulate a ‘code of conduct’ for evaluations.
- Carry out evaluations regularly (to achieve ‘economies of scale’ for the exercise, to accumulate knowledge for policy making, to influence administrative behaviour)

- Secure that evaluations are carried out in a pro-active manner and result in recommendations (without occupying the role of the policy decision-maker).
- Use evaluations cutting across different programmes for comparing the relative effectiveness of different policies
- Use a mix of methods tailored to the specific policy instrument (e.g. diffusion-oriented programmes, collaborative R&D programmes etc), the goal of the evaluation and the clientele.
- Use a 'portfolio'-approach rather than one focusing on individual projects
- Secure the provision of adequate data to allow the application of more advanced quantitative methods (especially micro-level/firm analysis)
- In evaluating the outcomes of technology programmes, take the 'soft factors' (intangibles, behavioural additionality etc) into account.
- Take into account the context (competitive environment of firms, history of the programme)
- Link evaluation to other sources of information gathering on technical change and the potential effects of policy: technology monitoring, technology assessment and forecasting. Integrate to produce appropriate information for strategic policy making.
- Secure feed-back mechanism into policy, either by public debate or by obligation to deal with evaluation results or both

**Annex: PROGRAM - OECD CONFERENCE ON POLICY EVALUATION IN INNOVATION AND TECHNOLOGY, Paris, 26-27 June 1997**

*Thursday 26 June 1997*

**Opening addresses**

*Ms. Joanna Shelton*, OECD Deputy Secretary-General

*T. Andersson*, Deputy Director, Directorate for Science, Technology and Industry, OECD

*D. Shand*, OECD Public Management Service (PUMA)

**Introductory paper**

*L. Georghiou* (Director, PREST, United Kingdom): "Issues in evaluation practices in OECD countries"

**Session 1: Evaluation of financial support to industrial R&D** (Chair: *Ms. J. Seyvet*, Deputy-Director ANVAR, France)

1. *R. Lattimore* (Productivity Commission, Australia): "R&D fiscal incentives in Australia: impact and policy lessons"
2. *H. Capron/B. van Pottelsberghe* (University of Brussels, Belgium): "Public support to business R&D: an integrated assessment scheme and some additional quantitative evidence"
3. *A. Hervik* (Norwegian School of Management, Norway): "Evaluation of user- oriented research in Norway: estimation of long run economic impacts"
4. *M. Dorsman* (Ministry of Economic Affairs, Netherlands): "Evaluation of industrial R&D support in the Netherlands: recent developments"

Discussant: *B. Hall* (UC Berkeley, United States)

**Session 2. Evaluation of large technology programs** (Chair: *Ms. C Stevens* (Head, Science and Technology Policy Division, DSTI, OECD))

1. *P. Cohendet* (BETA, France), "Evaluating industrial indirect effects of technology programs: the case of the European Space Agency"
2. *M. Sakakibara* (UCLA, United States), "Evaluation of government-sponsored research joint ventures in Japan"

3. *A. Link* (University of North Carolina, United States) and *J. Scott* (Dartmouth College, United States) “Evaluating technology-based public institutions: lessons from the National Institute of Standards and Technology”
4. *T. Luukkonen* (VTT Group of Technology Studies, Finland) “The increasing professionalisation of the evaluation of mission-oriented research: implications for the evaluation process”

Discussant: *R. Lyon* (Senior Economist, Office of Management and Budget, United States)

**Discussion panel #1: Evaluation techniques** (Chair: *D. Malkin*, Head, Economic Analysis and Statistics Division, DSTI, OECD)

*M. Pianta* (National Research Council, Italy): The use of innovation surveys for policy evaluation purposes

*A. Piric* (Ministry of Research, Science and Technology, New Zealand): Methodological issues in evaluating public investment in R&D

*R. Jarmin* (Bureau of the Census, United States) The use of micro-level data

*R. Lattimore* (Productivity Commission, Australia) Cost-benefit analysis

*J-P. Meurice* (Ministry of Industry, France)

**Friday 27 June 1997**

**Session 3. Evaluation of diffusion-oriented policies** (Chair : *D. Deniozos* (Head, European Framework Programs, Ministry of Education, Greece))

1. *R. Jarmin* (Bureau of the Census, United States) and *J. Jensen* (Carnegie Mellon University, United States): “Evaluating government technology programs: the case of manufacturing extension”
2. *P. Shapira* (Georgia Institute of Technology, US) and *J. Youtie* (Georgia Tech, US): “Evaluating technology deployment at State level: Methods, results and insights from the Georgia Manufacturing Extension Alliance”
3. *S. Arvanitis* and *H. Hollenstein* (Federal Institute of Technology, Switzerland) “Evaluating the promotion of advanced manufacturing technologies using micro-level survey data”
4. *E. Arnold* and *K. Guy* (Technopolis, UK) “Technology diffusion programs and the challenge of evaluation”

Discussant: *J. van Reenen* (Institute for Fiscal Studies, United Kingdom)

**Discussion panel #2: The role of institutions in policy evaluation** (Chair: *E. Ormala*, Chief Planning Officer, S&T Policy Council, Finland; Chairman, OECD TIP Group)

*L. Durieux* (Evaluation Unit, DG12, European Commission) on the scheme used for evaluation of European RTD Programs

*J-E. Aubert* (OECD) on OECD experience with evaluating innovation policies

*M. Stampfer* (Federal Ministry for Science and Transport, Austria) on S&T policy evaluation in Austria

*A. Birch* (Agency of Trade and Industry, Denmark) on the evaluation of the GTS-institutes in Denmark

**Session 4. Strategic evaluation and new trends and approaches** (Chair: *M. Bradbury*, Chairman of Evaluation Methodology Committee, Department of Trade and Industry, United Kingdom)

1. *S. Kuhlmann* (Fraunhofer-Institute Systems and Innovation Research, Germany), “Technology Policy Evaluation: Present State in Germany and European Outlook”
2. *P. Laredo* (Ecoles des Mines, France): “Evaluation of technology policies in France”
3. *R. McDonald* (Industry Canada) and *G. Teather* (National Research Council of Canada): “Science and Technology evaluation practices in the government of Canada”

Discussant: *L. Georghiou* (Director, PREST, United Kingdom)

**Policy roundtable : The feedback of evaluation results into policy design** (Chair: *J. Ryan*, First Assistant Secretary, Department of Industry, Science and Tourism, Australia; Chairman, OECD Industry Committee)

*D. Bureau* (Sous-directeur de la prévision, Ministère de l'économie, France)

*J. Gabolde* (Director, Strategy and Coordination, DGXII, European Commission)

*S. Wakimoto* (Director for Research Coordination and Planning, MITI)

*M. Bradbury* (Chairman of Evaluation Methodology Committee, Department of Trade and Industry, United Kingdom)

*R. Lyon* (Senior Economist, Office of Management and Budget, United States)

**Closing remarks** *R. Nezu*, Director, Directorate for Science, Technology and Industry, OECD

**Wolfgang Polt**

(Austrian Research Center Seibersdorf/ Department for Technology Studies)

## **RTD - Evaluations: Taking a look at the institutions**

### **Do institutional settings influence evaluation processes and the implementation of results? Remarks on the OECD Conference: Policy Evaluation in Innovation and Technology**

*Michael Stampfer*

Whenever evaluators congregate, methodological questions tend to dominate the discussion. Particularly the intricate questions concerning qualitative versus quantitative approaches and the right mix between them are always on top of the agenda. This is true for evaluations of programmes and institutions. Other equally important topics, if not completely neglected, appear only on the fringes of sometimes titanic efforts in methodological tournaments.

In its RTD Policy Evaluation Conference, the OECD secretariat tried to find an equilibrium between the above mentioned methodological and other, namely institutional questions. This conference was structured around 4 sessions covering different areas in the evaluation context: financial support to industrial RTD; large technology programmes; technology diffusion programmes; and strategic issues, together with two discussion panels. One panel covered the question of evaluation techniques, the other discussion panel, chaired by E. Ormala (SF, Science and Technology Policy Council), was devoted to the role of institutions in policy evaluations. A number of speakers in other sessions stressed the necessity to take an institution-oriented view in order to better shape and implement evaluations, to use their results and to feed them back effectively into the policy planning process.

In many cases the impact of institutional settings is a decisive factor for evaluation processes of projects, programmes or institutions themselves. Of quite considerable importance can be (inter alia):

- The existence of a mission statement and a strategy. If ministries, universities, funding organisations, research centers etc. are able to deliver a clear-cut medium term strategy, two things become easier: Firstly for the institution itself, the commissioning of an evaluation: what do we want to know? Secondly for the evaluators: What is the strategic background for the evaluation process? How does the evaluated part (a distinct programme, unit etc.) fit in the overall picture? The same is true with clear mission statements on the level of RTD programmes. The clearer the goals are set, the better an evaluation can mirror the outcomes. Programmes lacking quantitative or even qualitative goals are hard to evaluate.

- Distinct functions within an institution, discernible division of labour: Institutions on the ministerial or agency level generally need a strategic planning unit to shape the above mentioned mission statements and to deliver benchmarks. In addition, a small unit for organizational issues of evaluations is strongly recommended by some authors [Kuhlmann/Holland (1994), 201 ff]. Such in-house experts organise evaluations, set common standards, tackle methodological questions, help to build up a common understanding (an "evaluation culture"), and probably fund those evaluations. Two things can be guaranteed by such an organizational approach: drive and impartiality. On the other hand, a sometimes dangerous routine feeling and unduly standardized procedures can be pitfalls of such a model. This touches a question of general relevance: Even though evaluations should be comparable to a certain degree, the use of uniform methods and procedures does not take into account the huge differences between programmes, between institutions etc. The case of uniformity seems to be a "how

---

<sup>4</sup> Exclaimer: of course, this paper does not represent an official position of the OECD.

far shall we go?"-question, given the necessity to have comparable answers on the value-for-money quest [e.g. Cunion (1995), 367 ff]. Another important point is that no organisational unit can serve as a substitute for political will and an overall strategy.

- The Nordic example [Luukkonen, Hervik] shows the importance of an underlying culture providing openness and discussion. As long as evaluations serve as ammunition for intra-department feud, the most important goal is missed. Everybody knows this bureaucratic standard situation: taking point III.b.5 out of the summary to use it against a rival for scarce funds: "As I always said ... ". As in all institutional arrangements, the usefulness of the tool evaluation depends on the rules of the game. This has much to do with organisational aspects of institutions (never mix strategic with operative agendas too much and never help create hidden agendas), with deeply rooted socio-cultural habits, but also with learning processes. Monitoring groups or discussion circles assembling all the interested parties right at the beginning of an evaluation process help to form a common view and in the long term even common values. All those cultural factors generally need a long term approach. Luukkonen shows this in her paper, presenting a development over nearly twenty years in different stages, starting from a small traditional basis (peer reviews in basic research), and enlarging the evaluation agenda step by step.

- Cooperative behaviour is fostered by a climate open for discussion, for which publicity of evaluation studies is an important issue. The public and media interest may be small in the beginning but can grow over time. Media attention can be a problem, if negative findings lead to articles crying for radical solutions. Again, the Nordic example has shown that in the course of the years, evaluations become something normal instead of something sensational ("Secret expert paper reveals scandalous sleaze" etc.). Publicity on expert level can be achieved by regular publications, newsletters or discussion fora, where the findings and the methodic approaches are presented to researchers, policy makers and other interested parties.

- Are strict legal provisions necessary? Answers to this question seem to be rather country specific and depend on history in a broader context. If there exists a climate open for discussions together with value for money - mechanisms, informal arrangements will prove sufficient. There are also big differences in legal systems: The Anglo-American tradition does not rely that much on extensive codification; the opposite is true in some European countries like Germany or Austria with their rather obsessive reliance on legal regulations. Anyway, there are countries lacking both an open discussion climate and high procedural standards in policy areas relevant in our context [see also Georghiou's distinction in centralised, decentralised and pre-evaluation systems]. In such cases, explicit legal texts will help a lot. They force the policy makers to act along the established procedures - making these procedures normative and normal. This has of course a lot to do with cultural factors; the German legal term *Rechtsunterworfenen* says a lot in this context. So some clear cut statements about compulsory and regular evaluation processes of all major programmes and institutions can indeed change the situation.

- Ad hoc vs. institutionalised practice: Some papers show an evolution from the former to the latter stage in different contexts [Luukkonen, Laredo, for the EC Durieux/Fayl]. As mentioned before, three broad categories seem to exist: "that of centralised systems (the UK and France) in which the requirement to evaluate and guidelines for practice emanate from the centre of government, decentralised systems (the Netherlands and Germany) where evaluation was common practice but was not promoted or co-ordinated in a systematic way across departments and agencies, and a pre-evaluation stage (most of Southern Europe) where a legislative framework was normally in place but where practice had yet to develop in the majority of circumstances." [Georghiou, 18]. Many countries began with evaluations in basic research projects, using peer reviews; in a second stage best practices in programme evaluation were developed; the third stage is often an institutionalised and professional approach including evaluation of institutions, programmes and projects.

- Checks and balances: Never let an operative unit evaluate itself. It is even problematic, if (e.g.) a funding organisation commissions an evaluation (of their programmes etc.) with its own money and resources. In this case the evaluation contract exists between external evaluation experts and the operative organisation in their double role as evaluated institution and term-setting "employer". Preferable solutions include third parties to commission the study and to pay the bill. This can be parliament or a chamber of audit in the case of a purely quantitative evaluation or the responsible ministry in all kinds of evaluations. Some cases of specialised institutional settings like the French guarantor model [Laredo, see below] are worth being noted. In all these cases a common and cooperative approach with the evaluated institution is of course necessary.

In the Paris Conference, a number of country profiles were discussed:

In France it is "interesting to see how evaluation of S+T operators was developed to consider them not solely as agents of implementation of a policy defined at a higher level, but as mediators between research activities and policy making with their own strategic, political and managerial capabilities." [Laredo, 4]. Two important institutions act as "guarantors" by evaluating the "operators" (funds, research institutions). The guarantors are committees, established at the central political level by law, with a strong secretariat and own budgets; CNE (Comité national d'évaluation) is responsible for university evaluation; CNER (Comité national d'évaluation de la recherche) with the objective to evaluate research institutions and to identify strengths and weaknesses of the French RTD System - with the explicit task to foster the public debate on the base of robust and reliable data. The checks and balances are worth being noted: CNE and CNER are independent from the ministry of Research and report directly to the French President. While the advantages of the guarantor approach are obvious, some (but not structural) flaws lie in goal overload and a certain blurring of the guarantors' rationale.

In Germany the RTD system is much more decentralised, with many powerful actors on the federal and Länder level. An increased interest in RTD evaluation in Germany is stated, but "the question remains open whether this indicates a growing 'rationality' of the actors of the research system, or rather an increased sophistication of the player's battle for funds" [Kuhlmann, 3]. The set of actors is characterised as rather confusing, decentralised, negotiation- and network-oriented. The function of evaluations in this institutional interplay is described as broadly varying, from value-for-money to procedural fine-tuning. Evaluations are also utilized in policy moderation processes, aiming at greater transparency and common information pools. The author argues that evaluations can and should play a role in the negotiation processes between institutions by providing a kind of common ground. This on the other hand influences the shaping of evaluations; serving as instruments in broader processes, their language and terms must be common ones.

For the USA, the evaluation tradition is strongly based on accountability issues. One presentation gave a short overview on the development of the underlying legal provisions - from the 1921 Budget and Accounting Act to the 1993 Government Performance and Results Act (GPRA) forcing all relevant Federal Agencies to implement programme evaluations. The overall goal is performance accountability [Link, 3 ff]. This approach has been criticized for being inappropriate for basic research [Cozzens]. A different case study from the State level [Shapira/Youtie, 22] shows the different usefulness of indicators and methods for different stakeholders. Canada has traditionally an organised, central and methodic evaluation culture with a strong role for the Treasury Board (six government ministers responsible for the overall management of the federal public service) and Parliament (Committees and parliamentary Auditor General). Each major organisation is responsible for evaluating its own initiatives under common guidelines [McDonald/Teather].

The Dutch example is described in this issue extensively by Hutschenreiter. In our context it is interesting to mention the considerable role of the Dutch General Chamber of Audit in influencing the shaping of evaluations [of RTD tax allowance schemes, Dorsman, 3]. The Scandinavian approach was already mentioned above [Luukkonen]. The striking 'normality' of evaluations must be stressed again: No explicit legal provisions are needed, obviously all major institutions are accustomed to foreign

evaluators, public reports and open discussions. In a Vienna Plattform Technologieevaluierung - meeting some weeks ago, Luukkonen explained this somehow astonishing situation (the Finnish case) with socio-cultural factors and a longstanding, still growing evaluation culture. The European Union finally has an elaborated evaluation system containing project, programme and overall evaluations in nearly all stages. An overview [Durieux/Fayl] was given already in this Newsletter [Sturn et. al., 2 f] and is not to be repeated here. As the MONITOR/SPEAR programme tried to enhance methodological knowledge in the early 90ies, now at least one big project within the 4th Framework Programme deals with conceptual evaluation issues: Within the Targeted Socio Economic Research, the Advanced Science and Technology Policy Planning (ASTPP) network "investigates the ways and to which extent technology foresight, technology assessment and policy impact evaluations can be used and further enhanced in order to make the interfaces between the research system, industry, society and policy more "intelligent" [Kuhlmann, 13].

General conclusions for all countries seem difficult to be drawn. Finally, at least three institutional requirements have to be safeguarded [Georghiou, 18]:

- Evaluations have to take place on a programmed and properly resourced base;
- the independence of the evaluators must be guaranteed; and
- there must be mechanisms for feedback of the results into policy-making and hence for learning.

A detailed analysis of institutional impacts on evaluations was not brought forward in the Paris conference. Further research, e.g. case studies in a cultural context, will be necessary. Austria with its institutional settings, namely the independent funding organisations and the big but rather fragmented university system, seems to deserve a special look. It would be helpful to know better the cultural and institutional factors that have prevented the emergence of a regular evaluation system over the last decades.

The conference proceedings will be available at the OECD secretariat within a few months.

### ***Papers and Literature:***

Cozzens, S., U.S. Research Assessment: Recent Developments. *Scientometrics* Vol. 34. No.3 (1995), 351 ff

Cunio, K.M., UK Government departments experience of RT+D Programme Evaluation and Methodology. *Scientometrics* Vol. 34. No.3 (1995), 363 ff

Dorsman, R., Evaluation of industrial R+D support in the Netherlands. Conference Paper (1997)

Durieux, L., Fayl, G., The scheme used for evaluation of the European RTD programmes. Conference Paper (1997)

Fritz, O., Hutschenreiter, G., Sturn, D., Evaluierung von FTE-Programmen: Zwischen Best Practice Entwicklung und österreichischen Strukturen. Studie im Auftrag des BMWV (1997)

Georghiou, L., Issues in the evaluation of innovation and technology policy. Conference Paper (1997)

Hervik, A., Evaluation of user-oriented research in Norway: estimation of long run economic impacts. Conference Paper (1997)

Kuhlmann, S., Science and Technology Policy Evaluation: Present State in Germany and European Outlook. Conference Paper (1997)

Kuhlmann, S., Holland, D., Evaluation von Technologiepolitik in Deutschland. Konzepte, Anwendung, Perspektiven (1994).

Laredo, P., Evaluation in France: a decade of experience. Conference Paper (1997)

Link, A., Evaluating technology-based public institutions: lessons from the National Institute of Standards and Technology. Conference Paper (1997)

Luukkonen, T., The increasing professionalisation of the evaluation of mission-oriented research: implications for the evaluation process. Conference Paper (1997)

McDonald, R., Teather, G., Science and technology evaluation practices in the government of Canada. Conference Paper (1997)

Shapira, P., Youtie, J., Evaluating Technology Deployment at the State Level: Methods, Results, and insights from the Georgia Manufacturing Extension Alliance. Conference Paper (1997)

Sturn, D., Stampfer, M., Hutschenreiter G., Plattform Technologieevaluierung; Nr. 2, 1ff

**Dr. Michael Stampfer**

*Federal Ministry for Science and Transport , Technology Policy Division*

*Tel.: # 1/53464/3412 Fax: # 1/53464/2013 e-mail: michael.stampfer@bmv.gv.at*

## **Recent Experiences in the Evaluation of R&D Tax Incentives**

*Gernot Hutschenreiter*

A number of OECD countries are operating fiscal incentive schemes to stimulate R&D for some time now. In the course of time they have accumulated insight into the operation and effectiveness of their respective schemes. Some countries, such as Australia, have reformed their fiscal incentive schemes in the light of this experience while others, such as the Netherlands, have recently introduced such instruments. In Austria, recent technology policy proposals have argued for an extension and reform of the major tax incentive scheme ("Forschungsfreibetrag") currently in operation. It thus appears timely to observe more intensely the evaluation efforts conducted internationally.

At the OECD Conference "Policy Evaluation on Innovation and Technology" (Paris, 26-27 June 1997) two contributions specifically dealt with the evaluation of fiscal incentives for R&D:

R. Dorsman (Ministry of Economic Affairs, Netherlands): "Evaluation of Industrial R&D Support in the Netherlands" and

R. Lattimore (Productivity Commission, Australia): "R&D Fiscal Incentives in Australia: Impact and Policy Lessons".

These contributions vary considerably in scope and depth. The Dutch contribution provides information about an evaluation plan adopting a "more structured and systematic approach" to measure the effectiveness of a fairly new scheme. The work currently being drawn up is seen as a "pilot project". In contrast, the Australian paper is based on a decade of rather intense effort involving several research groups to shed light on the effects of the Australian R&D Tax Concession.

The Dutch contribution presents a current evaluation plan for the "Wage Tax and Social Insurance Allowance Act (WVA/S&O)" first introduced in 1994. This evaluation is expected to provide new insights after a preliminary and, obviously, not entirely satisfactory evaluation performed in 1995. The Dutch allowance scheme provides for a reduction of wage tax and social insurance contributions to be paid by the employer. It is thus independent of the current profitability of firms. The allowance amounts to 40% of the first Dfl 150,000 of R&D wage costs and to 12.5% of the amount in excess. The maximum yearly allowance per company (group) is Dfl 15 million. Project applications are evaluated -

in advance - by Senter, an agency of the Ministry of Economic Affairs (not handled ex post by the tax authorities).

The evaluation plan splits the evaluation in three separate issues:

the effect of the allowance on R&D efforts,

the effect of R&D efforts on innovative capacity and, most ambitiously,

the effect of innovative capacity on economic performance.

(Other topics to be studied include the participation of SMEs, contract research, self-employment, embedding/attracting R&D to the Netherlands, the effect of rejections, and subjective opinions of claimants.)

Considerations about populations/control groups to be surveyed as well as a set of indicators related to the three issues are given. It is not clear if it is intended to maintain the causal links between the three (certainly interrelated) issues quoted above and if so, how. In general, it would be worthwhile to learn more about the specific hypotheses to be tested. Certainly, the very short time elapsed since the introduction of the scheme together with the time lags involved pose a formidable challenge for deriving robust quantifications even if a panel data set is going to be used.

The Australian contribution, by contrast, is based on a comprehensive and sustained effort to gain insights with respect to the impact of R&D fiscal incentives in Australia. This emphasis is not coincidental. In 1985/86 Australia introduced one of the most generous schemes in the OECD area, the "R&D Tax Concession". Initially, the Tax Concession allowed firms to deduct a range of (R&D related) expenses at up to 150%. This level was later reduced to 125%. Other modifications included the permission of syndicates of firms in 1987 (and its withdrawal in 1996). Lattimore presents data according to which in 1994/95 the general (150%) Tax Concession had about 3,500 claimants accessing approximately \$ 550 million in R&D subsidies. The syndication program had just 103 claimants but was worth \$ 255 million in subsidies. Compared to these two programs all other Australian government R&D funding schemes were negligible in size. The main objective of the R&D Tax Concession consists in providing incentives to higher levels of business R&D.

In the 1990s several independent evaluations of the Tax Concession were undertaken combining qualitative and quantitative indicators of its impact and of the private and social benefits from induced R&D. Basically, social cost-benefit analysis was applied as a conceptual framework. Within such a framework, net social benefit may be defined as the sum of *induced private benefits plus spillovers* not captured by the private investor **minus** the *marginal excess burden of taxation, leakage to overseas shareholders, government administrative costs, business compliance costs and resource costs of rent seeking* (lobbying for more benefits or resist the erosion of existing benefits).

The evaluation studies carried out in Australia provide valuable information with respect to the various elements finally used as building blocks in the estimation of net social benefits. This is not the place to give a comprehensive account, however. Let us just mention that one of the most important elements in this context is empirical research on inducement, i.e. on the responsiveness of R&D to incentives. Here, inducement is measured as the share of policy induced current R&D activity in total current activity. A number of Australian studies based both on econometric methods and on subjective assessments of the impact of the R&D Tax Concession are surveyed.

In summary, Lattimore reports that the estimated social rate of return to the 125% Tax Concession is strongly positive. Interestingly, it also exceeds the rate of return to the 150% concession! This is accounted to the fact that the "bang for a buck" (i.e. the ratio of the inducement rate to the subsidy rate) is higher for the 125% Tax Concession. This result indicates a decreasing marginal inducement from an increasing subsidy rate. The results of the cost-benefit analysis also point to weaknesses of the Tax Concession such as a relatively low inducement of additional R&D. A large number of further results are reported. The clarity of the applied framework facilitates deriving options for redesigning fiscal

incentives. In particular, these proposals are aimed at increasing inducement (such as implementing support for *incremental* as opposed to total R&D).

Their conceptual clarity and empirical richness render the Australian studies particularly attractive and may serve as a valuable reference for future work to be conducted in Austria.

**Gernot Hutschenreiter**

*Austrian Institute of Economic Research (WIFO)*

*Tel +43-1-7982601-238, e-mail: hutsch@wsr.ac.at*

## **Impressum**

Die „Plattform Technologieevaluierung“ ist ein unregelmäßig erscheinendes offenes Forum zur Diskussion methodischer und inhaltlicher Evaluierungsfragen in der Technologiepolitik.

Für den Inhalt verantwortlich:

Bundesministerium für Wissenschaft und  
Verkehr - Dr. Michael Stampfer  
Renngasse 5, A-1010 Wien  
E-mail:michael.stampfer@bmv.gv.at

weitere Kontaktadressen:

Mag. Gernot Hutschenreiter  
WIFO  
Arsenal Objekt 20, PF 91  
A-1031 Wien  
E-mail: hutsch@wsr.ac.at

Dr. Oliver Fritz  
Joanneum Research  
Wiedner Hauptstraße 76  
A-1040 Wien  
E-mail: oliver.fritz@joanneum.ac.at