JOINT RESEARCH CENTRE
EUROPEAN COMMISSION

JOANNEUM RESEARCH

INSTITUTE FOR PROSPECTIVE TECHNOLOGICAL STUDIES

IPTS Technical Report Series

RTD Evaluation Toolbox

- Assessing the Socio-Economic Impact of RTD-Policies -

Strata Project HPV 1 CT 1999 - 00005

edited by:

GUSTAVO FAHRENKROG (JRC-IPTS)
WOLFGANG POLT (JOANNEUM RESEARCH)
JAIME ROJO (JOANNEUM RESEARCH)
ALEXANDER TÜBKE (JRC-IPTS)
KLAUS ZINÖCKER (JOANNEUM RESEARCH)

with chapter contributions from:

SPYROS ARVANITIS (ETH), MARC BODEN (IPTS/PREST),
SUSANNE BÜHRER (FHG-ISI), HENRI CAPRON (ULB),
MICHELE CINCERA (ULB), ROBIN COWAN (MERIT),
JONATHAN EATON (BOSTON UNIVERSITY), LUKE GEORGHIOU (PREST),
MAX KEILBACH (ZEW), EAMON KINSELLA (CIRCA),
STEFAN KUHLMANN (FHG-ISI), GEORG LICHT (ZEW),
PARI PATEL (UNIVERSITY OF SUSSEX), WOLFGANG POLT (JOANNEUM RESEARCH),
JOHN RIGBY (PREST), JAIME ROJO (JOANNEUM RESEARCH),
GIORGIO SIRILLI (CNRS), ELLIOT STERN (TAVISTOCK),
BIRGIT WOITECH (JOANNEUM RESEARCH)

August 2002
EUR 20382 EN

IPTS Technical Report Series, EUR 20382 EN

"**RTD Evaluation Toolbox** - Assessing the Socio-Economic Impact of RTD-Policies"
Editors: Gustavo Fahrenkrog (JRC-IPTS), Wolfgang Polt (Joanneum Research), Jaime Rojo (Joanneum Research), Alexander Tübke (JRC-IPTS), Klaus Zinöcker (Joanneum Research)

Authors: Spyros Arvanitis (ETH), Marc Boden (IPTS/PREST), Susanne Bührer (FhG-ISI), Henri Capron (ULB), Michele Cincera (ULB), Robin Cowan (MERIT), Jonathan Eaton (Boston University), Luke Georghiou (PREST), Max Keilbach (ZEW), Eamon Kinsella (CIRCA), Stefan Kuhlmann (FhG-ISI), Georg Licht (ZEW), Pari Patel (University of Sussex), Wolfgang Polt (Joanneum Research), John Rigby (PREST), Jaime Rojo (Joanneum Research), Giorgio Sirilli (CNRS), Elliot Stern (Tavistock), Birgit Woitech (Joanneum Research)

Seville, Spain, 2002

Published by:
EUROPEAN COMMISSION
Joint Research Centre
Institute for Prospective Technological Studies (IPTS)
Edificio Expo-WTC
C/ Inca Garcilaso, s/n
E-41092 Sevilla, España (Spain)
http:\\www.jrc.es

Printed in Spain

EUROPEAN COMMISSION

Directorate-General for Research

In the Directorate-General for Research, the Directorate "Knowledge-based economy and society" (Directorate K) contributes to the realisation of the European Research Area in the fields of the social sciences, economic, science and technology foresight, and the respective analyses. To this end, it monitors and encourages science and technology foresight activities, conducts the economic analyses necessary for the work of the Directorate-General, and co-ordinates policy as regards the relevant political, economic, human and social sciences. It prepares the European reports on science and technology indicators, and it contributes to the development and implementation of the Framework Programmes in these fields. It monitors the progress made in the implementation of the Lisbon strategy.

Unit K 2 - "Science and Technology Foresight, links with the IPTS"
The unit contributes to the Directorate General's policy reflections by conducting forward-looking analyses and syntheses and encourages, together with other Directorates, DGs and JRC/IPTS, co-operation between the actors and users of foresight in Europe. It is also responsible for the implementation of the related actions (5th and 6th Framework Programmes). The Unit manages also the **STRATA** action in FP5.
The objective of the **STRATA** activity is to enable considered reasoning by researchers and Research Technology Development and Innovation (RTDI) policy makers alike, via dialogue, to improve the efficiency of the European policy development process at local, regional, national and international level, as well as interactions with other related policies. It is intended to build dialogue on emerging and/or relevant academic work, to develop RTDI policy options and to allow good practice to emerge.

Director: Jean-François Marchipont
Head of Unit K2: Paraskevas Caracostas
Scientific Officer: Elie Faroult
E-mail: elie.faroult@cec.eu.int
www.cordis.lu/rtd2002/foresight/home.html

This publication presents results of the project called "Socio-Economic Evaluation of Public RTD policies (EPUB)", which has been awarded financial support by the European Commission under the 5th Framework Programme of the European Community for Research, Technological Development and Demonstration Activities (1998 to 2002), and its specific programme "Improving the Human Research Potential and the Socio-Economic Knowledge Base" ("**STRATA**- Strategic Analysis of Specific Political Issues"). The authors are solely responsible for this publication, which does not represent the opinion of the Commission. Neither the Commission nor any person acting on behalf of the Commission is responsible for the use which might be made of the information contained in this publication.

ABOUT THE JRC-IPTS

The Joint Research Centre (JRC) is a Directorate General of the European Commission, staffed with approximately 2,100 people, coming in the vast majority from the 15 Member States of the European Union. The Headquarters, the Directorate for RTD Strategy and seven Institutes located in five different countries compose the main organisational structure of the JRC (http//:www.jrc.org). The **mission of the JRC** is to provide customer-driven scientific and technical support for the conception, implementation and monitoring of EU policies.

The Institute for Prospective Technological Studies (IPTS) is one of the seven Institutes making up the JRC. It was established in Seville, Spain, in September 1994.

The **mission of the IPTS** is to provide prospective techno-economic analyses in support of the European policy-making process. IPTS' prime objectives are to monitor and analyse science and technology developments, their cross-sectoral impact, and their inter-relationship with the socio-economic context and their implications for future policy development. IPTS operates international networks, pools the expertise of high level advisors, and presents information in a timely and synthetic fashion to policy makers (http//:www.jrc.es).

The **IPTS is a unique public advisory body**, independent from special national or commercial interests, closely associated with the EU policy-making process. In fact, most of the work undertaken by the IPTS is in response to direct requests from (or takes the form of long-term policy support on behalf of) the European Commission Directorate Generals, or European Parliament Committees. The IPTS also does work for Member States' governmental, academic or industrial organisations, though this represents a minor share of its total activities.

Although particular emphasis is placed on **key Science and Technology fields**, especially those that have a driving role and even the potential to reshape our society, important efforts are devoted to improving the understanding of the complex interactions between technology, economy and society. Indeed, the impact of technology on society and, conversely, the way technological development is driven by societal changes, are **highly relevant themes within the European decision-making context**.

The **inter-disciplinary prospective approach** adopted by the Institute is intended to provide European decision-makers with a deeper understanding of the emerging science and technology issues, and it complements the activities undertaken by other institutes of the Joint Research Centre.

The IPTS **approach** is to collect information about technological developments and their application in Europe and the world, analyse this information and transmit it in an accessible form to European decision-makers. This is implemented in the following **sectors of activity**:

- Technologies for Sustainable Development
- Life Sciences / Information and Communication Technologies
- Technology, Employment, Competitiveness and Society
- Futures project

In order to implement its mission, the Institute develops appropriate contacts, awareness and skills to anticipate and follow the agenda of the policy decision-makers. **IPTS Staff** is a mix of highly experienced engineers, scientists (life-, social- material- etc.) and economists. Cross-disciplinary experience is a necessary asset. The IPTS success is also based on its **networking capabilities and the quality of its networks** as enabling sources of relevant information. In fact, in addition to its own resources, the IPTS makes use of external Advisory Groups and operates a number of formal or informal networks. The most important is a Network of European Institutes (*the European Science and Technology Observatory*) working in similar areas. These networking activities enable the IPTS to draw on a large pool of available expertise, while allowing a continuous process of external peer-review of the in-house activities.

About Joanneum Research

With more than 350 employees JOANNEUM RESEARCH is the largest independent R&D institution in Provincial ownership. As an innovation partner for business, industry and public administration we focus on applied research and technological development in cutting-edge key technologies. At present, we receive 40 % of our contracts from business enterprises and 44 % from public authorities. The percentage of research contracts from international organisations amounts to 16 %, accounting for 31 % of our operating performance.

The 21 JOANNEUM RESEARCH institutes are located in Graz, Vienna, Judenburg, Leoben, Niklasdorf, Frohnleiten, Weiz and Hartberg. Close co-operation with the Styrian universities allows us to concentrate on applied research and deal with our customers' wishes in detail. About a third of the JOANNEUM RESEARCH heads of institute are university professors or university lecturers. This close link provides access to fundamental research on the one hand and ensures a steady supply of young and highly talented university graduates on the other. Although we can draw on long years of experience in all our research divisions, we are an extremely young enterprise: the average age of our employees is 36 years and more than half of our institutes was established in the past eight years.

RTD Evaluation Toolbox
- Assessing the Socio-Economic Impact of RTD-Policies

TABLE OF CONTENTS

FOREWORD

The political objectives agreed at the European Council of Lisbon in 2000 face numerous barriers on the way from knowledge to value creation. Policy-makers have become aware that the "translation" of scientific and technological progress into economic growth takes place in a distributed context with complex interactions among the different stakeholders. This is accompanied by an increasing speed of change, global competition and uncertain impacts of policy-action. Taken together, these factors make it difficult to establish Science and Technology policies for the Lisbon process.

Not only within the EU policy-making process, but also at the level of many Member States, evaluation has widely been applied in a way that very much resembles what in this report is understood as "monitoring". For other steps of the evaluation process, peer reviews and expert groups have mainly been used. However, the analysis of the available evaluation techniques and experiences together with methodological advances show that the evaluation process can offer more to decision-makers than currently is made use of. The use of foresight results is highly valuable, especially during the ex-ante evaluation. The bigger size and wider impact of policies can only be captured and understood if a wide range of evaluation methodologies and indicators is at hand. At present, experience with such an interplay of evaluation methodologies is rather sparse.

Therefore, as a project supported by the STRATA programme, the Epub thematic network analysed the methodologies for evaluating the socio-economic impact of Research and Technological Development (RTD) policies during the last 28 months. The present toolbox provides policy-makers, scientists and practitioners with an overview of the main evaluation concepts and methodologies, outlines their strengths and limitations, and sets them in relation to the policy context. Emphasis is set on a practice-oriented presentation.

The editors hope that this toolbox raises awareness of the potential of RTD-policy evaluation among decision-makers, contributes to an integrated view of evaluation and thus increases its value for achieving the objectives of the Lisbon process.

Seville and Vienna, August 2002

EXECUTIVE SUMMARY

This report provides the state of affairs in methodologies for evaluating the socio-economic impact of Research and Technological Development (RTD) policies and their contribution to the policy-making process.

Evaluation is a systematic and objective process that assesses the relevance, efficiency and effectiveness of policies, programmes and projects in attaining their originally stated objectives. It is both a theory- and practice-driven approach, whose results feed back into the policy-making process and help formulating and assessing policy rationales. However, there is still a lot of confusion on what evaluation can realistically and effectively achieve.

The growing importance of science and technology-induced innovation, new forms of knowledge and their application, complex exchange processes, the increasing speed of change, and the distributed nature of innovation increase the complexity and uncertainties about the impact of RTD policies, which makes the establishment of systemic policies much harder. Further, additional inter-dependencies appear as the leverage of public resources is promoted and Foresight and Strategic Policy Intelligence flow into the decision-making process.

Both theory and practice of evaluation have co-evolved with the developments experienced in Science and Technology (S&T) policy. Evaluation tools have expanded from providing a means for quantification of policy impacts towards facilitate mutual learning from past experiences, supporting mediation, decision-making and policy strategy definition. The increasing complexities and uncertainties currently present in policy decision-making foster the emergence of tools like Strategic Intelligence. Such tools combine single methodologies like evaluation, Technology Foresight, Forecasting and Assessment in order to provide decision-makers with comprehensive, objective, politically unbiased, and independent information.

The report is structured as follows. The first chapter presents evaluation from a user perspective and highlights the sometimes conflicting expectations of the different actors. The second chapter describes aspects of evaluation in four broad policy areas, i.e. financing Research and Development (R&D), the provision of R&D infrastructures, technology transfer and the legal framework. The third chapter reviews eleven main evaluation methodologies, providing their descriptions, requirements for their application and good practice examples. The fourth chapter presents evaluation within a distributed network. The fifth chapter explores the role of evaluation for the policy instruments in the European Research Area (ERA) and outlines the synergies between the different evaluation methods. Chapter six provides the results of an expert-conference, which discussed the results of the work on this toolbox with respect to the future policy context. The concluding remarks in the last chapter give some indications on how to match evaluation methodologies and policy-instruments.

Evaluation processes can benefit from a combined use of the methodologies presented in this report. For example, quantitative evaluation methods combined with performance indicators permit to capture the dynamics involved in S&T, which provides good estimates of the outputs and impacts of public intervention. Policy-makers could make use of these impact estimates to legitimise and support policy intervention. Qualitative

evaluation methods offer more detailed insights on the multiple effects of policy intervention, which might help improving the processes and instruments of S&T policies.

Current EU RTD evaluation practices –comprising continuous monitoring, five year assessments and mid-term evaluation– are characterised by a strong focus on monitoring compared to impact assessment, on projects and programmes rather than the broad policy context, and a heavy reliance on expert panels rather than on studies. Also, there is a constraint imposed by the limited time and monetary resources devoted to evaluation. The increasing diversity of RTD policy instruments (e.g. funding of collaborative R&D, support of R&D infrastructures, measures for technology transfer and diffusion, standards and regulations, Intellectual Property Rights, networking…) at the EU level makes it necessary to apply a mix methodologies that accounts for the different kinds of instruments and their wide range of impacts.

This toolbox shows the basic scope and limits of various evaluation methods. From this, areas of synergies can be pointed out. An importnant finding is that the results of innovation surveys should be better linked to evaluation exercises at the macro level. Thus, econometric impact assessments could be used on a wider scale than at present. This holds true not only for the macro-level but also –and especially– for micro-econometric tools like control group approaches. In the future, also cost-benefit analysis might play a bigger role, not least in the ex-ante and ex-post evaluation of large-scale projects. Even with peer review, probably the most widely used approach in European S&T evaluation, it is possible to improve applicability through refined ways of panel composition, task allocation and decision making power.

However, probably as important as the suggestions and recommendations with respect to individual instruments and approaches is the perception that evaluation, in order to serve its purpose to empower policy learning, should follow some general good practice rules and ought to be embedded in a broader "system of distributed intelligence". Such system comprises other policy-making support tools as well, like benchmarking, Technology Foresight, Forecasting and Assessment. In this setting, the use of evaluation results will yield most benefits to the policy matter.

Although a variety of established evaluation methods is available, there is scope and need to look for further methodological improvements in evaluation. At present, consistent evaluations can be conducted at the project level, more thorough evaluations at programme or policy level will require advances in knowledge both in the causal relations between inputs and outputs as well as to arrive at meaningful ways to measure and to aggregate these outputs.

Moreover, the continuous evolution of science and technology policies in modern economies requires devising methods to evaluate the efficiency and efficacy of new policies. For instance, the ERA concept requires a better understanding of the interconnections and the integration of S&T organizations and other stakeholders. This would require implementing methods that allow evaluating institutional capacity within changing environments. Evaluation processes are thus increasingly required to capture the effect of S&T policy on the behaviour of actors and institutional change.

In total, the RTD-evaluation toolbox shows that evaluation can constitute an important support to policy-making, but only on the prerequisite of an adequate degree of policy-awareness. At present, the European Union and its Member States face a wide range of challenges, such as the enlargement of the union, the implementation of the European Research Area, the search for new modes of governance and improving the economic impacts of S&T. RTD-policy evaluation can provide an important input to the policy-making process, based not only methodological strength to address these issues, but also enough degrees of flexibility to link to other forms of interaction, to adapt to new governance models and be open to the rapid and unforeseen technological changes and societal developments. This set of instruments and methods might help policy-makers, evaluators and programme managers in accomplishing their tasks.

GENERAL INTRODUCTION

Authors: **Gustavo Fahrenkrog** and **Alexander Tübke** (JRC-IPTS), **Wolfgang Polt** and **Jaime Rojo** (Joanneum Research)

With the ever-growing importance of knowledge in the economy, science and technology-induced innovation will be decisive for meeting the Lisbon process target of making the European Union "the most competitive and dynamic knowledge-based economy in the world by 2010"[1]. While this statement is widely shared and easily formulated, it is a key-challenge to establish Research and Technological Development (RTD) policies that lead to this objective. Economic and social realities have changed significantly during the past decade and created a "policymaker's dilemma". New forms of knowledge and their application in products and services go together with complex exchange processes. Innovation occurs in networks and alliances rather than in an individual firm or Research and Development (R&D) lab, which increases the importance of innovation systems. This distributed nature of innovation creates a much more complex and volatile picture than the traditional view of a successful invention. Most policy relevant science and technology applications affect wider society. This increases the importance of policy impact assessment that is able to take account of the shared nature of innovation. In addition, the rising demand for greater transparency and participation in public decisions about Science and Technology (S&T) creates new patterns for S&T decision-making. Risks and decisions are increasingly negotiated between different stakeholders in society. This is exemplified by the increased profile of Science and Governance issues in recent years, and the demand for higher levels of participation and transparency in setting S&T agendas[2]. It also finds its expression in a reinforced public attitude that policy-makers be able to explain and justify their decisions at almost any time. Taken together, these trends increase the complexity and uncertainties about the impact of RTD policies even more. In addition, they raise the stakes for decision-makers because of the increased importance of providing framework conditions that are fit for the future. This makes the policy-maker's task of establishing systemic policies much harder.

In addition to the changing conditions of economic reality, recent trends in RTD policies further foster the need for a well-designed approach towards policy evaluation and impact assessment. At the European level, differentiated approaches and a systems view are emerging[3]. Foresight[4] and Strategic Policy Intelligence[5] are new tools that flow into the decision-making process. The European Research Area (ERA)[6] and new forms of governance[7] emphasise the networking of national act4ivities, benchmarking and the interaction of science, its application and evaluation. In addition, a systematic evaluation process can be interpreted as a good governance practice in itself. At the national and regional levels, direct subsidies decrease (except to SMEs). Start-ups, entrepreneurship and New Technology Based Firms are increasingly supported, and intermediaries, brokers and the public-private interface are fostered. National and

[1] See European Commission (2001) (COM (2001) 79 final)

[2] For EU-activities on Science and Governance see: http://www.jrc.es/sci-gov

[3] See http://www.cordis.lu/innovation-smes/src/policy.htm

[4] See http://futures.jrc.es/

[5] See http://www.jrc.es/pages/projects/stratpolint.htm

[6] See the ERA web-site http://europa.eu.int/comm/research/era/index_en.html and European Commission (2000) (COM (2000) 6 final) (http://europa.eu.int/comm/research/era/pdf/com2000-6-en.pdf)

[7] See http://www.jrc.es/sci-gov

regional governments promote the leverage of public resources, improve linkages between research in the public and the private sector and increase regional involvement. Evaluation of public RTD policies is thus confronted with additional inter-dependencies and secondary effects, which are difficult to forecast and need a systemic view.

Further, the speed of technological development and its role for the society and the economy leave less time for political decision-making. This changes the position of evaluation with respect to the policy-making process. In the future, evaluation activities will only be able to effectively contribute to the policy-making process if they are implemented as a continuous and multi-layered process that is independent from the interests of single stakeholders. If the position of evaluation with respect to the policy-making process is not strong enough, its results will not be "objective" enough to produce the necessary impact. In this sense, policymakers at all levels must be aware that, even if there might be no formal requirements for such a strong position of evaluation, not using it adequately might produce structural disadvantages on the longer-term. These would be difficult and costly to correct in times of global competition.

Against this background, the objective of this report is to provide an overview of the state of affairs in methodologies for evaluating the socio-economic impact of RTD policies. It is based on the toolbox produced by the EPUB thematic network[8] as a result of its work during the last 28 months. The aim is to provide policy-makers, scientists and practitioners with an overview of the main evaluation concepts and methodologies, outline their strengths and limitations, and set them in relation to the policy context. Emphasis is set on a practice-oriented presentation.

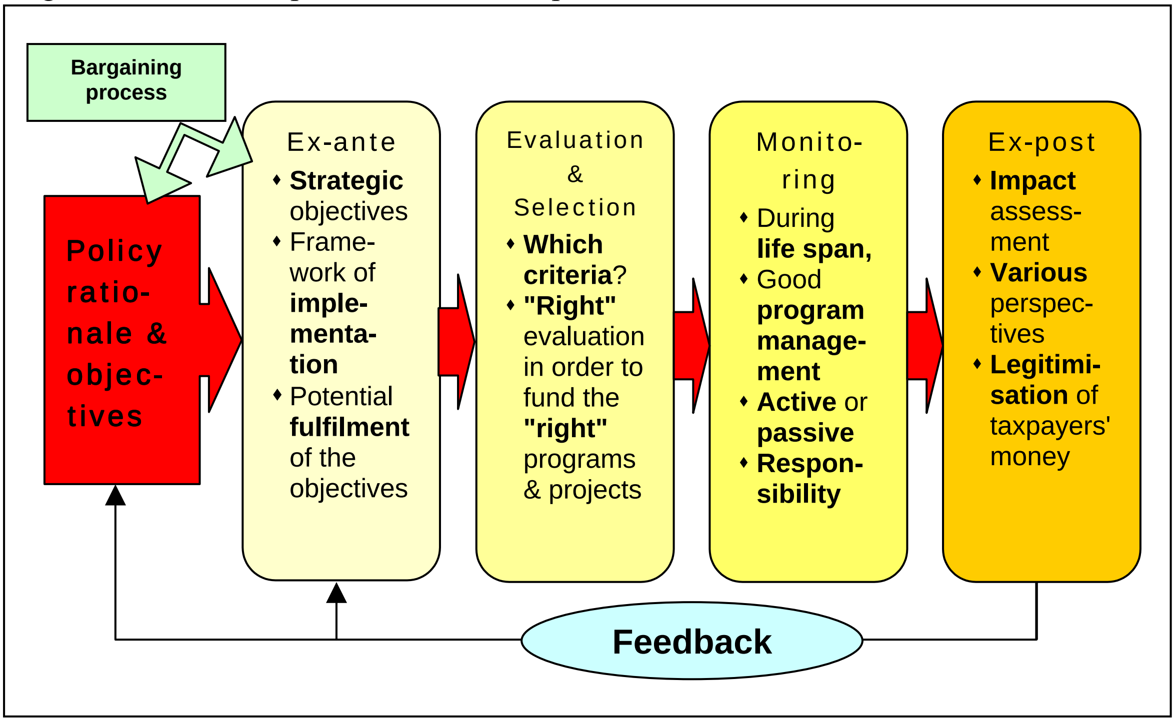
In this context, evaluation can be defined as a systematic and objective process that assesses the relevance, efficiency and effectiveness of policies, programmes and projects in attaining their originally stated objectives. It is both a theory- and practice-driven approach. Evaluation results feed back into the policy-making process, so that it is part of a continuous learning process. This brings transparency and accountability to the policy-making process and helps formulating and assessing policy rationales.

As presented in figure 1, the evaluation process has a direct relation to policy-making and is composed of four steps.

---

8 The "Socio-Economic Evaluation of Public RTD-Policies (Epub)" was a thematic network under the STRATA action (Strategic Analysis of Specific Political Issues) of the 5th EC RTD Framework Programme (see http://epub.jrc.es/). The Epub thematic network constituted a pilot action aiming at supporting European RTD policy making and improving efficiency in RTD policy actions. The network's main objective was to provide a better understanding of the dynamics of science and technology by helping to investigate the socio-economic impact of private/public RTD policies at a micro/macro level. A special emphasis was given to analyse and review the rationales, concepts, methodologies and indicators generally applied in RTD evaluation and to investigate the implications for evaluation of new S&T policy approaches. The network also facilitates the sharing and diffusion of knowledge, techniques and tools applied in RTD evaluation.

Figure 1: General shape of the evaluation process



Source: Own compilation

As shown by the horizontal arrows, the policy rationale and objectives determine the shape of the evaluation process from the beginning to the end. Especially for longer-term projects, this basic principle of evaluation has to be respected, which means that a consistent evaluation process can only be implemented with respect to the original policy rationale and objectives that shaped the process at the outset. This does not mean that the policy rationale and objectives cannot be revised, but the evaluation process is bound to the original rationale and objectives from the beginning. The results obtained during the evaluation are mainly valid with respect to these objectives. If the policy rationale and objectives are changed in the course of the evaluation process, the process does not only have to be reshaped, or eventually reinitiated, but the results obtained up to then will lose validity and might become meaningless.

In ex-ante evaluation, the strategic objectives of a policy intervention are formulated and the framework of implementation is defined. In addition, the potential for fulfilling the policy objectives is assessed, which also regards the expected efficiency and effectiveness of the intervention. The results of the ex-ante evaluation often flow back into the bargaining process and may lead to a revision of the policy rationale and objectives. Once the policy objectives are defined and the intervention is considered suitable for attaining them, the evaluation and selection phase regards the criteria that define who will participate in the intervention. They can concern individuals who benefit from project funding or programs and program lines that are eligible for policy support. Due to its excluding character, this phase is highly important as only the "right" selection leads to the funding of the "right" programs and projects. The selected programs and projects then pass to the monitoring phase, which covers their whole life span and thus resembles classical project management. In active monitoring, the evaluator intervenes with the project steering. In passive monitoring, the project's progress is observed and no intervention made. Important aspects of the monitoring phase are who takes which responsibilities during the project implementation and who receives the results of the monitoring. The ex-post phase assesses the impact of the policy intervention, in many cases with regard to a number of perspectives (e.g. political, economic or social). Its results provide a legitimisation for the use of the taxpayers' money. In comparison to the ex-ante phase, the results of the ex-post evaluation provide a feedback on the degree of accomplishment of the original policy objectives and rationale. This experience then flows back into the policy-making process. Especially in large programs, the evaluation process is often undergone a number of times and complemented by intermediate evaluations and feedback. However, also in these cases the evaluation process generally follows the above scheme.

The practice of evaluation has developed from purely quantifying policy impacts towards facilitating mutual learning in the policy-process. At the present point, however, the full potential of evaluation is not adequately exploited. On the methodological side, a comprehensive overview of the evaluation instruments and their interrelations is still missing. While evaluations at the project level are methodologically consistent, evaluations at the programme or policy level will require further insights into the causal relations between inputs and outputs and consistent ways of measuring and combining these outputs. The application of evaluation would benefit from a systematic connection to concrete policy instruments. Further, with regard to its position within the policy-process, there is a convergence between evaluation and other support tools, like Foresight or Strategic Policy Intelligence. Their methodologies and results can contribute especially to the ex-ante phase of evaluation, but also offer an opportunity for including forward-looking information into the whole evaluation process. This is the more important the bigger and the longer-term policy commitments become. Possible synergies among these tools should be further examined in order to inform policymakers in a comprehensive, unbiased and consistent way. This is especially necessary for the formulation of new policy instruments. Current EU evaluation practices strongly focus on monitoring and less on impact assessment, and rely mainly on expert panels. They also tend to centre on projects and programmes rather than on the wider policy context. However, the ERA concept will require more comprehensive evaluation concepts in order to assess the impact of policymaking on the networked economy and the behaviour of S&T organisations and other actors involved. This includes finding the right balance of evaluation methods for each policy instrument. A proposal for matching evaluation methods and selected policy instruments is undertaken in table 1.

Table 1: Matching policy instruments and evaluation methods

	Innovation Surveys	Econometric Models	Control Group Approaches	Cost Benefit Analysis	Expert Panels/ Peer Review	Field / Case Studies	Network Analysis	Foresight/ Technology Assessment	Benchmarking
Financing R&D	●●●	●●●	●●●	●		●●●		●	●
Provision of R&D infrastructure		●●		●●●	●●●	●●●	●●●	●●	●●●
Technology transfer and innovation diffusion	●●●	●●●	●●●	●●	●	●●	●●●	●●●	●●●
Legal frameworks (IPRs, standards and regulation)	●	●	●	●●●		●●●		●●	●●●
Integrated projects			●	●●●	●●●	●●●	●●●	●●	●●
Networks of excellence					●●●	●●	●●●	●●	●●

Legend: ●●● *Highly suitable* ●● *Suitable* ● *Less suitable*

Source: See Chapters 2.2 ("Mechanisms Supporting R&D") and 7 ("Concluding Remarks")

The table shows that the methodologies presented in the toolbox should not be perceived as substitutes but rather as complements. For example, a quantitative evaluation combined with performance indicators permits to estimate the output and impact of public intervention in dynamic S&T environments. Qualitative evaluation methods provide more detailed insights on the multiple and secondary effects of policy intervention, which may help improve RTD-policies.

With these aspects in mind, an important step for evaluation in the future would be to better link not only its methods but also the actors involved. Together with a systematic information exchange with other policy-support functions, like Strategic Policy Intelligence or Foresight, a distributed network could be established across Europe. This would enable a timely generation of information from independent and heterogeneous sources and allow covering a wide range of themes and demands. It might also build a more adequate link towards meeting the requirements of decision-makers in the knowledge-based economy. In the new role of evaluation, an important function of such distributed network would be quality-control, the evaluation of its members, and the monitoring and improvement of the methodologies. A distributed network would not only stimulate scientific and methodological progress in this area, but also lead to a sharper, more rigorous profile of evaluation and related areas.

However, the establishment of a new role of evaluation also requires some learning from its users. The systematic exchange of experience, the connection of stakeholders from different fields and methodological progress can only be achieved with the support of a clear mandate, combined with a certain degree of independence from single stakeholders. Evaluation exercises should be undertaken in an objective, continuous and multi-layered process. Its results are not to be taken as final judgements, but as one input amongst others to policy assessment. Evaluation thus needs to be embedded into the wider policy-making process. Its objective is therefore not to provide a positive or negative judgement, but rather to contribute, in a constructive way, to an informed, qualified and legitimate policy-debate.

The report is structured as follows. The first chapter presents evaluation from a user perspective and highlights the sometimes conflicting expectations of the different actors. The second chapter describes aspects of evaluation in four broad policy areas, i.e. financing R&D, the provision of R&D infrastructures, technology transfer and the legal framework. It reveals the main evaluation techniques applicable in these contexts and presents sample evaluations that have been conducted in the past. The third chapter reviews eleven main evaluation methodologies, providing their descriptions, requirements for their application and good practice examples. It shows the potential synergies emerging from the combination of different evaluation instruments. For example, the result of innovation surveys could be better linked to evaluation exercises at the macro level, and econometric impact assessments might be used on a wider scale than at present. This holds true not only for macro-, but especially for micro-econometric tools like control group approaches. Also cost-benefit analysis could play a bigger role in the evaluation of large-scale projects. Even the applicability of peer review, which is probably the most widely used approach in European S&T evaluation, should be improved through redefined ways of panel composition, task allocation and decision making power. The fourth chapter presents evaluation within a system of "distributed techno-economic intelligence". The fifth chapter explores the role of evaluation for the policy instruments in the ERA and outlines the synergies between the different evaluation methods. Chapter six provides the results of an expert-conference, which discussed the results of the work on this toolbox with respect to the future policy context. The concluding remarks in the last chapter give some indications on how to match evaluation methodologies and policy-instruments.

In summary, this RTD-evaluation toolbox is an attempt to provide a comprehensive overview of methodologies in a user-friendly presentation. It shows that evaluation can constitute an important support to policy-making, but only on the prerequisite of an adequate degree of policy-awareness. At present, the European Union and its Member States face a wide range of challenges, such as the enlargement of the union, the implementation of the European Research Area, the search for new modes of governance and improving the economic impacts of S&T. RTD-policy evaluation can provide an important input to the policy-making process, based not only on methodological strength to address these issues, but also enough degrees of flexibility to link to other forms of interaction, to adapt to new governance models and be open to the rapid and unforeseen technological changes and societal developments. It is hoped that this set of instruments and methods might help policy-makers, evaluators and programme managers in the accomplishment of their tasks.

1. USER PERSPECTIVES

Authors: **Mark Boden** (PREST) and **Elliot Stern** (Tavistock)

1.1. INTRODUCTION

Matching the requirements of policy makers with the skills and experience of evaluators can reveal crucial differences in perspectives. These may affect the delivery and implementation of evaluation studies that serve the desired policy purposes. These differences in perspective have been caricatured[9] as two gaps: the "delivery gap" between what policymakers want and what evaluators say, and the "customer gap" between what evaluators want and what policymakers say. While these caricatures combine and contrast features of ideal and real imagined situations, they set the scene for more serious consideration of matching user needs with available tools.

The delivery gap illustrates what policy makers would ideally want from an evaluation to inform policy decisions and what evaluators believe is actually feasible. In the real world, the diffusion of knowledge is a complex process, governed by interactions between various knowledge producers and users. While policymakers may need information to inform spending decisions, evaluators might remind them that research may require years to have effects. While evaluators want clear attribution of effects to investment, a linear perspective on funding and output is usually unrealistic and additionality is complex to assess. Independent evidence of research excellence may also be unachievable given the loyalty of peers to their subject field and international colleagues. Also, while indicators to monitor and benchmark research performance are highly desirable, a crude regime may distort performance and be open to manipulation. Table 2 briefly indicates the main issues contributing to the delivery gap.

Table 2: The Delivery Gap

What policymakers want	What evaluators say
• Information in time for spending decision	• Research may take years to have effects
• Clear attribution of effects to investment	• Linear model is a rare case and additionality is complex to assess
• Independent evidence of research excellence	• Peers defend their subject field & international colleagues
• Key indicators to monitor & benchmark	• Crude regime distorts performance & can be manipulated

In the converse situation, evaluators ideally need to have a clear, comprehensive and logical picture of the programmes they are evaluating, together with freedom and adequate resources. However, the real world of the policymakers is also complex and constrained. While evaluators want clearly defined and hierarchical programme objectives against which assess outcomes, policymakers may tell them that their programmes are a compromise involving multiple and conflicting objectives. While guaranteed independence might be desirable, this may also be compromised by the need for realistic recommendations within current policy constraints. Many evaluations do not have the luxury of time and resources for a full and thorough investigation, but face

[9] Georghiou, L. (2001) "The Impact and Utility of Evaluation", Conference on International best practices in evaluation of research in public institutes and universities, Brussels, 16.10.01

deadlines. Likewise, time constraints on programme management and participants are not conducive to giving evaluators full access to information and stakeholders.

Figure 2: A Framework for a Process of Informed Design

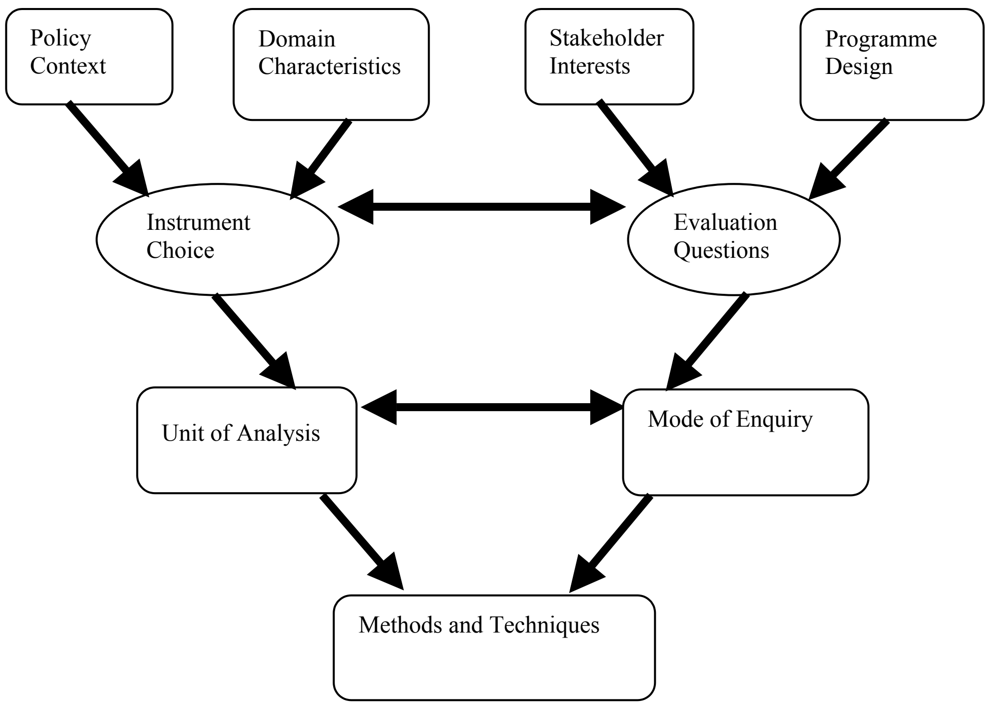


Table 3: The Customer Gap

What evaluators want	What policymakers say
• Clearly defined & hierarchical objectives	• Programmes are a compromise involving multiple & conflicting objectives
• Guaranteed independence	• Recommendations must be within realistic policy constraints
• Time & resources to do the job	• We need the results in three months
• Full access to information and stakeholders	• Everyone is overworked and busy

A longstanding weakness of many evaluations across many domains is their tendency to be method or technique led. The approach is determined by how rather than why or what questions. Particular experts with a commitment to their toolkit seek to apply their favoured approach. Any evaluation or analytic approach brings with it a certain logic, philosophical stance, range of possible outcomes and findings. New approaches to RTD policy evaluation need to be set into a framework that privileges questions of purpose, of ends rather than means. This introduction is intended to outline the kinds of questions evaluators and policy-makers should and do ask themselves when designing evaluations and as part of that design, before they come to select methods and techniques.
Figure 2 suggests a framework for a process of informed design. By going through the process outlined in this figure policy makers should be better able to select appropriate evaluation instruments, fit for purpose.

The structure of this chapter attempts to reflect this framework, examining first the relationship between evaluation and the programme context in which it may operate. This takes into account the nature of the policy context, the characteristics of the scientific and technological domains that are relevant to it and how together these shape the choice and mode of implementation of policy instrument. This is followed by consideration of the design and implementation of the instrument, stakeholder issues and how these then relate to the evaluation questions that may be posed.

Against the background of the relationship between policy and evaluation, attention is turned to the organisation of evaluation, matching the unit of analysis selected and the mode of enquiry to the methods and techniques outlined in subsequent chapters/sections of this report.

The preparation of this chapter has drawn on a combination of personal and documentary sources. Any adequate account of user perspectives requires consultation with users, and this has entailed a series of interviews at European and national levels, with policy-makers, programme managers and other stakeholders.

On this basis this chapter offers insights into the design, use and impacts of evaluation at both national and European Commission levels, and, in particular, considers implications for the construction and use of a toolkit of evaluation methods.

1.2. EVALUATION IN THE POLICY CONTEXT

This section is concerned with the context in which evaluation activities are conducted, both the policy and the scientific and technical determinants of the design and implementation of evaluation activity and the role of evaluation within the policy process. While context determines the form of policy instruments, and, in turn, the approach to their evaluation, so evaluation can shape the formulation of policy instruments.

1.2.1. The Policy Context

Overarching policy frameworks will affect many aspects of activity, including the design and implementation of research and technology programmes, and the criteria by which their success may be judged. A policy orientation towards improving national competitiveness, or wealth creation may place emphasis on the industrial relevance of scientific and technological research, and consequently, the type of programme, its stakeholders and its desired outcome. Science oriented policies, may place greater emphasis on research excellence for its own sake, as well as its economic potential.

Beyond the national policy context, European research policy constitutes a further set of requirements. In the last decade, various countries in Europe were in receipt of structural funds, which played a dominant role in research funding in those countries and led to the formulation of suites of national projects with a built in requirement for evaluation. However, these have largely come to an end leaving fewer such programmes to be evaluated.

In Ireland, for example, with the end of the structural finds, policy for and the structure and operation of national funding for science have grown in importance, with a shift in evaluation activity from project outcome towards assessment of national competence in science and technology to provide a basis for funding decisions. This funding in turn has renewed commitment to economic evaluation to ensure public accountability.

The coexistence of national and Framework Programme funding raises questions of additionality in evaluating the impact of either, with consequent policy implications. New European policies for the realisation of the concept ERA may have a more complex interaction with national level activities, and raise various issues for evaluation (see Chapter on Future Policy Instruments).

Policy Trends, such as the increasing outsourcing of science and technology, as exemplified in the United Kingdom with the privatization of Government laboratories in

the UK, increase the number of stakeholders and generate a greater need for accountability, transparency and evaluation. As the numbers of organisations involved in policy programmes increase, the evaluation issues can become more complex. In turn the demands for appropriate indicators of policy and programme effectiveness increase.
A further phenomenon that has been observed in various countries has been the development of an evaluation culture in policy communities. Policy trends produce waves of varying intensity of interest and direction, or "fashions" in evaluation.

1.2.2. Domain Characteristics

The nature of the scientific and technological communities, and the activities they conduct, interact with the policy context, and thus the type of instruments that may be designed and implemented for their advancement or exploitation. The characteristics of a domain will also influence the formulation and implementation of evaluation studies.
In addition to the scientific and technological complexity that policy intervention must contend with, the size, organisation and functioning of the relevant practitioner communities are also important considerations. That other institutions, such as government and industry bodies, may be involved also shapes thinking. Certain domains may attract additional attention through ethical, legal and political dimensions to their activities. For example, Rip[10] acknowledges the pressures imposed on the process of R&D evaluation through the need to address the new stakeholders for research and notes the growing interest in the ethical, legal and social implications of R&D.
In evaluation terms, certain domains give rise to their own problems. As Barber notes11, such aspects of the character of a scientific or technological domain may pose problems for those commissioning or organising an evaluation. Evaluation of activity in the domain may require evaluators, and their advisers, to be sufficiently expert to carry out the evaluation, while remaining independent of, and without connections to, the people or organisation responsible for the inception, running and conduct of the programme concerned. While the use of evaluators and experts from overseas is a solution in some circumstances, certain specialist or advanced areas may already transcend national boundaries, through both formal and informal networks of communication in the domain.
While this is particularly serious when the focus of the evaluation is on the *ex ante* or *post hoc* assessment of scientific quality, strategic policy review and economic impact evaluations both require informed insight into the nature of the domain. While it is clear that all types of evaluation carry costs in both time and resources, expertise requirements may add to these.

1.2.3. Programme Design

Against the intertwined backgrounds of the policy context and domain characteristics, there are two key issues pertinent to the relationship between the selection, design and implementation of policy instruments and evaluation. On the one hand, evaluation, particularly the results of previous studies, can inform programme design, providing

---

[10] Rip, A. Challenges for R&D evaluation. (in prep.) – based on a presentation at the European – US Workshop on Learning from Science and Technology Policy Evaluation, Bad Herrenalb, Germany, 11-14 September 2000.

[11] Barber, J. (2000) "Making Evaluation Useful for Decision Makers" a paper presented at the European – US Workshop on Learning from Science and Technology Policy Evaluation, Bad Herrenalb, Germany, 11-14 September 2000.

knowledge and learning inputs. On the other, evaluation, either real-time or post hoc can be built into programmes at the design stage, both for accountability and impact assessment purposes and as part of an ongoing process of policy and programme design learning through evaluation.
Certainly users have placed value in the communication of programme outcome to those designing similar or related programmes, in addition to the wider policy interest of findings. Examples can be found of where errors have been repeated in subsequent programmes where this did not happen. However, the need to learn about programme design from evaluation is recognised. As there are frequently limited resources for evaluation, the Department of Trade and Industry (DTI) in the UK tends to concentrate its evaluation efforts on novel programmes to maximise opportunities for learning.

1.3. THE PURPOSES OF EVALUATION

While the design of programmes may explicitly acknowledge certain evaluation issues, either during project execution or after completion, the full range of issues may not always be anticipated ex ante. The policy context may change during the lifetime of the programme, while the research may change the very nature of the domain in which it is operating. This section examines the various types of evaluation issues that arise, the audience for evaluation and other the stakeholder interests that may influence them.

1.3.1. The Audience for Evaluation

The audience for evaluation may include some or all of the following:

- Policymakers, including both politicians and officials in administration;
- Programme managers, who may be acting on behalf of policymakers;
- Participants in the programme, conducting and implementing the research; and
- Other stakeholders, such as those representing industry and consumers.

While an evaluation may focus on the needs of one section of the audience, it may need to be of wider relevance, and should recognise differences in the levels of analysis required and the specific issues of interest.

1.3.2. Stakeholder Interests

The combination of the policy context, the characteristics of the scientific and technological domain, and the architecture of a programme will determine the organisations involved in a research programme, either directly in its design and implementation or indirectly through the impact and utilisation of its results. These may all be stakeholders in evaluation, and each may have differing expectations of the design and outcome of an evaluation.
The widening of the stakeholder community is also an interesting trend noted in science and technology policy[12] particularly greater public involvement, and particularly through interest groups and consumer associations, in S&T policy setting and in the assessment of research outputs. This may be particularly significant where there may be ethical dimensions to the research or its exploitation. Organisations commissioning

[12] PREST (2001) "Measuring and Ensuring Excellence in Government Science and Technology: France, Germany, Sweden and the UK" – a Report Prepared for Science and Technology Strategy Directorate of Industry Canada in support of the work of the Canadian Council of Science and Technology Advisers (CSTA).

evaluation of medical and related research provide an example where the involvement of stakeholders is encouraged.

1.3.3. Evaluation Issues

The following four main categories of evaluation activity can be distinguished:

- The selection of projects
- Monitoring project progress
- Examination of the management of programmes
- The outcome and impact of funding

The issues arising in commissioning, managing and conducting each of them are examined in turn.

1.3.3.1. The Selection of Projects

This form of evaluation tends to be carried out *ex ante*, although longer term, multi-phased programmes may be able to benefit from the results of the evaluation and monitoring of earlier phases. The key issues are to appraise the potential scientific, social and economic benefit of a programme or project, to ensure that benefits should cover the costs, to agree objectives and targets, and to define the necessary monitoring processes.

1.3.3.2. Monitoring Project Progress

The monitoring of the progress of programmes may address both management and technical issues, and may constitute a form of intermediate evaluation. As such it may address the issues outlined below, and may contribute to the final evaluation. As well as essentially being a tool for the effective management of programmes in line with stated objectives, or a method through which those objectives could be modified, it can also support policy requirements for the accountability of funding bodies and the transparency of the funding processes.

1.3.3.3. Examination of the Management of Programmes

While evaluation may be part of the management process, the management process can also be subject to evaluations, particularly in mid-term evaluations, where findings can be usefully implemented or transferred from final evaluations to future projects. This supports the idea of the learning dimension of evaluation, and contributes to the culture of the evaluation. Evaluation not only studies impacts, but also has impacts. As above, the dimensions of accountability and transparency are key evaluation issues and are related closely to the ways in which programmes are run.

1.3.3.4. The Outcome of Funding

This is the broadest ranging category, and can be subdivided into a range of impacts and output that are of interest to various stakeholders. Table 4 below categorises various types of these[13].

Table 4: RTD Outputs and Impacts

Scientific & Technological Output	Papers Patents Prototypes Products Processes Services Standards Knowledge & Skills	**Scientific & Technological Impacts**	New Knowledge Exchange of Knowledge Culture of Collaboration Network Formation Scientific Reputation Community Development
		Economic Impacts	Economic performance Industrial competitiveness Organisational Innovation Employment
		Social Impacts	Quality of life Social development & Services Control & care of the environment
		Policy Impacts	Economic & industrial development Follow-on Projects Regulatory Change Contribution to policies

In seeking to ascertain the scientific and technological impacts of a programme, it is logical to start with the effects of its codified outputs, particularly as there are numerous tools available to capture and measure these. Measurements of output generally do reflect the generation of new knowledge and its impact on the scientific and technological communities. However, the full scientific and technological impact can only be captured by looking at the full range of activities entailed by the programme. Any collaboration, both formal and informal, and other communication with the wider scientific and technical community may advance the area.

The classification of output and impact is also an issue of some concern to users, with measurements of quality and excellence required to demonstrate the successful outcome of a programme. For programmes that specifically target the achievement of excellence (see Future Policy Instruments Section) this requires consideration of the concept and measurement of excellence and thus to the setting and adoption of standards.

In general terms, economic impacts result from the commercial application of a programme's research results, and the benefits accrue to the company exploiting programme results, for the industry, users and even economy as a whole. However, while such economic impacts can be measured in theory, it may be difficult in practice. Less direct commercial applications are even more problematic to assess. Attribution can be a problem: while specific achievements can be linked to programme research, in commercial applications they may be combined with existing product technologies, and the results of research undertaken elsewhere and funded from other sources.

Programmes may also cause changes in the behaviour and organisational competencies of the participants. These include both internal and external effects, including the creation of new organisational forms to implement programmes and to exploit results,

[13] Adapted from Boden, M. and Georghiou, L. (2000), "Evaluation of Socio-Economic And Technological Impacts of the Ministry of International Trade and Industry (MITI) Advanced Robotics Technology Project" a report to the Japanese government conducted on behalf of the Ministry of International Trade and Industry, Japan.

as well as new or improved networking activities. However, such impacts are difficult to assess.

Social impacts may be linked closely to broader economic impacts, such as improvements in the quality of life. However, there may be more specific impacts, such as safer working conditions, or improvements to the natural environment.

Policy impacts can be classified into specific and general policy impacts. The former relate to the implications of the programme for policy relating to the area of research funded. The latter relate to its implications for wider policy, and the formulation and implementation of research programmes more generally. A further distinction can also be made between direct impacts on, and consequent changes to, the policy process and to the lessons for future policies and programme design. Accountability and transparency are also important aspects of policy impact.

Of course, not all impacts of funding are positive, and users may be interested in negative programme effects. For example, the directions of funding are a concern, particularly if any resources are diverted from other trajectories. Also of concern is that participants do not have bad experience of the project management and organisation that may discourage them from similar activities in the future.

If at the completion of a project a significant number of former participants discontinue research in the area funded may point to a negative impact, and the potential waste of resources. However, this must be balanced against the more general knowledge and learning opportunities the project afforded. Evaluation studies may also try to hypothesise on what might have happened in the absence of the programme, particularly whether the situation could have improved without the funding. This may only realistically be possible when there is the opportunity to look at the activities of firms that did not participate in the project.

All impacts positive or negative arise from the interaction between programme outputs and the economy or society into which they are diffusing. The question of the extent to which observed socio-economic phenomena are attributable directly to specific policy instruments rather than to other activities and funding sources requires careful attention, particularly in the presence of other influences and with the passage of time. Impacts, particularly indirect benefits, are actually more likely to develop in the medium-long term after the conclusion of a project. Indicators must be treated with some caution.

Additionality is a core issue for users of impact evaluations. However, it is not simply a question of whether research would have taken place at all without the funding. There is a spectrum of additionality, with funding leading to faster progress than would otherwise have occurred (partial additionality by acceleration), or funding leading to progress on a larger scale than it would have done without support (scale additionality). In national and European research projects there is also the possibility that, although the research would have proceeded without a particular means of support, it would not have been collaborative. This might affect scale and/or acceleration or add a significant qualitative element.

1.4. ORGANISATION OF EVALUATION

The way in which evaluation is organised can be seen in terms not only of the process by which policy users and programme managers articulate policy and stakeholders' needs into an evaluation study, but how the relevant unit of analysis and mode of enquiry is determined, and, in turn, appropriate methods and techniques employed.

1.4.1. The Evaluation Process

The process of evaluation depends on the issues to be addressed, the timing of the programme and the role and experience of those involved in commissioning and implementing such studies. These may not, of course, be the same part of an organisation, or even the same organisation.
The establishment of "in-house" evaluation units in a number of countries (some following the lead of the Department of Trade and Industry (DTI) in the United Kingdom, the Economic and Social Research Council (ESRC), or the establishment of national committees to oversee evaluation In France) is designed to oversee the evaluation process rather than to conduct evaluation *per se*. Rather, such units assist research funders within a ministry to evaluate their research programmes, by advising on the format for the preparation of terms of reference, organising the tendering process, managing the selected contractors, and helping to interpret and diffuse the results.
Evaluation units may also run rolling programmes of evaluation ensuring that all research programmes funded are considered for evaluation within predefined time periods, or may allocate evaluation budgets in such a way as to enhance learning potential and impact. They may promote their wider dissemination and utilisation. A further role may be to help establish follow-up procedures to monitor the implementation of evaluators' recommendations.
While the growing use of evaluation can lead to so-called evaluation "fatigue" users have reported a recognition of the utility of evaluation.

1.4.2. Use of Evaluation

When considering the use of evaluation it is common to differentiate between 'process use' and use of results. Process use is increasingly recognised as an important form of evaluation use where the process of undertaking the evaluation helps parties to clarify their thinking and share information. At the final stage in the process of evaluation, there are opportunities to implement results. Successful implementation can be seen as related to the following three main factors:

- Absorbability -The dissemination of the content and findings of evaluation studies requires good levels of awareness of the study among the audience, particularly where it combines various elements. The final report itself must be digestible and target its recommendations at an appropriate level: not too specific and not too general. It should be delivered in time for follow-on decisions, it may need later validation, and may have objectives linked to its timing and relevance to programme and policy cycles.
- Credibility - The credibility of evaluators should be taken into account in awarding the original tender, and should not relate to technical ability and proven quality and cost effectiveness, but to fairness and independence, and reputation. The results produced should be of high quality and founded on solid evidence and sufficient depth and breadth of coverage.
- Steerability - Some policy and programme initiatives are more open to being 'steered' by evaluation than others. Evaluation is only one among many influences on policy decisions. In some policy domains a high proportion of implements is located with policy makers and politicians and in others less so. The extent to which evaluations can make a contribution towards programme development and policy decisions is a factor that needs to be taken into account when decisions are made about funding

evaluations. What is the added value of an evaluation also needs to be considered in these terms.

1.5. Modes of Enquiry, Methods and Techniques

The translation of issues through the types of organising structures and processes outlined above, into evaluation studies, leads to more detailed methodological consideration, and thus to the consideration of the utility of the tools and methods described in subsequent chapters/sections.
In research design terms, the mode of enquiry reflects a broad methodological orientation which is associated with particular methods and techniques. Modes of enquiry are of six main kinds[14]. The choice of mode of enquiry for research design follows from the consideration of the kind of question being asked and its underlying rationale or intent. They may be:

- *exploratory*, i.e. identify the main issues or factors in a domain where little is already known or where new theories or hypotheses are being developed.
- *descriptive*, i.e. define and elaborate the main issues or factors already identified possibly as an input into theory building or to provide examples.
- *predictive*, i.e. to make predictions that will usually test out a theoretically derived set of relationships and anticipate future difficulties.
- *explanatory***,** i.e. to establish the precise links between different factors and variables, usually within some theoretically based set of assumed relationships.
- *prescriptive*, i.e. to suggest a path of action usually as part of a management or planning (operational) purpose of evaluation.
- *critical*, i.e. to question the normative basis of a phenomena, possibly with the intention of identifying ways of introducing change.
-

These six categories of mode of enquiry can be mapped on to the types of methods commonly applied in evaluation studies. These can be grouped into the following categories:

- Experimental Methods
- Survey Methods
- Field study/social anthropology methods
- Modelling
- Interpretative
- Critical
- Participatory

There are various techniques associated with, and which exemplify, each of these, as Table 5 illustrates. It also shows the links with the methods discussed in subsequent sections

[14] This typology derives from *Evaluation Guidelines Handbook for Learning Technology Innovation,* Volume Two (of Three). The ARTICULATE Consortium, Coordinators Tavistock Institute 1995.

Table 5: Methodology Typology Categorisation

Type of Method	Associated techniques
Experimental Methods	Post-test control group Pre-test post-test control group Solomon four group Factorial Time series Non equivalent groups
Survey Methods	Face to face interviews Focus groups Mapping techniques Questionnaire surveys Criterion tests
Field study/social anthropology methods	Observation(participant/non participant) Protocol/ critical incidents analysis Ethnographic techniques Physical trace analysis Case studies
Modelling	Game simulation Economic modelling Systems analysis
Interpretative	Content Analysis Oral history
Critical	Discourse Analysis Critical ethnography
Participatory	Action research

Table 6 lists the six types of methods, the types of questions they involve, and their modes of enquiry.

Table 6: Methodology Scope Taxonomy[15]

Type of Method	Type/Scope of Question	Mode of Enquiry
Experimental Methods	Causal/Explanatory Instrumental	Explanatory Prescriptive Predictive
Survey Methods	Instrumental Open	Descriptive Explanatory Prescriptive
Field study/ social anthropology methods	Open Normative	Exploratory Descriptive Prescriptive Critical
Modelling	Causal/Explanatory	Explanatory Predictive
Interpretative	Open Normative	Exploratory Descriptive Prescriptive Explanatory
Critical	Normative	Critical
Participatory	Open Instrumental	Exploratory Descriptive Prescriptive

[15] Op cit. ARTICULATE / Tavistock Institute 1995.

1.6. THE USE OF EVALUATION TOOLS IN THE EU

This section is concerned with the use and prospective use of tools to assess socio-economic impact and draws on interviews with programme managers and evaluation specialists in DG Research and documentation provided by the DG[16].

Any new evaluation methods need to be considered alongside existing evaluation approaches among potential users of evaluation. Current DG Research evaluation consists of a number of separate and (partly) linked activities, including:

- ‘continuous’ monitoring, which is conducted by programme managers and supplemented by panels of experts at annual intervals and focuses mainly on implementation;
- five year assessments which are conducted by independent panels of experts and focuses on the achievement of objectives of the individual programmes and the framework programme overall; and (in the 5^{th} framework),
- a mid-term evaluation, which considers objectives set and achieved to date on the basis of self-evaluation by projects and programme managers.

The existing system of evaluation has the following characteristics:

- It combines monitoring and evaluation and probably errs on the side of monitoring.
- It focuses on projects and programmes – grossing up to the Framework level from pre-set programme indicators and annual continuous monitoring.
- It concentrates on what the projects and programmes achieve, rather than on the wider context within which programmes are set.
- It is to a large extent standardised, using the same questionnaires and rating scales across all programmes.
- It continues to rely heavily on expert panels rather than on studies.
- There is little impact assessment and little emphasis at the policy level.

According to those we interviewed in DG research, the current monitoring/evaluation system is demanding in terms of time and resources. There is little time for customised, or programme specific evaluations. The legal framework that specifies the obligation to conduct ‘continuous’ monitoring and five yearly reviews has become the maximum rather than minimum. This is reinforced by the difficulties of issuing calls for proposals and contracts, which are widely seen as complex, time-consuming and a barrier to customised evaluations.

Several programmes have their own evaluation traditions. These are often well developed but not necessarily consistent with most recent developments in evaluation practice. BRITE EURAM is an example of a well-established evaluation system that has the advantage of continuity and comparability across Framework programmes. It is however very focused on project performance, gives priority to 15 indicators used by the Commission across the whole Fifth Framework and relies on self-reports by project informants to account for policy impacts such as employment.

Against this background there appears to be no ‘user demand’ from RTD managers and evaluators in DG Research for new evaluation tools or methods. Indeed new evaluation

[16] This section is based upon interviews undertaken by both authors and kindly arranged by the Commission with its staff of evaluation specialists and programme managers during March and April 2002. A total of seven senior staff was interviewed during the course of the research for this section.

tools would only be welcome if they could produce results reliably and with minimum effort. On the basis of what we have seen we would suggest that:

- There is unlikely to be much user take-up for new and innovative methods at Commission level unless adequate and additional time and resources is made available to key managers and evaluation specialists.
- In order to shift demand and expectations in the Commission towards socio-economic impacts, there would need to be a new evaluation and monitoring framework that obliged a greater effort at impact level, rather than monitoring of outputs and impacts.

However irrespective of Commission internal arrangements, we are still constrained by the state of knowledge – and state of the art - within the evaluation community.

1.7. THE DEVELOPMENT OF AN EVALUATION TOOLKIT

Any toolkit for assessing impacts needs to start with what is known and understood within the particular evaluation community. What follows therefore is a brief listing of some of the problems that would need to be addressed for a toolkit project to be credible.

- Different evaluators are firmly embedded within their own paradigms. The models that they adhere to shape what are defined as impacts. Unless experts stop misleading policy makers by claiming the superiority of their models, they will continue to cancel each other out and undermine each other.
- Attribution problems are manifest in RTD impact evaluations. We cannot on the basis of beneficiary self-reports know whether employment, environmental, health and other socio-economic outcomes would not have happened anyhow. There is probably no adequate theory to offer explanations except in piecemeal parts of programmes.
- Changes in socio-economic outcomes are complex and cannot be simply understood in terms of programme effects. Contextual, comparative and sectoral studies are also needed.
- Time-scales of impact are often not consistent with the expectations and needs of policy makers for evaluation. Data needs to be collected and analyses undertaken over the long term.
- Choices of methods follow from the identification of evaluation questions and of the kinds of impacts judged important.
- We do not have sufficiently developed models that explain the relationship between RTD, other socio-economic processes and mediating factors and socio-economic impacts such as competitiveness, innovation, public health and quality of life.

Overall the most useful message that EPUB could communicate as an aid to greater understanding is that our ability to measure impacts is necessarily limited by:

- the limitations and incompleteness of the data we have available
- the limitations of the models and theories we can deploy
- the policy community’s wish for short-term answers to long term questions
- the indeterminacy and complexity of socio-economic phenomena
- the difficulties of grossing up from projects to programmes and to Frameworks of multiple programmes
- the discrepancy between the resources that would need to be deployed to begin to answer complex questions and what is practical to expend

The least useful message that EPUB could offer is that:

- impacts can be fully measured by the deployment of known methods
- that the problem is only the dissemination of methods rather than the simultaneous development of frameworks, theories and models
- that stakeholders are motivated by technical answers to questions that are essentially political and institutional
- that we can do evaluation on the cheap
- that we can answer policy questions by aggregating project and programme data.

Despite these caveats the remainder of this volume demonstrates the considerable body of expertise that has been developed within the evaluation community in general and in particular that part of the evaluation community that is concerned with RTD impacts. We believe that evaluation experts should be able to use this toolkit selectively to guide policy makers and programme managers through the labyrinth of what is and what is not possible to evaluate. The toolkit should also enable policy makers and programme managers to be made aware of what they can reasonable ask evaluators to deliver, taking into account the instruments that they are using, the policy context within which they are operating and the evaluation questions that they wish to ask.

2. POLICY INSTRUMENTS

Authors: **Wolfgang Polt** and **Jaime Rojo** (Joanneum Research)

2.1. INTRODUCTION

There is nowadays a broadly shared perception of technological progress being the main contributor to long-term economic growth and consequently to the improvement of living standards and the quality of life. In this respect, R&D activities are one major source for producing the knowledge and ideas required to sustain technological progress. Theoretical and empirical models have indicated the relevance of technological progress in productivity growth. However, there is still need to improve the current understanding of the process of knowledge production, the connections between R&D, technology and diffusion, the magnitude of the contribution of technology to economic growth and the role of public intervention in the advance of technological progress.

Government support to R&D is generally justified on the grounds of existence of a market failure which leads to an under-investment in R&D. The market failure appears in situations where firms would perform less R&D than is desirable from a social perspective, that is, in situations where due to imperfections in the market allocation mechanisms, the market will fail to allocate an efficient or socially optimal quantity of resources to the R&D. Theoretical and empirical research has shown the existence of a market failure in R&D provision. The main rationales for government support to R&D include:

- *Positive externalities or spillovers*: This results when the benefits from R&D activities are not fully captured by the R&D performer. Part of the benefits from performing R&D spillover to other individuals, firms or even across economies. This is due to the public good characteristic of knowledge and ideas, namely non-rivalry and non-excludability. In this respect, the empirical literature has found evidence on social rates of return to R&D exceeding the private ones and on the existence of significant spillovers to R&D.
- *Risk and uncertainty*: The large risk and uncertainty involved in R&D activities might deter private investment.
- *Network externalities*: The benefits of the technology increase with the pool of adopters.
- *Asymmetric information*: This situations appear in agent-principal relationships where one of the parts in the transaction the agent, has more information than the other, the principal.
- *Indivisibilities*: the large investments required to produce results might prevent firms from conducting R&D.
- *Evolutionary approaches*: technological progress benefits from enhanced competition and R&D diversity. Agents when making economic decisions are subject to bounded rationality which leads to the common pattern observed in R&D of path dependency and lock-in effects.
-

2.2. MECHANISMS SUPPORTING R&D

Most countries have introduced mechanisms to support R&D to tackle the different types of market failures. Although the ultimate purpose of most of the instruments is to enhance technological progress, living standards and quality of life, the available instruments differ on their direct pursued objectives. The policy instruments described in the toolbox are framed within the context of the methods used for the evaluation and assessment of their impact. They have been grouped into four broad categories:

- *Financing R&D*: These interventions aim to compensate firms and individuals conducting R&D for the spillovers to society these activities generate and that they cannot appropriate. Main interventions in this area include direct subsidies where the government has to select the type of R&D conducted and indirect tax incentives where the firms select themselves the type the of R&D they want to conduct.
- *Provision of R&D infrastructure*: The government uses direct intervention in those situations where the market incentives are weak and spillover benefits are likely to be large and pervasive across sectors. The most commonly used measures include direct support to R&D infrastructures and provision of government sponsored R&D to PPPs in the formatation of infrastructures (e.d. networks of excellence).
- *Technology Transfer and Innovation Diffusion*: The policy intervention intends to increase social welfare by stimulating the diffusion of knowledge and the transformation of research results into commercial products. The most commonly used measures are schemes stimulating co-operation, RJV, science-industry collaborations, mobility of researchers, spin-offs, etc.
- *Legal Framework*: Interventions in this field pursue reducing existing market failures in the provision of private R&D. In the field of intellectual property rights (IPRs), intervention enhances the private incentives for conducting R&D by allowing firms to exert partial excludability for the use of the knowledge produced in the R&D process. The regulatory framework and the setting of standards also affect the path of technological development.

The purpose of table 7 below is to assist the reader in the process of browsing through the contents of the RTD evaluation toolbox. It provides a simplified matrix matching the categories of public RTD policy instruments available with the methods adapted to evaluate their socio-economic impact.

Table 7: Evaluation Matrix: Matching policy instruments and methods

	Innovation Surveys	Econometric Models	Control Group Approaches	Cost Benefit Analysis	Expert Panels/ Peer Review	Field / Case Studies	Network Analysis	Foresight/ Technology Assessment	Benchmarking
Financing R&D	●●●	●●●	●●●	●		●●●		●	●
Provision of R&D infrastructure		●●		●●●	●●●	●●●	●●●	●●	●●●
Technology transfer/ innovation diffusion	●●●	●●●	●●●	●●	●	●●	●●●	●●●	●●●
Legal frameworks (IPRs, standards and regulation)	●	●	●	●●●		●●●		●●	●●●
Integrated projects			●	●●●	●●●	●●●	●●●	●●	●●
Networks of excellence					●●●	●●	●●●	●●	●●

Legend: ●●● *Highly suitable* ●● *Suitable* ● *Less suitable*

2.2.1. Level of Intervention

Although evaluation can be applied at any level of policy intervention, the following distinction is generally applied:

- Policy: a set of activities which may differ in type and may have different beneficiaries, directed towards common general objectives or goals.
- Programme: a set of organised but often varied activities bundled together –for example, projects, measures and processes– to achieve a common objective.
- Project: a single intervention with a fixed time schedule and dedicated budget.
- Thematic: is centred on a common objective pursued by several programmes.

2.2.2. Data Requirements

The objectives define the expected effects of the intervention. Once the objectives have been defined, indicators allow to evaluate the performance of the intervention and establish if the intervention is progressing towards meeting the defined objectives. An indicator is an objectively verifiable measurement which reflects the activity or effect being measured, allowing comparisons across different populations or individuals, and in time. Indicators are by definition imprecise measurements of the underlying concept of interest. The set of available indicators used in evaluation include:

- Input: are the resources consumed in the implementation of an intervention.
- Output: are the goods and services directly produced as a consequence of the intervention.
- Outcome: are the initial impacts of the intervention providing the reason for the programme. Outcomes tend to be less tangible than outputs.
- Impact: are the long-term socio-economic changes the intervention brings about.

Table 8 provides a useful illustration of the concepts and proxy variables available to capture the knowledge produced to measure the impact of R&D activities.

Table 8: Conceptual framework for knowledge measurement

Indicators	Concept	Proxies
Input	Persons-year, equipment and machinery-years	Expenditures
Output	Ideas, knowledge, invention	Publications, patents, prises
Outcome/Impact	Advance of knowledge base	Papers, citations expert evaluation
	Societal improvement	Surveys and case studies, life expectancy
	Reduction of costs	Studies, statistical analysis costs and expenditures
	Economic output	Profit, revenues growth, revenue from new products
	Performance improvement	Productivity studies

Source: Adapted from Adam Jaffe (1998) Measuring Knowledge in the Health Sector, OECD/NSF High-Level Forum

The design of the policy intervention is the framework that permits to explain the expected achievements of the intervention and how the intervention its supposed to achieve its pursued objectives. In the evaluation process there is a need to establish how the inputs lead to the output and how these outputs subsequently lead to the outcome and impact that is expected from the intervention. One relevant aspect of the evaluation process is to identify and analyse the implicit assumptions and causal linkages behind the policy intervention.

2.2.3. Operational Steps for Conducting an Evaluation

2.2.3.1. Ex-ante Evaluation

The key aspects to assess the potential socio-economic benefit of policy intervention include:

- Establish a firm justification for public intervention, identifying the rationale and market failures addressed, and discussing how and why public intervention is appropriate.
- Introduce the modelling approach and the assumptions on which the projections are made, as results might be very sensitive to the modelling assumptions.
- Analyse the counterfactual based on constructed scenarios. What would happen with and without the project.
- Present cost-benefit or cost-effectiveness analysis comparing the various alternatives available. Evaluate alternatives of with project scenarios and with-out project scenarios. Provide indications based on the scenario projection on the costs considered and the benefits of the project.

2.2.3.2. Ex-post Evaluation

The key aspects to conduct an ex-post evaluation of the socio-economic impact of a policy intervention include:

- Provide a clear specification of policy objectives.
- Define data collection at the programme design stage.
- Design the evaluation approach, with possible interaction with the evaluation sponsor.
- Control for the counterfactual, that is what would have occurred had the project not taken place.
- Compile evidence on success and failures.
- Provide clear results and recommendations when writing the evaluation report.

2.2.3.3. Evaluation Design

Evaluation design allows evaluators to quantify the magnitude of the effects of a policy intervention while eliminating other competing explanations of the experimented change which are not related to the policy intervention. The two main alternatives in evaluation design are:

- *True experimental designs:* although appealing for the straightforward interpretation of the provided results they are rarely used in policy evaluation. This approach uses random selection to define programme participants (treatment group) and non participants (control group), but its implementations often finds resistance due to ethical, political or operational reasons.

- *Quasi experimental designs:* Quasi experimental designs are commonly used in evaluation and provide confident conclusions on the effect of the programme when the selection method for the comparison group is carefully implemented. The application of the method tries to approximate as much as possible the true experimental conditions. Besides their difference relating to participation in the programme, the characteristics of the comparison group should be as similar as possible to the treated group.

2.2.3.4. Evaluation Implementation

The baseline data information is a key element in every evaluation. The baseline information requires to collect information referring to the situation before the policy intervention both on the participant and control (comparison) group. It will permit to measure with confidence the magnitude of change produced by the intervention. In simple before-after estimations, it allows to analyses the effect of the intervention by comparing post-intervention to pre-intervention status on selected indicators.
A relevant step in the preparatory procedure for conducting the evaluation is to establish the sample size and the sampling design method. The sample size is required because most projects are too large to permit the evaluation of the effects in all the participants. It will depend on the numbers of groups being studied, the amount of change expected in the indicator, confidence level required in the conclusions and the probability of detecting a difference in the indicators when one exists. The sampling method is required for selecting the individuals that will be used in the evaluation. The evaluation through the use of the control groups should also control for confounding factors, namely factors independent of the policy intervention that explain the changes observed in the supported individuals.

2.2.3.5. Data Collection Techniques

There are various methods for collecting data for evaluation purposes. Case studies and interviews, offer rich detailed data on the context in which an intervention is conducted and the problems it might solve. On the contrary, surveys provide quantitative data to measure the effects of the intervention. As reflected in table 9 each technique has its own strengths and limitations. The adoption of appropriate approaches in data collection might reduce the data collection bias.

Table 9: Strengths and limitations in data collection

Data method	Strengths	Limitations
Interviews and case studies	quick implementation and low costs provides rich contextual information reveal project issues originally not thought	difficulty to code and analyse responses to open-ended questions difficulty to compare across interventions conduct of interviews requires expert staff
Surveys	capture information on inputs, outputs and impact allows to analyse broad range of issues information provided is easy to analyse possibility to generalise to the population if sufficient information	oversimplification of the process as a result of closed-ended questions interview bias expensive and time consuming

2.3. FINANCING RESEARCH AND DEVELOPMENT

Authors: **Eamonn Kinsella** (The Circa Group), **Georg Licht** (ZEW), **Pari Patel** (SPRU) and **Giorgio Sirilli** (CNR)[17]

2.3.1. Introduction

This section refers to ex post evaluation of public funding of R&D, and excludes funding provided by the private sector, charitable institutions or foundations. It covers R&D funding provided directly and indirectly to the private sector, as well as to the public sector. The term "R&D" covers a wide spectrum from very basic research to innovation, the latter taking a product, process or management technique to the market place. Generally speaking, basic research is carried out primarily in universities and institutes, with the more applied R&D being carried out through increasing public-private collaboration or – more often – by private firms or private R&D partnerships[18].

The problem here is not only semantic: quite often those who finance R&D expect more or less implicitly that the R&D results will be very quickly turned into solutions to societal problems, and therefore may impose objectives which are unrealistic. This is often justified by the necessity to legitimise R&D expenses which, if not promising the achievement of socially valuable results, would not receive public support. As a consequence, the evaluation of R&D financing becomes more difficult and debatable.

The nature of the most appropriate financing instrument varies to reflect this progression: subsidies and grants are used for basic research, whereas more applied research may also be assisted with tax incentives. Expenditure on company R&D is deductible from corporate tax, as is any other expenditure, but tax incentives allow this to be deducted at a higher rate. (In some versions the pure tax incentive is combined with a subsidy element for those firms, which have negative profits (e.g. start-ups), or the tax credit is made tradable so that the tax credit can be sold to other firms). Clearly, universities and private not-for-profit organisations (PNPs) cannot benefit from tax incentives. Venture capital may be used to assist innovation and particularly risky projects, and there is public provision of venture capital. In addition, governments can give a guarantee to venture funds (reducing the risk to the venture fund), use an interest subsidy to venture funds (making refinancing for the venture firm cheaper) or implement favourable tax treatments of investment in venture funds by private households or firms (e.g. corporate venture funds). Sometimes, these subsidies only refer to venture funds primarily investing in high-tech start-ups.

The EU Trend Chart provides further extensive information about national programmes for financing innovation activities in the private sector. A large variety of programmes can be funded in the member states. In addition, more or less all government R&D programmes aimed at stimulating R&D and innovation in the private sector involve an R&D financing dimension (e.g. project grants to stimulate biotechnology research or implement new production technologies e.g. CIM). Although most of the following thoughts can be applied to a wide variety of programmes, we primarily have in mind

[17] The authors thank Nikolaus Gretzmacher and Andreas Fier from Joanneum Research for valuable inputs to this section.

[18] In more general terms, R&D is defined in the OECD "Frascati" Manual as "creative work undertaken on a systematic basis in order to increase the stock of knowledge, including knowledge of man, culture and society, and the use of this stock of knowledge to devise new applications". This means that various other activities which may – or may not - be involved in the invention/innovation process are outside the realm of R&D. The OECD "Oslo" Manual, which addresses the measurement of innovation in firms, envisages various activities as well as R&D: design, know-how, engineering, acquisition of machinery, training, and marketing. Furthermore, at the EU level the term RTD (Research, Technology and Development) is currently used: this acronym covers, stricto sensu, technology as well as R&D. It is not clear if, in practice, the two terms (R&D and RTD) coincide.

programs, which place special emphasis on the financing dimension. Even more, we primarily concentrate on financing R&D in the private sector. However, general remarks made also refer to evaluating financing R&D of public or semi-public sector research facilities.

Depending on the nature and objectives of each method of financing R&D, the impacts to be evaluated, and, to a lesser extent, the approaches to evaluating them, will vary. An issue in all financing of R&D is ownership and exploitation of Intellectual Property Rights (IPR) resulting from government funded projects. Revenues from IPR can be very significant, making R&D, which generates IPR, sometimes more important than bringing the product to the market place (e.g. in the case of many biotechnology start-ups). These revenues are an important impact of financing R&D[19].

2.3.2. Policy Objectives, Instruments and Outcomes

The public good nature of R&D investments has attracted economists and policy makers' attention for decades. That R&D activities provides outputs that are at least partially not excludable and non-rivalrous was pointed out by Arrow (1962) and recently stressed by Romer (1990) and Griliches (1998), among others. To a large extend under-investment in R&D occurs, because the social benefits from new technologies are difficult to appropriate by private firms bearing the costs of their discovery, and because imperfect capital markets may inhibit firms from investing in socially valuable R&D projects (see Lach, 2000). An aim of public R&D funding is to support public good research and to overcome market failures. There are arguments in favour of limiting public financing very strictly to those cases, as the internal rate of return on research funding is very high. In addition, a further rationale for government intervention in order to enhance financing possibilities of SMEs or young firms (as stated in some programmes) is imperfection in the capital markets resulting e.g. from asymmetric information. According to economic theory, there are many different options available to deal with market failure due to externalities, such as tax credits, subsidies, extending property rights and public demand.

Based on European experience in financing R&D, the funding areas of most member states are comparable. In general ministries and government departments use the following instruments to improve scientific-industry links or to stimulate private R&D activities:

(a) Project funding (selected R&D projects, thematic competitions)

For direct R&D project subsidies the administration chooses ex-ante the R&D project to support. In terms of technology policy this instrument qualifies as the one with the highest steering effect. The administration needs to have full market information to avoid influencing future technological standards. The problem of asymmetric information can be partly reduced by letting the firms bid for the indirect subsidy in a competitive environment to ascertain the firm's internal valuation of a project. Direct R&D subsidies in Europe are comprised by project grants, project loans at subsidised interest rates, loan guarantees and prizes.

[19] DG Research is about to commission a study on international practices in IPR issues.

(b) Basic funding (statutory) funding of institutions (including the funds allocated to federal institutions performing R&D)

Institutional funding is designed to provide a long-term funding for basic research. The allocated subsidies are not intended to support R&D on a project level, but rather to contribute to basis research in universities and non-university research institution. Tasks and goals are neither set by the policy maker nor by the administration. It is the decision of the institution which projects will be carried out. The business R&D sector benefits of know how transfer of universities and other research institutions as well as their technical capabilities (equipment and machinery).

(c) Tax incentives or subsidies for extramural R&D

In the last decade an increasing number of countries use tax incentives to stimulate R&D activities of private firms (e.g. Japan, USA, Canada, Australia, UK, France, Netherlands, Spain). In principle these tax incentive provide additional financial means to company when they perform and/or extend their R&D expenditure. The primary impact of tax incentives is that the cost of R&D is reduced absolutely and with regard to other types of investment by reducing the opportunity costs of financing R&D.

There are different variants of tax incentive schemes. In most countries tax incentives are implemented within the corporation tax. Tax incentives can take the form of (1) an extra tax allowance, which enables firms to deduct from the tax base more than 100% of their R&D expenditure (e.g. 150% in Australia, 120% in the UK), (2) a tax credit, which enables firms to deduct a percentage of the R&D expenditure from the tax bill, or (3) an accelerated depreciation of investment in equipment linked directly to R&D projects. Some countries also use a mixture of these measures.

In some cases the tax incentive is restricted to R&D expenditure, which are above a defined, company-specific R&D base (tax incentive refer to the increase in R&D expenditure only = incremental tax credit see US or France). In other cases the tax incentives refer to the total sum of R&D expenditure. Some countries implemented a more generous R&D tax treatment for small firms by restricting the firms that are eligible for the tax incentives (e.g. UK before 2002, Italy) or by setting a ceiling for the absolute values of the tax incentives (e.g. France). A special type of tax incentive is implemented in the Netherlands using the wage tax and social contribution system to profit financial incentives for R&D expenditure.[20]

At the first sight R&D tax incentives are attractive as they are reach a large number of firms with low administrative cost incurred by the government and the firms. However, tax incentives are also criticised with regard to a low degree of additionality and the openness with regard to redefinition of what is regard as R&D inside the firms. Hence, their is a strong need for comprehensive evaluation of this type of financial R&D support. Various evaluations of tax incentive were undertaken in different countries (e.g. US, Canada, France, Netherlands, Australia). The majority of these studies conclude that overall it is more likely that the net social benefit of tax incentives is positive than negative. But there seems to be a high degree of variability with regard to these positive outcomes. The impact of tax

[20] A recent EU study has compiled an overview of the different systems refering to fiscal incentives for R&D in member states and the US see EU Commission (2002): Corporation Tax and Innovation, Innovation Papers No. 19, Luxembourg.

incentives differs by firm size (In general small firms show a high responsiveness with regard to a tax credit) and other firm characteristics depend on the time period (e.g. business cycle), or the concrete details with regard to the implementation of the scheme. Sometime implementation details interact with firm characteristics resulting in additional complexity. Hence, there is a clear need for more careful evaluation of tax incentive. Especially, in the light of increased use of tax incentives as an this instrument for financing R&D in member states.

(d) National/international co-operation, research consortia

Today, research results and new technologies are developed in international networks or research institutions and enterprises, in a mixture of competition and co-operation. International research consortia not only increase efficiency, they also strengthen political coherence and understanding and support the integration of developing and newly industrialised countries into the global economy.

Tax concessions or general subsidies for extramural R&D strengthen existing internal or extramural R&D activities, but have small effects on R&D co-operations and start-ups. Because of this, a broad field of "innovation"-funding activities arises in recent years. To direct the research process towards the innovation goal and to create a co-operating network of researchers, developers and users, "competitions" and "networks" are sponsored. Moreover, the R&D-financing circumstances for start-ups and SMEs were improved: easier access to venture capital, guarantees for equity participations or seed capital programmes.

The set of objectives of any R&D activity, and the most appropriate instruments for financing that activity, vary with the level of the R&D, i.e. from basic to applied or targeted. Basic research is usually associated with public good activities or other market failures, and is not expected to directly produce added employment or increased industrial competitiveness. Applied research and technology transfer are closer to the market, and do affect competitiveness and employment. "Financing" is generally by subsidies or grants. In some cases governments take "failure-risks" if private companies invest in high-risky R&D projects or high-tech start-ups. If the project resp. the company-share fails, the government reimburse some percentage of the total expenses. In some countries financial support take the form of tax incentives. Hence, policy objectives and outcomes are associated with programmes not instruments, but instruments tend to be associated with certain types of objectives.

Funding levels are related to the nature of the R&D and to the type of organisation being funded. There are well-defined EU rules on this point, covering funding to 100% of additional costs plus 20% for overheads for universities, and to 50% of total costs for the private sector. These rules apply to EU and to national funding alike for subsidies and venture capital.

2.3.3. Evaluation Methodologies and Good Practice Examples

A financial policy instrument is administered under different programmes for industry, health, food, marine, agriculture etc, at regional, national or EU level. Some instruments, such as tax incentives, are not available to the EU, and the use of others is restricted at national level by the EU. It is important to evaluate not only the impacts of an instrument, but the design and operation of its host programme insofar as the evaluation of the instrument is concerned. This is because the effectiveness and efficiency of an individual instrument, such as an R&D subsidy, are affected by the overall characteristics of the programme under which it is carried out, such as scope, relevance, administrative procedures and "user friendliness". Some of the points for the evaluation include:

- The clarity of objectives and targets, and the extent to which they were met.
- The nature of actual or anticipated benefits to participants, at national or industry level and at EU level.
- The size and appropriateness of the programme's budget.
- The efficiency and effectiveness of the programme promotion, project selection process, implementation and monitoring, including ease and rate of draw down and "user friendliness".
- The extent to which deficiencies identified in precursor programmes were eliminated.

An evaluation at national or EU level must be built up from evaluations of individual projects at the firm level to evaluation at the programme or economic sector level. These may be incorporated to give national or EU evaluations. While the objectives of different programmes may be radically different, the evaluation methodology can be sufficiently robust to cope with many programmes with little change. Equally, different financial instruments, which are means to attaining objectives, can be evaluated in the same way.

The methodology for evaluating an instrument and its host programme will then come to four steps:

- Review of the programme by collecting, analysing and synthesising data under the indents above, for example for a Cost Benefit Analysis. This will involve interviews using structured formats with policy makers, programme managers and beneficiaries, face to face, by telephone or by post. The importance of control groups for the last category is discussed in toolbox section 4.6.
- Review of an alternative instrument or programme for attaining the same objectives for the same sector of the economy in the same field, if there has been one. For example, compare the effectiveness of subsidies with tax incentives in increasing employment.
- Review of published and grey literature from other countries in, for example, the OECD publications.
- Benchmarking the instrument or programme if feasible with one in another country from which lessons can be learned (see toolbox section 4.12 on benchmarking).

The evaluation report will comment on the success of the programme and the appropriateness of the instrument used. It may be that the instrument was well suited, but that the performance of the programme was limited by poor management. The report will make recommendations for the future, based on the foregoing.
Any attempt to evaluate a financial instrument or programme in isolation, without reference to controls and comparisons would have to be regarded as bad practice. In practice, the available budget may preclude studies of alternative instruments, benchmarking or even literature searches.
Evaluations of two publicly-supported R&D programmes for industry have been carried out in recent years in Ireland, but neither has been published. Both studies covered many of the aspects described above.

(A) In 1999 the benefits to industry and to the economy deriving from **Ireland's participation in the R&D programme of the European Space Agency** were examined. A control group approach was used. Particular attention was paid to:

- Identifying the full range of effects of participation on enterprises and the economy
- Defining a methodology for measuring quantifiable benefits and non-quantifiable effects on enterprises
- Estimating the costs to firms of participation
- Identifying and estimating fiscal benefits to the economy
- Advising whether the subscription to the ESA was at an appropriate level, and represented value for money, in comparison with investment in other national instruments promoting industrial development
- Comparison with the benefits of ESA participation in other countries
- The operation of the instrument, including overhead costs, the effectiveness of promotion, the relation between firms and the national administration, and the ESA
- The relevance of national policies and strategies in relation to the ESA R&D programme

(B) The **Research, Technology and Innovation initiative (RTI)** was evaluated in 2000, using a control group approach. The main points addressed were:

- The quality of the proposals received
- The profile of the successful applicants
- The nature of actual and anticipated benefits to the participants
- The efficiency of the initiative's operation
- The effectiveness of the initiative's operation
- Improvements over preceding instruments of a similar design
- Performance judged against the objectives of the instrument
- The role of the instrument in the development of Irish industry
- The appropriateness of the size of the budget, support rates and alternative funding opportunities

(C) As to the evaluation of tax concessions, the Australian Bureau of Industry Economics published an **Evaluation of the Research and Development Tax Concession (BIE Research Report)** in 1993

This evaluation of the Australian tax concession scheme is conducted by the Bureau of Industry Economics (BIE) a independent research body within the Australian Government. The BIE is a fore-runner of the Productivity Commission which is now the Australian Government's principal review and advisory body on microeconomic policy and regulation (see www.pc.gov.au). This report represents best practice with regard to various dimensions of the implementation of this evaluation studies. It clearly states the general and specific objectives of the 150% tax concession for R&D in Australia and provides an empirical operationalisation for these objectives. The report discusses the environment of the scheme which might affect the effectiveness of the tax concession (e.g. other R&D support schemes, development of R&D intensity, competitiveness issues). This evaluation uses different methodological approaches and also discuss in detail the pro and cons of these approaches with regard to their reliability towards reaching firm conclusions for policy. Finally, this evaluation report clearly states concrete suggestions for improving policy and, hence, makes obvious the value of evaluation for the development of R&D policy schemes.

The Australian tax incentives consists of a 150% tax concession on all forms of R&D expenditure (current expenditure as well as investment) as defined in the Frascati manual and comprise software R&D connected to software sold in the market place.

There was minimum threshold which R&D spending must surpass (20 000 Australian Dollar). The maximum concession is granted if R&D expenditure exceeds Australian Dollar 50000. Firms must register with a central office to be eligible for the tax concession. Australian and foreign firms can apply of the tax concession provided that the R&D activity is performed in Australia.
The primary goal of the measure is to increase business R&D spending in order to increase innovation in private firms, stimulate the collaboration with Australia's public R&D infrastructure, strengthening the absorptive capacity of Australian firm for using foreign technology, and hence, increasing the competitiveness and innovativeness of Australian manufacturing industry.

The evaluation is based on three different data set:

- Time series of R&D expenditure at the firm-level taken from the R&D survey of the Australian Bureau of Statistics (Frascati-type survey).
- A merged data set consisting of R&D survey information and data taken from registers about those firms which have been applied for the tax concession.
- A special survey conducted by BIE comprising firms registered for the tax concession

Statistical methods applied comprise (1) before-and-after comparison at the firm level in order to check the incentive impact of the tax concession as well as (2) comparison between a control group (firms which did not applied for a tax concession) and the treatment group (firm which use the tax concession). Finally, parameter estimates based on this two comparisons as well as parameter estimates about the magnitude of national and international spill-over effects of R&D is then fed into a cost-benefit analyses comparing the impact of the scheme with the marginal costs (excluding administrative burdens).
Empirical modelling proceeds in two steps. First, the investigation looks at the impact of the scheme on R&D expenditure at the firm level. Second, a relation between R&D input and R&D output (sales share of new products) and competitiveness (firm growth) is estimated.
The main results are: In response to 1 Dollar taxes forgone firm increase their R&D expenditure between 0.6 and 1 Dollar. Hence, the scheme contributes to increased R&D spending of the Australian economy. However, only a minority of firm in fact increase their R&D spending. In addition, the scheme do not contribute to a more permanent R&D behaviour of Australian firms.
Recommendation deals with the question whether a marginal tax credit will be more effective in terms of the transfer payment needed to reach the same increase of R&D spending. In order to implement a marginal system BIE suggest to introduce compulsory tax consolidation of Australian enterprise groups. BIE suggested that the government should launch a marginal tax credit as an experimental scheme. BIE also suggest to restricted the ability to carry-forward or carry-backward of unused tax concession. In order to increase the participation rates of SMEs BIE suggest to decrease the minimum threshold for eligibility for the tax concession. BIE also suggest not to extended the tax concession of non-R&D innovation expenditure because of the increase danger of redefinition of various non-R&D activities of firms and the large amount of taxes forgone when extending the tax concession.

2.3.4. Data Requirements

Data requirements depend strongly on the type and goal of the government intervention. The data requirements relating to the input indicators and to the programme or instrument descriptors can all be easily and therefore cheaply obtained from management files: if they are incomplete this is in itself a comment on the management. Some of the data requirements for the output indicators can be obtained from periodic reports by beneficiaries to the instrument or programme managers, but these are often incomplete. In those cases, and anyway for the non-quantifiable data, interviews with project leaders or the participants' financial officers are required. Dead-weight, which is an important indicator of the effectiveness of an instrument or programme, may be defined as support for projects which cannot, or do not, contribute to at least some of the objectives of the programme. Face to face interviews, perhaps with guaranteed anonymity, are generally necessary to determine the level of dead-weight attached to a programme. Interviews at project level are expensive, even when conducted remotely. The methodology selected to obtain the data will depend on the evaluation budget. This will determine the size and stratification of the sample chosen, if it is not the universe. Considerable attention to data collection formats can be justified.

2.3.4.1. Input Indicators

- Budget, of which % disbursed and % overhead (of which promotion, selection, management, follow up)
- Instrument or programme staff resources, individual work loads

2.3.4.2. Programme or Instrument Descriptors

- Participant typology by sector, size, ownership, area of RTD, region
- Numbers of first time RTD performers
- Type and size of project
- Total R&D budget and turnover of the firm or institution in order to determine the relative significance of the subsidy to the beneficiary

2.3.4.3. Output Indicators

- Publications
- Patents
- Income from IPR
- Numbers of trained researchers (PhDs)
- Staff exchanges, visits, new collaboration and JVs, public-private co-operation

2.3.4.4. Impact Indicators

- Employment created and maintained
- Changes in graduate employment
- Continuation of RTD after expiry of the financial support
- New products and processes (inside and/or outside the traditional product range), increased added value, shorter time to market
- Increased revenue, profitability

- Improved competitiveness, market position, reputation with peers (non-quantifiable, but very important)
- New markets, especially exports
- Change in total R&D budget of the R&D performer

The above information stems typically from the programme management. It must be recognised that there may be a considerable period between the outputs of an instrument or programme, and its impacts. Other data, basically of a statistical nature, such as those available at the OECD and Eurostat can also be used.
There is a clear need to reflect in the collection of indicators the goals of the programme. Depending on the evaluation approach used data are needed not only for programme participants but also for non-participants. In addition, data collection should be based on a conceptual approach to the innovation process which reflects a theory-based picture of the transformation of the government subsidy to an innovation output and/or the relation of public money to other sources of innovation financing.

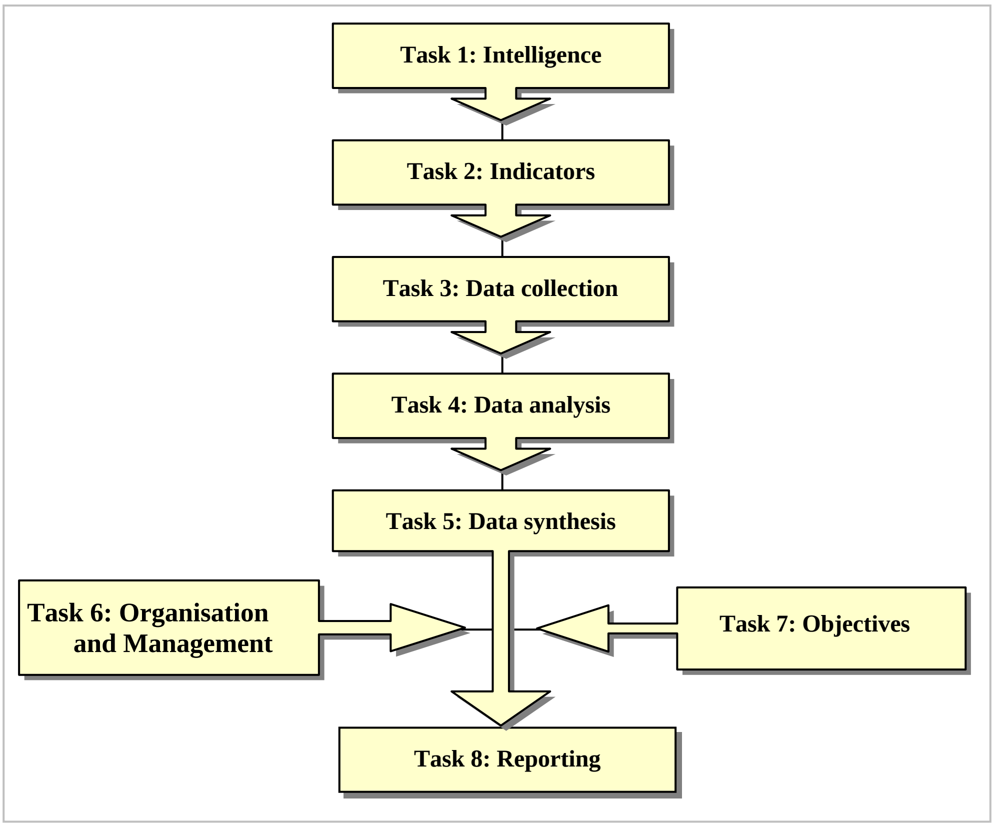
2.3.5. Implementation of Evaluation

In practice, the evaluation of a policy instrument might take the following path.

Task 1: Intelligence gathering in order to become familiar with the instrument, its input indicators, and management.
Task 2: Development of output and impact indicators, which are not the same
Task 3: Data collection
Task 4: Data analysis
Task 5: Data synthesis
Task 6: Evaluation of organisation and management
Task 7: Analysis of instrument or programme objectives, their continued validity
Task 8: Discussions with the client and reporting

Any benchmarking or comparison with alternative instruments, and literature surveys will contribute to the evaluation report.

Figure 3: Steps for the evaluation of policy instruments



2.3.6. Strengths and Limitations of Evaluation

Strengths:

- Bottom-up approach, using data acquired from participants at the project level, interviews with policy makers and programme managers.
- Ability to integrate project evaluations to industry level, to instrument level, to national level and finally to EU level.
- Assessment of important, if non-quantifiable, benefits.
- Confidence gained from comparative evaluations.
- Assessment of dead-weight.

Weaknesses:

- High cost, time required, especially for a full comparative evaluation. In addition, a long time may be required until the impacts of a programme are visible.
- Lack of participants' motivation to co-operate due to excessive similar demands and absence of contractual obligation to do so.
- Even more, a lack of willingness on the part of control groups to co-operate.
- Occasional lack of continuity in participants' staff.
- Programme managers' (in the public agency) potential self interest, which should be addressed before the data collection process starts.
-

2.3.7. Bibliography

BIE (1993), "R&D, Innovation and Competitiveness - An evaluation of the research and development tax concession." *BIE Research Report* 50. As to the evaluation of tax concessions, the Australian Bureau of Industry Economics published an Evaluation of the Research and Development Tax Concession (BIE Research Report) in 1993. This evaluation of the Australian tax concession scheme is conducted by the Bureau of Industry Economics (BIE) a independent research body within the Australian Government. The BIE is a fore-runner of the Productivity Commission which is now the Australian Government's principal review and advisory body on microeconomic policy and regulation (see www.pc.gov.au). This report represents best practice with regard to various dimensions of the implementation of this evaluation studies.

David, Paul A. and H. Hall, Bronwyn (2000), "Heart of Darkness: Modeling Public-Private Funding Interactions Inside the R&D Black Box," *Research Policy*, 29, 1165-1183. The paper is a first step to toward closing the analytical gap in the extensive literature on the results of interactions between public and private R&D expenditures and their joint effects on the economy. Within the context of their simple model it is possible to offer interpretations that shed light on recent cross-section and panel data findings at both (i.e., national) and low (specific technology area) levels of aggregation.

Guellec, D. and B. Van Pottelsberghe (1999), "Does Government Support Stimulate Private R&D?" OECD Economic Studies, No. 29, Paris.

Guellec, D. and B. Van Pottelsberghe (2000), "The Impact of Public R&D Expenditure on Business R&D," *STI Working Paper* 2000/4, OECD, Paris. Their work tries to quantify the aggregate net effect of government funding on business R&D in 17 OECD member countries over the past two decades. Grants, procurement, tax incentives and direct performance of research (in public laboratories or universities) are the major policy tools in the field. The empirical model is based on a macroeconomic approach.

Hall, Bronwyn and John Van Reenen (2000), "How Effective Are Fiscal Incentives for R&D? A Review of the Evidence," *Research Policy*, 29, 449-469. The paper surveys the econometric evidence on the effectiveness of fiscal incentives for R&D. The authors describe the effects of tax systems in OECD countries on the user cost of R&D. They describe and criticize the methodologies used to evaluate the effect of the tax system on R&D behaviour and the results from different studies.

Lach, Saul (2000), "Do R&D Subsidies Stimulate or Displace Private R&D? Evidence from Israel", *NBER Working Paper* No 7943. Using data on Israeli manufacturing firms in the 1990s Saul Lach finds evidence suggesting that the R&D subsidies granted by the Ministry of Industry and Trade stimulated long-run company-financed R&D expenditures. Although the magnitude of this effect is large enough to justify the existence of the subsidy programme, it is lower than expected given the dollar-by-dollar matching upon which most subsidised projects are based.

Lundvall, B.-Å. and Tomlinson, M. (2001), "Learning-by-comparing: Reflections on the use and abuse of international benchmarking", in Sweeney, G. (ed.), *Innovation Economic Progress and the Quality of Life*, Cheltenham: Edward Elgar, ISBN 1-84064-603-9).

OECD (1994), *The Measurement of Scientific and Technological Activities. Proposed Standard Practice for Surveys of Research and Experimental Development: Frascati Manual 1993*, OECD, Paris.

OECD (2001), "Public Funding of R&D: Emerging Policy Issues", Directorate for Science, Technology and Industry, Committee for Science and Technology Policy, DSTI/STP(2001)2.

OECD-EUROSTAT (1997), *OECD Proposed Guidelines for Collecting and Interpreting Technological Innovation Data: Oslo manual,* OECD, Paris.

Pavitt, Keith (2000), "Public Policies to Support Basic Research: What Can the Rest of the World Learn from US Theory and Practice (And What Should They Not Learn)?" *SPRU Electronic Working Paper* No. 53, Science and Technology Policy Research Unite, University of Sussex, England, October.

2.4. Provision of R&D Infrastructures

Author: **Jaime Rojo** (Joanneum Research)

2.4.1. Introduction

Research and development infrastructures are a key element in the definition of science and technology policy in advanced economies. R&D infrastructures constitute the required basic physical and organisational structures and facilities providing the supporting elements on which the factors of production interact to produce scientific and technological output. Two separated although related perspectives are addressed:

(i) provision of general research facilities in the form of universities and other institutional arrangements supporting research and technological innovation, and
(ii) provision of large research infrastructures for conducting research on an international scale. Table 9 provides an illustration of the different types of research infrastructures considered.

2.4.1.1. General Research Facilities: Institutions and Systems

General research facilities include basic institutional arrangements and facilities such as universities, government laboratories and Research and Technology Organisations (RTOs). Universities are evaluated on the extent that they are capable of fulfilling their mission as providers of a public good. This means one of the targets for their evaluation is to provide supporting evidence of their capacity to generate and extend the available knowledge base. Government laboratories and RTOs aim at matching the knowledge being generated to the needs of industry. However, the rejection of the linear model of innovation has lead to the neglect of the artificial separation existing between both types of research and nowadays policy intervention tries to stimulate stronger linkages between research laboratories, universities and industry aiming sustaining innovation and economic growth.

2.4.1.2. Large Research Infrastructures

Large Research Infrastructures (LRI) constitute a key component in the design and implementation of science and technology policy. Two embracing definitions of research infrastructures can be provided:

According to the European Commission

> Research Facilities are facilities and establishments that provide essential services to the research community. It covers for example large-scale research installations, collections, special habitats, libraries, databases, integrated arrays of small research installations, as well as infrastructural centres of competence which provide a service to the wider research community based on an assembly of techniques and know-how.

According to the European Science Foundation

> Large R&D Infrastructures[21] are facilities which relate to existing facilities or instruments, are of large capacity and trans-national relevance, require sizeable investment and, generally, have high operating costs, are unique or rare, and have a consequential impact on research at a European level. Issues of general research infrastructure (i.e. the concept of the well-founded laboratory) are excluded.

The role of research infrastructures in scientific and technological advance has progressively increased over time, experiencing a rapid transformation in the last two decades. In earlier times, the concept of research infrastructures was mainly associated with large scale, high cost facilities, mainly related to basic or fundamental sciences. Nowadays, the definition of R&D infrastructures has expanded to cover an increasing number and variety of scientific disciplines and involving a wider community of users including industry. Undoubtedly, strengthened co-operation in the provision of RTD infrastructure across Europe will play a fundamental role in the implementation of the European Research Area (ERA).

2.4.1.3. Evaluation of New Research Infrastructures

Interdisciplinarity is a increasingly relevant issue in which research infrastructures can play a leading role. Supporting diversity and multidisciplinarity in research activities instead of just excellence within already well consolidated research fields is needed to maintain the research capacity needed for scientific discovery. A growing number of the research questions faced at the scientific and technological frontier require an interdisciplinarity approach for their solution. This requires a new approach that breaks with the traditional approach of very compartmentalised academic disciplines existing in universities and applied by research funding agencies.
For instance, advances in the promising and growing area of nano-technologies require joint efforts from the fields of physics, chemistry, engineering, computer sciences and for the development of applications and devices with biotechnology, aerospace, mechanical engineering and clinical medicine. Other growing areas requiring multidisciplinary research include genomics, bioinformatics, climate change, ageing population and sustainable development. Given the difficulty of implementing multidisciplinarity, research infrastructures can play a leading role in this new policy approach, fostering co-operation between different institutes, grouping projects and creating interdisciplinary research centres. However, it is worth signalling that for some scientific disciplines such as oceanography, atmospheric sciences, earth observation and engineering science it has been practically impossible to establish European wide co-operation in infrastructure provision.

[21] For clarity in the exposition, in the text the term *large R&D infrastructure* has been used as synonymous to the term *large research facility* adopted by the European Science Foundation.

Table 10: Classification of research infrastructures

Institutions	Description	Examples
General university fund	pays for universities fixed costs and teaching as well as free research	University infrastructures
Research to support government research	wide coverage of scientific research fields group of well founded research laboratories customer-contractor relationships	grouping of independent dedicated laboratories specialised in various scientific fields (e.g. CNR, CSIC, CNRS, Max Planck) government laboratories university research, industrial research, policy support units
Large research facilities (big science)	usually on a very large physical and financial scale designed to solve a uniquely challenging and fundamental require high investment	particle accelerators neutron reactors oceanographic vessels telescopes
Medium sized facilities (small science)	solve wide variety of RTD problems	high power laser facilities, high intensity magnetic fields synchrotron radiation library networks museums databanks of EU scale communication networks

Evidence shows most research infrastructures lack a clear and well-established mechanism for their evaluation. This is even more worrying, given the rapid transformation experienced by research infrastructures and the growing relevance played by multidisciplinarity. New innovative approaches are being introduced such as benchmarking which is used to evaluate general research facilities as no single optimal efficiency measure exists to evaluate this institutional set ups. An example in this direction is the case of the research assessment exercise conducted by the UK Higher Education Councils which provides a ranking of the UK university departments based primarily on the use of peer review. Comparison with past experience is also a common approach. In the US, the implementation of the Government Performance and Results Act (GPRA) has enforced the introduction of efficiency and benchmarking criteria's in intermediate evaluations of research infrastructures and facilities.

2.4.1.4. Widening the Concept of Research Infrastructure

The central role of R&D infrastructures in the RTD system is strengthened by technological factors. Technology driven factors facilitating the internationalisation, rapid dissemination of results and a closer interconnection between users has strengthened the relevance of research infrastructures as a fundamental element supporting innovation. Advances in information and communication technologies have enabled the connection of medium size installations sharing their results with the larger research community.

The recent growth experienced in certain disciplines such as the life sciences and social sciences and the improvement of computation power have created the potential for the development of large databases and laboratories for services provision that could operate at a real European transnational scale. In this direction, the completion of the sequencing of the human genome makes possible the development of large databases that can be shared for scientific purposes. In the same spirit, the recent advances in

information and communication technologies allow the development of large computer networks and platforms on micro/nano electronics.
There is an identified need to expand the traditional evaluation capacity available in hard sciences to cover also the "soft sciences" which have different research traditions and their own performance indicators.

2.4.2. Policy Objectives, Instruments and Outcomes

Outcome and impact indicators are the social and economic objectives intended through the intervention and occur after the outputs interact with society and the economy. In their mission statements research infrastructures often refer to economic and social objectives. The main objectives of research infrastructures include:

1. *Improvement of efficiency of R&D and the scientific and technological knowledge base:* research infrastructures jointly with universities constitute the backbone of the research and innovation system. The public provision of research infrastructures solves the market failure problem arising in (fundamental or basic) R&D. Following a European approach in the design, construction and management of R&D infrastructures can bring efficiency gains in the form of economies of scale and stronger competition.
2. *Enhancement of economic performance and productivity growth*: research infrastructure affects the efficiency achieved by the science and technology system. Good research infrastructures *ceteris paribus* spur scientific performance (e.g. the building up of high bandwidth network infrastructure facilitates the development and spread of the Internet, leading to the appearance of new products and services and therefore to improved economic performance). Additionally they enable the absorption of technology produced elsewhere.
3. *Improvement of quality of human resources:* A key element of the contribution of research infrastructure to society is provided through the education and training of scientists and engineers and the knowledge transfer that occurs through strengthened collaboration with industry and university. Maintaining a research capacity at the technological frontier permits to improve the level of human capital available.
4. *Promotion of social cohesion*: research infrastructures contribute to the training of young researchers, facilitating integration of young researchers in the labour market. They also facilitate human capital mobility by strengthening collaboration links between scientists from different nationalities including less favoured regions and support the increased participation of woman in science.
5. *Facilitation of scientific knowledge and technological diffusion:* Research infrastructures constitute one of the policy instruments available to facilitate the transfer of codified or tacit knowledge and research results.
6. *Generation of employment:* Research infrastructures generate direct and indirect impacts on employment. The direct employment effects of the research infrastructure are the jobs required in the construction and operation of the infrastructure. The indirect effects are measured taking into consideration the externalities and spillovers generated by the research infrastructure and they remain more difficult to quantify.
7. *Reduction of transaction costs:* research infrastructures improve networking capabilities by facilitating the provision of the critical mass required for the development of certain scientific disciplines which involve too much risk to be taken up by the private initiative. Research infrastructures also facilitate the linkages and collaboration existing between actors –government, industry and research–, therefore reducing the level of asymmetric information existing between the parts.
8. *Improvement of quality of life:* The outcome of research infrastructures sometimes are in the form of intangible goods or services that improve the welfare of society. These ‘so-called’ public goods provided by research infrastructures include issues such as improvements in health, safety, consumer protection and preservation of the environment.
9. *Transformation of the research infrastructure:* The public intervention might also transform the nature of the research infrastructure as the introduction of the Government Performance Results Act (GPRA) and the Bay-Doyle Act in the USA have demonstrated.

2.4.3. Evaluation Methodologies

Evaluation of research infrastructures requires measuring output and performance. However, the difficulties in measurement are large due to the intangible character of output and dynamic effects being all common elements that should be taken into account.

Furthermore, a real improvement in evaluation and assessment of research infrastructures requires a better accounting of R&D resources and output.

In measuring output of research infrastructures one can distinguish between first order impacts for example, journal articles, presentations, patents, collaborations and second order impacts identified through the analysis of citations to journal articles attributed to the infrastructure. The wider economic impact involves measuring the effect of the knowledge generated through the infrastructure on industry revenues and productivity.

As indicated in the preceding section research infrastructures cover multidimensional objectives, therefore their evaluation generally requires the use of a combination of various methodologies and indicators. Moreover, evaluation and monitoring practices require adaptation to cope with the wide and rapid changes experienced by research infrastructures, facilitating the selection of those that provide high quality services to science and technological research. Research infrastructures appraisal faces similar limitations to those faced in the evaluation of RTD programmes (Kuhlmann *et al*, 1998, Arnold and Guy, 1998, OECD, 1997). The uncertainty, time lags, and externalities that characterise RTD processes, restrict the application of simple rate of return calculations to assess the value of RTD infrastructures. It is worth signalling that the methods available for the evaluation of research infrastructures should be used in combination as no method applied in isolation can provide an optimal assessment.

The social returns to R&D can be measured by measuring the benefits an innovation produces in a certain industry. For example if a new product incorporating the innovation reduces the cost of a particular industry that uses the innovation then the social benefits can be measured by adding to the profits achieved by the innovator through the innovation (producer surplus) the benefits to consumers (consumer surplus) due to the reduction in the price of the innovative good (Mansfield et al., 1977).

Although there is a general consensus found in the empirical literature on the positive contribution of R&D to productivity growth, finding the empirical support for the contribution of basic R&D to productivity has been more difficult. However, certain studies have been able to demonstrate the large contribution of basic R&D to productivity growth (Mansfield, 1980; Griliches, 1986). These studies estimate cross section production functions for firm level data adding as explanatory variables, besides capital and labour, the ratio of basic R&D to total R&D. The estimation is repeated for the production function expressed in growth rates. Patent citations have been also useful to show the quality of research output. Studies using patent citations permit to demonstrate the larger value in patents produced in academia or research infrastructures by showing that academic patents are more cited overtime and more widely across fields.

The indirect effects in the form externalities and spillovers generated by research infrastructure are more difficult to estimate. However an increasing number of studies have shown that research laboratories and academia produce significant R&D spillovers. Proximity to science generates geographical spillovers with industry benefiting from locating close to where science is produced. The use of indicators on patent and bibliometric citations permits to demonstrate the existence of spillovers from location and that location also matters for innovation (both in number of products available and patenting level). Knowledge spillovers occur because a large share of informal knowledge is transmitted in the form of tacit knowledge. Industry by locating close to the research infrastructures can benefit from enlarged access to a common pool of knowledge and educated human resources.

The main methodologies (qualitative and quantitative) adapted to the evaluation of research infrastructures include:

- *Detailed cost benefit analysis:* This methodology provides an ex-ante quantification of the (private and/or social) rate of return to build a research infrastructure and offers an estimation of the opportunity cost of the investment. It provides best results when used in comparison with alternative investment opportunities. The results achieved are very sensible to the discount rate applied in the calculation. The higher the discount rate applied the lower the valuation given to future incomes. More detailed evaluation approaches have divided the potential for value added generation into technological, commercial and work factor effects. Most of the tools used to quantify the economic return of research infrastructures constitute adaptations of conventional methods and measures applied in the economics and corporate finance disciplines. The strength of these methods resides in the existence of a learning curve in their application and the higher relevance of results as more data becomes available. The major drawback is the high dependence on the quality of data. The most widely formulas used in cost benefit analysis include:
- *Net Present Value (NPV):* requires discounting benefit and cost data over a time series to the reference year and subtracting present value of costs from present value of benefits.
- *Internal Rate of Return (IRR):* calculates the discount rate that brings the NPV to zero, i.e. the IRR is the discount rate needed to reduce the time series of net benefits realised by an research infrastructure to zero.
- *Benefit-Cost Ratio:* calculates the ratio of discounted benefits to discounted costs.

The approaches to quantify the socio-economic gains of a policy instrument include contingent valuation studies, simulating the existence of a market for a non-marketed good for example a health disease treatment or clean water. These studies generally adopt questionnaires incorporating willingness to pay schemes to try to infer the price a certain public good is worth to the respondent. Other approaches include the use conjoint analysis in surveys to determine the price users place on the attributes or features of goods and quality adjusted hedonic pricing for new or improved goods.

- *Econometric analysis*: This methodology is a combination of mathematical, statistical and economic theory to model and describe economic relationships and to test the validity of formulated hypothesis and estimating the parameters to measure the strength of the relationship between the variables of the model.
- *Technometrics*: This methodology analyses technological innovation in the development phase using output indicators derived from the technical specifications of products or processes. The method uses a combination of R&D output indicators, sociological and economic information.
- *Bibliometrics*: This ex-post methodology provides an objective assessment of the scientific quality of the output of the research infrastructure being evaluated.

Bibliometrics is the application of quantitative methods to the analysis of scientific literature. Impact factors are among the most widely used bibliometric indicators used to evaluate the quality of a research institution. They are constructed retrieving journal articles citations, available accessing the databases produced by ISI (Institute for Scientific Information), the SCI (Science Citation Index) and the SSCI (Social Science Citation Index).

- *Case studies*: This methodology involves examining a limited number of specific cases that the evaluator perceives will be revealing to better understand the dynamics applying within a specific setting. The case studies generally use telephone and face to face interviews with key staff.
- *Co-word analysis*: This methodology is used for mapping scientific fields and for detecting new emerging fields. It identifies keywords and relates the contents of papers with other scientific publications, grouping papers to show the structure and dynamics followed by science and technology.
- *Peer evaluation*: This methodology is based on the perception scientists have of the scientific contributions made by other peers. The outcome is significantly influenced by the quantity and the quality of the contributions. Peer review is the most widely used method for the evaluation of the output of scientific research. Peer review in large research organisations is structured panels of independent experts divided by in scientific disciplines.
- *Modified peer review*: This methodology constitutes an expanded version of traditional peer review incorporating the inputs of the potential users of scientific and technological research. The introduction of expert panels from outside the community being evaluated into the evaluation process ensures the social acceptability of the initiative.
- *Expert or Review Groups*: This methodology is a judgement tool which brings together a group of independent eminent scientists and/or research managers with broad views & expertise, not necessarily linked to the type of facility to be assessed to produce a value judgement on a research facility and its effects. It is generally used in ex-ante evaluation, e.g. for conducting a feasibility study for developing a research infrastructure.
- *Patents*: This methodology is useful to map the technological capacity of research infrastructures. It uses similar mathematical and statistical methods as bibliometrics but applied to patent counts and patent citations. New indicators are being developed to measure the quality of patents (e.g. patent families).
- *Questionnaire methods and surveys*: This methodology mixes qualitative and quantitative information. Once the model is developed, the survey allows to perform hypothesis testing and detailed exploration of process and impact.
- *Foresight and Delphi*: This methodology using experts and user moderated panels can be used in the ex-ante and intermediate evaluation of research infrastructures to identify weakness and strength areas and streamline its activities.

2.4.4. Good Practice Examples

2.4.4.1. The complete evaluation cycle: The case of the action on access to research infrastructures (ARI) of the EU RTD Framework Programme

The evaluation of the capacity and quality of the research infrastructures takes place in three different stages: ex-ante, monitoring and ex-post impact evaluation (European Commission 1999a):

Ex-ante evaluation: After the call for proposals, to establish the best infrastructures to serve as hosts for European researchers. The evaluation procedure takes place using peer review methods, in order to select excellent facilities which require transitional access. The ex-ante evaluation of proposals for supporting research infrastructures discloses the identity of the institution submitting the proposal. It is exactly the reputation and the excellence of the facility that needs to be judged. The evaluators need to know the organisation and the staff involved in order to correctly assess the appropriateness of the research facility.

When evaluating research infrastructures, the large variation on output indicators existing across scientific disciplines should be specifically taken into account. For example in engineering, publication in referred journals tends to be relatively low. In this field, dissemination is basically achieved through the publication of reports or through direct technology transfer.

Monitoring evaluation: After receiving the support, there is a need to evaluate whether the facilities are providing what they promised in the proposal. A monitoring system is needed, that could assist the learning process for the Commission to know how to adjust the programme to deliver the most efficient results.

The mid term evaluation of the ARI initiative for the period 1994-1998 was carried out by a panel of independent experts. The evaluation analysed the impact and the effectiveness of the activity and provided recommendations for improvement. The inputs to the evaluation included a technical audit of the contracts against its stated objectives, hearings of the contract managers conducted on-site and a questionnaire survey of the users who benefited from the access provided. The users' survey included questions relating to the following set of topics:

- perception on the need of the support action for carrying out the research.
- assessment of the services provided by the facility 0with respect to: technical, logistic support, intellectual environment(good-average-poor).
- indications if the work at the facility generated any output in terms of journal articles, patents, conference presentations.
- perception regarding improvement of own scientific knowledge and career from the stay at the facility.

Aiming at facilitating the evaluation procedure, before leaving the facility all users should be requested to provide the management with a brief technical report on the research carried out and any difficulties encountered.

Ex-post evaluation: When the visits are finished, the performance of the infrastructure and the researchers as well as the value added to the research community needs to be demonstrated to justify the action. There is also an increasing pressure to demonstrate the impact on competitiveness and on socio-economic issues.

Since the initiative on ARI provides funding for external researchers, evaluating its impact requires measuring the marginal increase of scientific production generated by this action. This requires isolating the contribution to scientific production of foreign visitors and measuring this part of productivity separately. Wider geographic effects should be taken into account in the evaluation of this action including the facilitation of researcher mobility across EU countries and in particular the impact of this mobility on cohesion and technology diffusion (e.g. training of researchers from accession countries or less favoured regions).

In terms of measurement of regional economic effects, the research facility could establish links with local incubators and start-up initiatives. This is however a task more of the national and regional authorities rather than Commission policy, since international visitors supported by the initiative will be returning in principle to their national host institutions. However, synergies with structural funds could be sought in LFR to set up spin off activities around research infrastructures.

Main tasks of a research infrastructures monitoring and evaluation system (Panel D Report, Conference on Research Infrastructures, Strasbourg 2000):

- Selects only those infrastructure which address the user communities' needs.
- Prioritises research areas in terms of their relevance.
- Is capable of dealing with the larger concept of research infrastructures.
- Takes into account the different objectives and operations of research infrastructures.
- Maintains scientific excellence as the main scientific criterion.
- Organises evaluation panels to ensure high quality, fairness and user friendliness

Identified issues for improvement in the prospective assessment of ARI include:

- Increase evidence of the needs for international use of the contractors' infrastructure allowing evaluators to better assess the European value added of the proposal.
- Incorporate a combination of methods dealing with science and technology and societal issues complementing the results provided by peer review methods.
- Introduce comparison exercises between proposals in similar research areas.
- Devise mechanisms to identify medium size infrastructures introducing incentives to newcomers such as feasibility studies or support for preparatory meetings.
- Strengthen the character of mid term review panels due to their usefulness as a tool to support the work of evaluators and the research facility managers.
- Engage proficient evaluators providing attractive working conditions that allow to asses research excellence.
- Rationalise and homogenise the tools and methodologies used in evaluation, providing a more rigorous approach to impact assessment.

2.4.4.2. Basic criteria applied by the European Science Foundation in the evaluation of a Large R&D Infrastructures

As a consequence of the increasing costs of R&D, the tightening of national research budgets and the globalisation of science, research infrastructures have become an issue of international dimension. The recent progress experienced in the *networking of European research infrastructures* has increased the overall efficiency of the RTD system and has undoubtedly fostered European co-operation in science and technology. In this respect, Europe has been capable of developing a well-diversified and unique network of research infrastructures responding to the demands of the research community (e.g. CERN, EMBL, ESA, JRC, JET, Synchrotron ESRF, ILL).

Large R&D infrastructures can be closely connected to the concept of networks of excellence envisaged in ERA. In every sense, European large R&D infrastructures compete for scientific excellence with other large R&D infrastructures from US and Japan. Knowledge accumulation and diffusion is facilitated through training and education of young researchers and through the large turnover and researchers from academia and into industry.
The European Science Foundation has been commissioned the role to design and co-ordinate the evaluation of projects to support large R&D infrastructures. This task is performed considering two major interrelated issues: scientific and operational issues and strategic and policy issues. The main criteria's used in the evaluation of each of these issues are indicated below:

Scientific and operational issues

- analysis of the scientific goals to be (already) achieved through the realisation/operation of the large R&D infrastructure under study.
- strategic scientific value, impact and timeliness of these goals for the advancement of science, research, and technical development.
- methodological-technical performance of the facility, its uniqueness or complementarity, and its timeliness with regard to the projected goals and function of the facility, and providing a comparison with the state of the art in Europe and internationally.
- size and standing of the envisaged users community of the large R&D infrastructure within the related science and research communities and in a wider context, now and in the foreseeable future.
- format and site(s) of the project.
- positioning and costs of the large R&D infrastructure under study within the European (or possibly global) landscape of research facilities in the specific disciplinary domain(s) and in a wider context.

Strategic and policy issues

- trans-disciplinary impact of the large R&D infrastructure under study.
- educational value and impact of the large R&D infrastructure.
- profile, role and impact of the large R&D infrastructure in its trans-national, regional, European and global context.
- 'geographical issue' of the large R&D infrastructure, recognising the possibly differing situation, and differing needs and interests, of the research communities and funding agencies of smaller countries vis-à-vis those of larger countries.
- balance anticipated between the experimental research at the large R&D infrastructure and the scope and size of theoretical research in related fields needed to underpin it.
- average costs anticipated per experiment/investigation at the large R&D infrastructure derived from technical case and cost analysis of the large R&D infrastructure (and also the average number of researchers/authors per experiment/investigation), by comparison with the respective figures of 'complementary facilities and methodologies'.
- operational/managerial concept of the large R&D infrastructure, in particular for serving 'professional/expert users' and/or 'non-professional/non-expert users' and 'applied-research/commercial users'.

2.4.4.3. Evaluation of a Large Research Infrastructure: the Joint Research Centre

The Joint Research Centre (JRC) of the European Commission is the scientific and technical research laboratory of the European Union. It consists of seven institutes located in different European Member States employing some 2200 researchers and technicians and collaborating with over 2000 research groups world-wide and within 150 trans-national networks.
The JRC has established overtime a regular and well structured procedure for the evaluation of its activities based on the work of expert panels and visiting groups. The evaluation is performed at an individual institute level and then consolidated into an overall assessment for the whole organisation. The main operational inputs to the evaluation are the *scientific audit* and the *five years assessment*.

The Scientific Audit of JRC Activities

The Scientific Audit (European Commission, 1999b) is aimed at ensuring the quality of JRC's output by evaluating the compliance with the objectives defined in the research programmes of each of the JRC institutes. The main recommendations to the JRC management include:

- Support the institutional mission for the provision of scientific advice and technical input for the formulation, monitoring and implementation of EU policies.
- Enable the development of a European science and technology reference system, capable of underpinning future policy making and ensuring that scientific advice is acceptable to the citizen.
- Recommend the strengthening of the networking with Member States' laboratories.
- Recognise the relevance of the scientific base to support the mission of the JRC providing a balance between services and research.
- Signal the importance of instruments such as project clusters for inter-institute collaboration (e.g. refer to the grouping in programme clusters of areas of common research interest at http://www.jrc.cec.eu.int).
- Recognise the importance of provision of research on nuclear safety technology requiring a maintenance of the matching funds available.
- Indicate the need of providing a wide recruitment strategy to support long term scientific capacities.
- Recommend the strengthening of joint actions with industry and technology transfer actions.

The evaluation process is perceived as a process of mutual learning and consensus building for the organisation. Much of the process has to do with learning to formulate the right questions, reporting back and providing time to discuss the answers received helping to improve the efficiency of the organisation.

Five Years Assessment

The Five Years Assessment Report of the JRC (European Commission, 2000) provides an assessment of the activities carried out by the JRC for the period 1995-1999. The results of this overall report are based on individual reports produced by the visiting groups that reviewed separately the activities of each of the JRC institutes. The main recommendations for the JRC extracted from the overall assessment report include:

- Importance of creating advisory groups to support decision making in new areas of research.
- Facilitate the integration between institutes to promote the image of JRC as a single corporate organism.
- Further develop the Cluster concept and allocate some funding for certain networks of projects with JRC-wide implications.
- Preserve and develop the competence pillars as means of matching core competencies to present needs and future demands. A sufficient amount of high-level research is important to maintain existing competencies and to extend them into forthcoming areas.
- Apply and develop further the traditional JRC skills of measurement reliability, validation expertise, the analysis and understanding of complex processes and in technical inspection.
- Develop exploratory research into a forward-looking tool accompanied by suitable level of funding. Improve the awareness of the need for networking and identify knowledge gaps and outside opportunities for the benefit of supporting policies, especially considering a developing ERA.
- Strengthen the collaboration linkages with Universities. Try to increase the level of PhD students, post-doctoral fellowships and visiting scientists. Try to improve collaboration of staff with local universities and consider sabbatical detachment.
- Improve the visibility of the JRC and its achievements. Promote technology transfer of certain JRC results.
- Allocate more resources to the management of interfaces, particularly regarding the institutional customers, the various services of the Commission.
- Increase the perception of project managers of the outside world and what it can offer in terms of knowledge or provide as deliverables.
- Promote project managers’ skills and training to deal with their customers and manage their projects.
- Develop effective project management tools and task break-down of programmes for identifying resources and milestones for project control and for guiding the overall research plan.
- Identify “best practices” and, where appropriate, apply over the whole organisation.
- Strengthen the “independence” character of the research carried out by the organisation, based on the highest quality and integrity, as well as timeliness, of its scientific and technical output.

2.4.4.4. Ex-ante Evaluation of Construction of new Synchrotron Radiation Facility

The appraisal for the new site of the new synchrotron facility studied the location in three alternative sites, two for whose location was already decided and one ‘best likely case’ green-field site (OST, 1999). The appraisal used a mixture of criteria’s for the selection of the site. First, it analysed whether the proposed facility met the necessary requirements for the site. Second, provided supporting evidence that the facility generated large benefits for the science community. Third, selected that facility with the lowest net present cost. For the site selection three types of factors were taken into account: i) financial factors including land price, infrastructure, cost of capital, operating cost, combined operating saving, cost of living, net present cost of closing one of the facilities, ii) essential requirements including local and environmental considerations, accessibility and services iii) beneficial factors including availability of staff, impact of the construction of the project, local scientific culture, location convenience.

In the appraisal, the net present costs for the three potential sites were calculated applying cost-benefit analysis adopting three different scenarios: a realistic, a pessimistic and an optimistic. The annual discount factor applied in the cost-benefit analysis for the three sites was a 6%. A determinant factor in the determination of the net present costs was the saving generated by the closing of one of the proposed facilities.

2.4.4.5. Quantitative evaluation of the economic impact of research infrastructures

Econometric approaches constitute a useful methodology to evaluate the economic impact of research infrastructures. This methodology allows to capture the productivity of research infrastructures. In this approach, research output is expressed as a function of R&D expenditures adopting generally a simplified version of the linear model of innovation. Control variables are generally included in the estimated equations (for a good overview compilation of work in the field see Griliches 1998). Output is generally measured in terms of journal articles or patent counts or citations and input is captured by the amount of R&D expenditures. Adams and Griliches (1996) analyse the productivity of university research controlling by scientific discipline or by institution. The former includes the impact of research spillover while the latter excludes the effect of research spillovers. Their estimated regression takes the form

$$y = \alpha + \beta W(r) + \gamma X + u$$

where y is the logarithm of the intermediate output (papers or citations), $W(r)$ is a the logarithm of the distributed lag function of real past R&D expenditures, X is a vector of control variables (including dummies for changes in research production system and type of infrastructure), and u is a random disturbance. The coefficient β determines the existence of diminishing (constant or increasing) returns to R&D production depending on whether $\beta < 1\,(\beta \geq 1)$.

The analysis also permits to differentiate impacts between funding sources. This is possible by estimating the equation

$$y = \alpha + \beta \log\left[W(r_A) + \delta W(r_B)\right] + \gamma X + u$$

where $W(r_A)$ is the arithmetic distributed lag of R&D funded through type A funds and $W(r_B)$ is the arithmetic distributed lag of R&D funded through type B funds. Under this specification the effects of both types of funds are allowed to differ. A coefficient $\delta < 1\,(\geq 1)$ indicates the effects of funds of type B are smaller (equal or greater) than the effects of funds of type A.

2.4.4.6. Evaluation of the German Research Infrastructure System

In December 1996 the State and Federal Authorities in Germany decided to commission a series of studies to evaluate all jointly funded research institutions. Evaluation of research activities has been conducted in four of the major publicly supported research organisations in Germany*, The Max Planck Society* (MPG), the German Research Association *Deutsche Forschungsgemeinschaft* (DFG), the *Hermann von Helmholtz German Association of German Research Centres* (HFG) and the *Fraunhofer Society*

(FhG). The evaluation approach followed focuses on conducting a *systemic evaluation* instead of carrying single independent evaluations for each of institutions conforming the system. The broad criteria applied in the evaluation include an assessment of the internal efficiency, the scientific excellence and the co-operation capacity of the system. The common parameters emerging from the evaluation of the German system of large research infrastructures include:

- Favour interdisciplinarity and facilitate interinstitutional forms of research.
- Provide enough flexibility in workprogramme definition to make possible the support of new areas of research.
- Enable the long term guaranteed financial framework and solid infrastructures required to maintain and strengthen research capacities.
- Incorporate the new challenges identified in exchanges with the wider community (e.g. making use of surveys, enquiries, science shops, foresight exercises).
- Increase the linkages with the international research community ensuring high standards of achievements.
- Facilitate the training of (young) researchers and ensure the transfer to industry and society of the accumulated expertise.
- Enable the diversification of funding sources to facilitate the independence of the system.
- Improve competition for funding among research organisations.

2.4.4.7. Evaluation of the Institutional System of Research in the Netherlands

The Netherlands is a country with a long tradition in evaluation that has set-up a well developed system for the evaluation of its research organisations. The Netherlands Organisation for Scientific Research (NOW) mission is to promote scientific research at universities and research institutes in the Netherlands and research quality promotion. Relevant aspects of the NOW evaluation process include (OECD, 1997):

- Peer review, sometimes backed up by bibliometric analyses.
- Evaluation set-up may differ for organisational levels and vary according to institute's mission and its life-cycle phase; the following characterisation applies as a rule.
- The scope covers aspects like: vision; strategy; scientific quality; R&D; human resources and financial management; quality and coherence of the institute programme; infrastructure and housing; position of the institute; size and quality of so-called third-party funding; follow up of previous recommendations.
- Site visits are standard procedure.
- Review committees: consist of at least three peers; the chairman or at least one of the peers is Dutch.
- Evaluations start off by self-assessment and may address past performance as well as the expected research potential.
- Follow up is embedded in managerial processes and discussions on the level of the executive board of institutes organisation. Evaluation Reports may imply funding reallocation.
- Evaluations are carried out every five-years.
- Sometimes ad hoc evaluations are initiated.
- Almost every institute is counselled by an external committee.

The institutes of the Royal Netherlands Academy of Arts and Sciences (KNAW) are engaged in basic and strategic research, scientific information services and biological collection management. The evaluation of KNAW institutes is similar to the NWO institutes, although evaluation of KNAW institutes:

- Focuses more on scientific quality than on societal relevance related criteria.
- Uses peer review committees, chaired by a national , which are expected to use a fixed-interval assessment scale.
- Sends summaries of its evaluation reports to the minister (for information purposes only).

2.4.5. Data Requirements

Indicators in evaluation are used to measure the progress towards predefined objectives. The indicators available in impact evaluation of research infrastructures include input indicators and output indicators.

Inputs

Inputs are the resources consumed by the operation of the research infrastructure.

- Expenditures in equipment-years (fixed costs);
- Expenditures in person-years (variable costs).

Outputs

Outputs are the direct product of the instrument.

- scientific publications counts;
- patents counts;
- invention disclosures;
- prizes;
- citations;
- investment level;
- services provided;
- employment generation;
- revenues;
- value added;
- amount of funding obtained from third parties (e.g. competing for EU funding);
- collaborations, industry–science relationships, research joint ventures;
- licenses;
- training of researchers (by categories).

In supporting certain research infrastructures it is relevant to capture the incremental value added generated. This is the case of the European Commission Framework Programme action line on access to research infrastructures (ARI) which provides additional funding to support existing national infrastructures to extend their use beyond the usual group of users (European Commission, 2001). The inputs are measured in terms of additional use of equipment or instrument per unit of time –for example telescope hours, instruments day, magnet hours, etc–. Output indicators capture additional value added generated by the initiative, for example, number of visiting scientists and secondment experts, number of PhD students and post-docs, number of experiments, visitors day, number of action related publications compared to total number of publications.

2.4.6. Application of Evaluation

As an illustration, the design an implementation of a monitoring and ex-post impact evaluation of research infrastructures should ideally include the following steps:

1. Provide the terms of reference for the evaluation.

- Describe the scope and objectives of the research infrastructure.
- Include the guidelines for the reporting process in the terms of reference.
- Specify the requirements for data compilation, progress, mid-term assessment and final reports.

2. Clarify the objectives and goals of the research infrastructure.

- Formulate several questions on the infrastructure meeting its objectives.
- Provide through the evaluation reasoned answer to those questions based on all the evidence available.

3. Determine the data availability and the requirements for data collection and analysis.

- Select indicators for measuring research infrastructure effectiveness (qualitative and quantitative indicators).
- Control possible selection bias through data collection.
- Decide on the need to launch supplementary surveys or questionnaires.

4. Establish the evaluation approach and applied methodologies.

- Plan the evaluation.
- Compile the complementary field work available that could assist the evaluation. process for example self evaluation reports and technical reports.
- Establish the applied evaluation methodologies, for example, case studies, surveys cost benefit analysis, econometric analysis.

5. Determine the sample size and establish a control group when required.

6. Design the structure of the evaluation.

- Refer to the design and structure which might include a combination of peer groups and expert panels.
- Include a good balance between scientists, industrialists and senior policy officials.
- Analyse the wealth of information available.

7. Write the results and discuss them with the interested actors.

- Provide clear and effective findings and recommendations.
- Ensure appropriateness of the methodologies applied.
- Guarantee the credibility of the results.

8. Feed the results into the policy making process.

- Ensure from the beginning a strong commitment from policy makers for taking up the recommendations.
- Embark senior science officials in the process to ensure impact of evaluation results.

2.4.7. Bibliography

Commented Bibliography

Large Research Infrastructures

European Commission (2001), *European Research Area for Infrastructures* [SEC(2001) 356]. This report provides an updated overview of the main achievements and shortcomings of the current EU approach to research infrastructures.

Proceedings on International Conference on Research Infrastructures, (2000) Strasbourg, 18-20 September 2000. The proceedings of this conference offer an overview of the current issues of debate on research infrastructures. The conference included a panel dedicated to the evaluation of research infrastructures.

Link, A. (1996), *Economic Impact Assessment: Guidelines for Conducting and Interpreting Assessment Studies*, NIST: US Department of Commerce. This report provides a well documented overview and a practical example on how to conduct economic impact assessment studies.

Office of Management and Budget (1992), *Guidelines and Discount Rates for Benefit-Cost Analysis of Federal Programs*, Circular No. A-94 Revised (Transmittal memo No 64). This circular describes the general principles and rules required in applying benefit-cost analysis to US federal government supported programmes.

OECD (1995), *Very Large Scientific Facilities in Europe: Analysis of Legal Texts Governing Institutional Co-operation,* Megascience: the OECD Forum,. This report provides a synthetic introduction to the legal instruments governing scientific co-operation in Europe for the major large scale projects and programmes in Europe making special emphasis on their objectives, organisation and economic and industrial returns.

OECD (1998), *Evaluation of the OECD Megascience Forum*. In this report an independent expert panel provides an evaluation of the impact of the recommendations of a consultative forum on large-scale scientific collaboration in government policy making.

European Commission (1999), *Mid Term Review of the Access to Large Scale Facilities Activity*, This report provides an expert panel lead detailed review of the EU RTD Framework Programme activity on access to large-scale facilities. The evaluation focuses on access practices, scientific output and value-for-money. The supporting material used for conducting the evaluation included contract review reports, user's surveys and personal hearings of contract managers. The division of the panel into expert-groups allowed further comparisons across scientific and technical areas. http://www.cordis.lu/tmr/src/alsf1.htm

General Research Facilities

US Congress (1993), *Government Performance Results Act (GPRA) of 1993*, One Hundred Third Congress of the United States of America. This law enacts the establishment of strategic planning and performance measurement in the Federal government, enabling the evaluation of the relative efficiency and effectiveness of US federal agencies' programmes. Under current trend of budgetary tightening of

public support to R&D, GPRA can be used in the process of project funding prioritisation. http://www.nsf.gov/od/gpra/law.htm

National Institute of Standards and Technology, *Strategic Planning at the National Institute of Standards and Technology*. Provides on-line a series of economic impact assessment studies. http://www.nist.gov/director/planning/strategicplanning.htm

Committee on Fundamental Science: National Science and Technology Council (1996), *Assessing Fundamental Science*, Washington. Provides a description of the assessment process enacted by the GPRA including a review of the performance measures it introduced.

Arnold E. and K. Balazs, (1998), *Methods in the Evaluation of Publicly Funded Research*, Technopolis. An exhaustive review of the main methods applied in the evaluation of publicly funded research.

Peterson, J. and M. Sharp (1998), *Technology Policy in the European Union*, MacMillan Press Ltd.: London. Chapter 9 provides a critical review of the achievements of European-wide technology policy initiatives aiming at strengthening collaboration including the Framework Programme and EUREKA.

OECD (1997), *The Evaluation of Scientific Research: Selected Experiences*. 1997. The document provides an overview on OECD country experiences on the evaluation of scientific research. It offers an overview of the evaluation of research in the CNRS in France and the blue list institutes in Germany.

Office of Science and Technology (2001), *Study of Science Research Infrastructure*, December 2001. The report provides a review assessment exercise on the past investment in research infrastructure in the higher education sector in the UK. http://www.dti.gov.uk/ost/sosri/Summary.pdf

Technopolis, *Evaluation of the Research Council in Norway*. 2001. This large evaluation exercise with its integral series of reports provides an example on how to perform a detailed and exhaustive evaluation of a large funding institution.

References

Adams, James and Zvi Griliches (1996), "Research Productivity in a System of Universities", *NBER Working Paper* No 5833.

Arnold, Erik and Ken Guy (1998), *Uses and Limits of Cost Benefit Analysis in Evaluating Industrial Modernisation Programmes*.

European Commission (2001), *Annual Summary Report on the Coordination of Activities in Support of Research Infrastructures*. Brussels.

European Commission (1999a), *Mid-Term Review of the Access to Large-Scale Facilities Activity*. Brussels.

European Commission (1999b), *The 1999 Scientific Audit of the JRC Institutes*. Brussels.

European Commission (2000), *Five-year Assessment of the European Union Research and Technological Development Programmes*, 1995-1999, Brussels.

Griliches, Zvi, (1986), "Productivity, R&D, and Basic Research at the Firm Level in the 1970s," *American Economic Review*, 76, 1,141-154.

Kuhlmann, Stefan; Airaghi, Angelo; Busch, Niels; Georghiou, Luke; Ledoux, Marc; Van Raan, Anthony and Jose Viana Baptista, (1998), *Options and Limits of Evaluating RTD Programmes*,.

Mansfield, Edwin (1980), "Basic Research and Productivity Increase in Manufacturing," *American Economic Review*, 70, 863-73.

Mansfield, Edwin, Rapoport, John, Romeo, Anthony, Wagner, Samuel, and Beardsley, George (1977), "Social and Private Rates of Return from Industrial Innovations," *Quarterly Journal of Economics*, 91 (2), 221-240.

OECD (1997), *The Evaluation of Science and Research: Selected Experiences*. Paris.

Office of Science and Technology (1999), Investment Appraisal of Options for the Site of the Proposed New Synchrotron Radiation Facility. London.

2.5. TECHNOLOGY TRANSFER

Authors: **Pari Patel** (SPRU) and **Robin Cowan** (PREST)

2.5.1. Introduction

From a public policy perspective one of the key interactions in a national system of innovation is that between publicly funded basic research conducted in universities and other public institutions (henceforth referred to as PRIs[22]) and innovative activities of business firms. This has long been recognised in public policy debates but has assumed increasing importance in the past decade. Thus one of the important current policy priorities within many countries (both within the EU and more generally) is to encourage intensified co-operation between public sector research institutions and private companies in order to convert research results into commercial products.

In order to understand the rationale underlying such policy concerns we first need to examine the contributions that publicly funded research makes to industrial innovation. Recent research[23] has shown that such contributions can be either *direct* or *indirect*. Amongst the former are scientific discoveries leading directly to new products and processes, engineering research techniques (such as computer simulations) and new scientific instrumentation. Some of these are achieved by means of setting up spin-off companies, others by collaborating with established firms. However other contributions of public research are *indirect*, for example, the training of graduates and other researchers, creation of background knowledge and professional networks contribute to business firms' own problem-solving activities. These are particularly relevant to experimental engineering research, design practices, production and operational activities located within firms. One danger in both policy making and analysis is giving excessive attention to the direct contributions of publicly funded research to technology, to the neglect of the indirect contributions that are often more highly valued the business practitioners themselves.

These considerations have given rise to a number of policy initiatives that can be grouped under the following headings:

- Policies aimed at encouraging collaborative research.
- Under this heading are schemes aimed at providing: (a) subsidies for firms to contract out research to PRIs; (b) incentives for PRIs to supply consultancy and technical services to industry; (c) funding for collaborative programmes of research in specific areas of technology; (d) facilities in a specific location for interaction to take place (science parks).
- Policies aimed at developing co-operative training and education programmes.
- These schemes are aimed at setting up training and education programmes within PRIs that are more suited to the skill needs of industry. These include providing short courses for professionals as well as co-financing of post-graduate students.
- Policies aimed at increasing the mobility of research and technical personnel.
- Some schemes under this heading are designed to encourage PRI staff to take up temporary or permanent positions in industry and vice-versa. Another set give incentives to hire recent graduates to work on a specified R&D project within firms.
- Policies aimed at encouraging direct commercialisation of research results through the creation of spin-offs and licensing of IPRs.

---

[22] Public sector Research Institutions

[23] For reviews see Martin and Salter (1996) and Salter et. al (2000)

These include: (a) changing regulations to make it easier for PRI employees to set up new businesses; (b) changing IPR regulations making it easier both for PRIs to obtain patents and for newly set up firms to licence technology developed within PRIs; (c) providing financial support to investigate the feasibility of setting up a new business; and (d) providing seed capital to enable commercial initiatives to be developed from university research.

Finally there are a range of other policy measures that have an important effect on the linkages between PRIs and firms: those related to the stimulation of innovative activities within enterprises. The underlying rationale being that to a large extent such linkages are shaped by the nature and extent of the demand from private firms. For firms to benefit from the results of public sector research, they need to have in-house competencies and skills to be able to absorb such results (*absorptive capacity*).
Furthermore, the impact of the legal framework on the demand side should not be underestimated. "Changes in competition policy to enable networking and co-operation in pre-competitive research" can undermine the stimulation of innovative activities[24].

2.5.2. Policy Objectives, Instruments and Outcomes

The ultimate objective of public policies towards technology transfer is to increase productivity and enhance economic growth. The rationale being that combining the skills and competencies that exist within the public sector research system with those of leading edge business firms will result in higher productivity levels and greater market share in a wide range of high-tech product groups and services. Further motivation for technology transfer policies is the need to obtain better value for large public investments in research.
In addition though, technological leadership does not in itself lead to better economic performance. Also central is the level of diffusion of new technology. Policies to increase diffusion aim directly at the adoption process as well as indirectly at enlarging economic capabilities. These policies pay special attention to the broader social and institutional settings, as they have an important impact on the development of diffusion.

2.5.3. Policy Evaluation

Evaluations of policies can be undertaken at least two levels of aggregation: at the level of specific schemes or at the level of broad areas of policy, where schemes with similar aims are combined (for example the 4 areas outlined above). Each of these involves use of different sets of data and techniques. The first requires detailed information from administrative records of the scheme and from the participating institutions. This includes the amount of resources devoted to the scheme, some information about the characteristics of participants (in the case of firms this could be their size, sector, ownership etc.) as well as their opinions about the nature and effectiveness of the scheme.
The latter could also include some measures of specific benefits attributable to the policy. To be useful for evaluation purposes, this information needs to be combined with information from non-participants in order to employ 'control group' approaches. In practice implementing such an evaluation strategy is extremely time consuming and expensive.

---

[24] http://www.oecd.org/dsti/sti/s_t/prod/Outlook2000/OutQues.pdf

Individual evaluation tools such as case studies, peer review and user summaries can be employed at the level of specific schemes.[25] Case studies have the advantage that they help to understand complex processes and explore situations where interesting variables are not predefined. The evaluator is able to learn as the work is progressing, gradually building up a model. The studies can be structured to facilitate cross-case analysis, and they can be a source of 'how to' understanding. The disadvantages of such an evaluation procedure are not only the high costs of information gathering but also the high dependence on the skill and experience of the evaluator and the difficulty to incorporate it into routine monitoring.

Peer review on the other hand is systematised, checked and analysed to increase the confidence in results and it remains the most reliable approach to establish scientific quality. A peer review can be characterised as an informed, 'rounded judgement'. The qualitative and judgement basis is a point of criticism as well as the problem of criterion referencing and the differing in cultural behaviour. 'Group think' and social dominance can also be identified as a disadvantage of the peer review evaluation as well as the risk of facing a 'prisoner's dilemma' behaviour by peers. Also is this evaluation method hard to apply in a commercially-sensitive work.

User surveys can provide a nuanced, quantified understanding of programs, collect direct process experience as well as indicators, test and generalise case studies and other findings, enable estimation and description of key impacts and provide quality control of program management. Once a model is built to evaluate the intervention, user surveys allow hypothesis testing and detailed exploration of both process and impacts. The disadvantages are that user surveys are subject to bias; reflect users' appreciation to receive resources and optimism about impacts.

Evaluations at the level of broad policy areas can proceed on the basis of publicly available information. In the case of technology transfer policies this includes data obtained from R&D surveys, CIS (Community Innovation Surveys), ad hoc country level surveys, as well as bibliometric information on publications and citations, and patent data. The main advantage is that such data are readily available at low cost. One way of partially assessing policy impact would be to examine the relevant data series over time in order to identify structural breaks and to see if these coincided with the introduction of the policy. The problem with such an approach is of course that the secular changes in the data series may be driven by factors other than the introduction of the policy. Another major difficulty is obtaining information on the total amount of resources devoted to major groups of policies.

2.5.4. Data Requirements

This section concentrates on data requirements for evaluation of broad sets of policies rather than specific schemes. In particular it highlights indicators that can be obtained from internationally comparable surveys, those that are only available from ad hoc country surveys, and contains some suggestions for further data collection. The aim here is to highlight the types of data that are either readily available or can be obtained at low cost.

[25] See http://www.oecd.org/dsti/sti/stat-ana/prod/arnold.pdf

2.5.4.1. Policies Aimed at Encouraging Collaborative Research

Output/Activity Indicators:

- Per cent HERD and GOVRD funded by industry over time: total and by fields of science. Source: OECD Basic Science and Technology Statistics
- Per cent total income of PRIs funded by firms according to field of science. Source: Ad hoc country level surveys, for UK example see: J. Howells et al. (1999), "Industry-Academic Links in the UK: A report to the Higher Education Funding Councils of England, Scotland & Wales", PREST, Manchester.
- Number of contracts with firms per research staff. Source: Data to be collected from PRIs.
- Numbers and types of firms co-operating with PRIs (according to size, sector and level of innovation). Source: CIS 2
- Number and types of firms located at Science Parks. Source: Data to be collected from National funding agencies.
- Numbers and types of PRIs engaging in collaborative research. Source: Data to be collected from PRIs.
- Per cent of all publications that are written jointly between PRIs and firms by fields of science. Source: Bibliometric databases such as ISI
- Per cent of all patents that are jointly assigned to PRIs and firms by area of technology. Source: EPO databases

2.5.4.2. Policies Aimed at Developing Co-operative Training and Education Programmes.

Output/Activity Indicators that could be collected from the records of PRIs involved in co-operative training and education programmes:

- Per cent of total fees derived from vocational training courses, joint studentships etc. at higher educational establishments.
- Per cent of total post-graduate students paid for by firms according to area of science. Source:
- Per cent of all courses developed jointly with firms.
- Number and types of PRIs engaged in developing co-operative training and education programmes.

2.5.4.3. Policies Aimed at Increasing the Mobility of Research and Technical Personnel.

Output/Activity Indicators that could be collected from the records of PRIs involved in post-graduate training:

- Per cent of total research staff at PRI's taking up a position in industry.
- Per cent of all post-graduate students on industrial placements.
- Per cent of all post-graduate students finding employment in industry.

2.5.4.4. Policies Aimed at Encouraging Direct Commercialisation of Research Results through the Creation of Spin-Offs and Licensing of IPRs.

Output/Activity Indicators that could be collected from the records of PRIs:

- Numbers of spin-offs[26] created by PRIs by industrial sector. Source: Ad hoc country surveys reported in OECD (2000).
- Per cent of total PRI income attributable to spin-offs. Source:
- Age and survival rates of spin-offs.
- Per cent of total patenting attributed to PRIs by area of technology. Source: EPO data.
- Per cent of total PRI Income gained from licensing.
- Per cent of all PRI patents that are licensed.

2.5.5. Application of Evaluation

Evaluation of the UK Teaching Company Scheme (TCS).[27]

This long running scheme provides access to technology to firms and facilitates academic-industrial technology transfer by employing graduates to work on a specific R&D project of relevance to the firm. Between 1992 and 1994 SPRU[28] undertook an evaluation of the TCS aimed at investigating the effects of the scheme on industry and on academia. The aim of the evaluation was twofold: to assess the economic impact on the firm and to assess the academic impact on the participating academic institution.
The methodology consisted of:

- Collecting quantitative data on a selected number of programmes from the detailed records of kept by the scheme mangers. These consisted of end of award reports filed by the firm, the academic institution, the graduate employee, and the assessment of the Teaching Company Executive. Out of 500 programmes 95 were selected on the basis of ensuring representativeness of the full range of programme types, academic disciplines, firms of different sizes and industrial sectors, and universities from different geographic areas.
- Supplementary information obtained directly from 40 academic departments by means of questionnaires and interviews.
- Detailed case studies of 15 programmes.

---

[26] OECD (2000) defines spin-offs as: (i) firms that are founded by public sector employees in universities and other higher education institutions – including staff, professors and post-docs; (ii) start-ups which have licensed public sector technologies; and (iii) firms in which a public sector institution makes an equity investment or which are directly established by a public research institution.

[27] For more examples see Erik Arnold and Ken Guy; "Technology Diffusion Programs and the Challenge for Evaluation" in *Policy Evaluation in Innovation and Technology Towards Best Practices*, proceedings of an OECD Conference (1997), available at http://www.oecd.org/dsti/sti/stat-ana/prod/arnold.pdf They explain briefly: Inter-Regional Information Society Initiative (IRISI) giving advice to the European Commission; Specific Projects Action Line (SPAL) dealing with technology transfer; TEKES in Finland dealing with technology transfer involving Nokia; Technology Transfer and Partnership Programme (TTPP) in Ireland; KFB Biofuels in Sweden dealing with policy advice.

[28] Senker and Senker (1994), " Transferring technology and expertise from universities to industry: Britain's Teaching Company Scheme", *New technology, Work and Employment*, Vol. 9, pp. 81-92; Senker and Senker (1997), "Implications of industrial relationships for universities: a case study of the UK Teaching Company Scheme", *Science and Public Policy*, vol. 24, pp. 173-182

The most difficult part was obtaining the relevant quantitative information. For example the scheme records contained very little information on economic outcomes (only 38 out of the 95 programmes had data on cost savings achieved). Further there was no information on the indirect costs and benefits. Thus the message from the evaluation team was that the direct comparison of the costs and benefits of the scheme had to be treated with extreme caution.

Another part of the evaluation consisted of firstly outlining the main qualitative factors affecting the outcome of programmes in a systematic way. Each of these factors was then awarded a subjective score (on a scale 1-8) based on the information contained in the programme reports. The scores were awarded by at least two researchers working independently. The aim was to identify the main factors that determined the success of a programme, where success was defined by the various reports on each programme. The following factors were significant determinants of success:

- The closeness of the relationship between the various actors: the Company, University and the Teaching Company Associate (i.e. the graduate employee).
- The Company's commitment to the programme.
- The clarity and cohesiveness of the objectives.
- The efficiency of monitoring progress against agreed targets.
- Expertise of the academic partner relative to the company's existing capability.

The impact on academic institutions of participating in the scheme was assessed by means of 'matched pairs'. This consisted of identifying 17 specific departments who had a high participation rate in the scheme and matching them with 17 departments in similar disciplines with a low or non-existent involvement. Questionnaires were sent to individual academics within the 34 departments, with the aim of identifying the effects of participation on departmental performance, on individuals' performance and careers and on their attitude to industry. The topics covered included research grants, teaching activities and other industrial linkages, together with other wider benefits of participation in government sponsored collaborations with industry, such as new course development, industrial contracts, equipment acquisition, and student placements.

The number of usable replies was 428, representing a rather low response rate (24%). The analysis indicated that participation in the scheme had generally positive effects on departments' and individuals' involvement in activities such as EU programmes, industrial consultancy or research contracts, patents applied for and granted, and academic research grants. However these effects could not be attributed exclusively to participation in the scheme as they also occurred in departments that had strong links with industry not related to the scheme. The main message from this evaluation was that it is extremely difficult to isolate the effect of a single stimulus from a wide array of other variables which can have an impact on the participants.

2.5.6. Bibliography

A critical review of the literature on benefits of publicly funded research:

Martin, B., A. Salter with D. Hicks, K. Pavitt, J. Senker, M. Sharp, and N. Von Tunzelmann. 1996. *The relationship between publicly funded basic research and economic performance. A SPRU review.* Report prepared for HM Treasury: SPRU, Brighton

Salter, A., P. D'Este, B. Martin, A. Geuna, A. Scott, K. Pavitt, P. Patel, and P. Nightingale. 2001. *Talent, not Technology: Publicly funded research and innovation in the UK.* London, Committee of Vice-Chancellors and Principles and the Higher Education Funding Council of England.

A US study tracing the effect of the Bayh-Dole Act on the evolution of patenting by US universities:

Mowery, D., R. Nelson, B. Sampat and A Ziedonis. 2001. The effects of the Bayh-Dole Act on US Academic research and technology transfer. *Research Policy*. 30: 99-120.

Surveys of technology transfer activities of UK Universities:

Charles, D. and C. Conway. 2001. *Higher Education-Business Interaction Survey*, report n°1/68 to the UK HE Funding bodies, CURDS, University of Newcastle. www.hefce.ac.uk/pubs/2001/01_68.
Howells, J., M. Nedeva, and L. Georghiou. 1998. *Industry-Academic Links in the UK: A report to the Higher Education Funding Councils of England, Scotland & Wales*, PREST, Manchester. (http://les.man.ac.uk/PREST/Publications/)

Survey of Commercialisation activities of US Universities:

Massing, D. 2001. The AUTM survey: its development and use in monitoring commercialisation in North America. *OECD Science, Technology and Industry Review* 26: 58-73.

A major EC report on Industry-Science relationship

Polt, W., C. Rammer, H. Gassler, A. Schibany and D. Schartinger. 2001. *Benchmarking Industry-Science Relations : the role of framework conditions.* Report to the Federal Ministry of Economy and Labour, Austria and to the European Commission (Enterprise DG). Vienna/Mannheim

2.6. LEGAL FRAMEWORKS

Author: **Robin Cowan** (MERIT)

2.6.1. Introduction

Recent focus of RTD policy has been on legal frameworks surrounding outputs of the research process. In particular, focus on ways to improve co-operation of RTD actors, and improving interactions and interchanges between producers and users of new knowledge. Regulations and standards restrict behaviour and thus may restrict innovation, but on the other hand, regulations and safety standards protect social welfare against negative externalities while technical or interconnections standards permit capturing network externalities. A key to generating innovation may lie in encouraging co-operation in research among different actors. This can only happen if the regulatory framework permits it, so one avenue of policy has been to try to create a framework that is conducive to innovation.

It is a challenge to preserve an important balance: on the one hand giving statutory expression to the moral and economic rights of the creator over his or her creation, and on the other hand giving public access to and use of information. Consequently policies, legal and otherwise, exist to promote creativity and its dissemination and application, and to establish fair trade which in its turn contributes to economic and social development.

For example, intellectual property rights such as patents, trademarks and copyrights provide a formal framework which aims to optimise the return of innovation to society. Copyright laws protects expression, but not the underlying ideas. EU Data Base law protects a collection of data; patents permit their holders to exclude others from making use, offering for sale, or importing the claimed invention for a limited period of time. These legal frameworks can have a very strong impact on the competitiveness and the innovativeness of an economy and thus is an important aspect of public policy.

Policies concerning the legal framework surrounding research and technology development can be summarised in the following categories:

2.6.1.1. Policies Focussed on Public-Private Co-operation.

Recently there has been considerable pressure to improve the networking among different parts of the innovation system of Europe. One difficulty being addressed is that of the interaction between different types of institutions, each having its own research culture, incentives and aims. Research co-operation between public and private institutions is one goal. In the United States, the Bayh-Dole act eases legal restrictions on the appropriation of the results of research financed with public funds. This permits public institutions to move towards the culture of private institutions such as firms who aim to profit financially from their research. Early indications are that the Bayh-Dole act has at least made it possible for universities to attempt to profit from the research conducted by their faculty members. So far, however, no legislative initiatives in Europe have been taken to come up with legislation comparable to the Bayh-Dole act e.g. regulation affecting deployment of publicly funded research.

2.6.1.2. Policies to Encourage Collection and Dissemination of Information and Data

The goal here is to provide incentives for agents to create widely-accessible data sources which can be used as research tools. Within the EU, the EC Database Directive was enacted in 1996, though there is so far no legislation on this issue in the US. The

Database Directive provides two kinds of protection; copyright in original compilations and a sui generis "database right" for non-original contents of a database. Rights are provided "for the maker of a database which shows that there has been …substantial investment…, to prevent acts of extraction and/or re-utilisation of the whole or a substantial part…of the contents…" An EU database reassessment is planned to look at fair use, to encourage cross licensing practices, and to discuss compulsory licenses. Means of protecting the free dissemination of public domain data are for example, lower prices for private database access by academics, or category exemptions in IP laws to allow access to certain types of users.

2.6.1.3. Policies to Raise Awareness of Regulations

Within the EU, Research and Development (R&D) and Research and Technological Development and Innovation (RTD) are supported through the 5th and 6th Framework Programme in which guidelines to research organisations and also policy recommendations for governments are proposed. Furthermore the working papers are used as initiatives to bring awareness to the member states of how to use directives and regulations. (See for example various projects within the 5th and 6th framework such as "IPR Aspects of Internet collaborations" (working paper) and workshop report on "Managing IPR in a knowledge-based economy – Bioinformatics and the influence of public policy".)

2.6.1.4. Policies to Encourage Inter-Firm Interactions

These policies aim to reduce obstacles to firm interactions in pre-competitive research. They are based on the belief that part of the Japanese success in catching up was driven largely by inter-firm collaboration in pre-competetive research. This issue involves making it possible for firms to collaborate. This may involve addressing anti-trust issues. In Europe, it may also be affected by the issue of national borders and that different countries in Europe have different standards and regulations in this regard. Consequently, a big part of policy here is to create a harmonised regulation system within which collaboration can take place.

2.6.1.5. Policies Encouraging the Use of Patents

These arrangements aim to reduce the obstacles for companies applying and owning patents and trade marks. The measure especially matters for SMEs. Policies include a) reduction of the application fee b) simplification of the legal framework, intending to improve the understanding of the regulations and its impacts in order to reduce reservations about the patent system c) support for the maintenance of patents d) promotion of the benefits of patents, e) standardisation and unification of European patent regulations to encourage use of the patent system to reduce financial and administrative complexity.

2.6.1.6 Policies Aimed at Creating Incentives for Public Institutions to Use Patents

The commercial gains from patent ideas developed in public research institutions are often foregone due to administrative or economic considerations. Policies designed to increase the return on public investment in these institutions are manifold and include measures such as a) simplifying the regulative framework b) employing an incentive scheme for institutions which rewards innovative activities with financial means c) increasing the information flow about the features of the patent system and the potential benefits for these institutions.

2.6.1.7. Policies Designed to Increase the Diffusion of Proprietary Information

The diffusion of knowledge and information is very important for an economy to maximise benefits from innovative activities. Diffusion is of crucial importance in the knowledge-based economy and thus receives special interest. The patent system is only one means to encourage faster disclosure of information about inventions. This permits further innovation and development of the idea by agents other than the original inventor. Furthermore, public access to patent databases, and other sources of technical knowledge decreases the potential for inefficiencies arising from duplication of R&D efforts.

2.6.2. Policy Objectives, Instruments and Outcomes

The main objective of public policies towards intellectual property rights is to enhance the process of innovation thus to improve economic performance and competitiveness. The provision of intellectual property rights protection aims to create incentives that maximise the difference between the values of the intellectual property created and used, and the social cost of its creation. The intermediate goal is to increase use of the patent system for IPR protection and for information diffusion.

Ultimately there must be a balance between the dead-weight loss of the temporary monopoly on an invention and the benefits to society from the invention and the fast disclosure of information about it. Moreover, it is of the utmost importance that firms as well as for public institutions see enough economic evidence to compel them to engage in research and to use the legal framework. The connection between IPRs and the rate of innovation, and especially economic performance in general, is confounded with many other factors. Thus direct evaluation of these final goals is typically eschewed in favour of evaluation in terms of the intermediate goals of participation in, and use of, the IPR system.

2.6.3. Policy Evaluation

In the literature on policy evaluation the subject is subdivided into two approaches. The first evaluates the direct impact of a policy on economic performance. The conclusions from this methodology face the limitation that different impacts of relevant variables are difficult or even impossible to separate and interpret. A second method mentioned identifies the rationale for a policy. Capturing the impact of a certain policy entails the difficulty that the separation of effects triggered by politics is hard to distinguish from the influence of other statistically significant variables.

Policy evaluation tends to be one of two types: statistical analysis and analysis using surveys of users or peers. The first category, which tries to capture the impact of a policy ex-post, is recognised as the traditional method involving econometric analysis of changes in, for example, patent counts which can be attributed to policy initiatives. The difficulty in this approach lies in the modelling, and ensuring that all causal factors are included in the model, so that policy initiatives are not erroneously attributed effects which are in fact due to other factors such as global changes in economy, society or politics for example. In the second type of evaluation, involving user surveys and peer review, personal opinions and experience shape the outcome. Even though these techniques enable research on specific questions, they may lack objectivity and may not be universally valid. In both cases, for policies that are applied at the firm level (e.g. subsidies given to firms that satisfy particular conditions, as opposed to a general reduction in patent applications fees) it is important to include data on non-qualifying and/or non-participating firms to serve as a control.

2.6.4. Data Requirements

2.6.4.1. Policies Focussed on Public-Private Co-operation

- Data on participation of public-private consortia in Project are undertaken in the 5th and 6th Framework Programmes. Source: EU.
- Data on royalties earned by university and public research institutions for IPR rights. Source: surveys on revenue of public-private collaboration.
- Data on public-private licensing. Source: Surveys and firm press releases.

2.6.4.2. Policies to Encourage the Creation and Dissemination of Information and Data

- Numbers of firms using commercial data base provision. Source: surveys on potential users.
- Revenue from data sales. Source: surveys.
- Prices for data acquisition. Source: surveys of data providers.

2.6.4.3. Policies to Raise Awareness of Regulations

- Numbers of hits on websites offering information about regulations. Source: surveys before and after policy implementation.
- Number of legal cases contesting regulation. Sources: European Court of Justice and National Court records.

2.6.4.4. Policies to Encourage Inter-Firm Interactions

- Time series on numbers of RTD alliances in Europe; by sector, industry, location *Sources: MERIT-CATI database.*
- Time series on numbers of RTD alliances outside Europe; by sector, industry, location, Sources: MERIT-CATI database; surveys.
- Time series on numbers of RTD alliances between European and external firms; by sector, industry. Sources: surveys of firms; MERIT-CATI database.
- Data on dates of harmonisation regulation in countries in Europe. Sources: National regulatory, or legislative bodies.
- Numbers of anti-trust cases involving RTD activities. Sources: National Court records.

2.6.4.5. Policies Encouraging the Use of Patents

- Number of Patent application over time; Number of patents granted over time; Number of patents by class. Source: WIPO; EPO; National patent offices.
- BRD time series by patent class/ fields of science. Source: OECD Science and Technology Indicators
- Survey data on firm use of patents, and explanatory variables. Sources: CIS II; Yale survey; European versions.

2.6.4.6. Policies Aimed at Creating Incentives for Public Institutions to Use Patents

- *Number of patents applied for by public institutions.* Source: EPO; CIS II; OECD Basic Science and Technology Statistics.
- Percent of total income generated by licensing from innovations patented by public institutions. Source: EPO; WIPO.
- Number of licenses negotiated by public institutions. Source: EPO; WIPO.
- Percent of total patenting attributed to PRIs by area of technology. Source: EPO data.
- Percent of all PRI patents that are licensed. Source: dedicated surveys
- User surveys to investigate patents and other means to transfer knowledge. Source: CIS II.

2.6.4.7. Policies Designed to Increase the Diffusion of Proprietary Information

- Data on the use of the patent database. Source: CIS II; User surveys.
- Percent of patents licensed over time. Source: EPO; WIPO.
- Patent citation; cross references over time. Source: EPO; WIPO; National patent offices.
- Time series on the use of patent databases. Source: EPO.

2.6.5. Application of Evaluation

Few detailed evaluations of IRP policies exist, but the European Commission has done one on Quick Scan, part of the INNOVATION programme under DG XIII-D1. Quick Scan is a "novelty search" facility, a project funded by the INNOVATION programme of the European Commission. Its main objectives were a) to avoid wasting scarce research funds and b) to provide added-value to contractors[29].
The service was to provide information on the novelty of potential research projects. This supportive measure of the patent database access system was offered to 100 project contractors in 1995. The database was queried regarding the contractors proposed projects. The contractors received a report including full text copies of all relevant documents, as well as some background information and guidelines on possible responses to the search results. This information was used to evaluate the novelty of the proposed project, as background information, and as a source of potential partners for the project.
Quick Scan was evaluated as a project, and the report is available at http://www.cordis.lu/ipr/src/scan.htm and http://www.cordis.lu/ipr/src/scan1.htm.
The evaluation uses two techniques — survey of users and cost-benefit analysis — and the report includes three (one paragraph) cases illustrating the usefulness of Quick Scan. The cases are not, in fact, useful as anything more than an illustration of the kinds of things that might happen when Quick Scan is employed. More useful is the detailed questionnaire, which produced 9 pieces of data. Participants were asked about the effects of Quick Scan on their implementation of the relevant project. The report presents numbers of firms which: found new markets or applications; identified new partners; gained insight into the competitive situation; found licensing opportunities; found solutions from elapsed [sic] patents; found solutions relevant to the current project. (Note that the last two almost certainly contain the same information.) While all numbers are positive (and fully 50% of firms gained insight into the competitive

[29] See http://www.cordis.lu/irp/src/scan.htm

situation) it is difficult to evaluate the programme based on them since there is no comparison. There are many sources of information of the type listed here, and there is no way to gauge the effectiveness of Quick Scan relative to other sources. Probably it was relatively effective, but there is no way to tell. Adding a control group of firms that did not use the service, and asking whether, over the same time frame they had from some source, found new markets or applications for their research, found new partners, gained insight into the competitive situation etc. would have permitted a better evaluation of how much value Quick Scan added over the status quo.

More useful is the cost-benefit analysis. Here we see that costs were roughly 50% of the benefits. This is a very high rate of return, so regardless of the effectiveness of other information sources, Quick Scan looks valuable. No details of the cost-benefit analysis are given, simply the final figures. Because the analysis here was very simple, this may be enough in this case, but it should not be taken as an exemplar.

One serious shortcoming of this evaluation, or at least in the report published on the web, is that the authors of it are not named. (This is in contrast to the five year assessment of the INNOVATION programme in which the panel is named on the title page (See reference below).) Thus the reader has no idea whether there is likely to be bias in it or not. One suspects that this evaluation was performed by DGXIII-D1 itself, which raises the natural questions about the conflict of interest. On the other hand, the brevity of the report is exemplary.

2.6.6. Bibliography

Robert J. W. Tijssen and Joke C. Korevaar, "Unravelling the cognitive and interorganisational structure of public/private R & D networks: A case study of catalysis research in the Netherlands", Research Policy, Volume 25, Issue 8, January 1997, Pages 1277-1293

The study is aimed at identifying all public and private sector research organisations involved in the network, characterising their R&D output in terms of international scientific papers and patents, and describing and analysing relational and positional dimensions of their interorganisational network. The results provide an overview of Dutch activity within the world-wide cognitive landscape of catalysis R & D - from both a scientific and technological perspective.

Pari Patel, "IRP Policies and Innovation Indicators", Draft Report prepared for the Benchmarking Workshop on IPRs, TREND CHART project, SPRU, First Draft, February 2001.

This report outlines the rationale for public policies concerning IPRs; reviews the main policy within the EU; and investigates the relevance of innovation indicators research in relation to these policies.

Quick Scan - a novelty search service in the Framework of EU-R&D Programmes (1998), www.cordis.lu/ipr/src/scan1.htm#6.

Report provided on the homepage of CORDIS, the Community Research and Development Information Service, which is a free service provided by the European Commission's Innovation/SMEs programme. The report presents a detailed description about the nature and the results of the Quick Scan project.

Christoph Mandl , Antti Kiikka, Dan Maher, Giorgio Tuninetti, George Kakouros, "Five year assessment report related to the specific programme INNOVATION and INNOVATION-SME covering the period 1995-1999" (May 2000), www.crue.org/europaid/documentacion/docs/Otros%20documentos/5yasses_innov.pdf

This report concerns the external evaluation of the activities carried out within the INNOVATION Programme, as well of those carried out within the INNOVATION-SME Programme.

European Commission "Working paper on IPR Aspects of Internet Collaborations" research within the Fifth Framework Programme. This paper is based on a workshop report of Strata, Improving Human Potential Programme, held in Brussels, Belgium on 22-23 January 2001.

European Commission "Managing IPR in a Knowledge Based Economy: Bioinformatics and the Influence of Public Policy", research within the Fifth Framework Programme.

This paper is a workshop report held in Brussels, Belgium on 11-12 September 2001.

OECD "Policy Evaluation in Innovation and Technology, towards best practices"
www.oecd.org/dsti/sti/stat-ana/prod/piric.pdf

This Internet page contains a report, that brings together presentations by a range of international researchers and policy makers at an OECD Conference held on 26-27 June 1997.

europa.eu.int: Homepage of the European Union, providing a section about the information society. This section contains European Union law, implementation of policies and sources for information retrieval about the information society.

3. EVALUATION METHODOLOGIES

Authors: **Wolfgang Polt** and **Jaime Rojo** (Joanneum Research)

3.1. INTRODUCTION

The set of evaluation methodologies reviewed in the toolbox constitute recognised techniques particularly adapted to evaluate the socio-economic impact of RTD policies. Each of the evaluation methodologies presented has its own strengths and limitations, and the evaluator has to select them according to the particular requirements of the evaluation.

Evaluation methodologies have been adapted to capture the benefits of policies not directly measurable in monetary terms–for example, in the areas of health, environmental sustainability, consumer protection–. The techniques developed to measure the returns in these fields include contingent valuation and revealed preferences. Other techniques, such as hedonic prices, have been developed to adjust prices series to reflect changes in the quality of products, particularly relevant in high technology products subject to rapid quality changes.

Evaluation methodologies rely on modelling frameworks that describe the process of research and technological innovation. The design of evaluation methodologies based on economic theoretical foundations permits to relate output measures with the broader impact of the policies on economic welfare. Most of the times, with the purpose of simplifying the complex dynamics of scientific and technological development, evaluation approaches adopt the operation of the linear model of innovation. However, for some purposes and within certain contexts more complex models of innovation are used in evaluation. For instance, the measurement and description of the creation of knowledge clusters, industry-science relationships and private-public collaborative agreements, use more complex approaches such as national systems of innovation, chain linkage models of innovation and models of creative destruction.

The variables of interest in evaluation, outcome and impact, are captured indirectly by analysing how the outcomes of intervention change as a result to changes in the relationships and quantities of inputs and outputs. As evaluation approaches recognise, institutional variables are a fundamental factor shaping the effects of policy interventions. The different evaluation methodologies reviewed specially the qualitative ones indicate the relevance of the context dependence in evaluation.

Methodological approaches have been developed to deal with uncertainty and risk considerations for example, use of decision theory and expected return calculations in a cost benefit analysis context. Uncertainty considerations can be introduced in ex-ante evaluation to provide a more accurate assessment of the return on R&D investment.

3.1.1. Ex-ante Evaluation Methodologies

The methodologies employed in ex-ante evaluation of programmes and policies include:

- *Foresight studies*: this structured consensus building methodology based on experts judgements permits to anticipate social, economical and technological development opportunities in policy planning.
- *Modelling and simulation*: this quantitative methodology uses scenario modelling to estimate the socio-economic impact of policy.
- *Cost-efficiency techniques*: this judgement methodology quantifies the costs and benefits associated with the policy intervention.

- *Cost-benefit techniques*: this judgement methodology compares in monetary terms all social and private cost and benefits of a programme to establish whether the benefits exceed the costs. The technique can be adapted to incorporate uncertainty and risk.

- Ex-ante economic evaluation methodologies can be a successful mechanism to:

- improve the performance of interventions.
- allow funding in relation to merit.
- ensure that the rationale of the funded interventions has been analysed.
- analyse the expected benefit of an intervention.

3.1.2. Monitoring and Ex-post Evaluation Methodologies

Monitoring and ex-post evaluation uses a combination of qualitative, statistical and econometric techniques to analyse the effects of the policy intervention. The diversity of methodologies available for performing an evaluation are a signal of the multiple dimensions in which the impacts of policy intervention might manifest themselves. For this reason, no single best evaluation methodology exists. Each methodology will be fitted to analyse particular dimensions of impacts, but the best evaluation approach would require a combination of various evaluation methodologies possibly applied at various level of data aggregation. Proceeding this way will allow cross checking the robustness on the observed effects of the intervention.
In the evaluation of the impact of RTD policies, there exists an added difficulty requiring to devise methodologies which allow to capture the effect produced by what is inherently an intangible good, the production and diffusion of knowledge. Most of the methodologies therefore focus on capturing particular relevant aspects of the RTD process. For instance, microeconometric methodologies allow to evaluate the existence of an additionality effect in public support on private R&D and to capture the private rate of return on R&D. Macroeconomic methodologies are better fitted to capture the generated R&D spillovers and longer term wider effects of the policy intervention on productivity and economic welfare. At a more disaggregated level, cost-benefit analysis, although not without difficulties, permits to transform all the benefits and cost of a project or programme intervention into monetary values and compare them. Network analysis and case studies are better fitted to capture the richness of impacts enabled in the RTD process, allowing to frame the evaluation in the broader socio-political context, and analyse accordingly the relevance of the social connections in the process. These methodologies are particularly relevant for the evaluation of institutions and systems.

The methodologies employed in ex-post evaluation of RTD programmes and policies include:

Statistical data analysis

- *Innovation Surveys*: provides basic data to describe the innovation process, summarised using descriptive statistics.
- *Benchmarking* allows to perform comparisons based on a relevant set of indicators across entities providing a reasoned explanation their values.

Modelling methodologies

- *Macroeconomic modelling and simulation approaches*: allows to estimate the broader socio-economic impact of policy interventions.
- *Microeconometric modelling*: permits to study the effect of policy intervention at the level of individuals or firms. There are mechanisms to control for the counterfactual by specifying a model which allows to estimate the effects on the outcome of the participant had the programme not taken place.
- *Productivity analysis*: permits to assess the impact of R&D on productivity growth at different levels data aggregation. This is particularly relevant to analyse the broader effects of R&D on the economy.
- *Control group approaches*: allows to capture the effect of the programme on participants using statistical sophisticated techniques.

Qualitative and semiquantitative methodologies

- *Interviews and case studies*: uses direct observation of naturally occurring events to investigate behaviours in their indigenous social setting.
- *Cost-benefit analysis*: allows to establish whether a programme or project is economically efficient by appraising all its economic and social effects.
- *Expert Panels/Peer Review*: measures scientific output relying on the perception scientists have of the scientific contributions made by other peers. Peer review is the most widely used method for the evaluation of the output of scientific research.
- *Network Analysis*: allows to analyse the structure of co-operation relationships and the consequences for individuals decisions' on actions providing explanations for the observed behaviours by analysing their social connections into networks.
- *Foresight/ Technology Assessment*: used to identify potential mismatches in the strategic efficiency of projects and programmes.

Tables 11 and 12 attempt to briefly summarise the wealth of information contained in each of the evaluation methodologies modules. For a deeper appreciation on the methodology of interest it is advised to refer directly to the information provided in the detailed modules.

Ex-post economic evaluation methodologies have proved a successful mechanism to:

- determine the efficiency and efficacy of the intervention (e.g. macro economic simulation and productivity studies).
- provide a quantitative estimation of the impact of the intervention (e.g. microeconomic evaluation studies).
- quantify the various dimensions in which returns should be considered within a defined framework.
- assess environmental sustainability and health issues (e.g. cost-benefit analysis), organisational impact (e.g. case studies, network analysis, innovation studies), strategic impact (e.g. foresight).

3.1.3. Modular Structure of the Toolbox

Each of the eleven evaluation methodologies described in the evaluation toolbox is structured to cover a common set of topics with the aim of facilitating comparisons across methodologies.

- *Introduction*: The section provides a general description of the evaluation methodology.
- *Application to Policy Instruments and Interventions*: The section discusses the policy instruments evaluated with the methodology.
- *Good Practice Examples*: The section illustrates with practical examples how the methodology as been used in RTD evaluation at different levels (firm, industry, aggregated). Examples of good practices of method application.
- *Conditions for Application*: The section provides indications on the costs of implementation, complexity of use, phase of application (i.e. ex ante, intermediate, ex-post evaluation), degree of methodology acceptance.
- *Steps for Application*: The section provides an illustrated example on how the method is operationally used in evaluation and comments the relevant concepts generally addressed (e.g. additionality, spillovers, cluster formation, rate of return, etc) and on particularly relevant methodological aspects to consider (e.g. panel data, non-linearities, time series)
- *Data Requirements*: The section provides an indication of the input, output, outcome/impact indicators available to measure the output and the outcome of the policy instrument.
- *Scope and Limits*: The section discusses the scope and limitations of the methodology use in evaluation.
- *Bibliography*: The section provides references for obtaining further insights on the methodology.

The following tables provide an overview of the methodologies addressed in the toolbox.

Table 11: Evaluation Methodologies

Methodology	Type / Use	Data Requirements	Strengths	Limitations	Good Practices
Innovation Surveys	Semi-quantitative Quantitative *Monitoring* *Ex-post*	Micro data Expenditures Profits Patents, Innovation	Detect innovation trends and insights on the soft side of innovation. Findings from interviewed sample can be generalised to the population Permits to identify size and distribution of impacts Provides groups comparisons and changes over time	High cost and time consuming Processing and analysis of data requires large human resources Some types of information are difficult to obtain Long time series generally not available	Analysis of the innovation process using data on the EU Community Innovation Survey
Micro Methods	Quantitative qualitative categorical data *Monitoring* *Ex-post*	Micro data Expenditures Profits Patents	Results based on explicit formulation of theory based causal relationships R&D Additionality Control for different effects: firm size, expenditures, innovation capacity	Quality of data Persuade participant and non participant entities to disclose information Only private rate of return to R&D	Effects of public R&D subsidies on firms in West Germany Evaluation of the ITF Programme FlexCIM Effects of R&D subsidies in Spain Promotion of AMT technologies based on Swiss Micro data
Macro Methods	Quantitative modelling methodology *Ex-ante (simulation)* *Monitoring* *Ex-post*	R&D Expenditures R&D output Macroeconomic data	Social Rate of return to R&D Capture R&D Spillovers Estimate long term policy intervention impact Scenario simulations for policy supported geographical areas	Average returns Robustness of results Time lags for observation of the effects	Modelling approaches: OECD Interlink, IMF Multimod, EU Quest. R&D Spillover studies: Jaffe, Nadiri International spillovers: Eaton and Kortum, Mohnen, Evenson
Productivity Studies	Quantitative modelling methodology *Monitoring* *Ex-post*	Micro data Expenditures Profits R&D, Patents	Estimation of effect of R&D on productivity Estimate the rate of return to R&D	Quality of data Deflation of series Required assumptions for measurement of stock variables	Productivity studies (Van Ark) Growth accounting (Griliches, Jorgenson) Micro datasets: French INSEE and US Census of Manufacturers
Control group approaches	Quantitative *Ex-post*	Micro data Expenditures Profits Patents	Capture the impact of policy intervention on the programme participant entity	Requires high technical capacity High Implementation Cost Data Demanding	Collaborative industrial Research between Japan and US Evaluation of RTDI instruments in Ireland Participation of Ireland in European Space Agency
Cost Benefit Analysis	Quantitative (with qualitative elements) *Ex-ante (especially)* *Monitoring* *Ex-post*	Micro data Profit & cost estimates	Provides an estimate of socio-economic effect of intervention Good approach to assess the efficiency of an intervention Addresses by making them explicit all the economic assumptions of the impact of the intervention	Requires high technical capacity Some degree of judgement and subjectivity, depends on largely on assumptions made Not easily comparable across cases Careful interpretation of results when benefits are not easily quantifiable in monetary terms	US Advanced Technology Programme US National Institute of Standards Methodology
Expert Panels /Peer Review	Qualitative Semi-quantitative *Ex-ante* *Monitoring* *Ex-post*	Project programme data	Evaluation of scientific merits Flexibility Wide scope of application Fairness	Peers independence Economic benefits not captured	Evaluation of Large Infrastructures Evaluation of EU Programmes
Field /Case studies	Qualitative Semi-quantitative *Monitoring* *Ex-post*	Project programme data	Observation of the socio-economic impacts of intervention under naturalistic conditions Good as exploratory and descriptive means of investigation Good for understanding how contexts affect and shape impacts	Results not generalisable	Telematic innovation in the health care sector. Evaluation case studies reviewed in Georghiou and Roessner (2000)
Network Analysis	Qualitative Semi-quantitative *Ex-post*	Project programme data	Comprehensive empirical material Compilation for policy purposes Co-operation linkages	Time involved in collecting the survey information Persuasion requirements	RTO systems Interdisciplinary centers of medical research
Foresight/ Technology Assessment	Qualitative Semi-quantitative *Ex-ante* *Monitoring*	Qualitative data Scenario	Consensus building to reduce uncertainty under different scenarios Combination on public domain and private domain data Articulation and road mapping of development of new technologies	Impossibility to detect major RTD breakthroughs	Benchmarking of ISI/FhG capacities against Foresight results
Benchmarking	Semi-quantitative *Ex-post* *Monitoring*	Science and technology Indicators	Comparison method across different sectors Support to systemic evaluation of institutions and systems	Data detail requirements Non transferable	EU Benchmarking national policies Innovation Trend Chart Science-industry relationship

Table 12: Evaluation Methodologies II

Methodology	Data application level	Areas of application	Output	Outcome	Impact
Innovation Surveys	Firm Industry Economy-wide	Innovation IPRs Technology transfer Research collaboration	New products and processes Increase in sales Increase in value added Patent counts, IPRs	Creation of new jobs Innovation capacity building	Enhanced Competitiveness Institutional and organisational efficiency, Faster diffusion of Innovation Employment
Micro Methods	Plant Firm Industry Economy-wide	Sectoral Returns to R&D	Output and value added (collect baseline info for before-after comparisons)	Sectoral productivity industry sectoral spillovers Additionality, Leverage effects	Firms competitiveness
Macro Methods	Firm Industry Economy-wide	Sectoral Regional Economy-wide	Output and value added	Change in R&D Capital, Human capital, Social capital International R&D Spillovers	Regional, country productivity Employment, Good governance Economic and social cohesion
Productivity Studies	Plant Firm Industry Regional Economy-wide	Sectoral Regional Economy-wide	Output and value added	knowledge, geographical and International R&D Spillovers	Regional, country productivity Employment Economic and social cohesion
Control Group Approaches	Firm Industry	Technology implementation Innovation	Output and value added (on supported and non supported firms)	Additionality Rate of return to R&D	Firm, industrial competitiveness
Cost Benefit Analysis	Firm Industry	Health Environment Energy Transport	Value added benefit-cost ratio IRR Consumer surplus	Health improvements Consumer protection Environmental sustainability	Quality of life Standard of living
Expert Panels/ Peer Review	Firm Industry Economy-wide	Scientific merit Technological capacity	Publication counts Technological output	Scientific and Technological capabilities	R&D performance
Field/ Case Studies	Firm Industry	Science-industry relationships	Detailed inputs and outputs	firms RTD capabilities on the job-training educational schemes	Industrial competitiveness Quality of life Organisational efficiency
Network Analysis	Firm Industry Regional	RJVs, cooperation science industry Clusters	Co-operation linkages	Co-operation in clusters Social embededness	Efficiency of institutional relationships
Foresight/ Technology Assessment	Institution Regional Economy-wide	Technology Trends	Identification of generic technologies Date of implementation	Technological capacities	Technological paradigms shifts
Benchmarking	Firm Industry Economy-wide	Efficiency of technology policy	S&T indicators	Technology capabilities	Industry competitiveness Good governance

3.2. INNOVATION SURVEYS

Authors: **Georg Licht** (ZEW) and **Giorgio Sirilli** (CNR)

3.2.1. Introduction

Over the last three decades an enormous effort has been undertaken to generate new measures of innovation. International standardisation and harmonisation of innovation measurement evolved. OECD issued several manuals dealing with the measurement of various aspects of the science and technology which are now called the FRASCATI-family of manuals (i.e. FRASCATI manual, patent manual, TPB-manual, OSLO-manual). These manuals extensively discuss the measurement techniques and interpretation of various indicators. Moreover, data gathering efforts increased and, now, a huge stock of internationally comparable data are available. Innovation Surveys are conducted regularly in most OECD countries since the beginning of the nineties. EUROSTAT took the lead role within the EU and co-ordinates national endeavours to improve innovation measurement launching the Community Innovation Surveys (CIS). Up to now three harmonised innovation surveys have been conducted comprising more or less all member states as well as some non-EU countries (e.g. Hungary, Slovenia).

With the multi-annual programme for the production and development of Community statistics on science and technology the EU commission supports the conduction of regular surveys on innovation. The purpose of this programme is to support and monitor the science, technology and innovation policy in the Community and to analyse the effects on competitiveness, employment, economic growth and patterns on trade. This programme aims at establishing a common framework for the collection and compilation of harmonised data on innovation in the EU member states. The programme also sets standards for the production of statistical variables, development of new indicators and aspects related to quality in the data and survey methods.

Innovation data allows to shed light on various aspects of the process of innovation at the enterprise level as well as at the industry and the economy level. As innovation is hard to grasp various measurement approaches should be used simultaneously. In addition, recent literature suggests that we should view innovation as an interactive process with various feed-backs loops which implies that there are sometimes problems of what are the inputs and what are the output of innovation processes. Moreover, the time structure of inputs and through-puts of innovation processes depend on the nature of the innovation itself. E.g. patents can be viewed as a measure of through-put when we look at an elementary stage of innovation whereas patents are maybe an input in later stage of innovation. Having in mind this general caveat we can sort innovation indicators by their typical function within innovation processes at the firm level. In addition, innovation is embedded in a complex system.

Hence, innovation surveys are primarily a data collection tool and not an evaluation method per se. However, innovation surveys have recently been used in recently by various researchers to look at several aspects of the impact on public R&D policy measures on innovation processes. This efforts can be viewed as starting points for using innovation surveys as a evaluation tool. And we should expect that the increasing availability of innovation survey data will lead to an increasing use of innovation survey data for assessing public R&D policy.

3.2.1.1. Innovation Surveys Based on the OSLO-Manual

At the end of the seventies research groups in various countries start to gather statistical information on innovation (see Archibugi and Pianta 1996, Smith 1989 for early developments). The early data collection efforts take two routes: First, highly significant innovations (primarily products) are identified and then augmented by data of the firms. This approach, called the object approach, focuses on individual innovations. Second, data are collected on the level of firms to identify firms with and without innovations. This approach – the subject approach – focuses on innovating firms.

Based on this early experience the OECD took the lead in standardising the existing endeavours of various research groups. This first standardisation attempt resulted in the OSLO-Manual which focuses on the subject approach which was later on been adopted by an increasing number of countries. In 1992 a first harmonised innovation survey in the EU took place which was based on the first edition of the OSLO Manual. Since, then two additional EU-wide innovation surveys took place (CIS II for 1996 and CIS III for 2000). Experience with these surveys shows that surveys on innovation are not only feasible but yield extremely interesting and useful results for research in innovation as well as for innovation policy making.

At present most innovation surveys are based on the first revision of the OSLO manual which was published in 1997. The OSLO manual focuses on innovation at the firm level. The construction of this measurement approach is highly influenced by Schumpeterian ideas and the chain-link model of innovation which views innovation as the result of market opportunities, appropriability conditions, technological opportunities and the firm's internal capabilities base and external linkages. In recognising the increasing importance of innovation in services the 1997 revision of OSLO manual also deals explicitly with collecting data on innovation in services.

For each round of the community innovation surveys (CIS) a harmonised core questionnaire was developed, it comprises a set of "core" questions which are used by all countries along with some optional questions to be implemented in the national surveys. Data are collected for technological innovation (= new products and processes) only. The survey does not attempt to collected data on organisational innovation as it is considered to be another type of innovation even though the two go often hand in hand. Furthermore, organisational innovation is hard to define in broad based survey addressing firms from all sectors and size classes. It should be pointed out that in the questionnaire of the CIS III a question has been introduced on strategic and organisational changes introduced by innovating firms (strategy, management, organisation, marketing, aesthetic changes).

An innovation survey is an instrument to collect data about innovation (products and new processes new to a firm) in the business enterprise sector. In principle all firms, organisations and institutions whose primary activity is the market production of goods and services for sale to the general public at an economically significant price. Firms are considered innovative if they have introduced a new or improved product or process over a three-year period. The product or process introduced should be new to the firm, and not necessarily to the market: this means that the survey covers also the diffusion of innovation.

The sectors to be covered varies between national surveys. However, most of the surveys adopted the statistical classifications of economic activities in the European Community (NACE Rev.1). The statistical unit to be used is the enterprise or the kind-of-activity unit (KAU). The statistical unit is in principle the smallest legal unit. Data are collected by mail surveys. Depending on the national legal environment these surveys are either compulsory or voluntary. Surveys are typically carried out on stratified samples – in some cases all large firms are covered.

3.2.1.2. Key Issues of Innovation Process Addressed by Innovation Surveys

Key issues on which data are collected regularly comprise (for more details see OSLO Manual and the EUROSTAT core questionnaire for CIS III which can be found in the appendix):

- factors influencing technological innovation,
- innovation activities and expenditures,
- characteristics of innovating firms and
- impacts of innovation.

Factors influencing technological innovation comprise data on corporate strategies (objectives of innovation), sources of information (e.g. scientific institutions, competitors, etc.) and obstacles to innovation. These are key areas on which innovation surveys collected data. As a rule questions on these topics comprise a long list of items where firms should give their views based on a binary or a Likert scale (range of importance e.g. high, medium, low, and not relevant).
Innovation activities and innovation expenditures comprise all those scientific, technological, organisational, financial, and commercial steps which actually, or are intended to lead to the implementation of technological new or significantly improved products or processes. These activities comprise R&D, acquisition of knowledge (through licences, technological co-operation, technical services, etc.), acquisition of machinery and equipment, various other preparations for production (including tooling up, training of staff) and also marketing efforts. In addition, innovation surveys also seek to collect data on how firms protect their knowledge assets by e.g. intellectual property right or secrecy.
Characteristics of innovating firms comprise size (measured by sales, employees), sector of economic activity, ownership, linkages with other firms and public sector institutions, investments in tangible goods.
Impacts of innovation are probably the most important topics of innovation surveys. This impacts are measured by sales due to new products as well as sales due to products which are not only new to the firm, but also new to the market. Innovation surveys also collect data on export activities in order to analyse the impacts of innovation in international competitiveness at the firm level. Finally, some innovation surveys also collect data on the impact of innovation on employment or the skill structure of employees.
Finally, innovation surveys increasingly often collect data on the usage of *government innovation policy schemes*, the appraisal of these schemes and the impact of participation in government policy programmes on innovation inputs and impacts of innovation.

3.2.1.3. Using Innovation Survey Data for Evaluation of Public Innovation Policies

In constructing innovation indicators the information needs of policy makers and analysts are of paramount importance. Innovation surveys are part of a large system of data collection efforts addressing policy makers and policy analysts information needs. Innovation surveys see private sector innovation activities as a complex, systemic phenomenon with a special emphasis on the interplay between various institutions, the interactive nature of innovation process both in the creation of knowledge and in its diffusion and application. However, innovation surveys are not

built upon a particular theoretical model of the innovation process but are based on a very loosely defined picture of the innovation process within and between companies.
After nearly twenty years of development, innovation surveys are now an established tool for addressing information needs of policy makers and policy analysts. Innovation surveys can be used in a variety of ways in the context of evaluation of public innovation policies. The different ways will by shortly described in the next paragraphs. For this purpose we assume the innovation survey contains at least questions on whether a firm took part in Government programmes or not. This is the case e.g. in the third community innovation survey which distinguishes government innovation policy at various levels of government (EU innovation policy schemes, country-wide schemes, local/regional schemes). Some innovation surveys (like the Italian CIS-I survey) contain much more detailed information on government policy programmes. As these more detailed questions are not wide-spread in innovation surveys we omit possible ways to use this information for the evaluation of public policy. However, we should note that existing innovation surveys have shown that it is possible to collect even more detailed information about government innovation policy than just the fact of firm's participation in government programmes.

3.2.1.4. Analyses of the Structure of Participants in Government Programmes

Data from innovation surveys allow to uncover the structure and characteristics of participants in government innovation policy programmes. Hence, they can provide answers to questions who profits from government R&D subsidies etc. E.g. it allows to answer the question whether government R&D policies reach SMEs or whether the bulk of R&D subsidies favours large firms. The advantage of innovation surveys in this context is that they allow an assessment of the system of government policy measures as a whole and not only at the individual programme level. If regional variations in the "supply" of government innovation policy schemes existed innovation surveys can shed light on the relation of this regional variation on the one hand and the degree of participation and various structural characteristics of participants on the other. The simple comparison of participants and non-participants also allows the analysis of inter-industry difference in the spread of innovation policy measures.
More importantly, a detailed analyses of whether government programmes primarily reach the better firms (e.g. those with a higher productivity or higher export share) or whether government programmes is primarily used by inferior firms. One also could distinguish participants and non-participants with regard to the importance of various innovation obstacles. This allows some rough answer to the question whether government support reaches those firms with a low internal resource base or limited amount of financial resources or not. This type of information on the structure of participants can be of value when designing additional government support measures.
These types of analyses can be performed using simple cross-tabulation techniques or more refined multivariate econometric methods like probit models.

3.2.1.5. International Comparisons of Participants in Government Programmes

Due to the international harmonisation of core questionnaires innovation surveys allow an international comparison on the structure of participants in government innovation programmes. So, this data enables some first insights about the relation of national support structures and innovation activities of private firms (e.g. which structure of government support measures stimulates SMEs participation in government fostered R&D).

3.2.1.6. Using Innovation Survey Data for Econometric Analyses of the Relation of Government Support and Innovation Input, Innovation Processes and the Impact of Innovation

More sophisticated and often more interesting insights into the impact of government support on innovation inputs, innovation process and innovation outputs can be derived if we apply econometric methods (for more details see the chapter on econometric evaluation methods in the manual) when using innovation survey data. Large scale innovation surveys address not only innovative firms but also non-innovative firms. Innovation surveys cover firms with and without government innovation subsidies. Especially, as innovation surveys are not conducted to evaluate a specific programme, it is unlikely that strategic answers by respondents toward the impact of government programmes are provided. E.g. it is less likely that programme participants rate a programme as successful just because of the fact that they have received government money for a specific project. Hence, innovation surveys probably give a more realistic picture of the impact of governmental innovation support measures. Even more, we are not only restricted to a few indicators of innovation activities. Innovation survey data allow a more holistic analysis of the impact of government programmes as we can analyse the impact on innovation inputs, networking activities (sources of knowledge, R&D co-operation) as well as on innovation outputs and employment effects.

3.2.1.7. Combining Innovation Survey Data with Other Data Sources

Despite the above mentioned “virtues” of the innovation survey data for evaluating government support there is probably a severe disadvantage. Policy makers are less interested in evaluation results concerning the whole structure of government support but need results for specific innovation policy programmes. One way to overcome this shortage is to extend the information about government support in the questionnaire. However, this will increase response burden of firms and hence reduce response rate to the survey. Hence, a more promising approach is to combine innovation survey data and data from government records about innovation support at the firm level. Merging this type of innovation will probably the best way to exploit the full potential of innovation surveys for evaluation purposes.

However, there is a caveat here. Evaluating the specific effect of a programme on the firm’s performance is intrinsically very difficult, but, more important, it implies that the respondent adopts an approach which is different from the one of the innovation survey - which looks at the innovative strategy and performance of the whole firm over a three year period.

3.2.2. Application to Policy Instruments and Interventions

As it follows from section II, innovation survey data can be used to evaluate several dimensions of government impact. In principle, innovation survey data can be used to evaluate R&D subsidies (e.g. tax credit), specific project based programmes, programmes to foster networking and co-operation between public research and private enterprises (e.g. technology transfer measures), and – depending on the questionnaire – allow to shed some light on the impact of various support measures on economic variables (e.g. exports). Hence, a wide range of policy interventions that target firms are possible evaluation candidates. A properly conducted innovation survey not only allows an analysis of the impact on participating firms but provides an opportunity to evaluate the impacts of government programmes on non-participating firms (evaluating the impact of spill-overs). However, most existing innovation surveys are not specifically designed as an evaluation tool but rather as a tool for providing information about the structure of national systems of innovation. In addition, as a rule innovation surveys address the system of government support and not specific programmes. As far as the analyst wants to shed light on a specific programme innovation survey will be less useful. Yet, innovation surveys can be refined to be a more powerful instrument for evaluation in the future. Presently, the information contained in standard innovation surveys should be best viewed as augmenting specifically designed surveys to evaluate single innovation policy instruments.

3.2.3. Good Practice Examples

We selected five studies which show the range of possible applications of innovation surveys to evaluate government innovation policies. These studies comprise only examples from EU member states as innovation surveys have been conducted most often in Europe mainly due to EUROSTAT endeavours. Our examples should demonstrate that innovation surveys are not a evaluation tool per se but give useful information which can be exploited by various statistical techniques.

3.2.3.1. M. Pianta and G. Sirilli (2001): The Use of Innovation Surveys for Policy Evaluation in Italy

This study uses data collected 1993 within the CIS I framework. However, the Italian innovation survey extended the core questionnaire of the CIS I surveys and included a question with respect to a variety of government intervention mechanisms (Government funds, EC funds, Indirect financial incentives, Public procurement of research and investment goods, Public R&D services etc.). Firms were asked to rate the usefulness of these types of government intervention on a five point Likert scale ranging from “little importance” to “crucial”. The paper also looks at various cross tabulations of the use and the usefulness of government intervention by size class, industry, etc. This study also shows the correlation between the use of government incentives and the importance on the one hand and sales and employment growth on the other hand. The statistical methodology applied was a simple cross-tabulation and univariate distributions. The paper shows that most public actions and funds have been targeted to the largest firms in high-tech sectors. The authors conclude that existing policy tools are not adequate for supporting the innovation needs of smaller firms. In addition, the authors argue that government incentives should be redesigned in order to focus more directly on product innovation addressing capacity expansion and global markets.

3.2.3.2. H. König and G. Licht (1997): Participation in Technology Policy Programmes in Germany

This study employs innovation survey data for the year 1994 from Germany. This innovation survey largely follows the CIS-I methodology but extends the CIS I core questionnaire by some questions on whether firms received government R&D

subsidies. This survey distinguishes R&D subsidies by the government level which runs the subsidy programme (EU, federal and regional). The paper looks at R&D performing firms only. The paper applies multivariate methods to analyse factors influencing the participation of firms in government R&D programmes as well as possible effects of R&D subsidies on the outcome of innovation processes.
It is shown that the are significant differences between government programmes with regard to the structure of participants. EU and programmes ran be the Federal government are highly selective and mainly firms with a large internal knowledge base and specific skills regarding the conceptualising R&D projects and project proposals are able to take part. Regional R&D subsidies address different firms and are less selective with regard to firm internal capabilities. As a consequence the simple comparison of participants and non-participants in government programmes will probably lead to wrong conclusions. In addition, the paper shows that government grants also have some impact on the innovation output (measured by labour productivity or new product introduction).

3.2.3.3. M. Almus and D. Czarnitzki (2001): The Effects of Public R&D Subsidies on Firms' Innovation Activities in a Transition Economy: The Case of Eastern Germany

This paper applies econometric evaluation methods to innovation survey data from Germany. The paper does not evaluate a specific programme but looks at the total structure of innovation support measures of firm's innovative activities in Eastern Germany. The paper is based on the presumption that public R&D schemes are intended to stimulate the privately unprofitable R&D projects which would not be carried out without public funding. The paper wants to establish whether firms that receive funds may simply substitute public for private investment, thus leading to an inefficient resource allocation. So, the paper asks whether as a whole public support is crowding-out private R&D money. The paper uses data come from different years of the German Innovation survey which is conducted on a yearly basis as a panel study. As such there is a systematic difference to other European innovation surveys which are conducted on a two-years or a four years basis.
The variable of main interest in the paper is innovation intensity (ratio of innovation expenditure to sales). The paper uses a econometric matching approach, comparing the innovation expenditure to sales ratio for firms with and without a public subsidy.
The paper concludes that firms that received public funding show a larger innovation expenditure to sales ratio than firms without a subsidy. It is also concluded that the positive impact of government support on innovation expenditure declines with firm size.

3.2.3.4. I. Busom (2000): An Empirical Evaluation of the Effects of R&D Subsidies"

This paper looks at the impacts of EU-R&D subsidy programmes on the R&D activities of the supported firms at the year 1988. The data set combines innovation survey like data and data coming from the "Centro para el Desarrollo Tecnologico e Industrial (CDTI)", an agency of the Spanish Ministry of Industry. The key variables of interest are the absolute and relative effort in research and development (R&D expenditure, R&D personnel; ratio of R&D expenditure to sales, ratio of R&D personnel to total employment). In econometric terms the paper employs a two-equation framework where the first equation explains the probability of participating in the R&D programme and the second equation examines the level of firms' absolute and relative R&D effort respectively as a function, among other things, of programme participation. The participation equation contains following explanatory variables: firm size (employment), firm age, exports share of sales, type of ownership (public, foreign), pricing behaviour (regulated prices, monopoly, etc.), type of R&D strategy (increase of a firm's own R&D in response to a rival's increase), number of patents obtained in the last ten years and industry dummies. The equations of R&D effort included, besides most of the independent variables of the participation equation, a dummy for participation in R&D programmes of the European Union.

Surprisingly, the study finds that smaller firms have a higher chance of participating in the programme than larger ones. The same holds also for foreign firms in comparison to domestic ones. On the whole, public funding had a positive impact on R&D efforts of the programme participants, but for a sizeable portion (about 30%) of supported firms complete crowding-out effects between public and private investment cannot be ruled out.

3.2.3.5. Czarnitzki, D. and A. Fier (2001): Do R&D Subsidies Matter? – Evidence for the German Service Sector

The study uses data from several waves of the Mannheim Innovation Survey which corresponds the three German DIS surveys. The paper concentrated on service sector firms. In order to overcome the problem that innovation surveys do not contain sufficient information of government R&D incentives innovation survey data and data coming from government records about project grants are merged at the firm level. Based on a merged sample the authors use sophisticated econometric techniques to establish whether private firms reduce their financial contribution to their R&D budget when they receive government R&D grants. The paper found out that government R&D grants show a significant incentive effect i.e. firm's increase their own financial contribution to R&D when they receive government money. The increase of the firms internal R&D budget is much larger than the necessary matching fund which is a standard requirement when the government is co-financing R&D projects (e.g. 50% of project costs must be covered by the own fund). Hence, government support in Germany shows a high degree of additionality.

3.2.4. Data Requirements

Innovation surveys should be based on data coming from a general firm register which is available for most EU countries. If firm registers are not existent or not available, researchers should use a data base which comes close to firm register data. In any case the main advantage of an innovation survey is that it addresses the universe of firms and not only firms undertaking specific innovations (e.g. R&D performing firm only).

Innovation surveys should be accompanied by a short non-response survey for those (randomly selected) firm which do not take part in the main survey. Non-response data can uncover selectivity in response behaviour of firms.

To exploit the full potential of innovation surveys evaluators/analysts need access to micro-level data and not only to aggregate tabulation of innovation survey results.

Innovation surveys should either contain specific questions on government incentives or it has to be possible to add information about government incentives to innovation survey data at the micro-level.

With respect to the indicators required, the following aspects have to be outlined:

- Innovation surveys offer a wide-range of innovation indicators. The selection of a certain indicator as evaluation tools depends on the government intervention (its main goals and side-effects) of interest.
- Innovation surveys offer a wide range of indicators for inputs into innovation process, the organisation of the innovation process, through-puts and outputs of innovation processes. Existing government interventions are designed to affect at least one of these dimensions of innovation. Often government interventions aim at influencing more than one dimension. E.g. the "PRO-INNO" program in Germany aims at extending the resources of a firm available for innovation and also tries to improve linkages between firms as well as between firms and public research institutions. Hence, we could use the innovation intensity or R&D intensity (measured as (1) expenditure for innovation or R&D expenditure to sales ratio or (2) R&D employee to total employment ratio) and the importance of public research institution as a knowledge source (respectively its change) as goal variable in econometric evaluation models.
- Other possible indicators developed by innovation surveys are (see core questionnaire for CIS III for details given in an appendix):

Innovation output indicators: Probability of market novelty; sales share of market novelties; sales share of new or significantly improved products; export to sales ratio; change in sales or change in employment; impacts of innovation;
Innovation input indicators: Expenditure on innovation; structure of the expenditure of innovation; R&D expenditure; number of R&D employees;
Through-put indicators: Number of patents; probability of patent application;
Innovation process indicators: Sources of knowledge, pattern of co-operation in innovative processes; Productivity of the innovation process measured by e.g. patent to R&D ratio; Sales to market novelties to R&D.

3.2.5. Conditions for Application

Innovation surveys should be based on random samples drawn from the universe of business sector firms.
Innovation surveys should contain some questions on government intervention or it has to be possible to merge information about participation in government programs to innovation survey data at the micro-level.
Size of the sample available for evaluation has to be large enough. However, there is no general rule on this topic. The use of innovation surveys is most appropriate when the number of participants in government programs is large and when specific information about a programme is available. The sample site needed depends primarily on the statistical / econometric method to be used for evaluation.

3.2.6. Steps for Implementation

We distinguish the following operational steps:

- *Construction of questionnaire* (including the definition of questions on public support or on the link to external data sources);
- Definition / Construction of target population in terms of sectors, size classes, and regions

- Collection of survey and preparation of data set;
- *Definition of goal variables of evaluation*: depending on the nature of the policy measures to be evaluated;
- *Construction of an econometric model* based on theoretically defined relation of government invention and the goal variable;
- Statistical or econometric *data analysis and interpretation* of results.

3.2.7. Scope and Limits

First of all we should recognise that innovation surveys are generally not designed for evaluation proposes. Innovation surveys are, like R&D surveys, are primarily an instrument to characterise innovative activities with respect to sectors, size classes, regions or countries. Hence, they are primarily an instrument to learn about innovation processes and their impact and the subjects (firms) performing innovation. There is only a limited experience about the use of innovation surveys for evaluation proposes. Hence, this general assessment mainly depends on an evaluation of the general concept of innovation surveys and not on experiences in use. So, there is clearly an urgent need for more experience in this respect and the following conclusions are offered for consideration:

- The main advantage of innovation surveys is that innovation surveys generally address the universe of firms and not a specific target population of a specific programme. Hence, strategic behaviour in answering the questionnaire with respect to participation in a specific program is unlikely.
- Innovation surveys potentially give information about the whole structure of government intervention towards fostering innovation in private sector firms. Hence, they are able to provide unique data in this respect.
- Innovation surveys address innovative as well as non-innovative firms.
- Depending on the size of the innovation survey, there should be the possibility to address the impact of government intervention not only on programme participants but also on non-participants. Hence, innovation surveys can shed some light on indirect (spill-over) effects. So, it is possible to look at the impact on government programmes which address public research institutions on private firms.
- Innovation surveys offer a wide range of innovation indicators comprising input, process, through-put and output indicators.
- Innovation survey can be combined with a wide-range of statistical and econometric evaluation techniques. As they sample non-innovative, as well as innovative firms which do not take part in government programmes, firms which take part in the specific programs selected for evaluation as well firms which take part in other programs one can construct complex comparison groups.
- If innovation surveys are based on a panel approach one can also gain information about the impact of government interventions over time.

However, there are also severe drawbacks of innovation surveys in relation to surveys which are conducted specifically for evaluation of a certain government intervention.

- Depending on the nature of innovation surveys information about the government program of interest is - as a rule - rather spare. However, to improve the design of existing programmes one needs specific information about the programme of interest.

- The relationship between the program to be evaluated and the outcome is often vague. Hence, the researcher needs to employ sometimes complex methods to overcome this shortage of information. E.g. if the size of the programme is small in relation to the total innovation activities of a firm it is unlikely to find an impact of the program on certain innovation indicators even though they have produced an effect/outcome.
- It normally takes some time for government intervention to impact firms' innovative activities. The way innovation survey are constructed does not allow to take into consideration the time lag between the intervention and its impact unless there is a possibility to merge external data about government intervention to the survey population at the micro-level.

So, innovation surveys are probably not the best tool to overcome the data availability problem of evaluation researchers. However, they can be further improved to become a more reliable instrument for evaluation of large government programs. Data from innovation surveys also can be used as a benchmark for interpretation of data resulting from surveys conducted specifically for evaluation of a policy tool. Finally, innovation surveys might help to identify needs for policy invention and to provide ex-ante assessments of the impacts of policy measures.

3.2.8. Bibliography

The most comprehensive information about innovation survey is the OSLO-Manual:

OECD-EUROSTAT (1997), Proposed Guidelines for Collecting and Interpreting Technological Innovation Data – OSLO Manual, Paris.

There is only a limited number of research papers using innovation survey data available:

Almus, M. and D. Czarnitzki (2001): The Effects of Public R&D Subsidies on Firms' Innovation Actvities in a Transition Economy: The Case of Eastern Germany, *ZEW Discussion Paper No. 01-,* Mannheim.

Archibugi, D. and M. Pianta (1996): Innovation Surveys and Patents as Technology Indicators: The State of the Art, in: OECD (Ed.) (1996): Innovation, Patents and Technological Strategies, Paris.

Busom, I. (2000): An Empirical Evaluation of the Effects of R&D Subsidies, *Economics of Innovation and New Technology*, 9, 111-148.

Czarnitzki, D. and A. Fier (2001): Do R&D Subsidies Matter? – Evidence for the German Service Sector, ZEW-Discussion Paper 01-09, Mannheim.

König, H. and G. Licht (1996): Participation in Technology Programmes in Germany, in: Wilhelm Krull, Frieder Meyer-Krahmer (ed.): The Cartermill Guides. Science and Technology in Germany, p. 82-98, London.

OECD-EUROSTAT (1997), Proposed Guidelines for Collecting and Interpreting Technological Innovation Data – OSLO Manual, Paris.

Pianta, M. and Sirilli, G. (2001): The Use of Innovation Surveys for Policy Evaluation in Italy, in: OECD (ed.), Policy Evaluation in Innovation and Technology – Towards Best Practices, Paris.

Smith, K. (1989): Technological Innovation Indicators: Experience and Prospects, Science and Public Policy, 19, 383-392.

3.3. Econometric Models: Macroeconomic Modelling and Simulation

Authors: **Henri Capron** and **Michele Cincera** (ULB)

3.3.1. Introduction

The evaluation of S&T policies has generally been focused on technological objectives, e.g. additionality of public subsidies on private R&D, leaving out the crucial issue of the socio-economic impacts of implemented policies. However, welfare improvements, which are often concretised through socio-economic impacts, can be expected to be the ultimate goal of S&T policies.

Macroeconometric models based on sound economic theory principles are useful to assess the effects of innovation policy on economic performance. They are extensively used by policy makers to analyse the alternative expected impacts of policy choices. Said differently, this approach looks at the impact of using different policy instruments on the relevant variables of the model. In general, the impact is measured in terms of changes of macroeconomic aggregates due to the policy implemented against a reference base scenario. The use of simulations is justified due to the non-linearities, complexities and feedback mechanisms linking innovative activities, productivity, specialisation patterns and the economy. In practice, most policy organisations use a range of models designed to analyse a range of problems by focussing on their most features. Thus using macroeconometric models to evaluate the impact of S&T policy on economic variables as employment, competitiveness and economic growth requires to have models that endogenise to some degree and in an appropriate way for policy concerns technological progress (human capital, knowledge formation, technology spillovers…).
For example, employment generation is an objective to all areas of public intervention. So, the Lisbon Council in 2000 made a link between the establishment of a European research area and its effects on job creation and economic growth. Several quantitative methodologies have been applied to evaluate the impact of policy instruments on employment. As emphasised by the OECD (1998), when jobs impacts of technology policies need to be identified, "there is a need for more sophisticated approaches where information from surveys is validated by independent quantitative estimations and complemented by quantitative tools that capture economy-wide effect (i.e., input-output techniques, macroeconometric modelling or general equilibrium approaches)". More generally, to provide evidence of the meso- and macro level impact of the Framework Programmes as a complement to the bottom-up information collection of the other evaluation activities, econometric studies of R&D in Europe can be viewed as an important element of assessment (European Commission, 1999).
The measure of the impact of S&T policy based on reduced forms, as it is the case for the analysis of the impact of R&D expenditures on total factor productivity, only gives a limited view of the linkages between S&T variables and macroeconomics. There is a need to clarify the chain of causal effects between technological change and the main macroeconomic variables such as productivity, production, employment, investment, profitability, and exports.

In reviewing the methodologies used for modelling the impact of technological change on economic performance two broad types of models can be distinguished (see Bradley and Whelan, 1992): Neo-Keynesian Macroeconomic Models (e.g.

HERMES/HERMIT and INTERLINK) and Computable General Equilibrium models (e.g. IMF Multimod model, European Commission Quest Model)[30] :

- A macroeconometric model is a set of structural equations based on the economic theory designed to explain the economy or some parts of the economy. It is generally divided into several blocs (production, demand, world trade, prices-wages, monetary and financial conditions). There are two types of equations: behavioural ones and identities. Stochastic behavioural equations are estimated from the historical data. Identities are equations that hold by definition. There are two types of variables in macroeconometric models: endogenous and exogenous. Endogenous variables are representative of the structure of the economy and are explained by the equations, either the behavioural equations or the identities. Exogenous variables are not explained within the model, they are mainly linked to the international environment or the economic policy. They are taken as given from the point of view of the model. Initially strongly structured according to the Keynesian theory (demand-side economy and short and medium term dynamics), the emphasis has been put these last years on endogenising the supply-side of the economy and improving their long term properties (neo-classical theory). Recently, the stress has been put on the features of the endogenous growth theory (endogenising the process of technological change).

- A computable General Equilibrium (CGE) model is an integrated system of equations derived from microeconomic theory of the behaviour of all economic agents and built on intertemporal market clearing behaviour, whose simultaneous solution uses a numerical database to determine values of the endogenous variables (rather than formal estimation). Its construction relies on an accounting of economic transactions, the social accounting matrix which describes the flows between agents of commodity and factor markets and institutions. A CGE model allows one to simulate the working of a market economy in which prices and quantities adjust to clear all markets. It specifies the behaviour of optimising consumers and producers while including the government as an agent and capturing all transactions in circular flow of income. By simulating the effects of policy, a CGE model is a useful tool for economic impact analysis. Although it is theoretically sound, it is not clear whether its quantitative predictions are superior to alternative models. Most of the specification problems in CGE analysis emanate from its reliance on exogenously estimated coefficients as well as one benchmark year of data to implement the calibration process. Compared to macroeconometric models, their main advantage rests on their very well grounded theoretical background.

[30] The vector auto regressive modelling approach (VAR models) is not discussed here. It has emerged as a consequence of the Sims (1980)'critiques against the identification process in models. A main argument is that the importance of expectations effects and the interaction of policy regimes and agents expectations make identification difficult. VAR models have proven suited to investigate and forecast macroeconomic activity and provide a relevant statistical point of departure for modelling (including macroeconometric modelling). However they are not well tailored to follow the policy process despite the fact they allow to analyse the impacts of policy changes and external stimuli to the economy (Smith, 1998). As all variables are endogenised, which prevents to make assumptions as regards the exogeneity of the contemporaneous variables, there is a problem with the identification of structural shocks and the absence of theoretical background can lead to biased results (Hall, 1995).

The former type of model consists in disequilibrium models where in the absence of well defined market clearing conditions, long run properties are difficult to rationalise and interpret. The latter approach is built from microeconomic foundations which determine the full dynamics of the model. The two approaches must be viewed as complementary because it is neither feasible nor desirable to estimate, as a system of simultaneous equations in a large-scale macroeconometric model, the full set of conditions describing a multisector economy model.

In many cases, general equilibrium analysis borrows parameter estimates from partial equilibrium econometric studies or macroeconometric models. In the last years, there has been an increasing convergence between the two approaches: macroeconometric models draw increasingly on theory (better understanding of the long term properties of models and integration of supply-side components) and the CGE models increasingly are based on econometric techniques (estimation of coefficients, dynamic effects)[31]. Both types of models can cohabit as it is shown by the Central Plan Bureau (CPS) in the Netherlands. Indeed, four models are actually used by the CPS: two econometric models and two CGE models (Broer, de Mooij et al., 1998). The models differ in their methodology, degree of disaggregation and policy focus.

The empirical evidence at the macro level supports the large contribution of R&D to productivity and economic growth complementing the microeconomic evidence on the positive contribution of R&D to economic performance (large payoffs to society, lower returns to the innovator, relevance of commercialisation of research). As illustrated by Richards (2000), with endogenous technological advance, the projected rate of productivity growth is much higher than in simple deterministic calculations. His simulation exercises for the US economy show that deterministic calculations substantially underestimate potential, essentially because the disembodied rate of technological advance is held to a fixed number.

Macroeconomic modelling and simulation exercises signal the non-linearities, complexity, and feedback mechanisms characterising R&D and innovation process. The existence of labour supply rigidities indicate the relevance of integrating the functioning of the labour markets in R&D models. Some recent studies show that simulation exercises permit to study and measure the impact of international spillovers in RTD (Eaton and Kortum, 1997).

On the whole, direct government funding and tax incentives have a positive impact on private R&D but the stimulation of government funding is non-linear. There are substitution effects between direct government funding and R&D tax incentives as well as between private R&D and public financing. Furthermore, the stability of policy tools over time is important. Public R&D support influence TFP. Subsidies to private R&D have a positive effect on targeted firms, but do not reduce market failure. Some studies give evidence of under-investment in R&D in US (optimal level of R&D investment four times larger than current investment).

At the European level, it is highly relevant to be aware that knowledge spillovers are geographically localised and are broader between regions with similar or complementary technology specialisation patterns. The measured social rates of return to R&D differ for different levels of geographic aggregation. Important technology

[31] See, for example, McKibbin and Wilcoxen (1999).

barriers remain in Europe, e.g. country borders significantly hinder knowledge spillovers. There are strong positive impacts of university research on private R&D and patents. The impact of the R&D performed by multinational corporations is important for the development of local markets. Differences across regions in learning capacity and as regards the knowledge stock growth persist over time. Consequently, country as well as regional patterns should be taken into account in any macro modelling exercise.

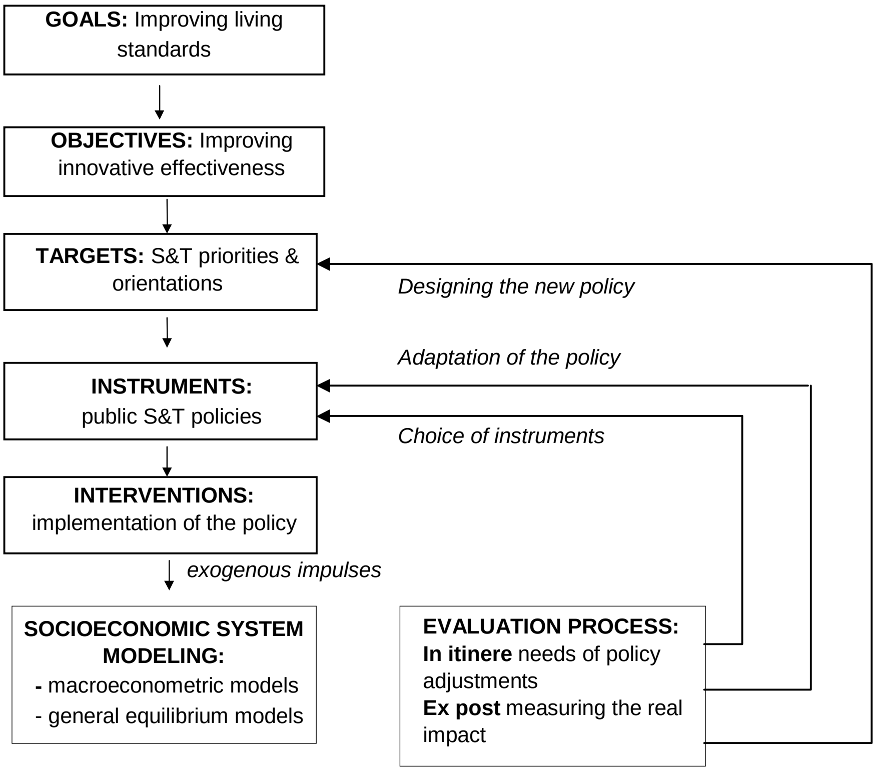
3.3.2. Application to Policy Instruments and Interventions

As a matter of fact, each policy, even if it is not directly targeted towards S&T development, can have strong implications on the innovative activities of economic agents.

Figure 4 illustrates the scheme generally in use for the evaluation of public action in the field of economic policy.

As a whole, the *goal* of government action in public life is to upgrade social welfare. The development of S&T is one of the means that allows the society to improve its social welfare. It is the efficiency and the effectiveness in producing, diffusing and exploiting useful knowledge, which allows improving the well being.
With reference to that, generating, acquiring and diffusing knowledge can be identified as the main *objectives* of government action. The creative, transfer and absorptive capacities have been viewed as the main characteristics which allow one to appreciate the efficiency of technological activities and flows at the source of knowledge accumulation. The process according to which the objectives can be achieved is determined by the policy mix implemented. The observation of market failures gives the economic foundation of technology policy.
However, this approach does not provide a sufficient foundation for implementing strategic technology policy. It is the comparative effectiveness between government intervention and market mechanisms which must guide policy choices. Their must take into account the socio-economic situation as well as the technological position of the country in order to define on which components of the innovation system the policy stress will be put. On the basis of these elements, public policy-makers identify *targets*, which define the technology orientations and priorities through which technology flows should mainly operate and new knowledge stimulated.
The targeting of public intervention leads them to implement *instruments* (i.e., education budget, R&D incentives, public R&D infrastructure…) which are mainly concretised through the institutional device aimed at improving the effectiveness of the knowledge-producing mechanisms. In practice, governments select the instruments and mix them according to the targeted policy.

Figure 4: Impact Assessment of S&T Policy



The choice of instruments will lead to the design of quantitative and qualitative *interventions*. Interventions are the channels through which instruments are made operational. They (as well as instruments) are not independent and vertical devices but are intertwined by a game of hierarchical and/or causal relationships and interdependencies. The implementation of the policy mix so obtained is expected to shape through the socio-economic mechanisms the whole process of economic and social growth, including technological change. The impact of such interventions can be assessed by socio-economic system modelling.

So far two methods can be used: macroeconometric models which are based on a set of econometrically estimated structural equations and computational general equilibrium models which are a mix of the input-output models and the Walrasian approach. These models are used for the assessment of both the economic development in the short and medium term, mainly the first ones, as well as long term, mainly the second ones, and the effects of alternative economic policies. Yet, some efforts should be devoted to model the driving forces behind the process of economic development along the lines defined by the endogenous growth theory.

The effectiveness of the S&T policy can be measured by the improvements observed in the economic effectiveness, the social equity and environmental sustainability. The performance observed leads decision-makers to adjust their policy (i.e. the instruments, in order to take into account the new socio-economic environment and improve the effectiveness of the S&T policy). In fact, we face a complex problem because there are strong interdependencies and interactions within each category of instrument. A main question of S&T policy is to realise the fine-tuning between the instruments available: How to allocate R&D resources among the different types of research? What is the most appropriate policy-mix to promote an efficient distribution of knowledge?… Sound macroeconometric models can only bring a partial contribution to the debate in drawing the attention of the policy makers on the incidences of policy choices and giving a normative analysis of policy options.

3.3.3. Good Practice Examples

Bayoumi, Coe et al. (1999) use an aggregated structural model of endogenous growth to simulate the effects of spillovers on productivity growth and consumption [32]. In these model R&D expenditures, R&D spillovers and trade endogenously determine TFP. In their simulation exercise an increase of R&D investment in US of 0.5% of GDP raises real output in the long run by more than 9%. At the same time domestic R&D spending generates significant spillovers to output in other countries. An increase in R&D spending of 0.5% of GDP, the increase in output in terms of US GDP is 50% larger than the case in which only US R&D spending augments.

Eaton and Kortum (1997) perform a simulation exercise using a multicountry model of international technology diffusion to study productivity growth differences in manufacturing. The state variables used are the productivity levels the national and foreign pool of ideas that countries have to adopt.

Jones and Williams (1999) propose an endogenous growth model incorporating four important distortions to R&D to investigate whether a decentralised economy undertakes too little or too much R&D. The distortions are the surplus appropriability problem, knowledge spillovers, creative destruction, and congestion externalities. A robust result achieved after model calibration is that the decentralised economy generally under invests in R&D relative to what is socially optimal. The only exceptions to this conclusions occur when both the congestion externality is extremely strong and the equilibrium real interest rate is very high.

Tancioni and Simonetti (2002) design a macroeconometric model to analyse the impact of technological change and trade on employment and apply it to Italy and United Kingdom. The results put forwards important differences between these two countries in the reactions of the institutional structure to supply-side shocks that emanate from both product and process innovation. These differences affect the efficiency of the various compensation mechanisms in the labour market and need to be taken into account in any policy formulation. The authors also emphasise that though macro models lack the detail of micro analysis, they capture feedback effects that involve other sectors of the economy.

Several recent studies (Goolsbee, 1998; David and Hall, 2000) have criticised the R&D and technology literature for concentrating on R&D spending assuming the private R&D performing sector is price and wage-taking. Public support can increase the average and marginal cost of private R&D performance by driving up the prices of R&D inputs. Therefore studying the responses of increases in public support to R&D on the flexibility of the scientific labour supply constitutes a highly relevant issue for policy making.

[32] They use a modified version of MULTIMOD, the model used for making the projections of the World Economic Outlook publication of the IMF.

3.3.3.1. S&T Policies, R&D, and Economic Growth (Guellec and van Pottelsberghe)

- **Policy instruments/interventions to be evaluated**

Guellec and van Pottelsberghe (1999, 2000, 2001) estimate the contribution of various sources of knowledge (R&D capital stocks performed by the business sector, by foreign firms, and by public institutions) to productivity growth as well as the determinants of privately funded and performed R&D.

- **Operational steps for method implementation**

Goal variable(s): Private R&D and productivity growth.
Econometric model: The contribution of various sources of knowledge to productivity performance is quantified by means of a Cobb-Douglas production function. The dependent variable is the multifactor productivity growth (*MFP*) of the industrial sector (computed under the hypotheses of perfect competition and constant returns to scale).

$$MFP_{it} = \exp\left[\phi_i + \varphi_t + \mu_{it}\right] \; SRP_{it-1}^{\beta_{rp}} \cdot SFR_{it-1}^{\beta_{fr}} \cdot SRHEGOV_{it-2}^{\beta_{hegov}} \cdot U_{it}^{\sigma_U} \cdot G^{\sigma_G} -$$

where: *SRP* is the stock of business performed R&D;
SFR is the stock of foreign business performed R&D;
SRHEGOV is the stock of publicly performed R&D;
i refers to the country and *t* to the period.

Country dummies, time dummies, employment rate (*U*, controlling for business cycle effects), and a dummy for the German unification (*G*) are included as control variables.
In order to assess the determinants of privately funded and performed R&D, a R&D investment model that considers business-funded R&D (RP) as a function of output (proxied by value added, VA) and several policy instruments, i.e. government funding of R&D performed by business (RG), tax incentives, government intramural expenditure on R&D (GOV), research performed by universities (or higher education, HE), time dummies, and country-specific fixed effects is estimated.

$$\Delta RP_{i,t} = \lambda \Delta RP_{i,t-1} + \beta_{VA} \Delta VA_{i,t} + \beta_{RG} \Delta RG_{i,t-1} + \beta_B \Delta B_{i,t-1} + \beta_{GOV} \Delta GOV_{i,t-1} + \beta_{HE} \Delta HE_{i,t-1} + \tau_t + e_{i,t}$$

Data base: The data consist of a panel of 16 and 17 OECD Member countries over the period 1980-1998.
Econometric method: Both equations are estimated through an error correction model that allows to separate short term and long term effect of the right-hand side variable. The econometric method was a three stages instrumental variable least squares that takes into account the presence of the lagged dependent variable among the explanatory variables and corrects for contemporaneous correlation of the error term.

- **Results of policy evaluation**

Public R&D and productivity: public R&D includes R&D performed both in government laboratories and in universities. The elasticity of government and university performed research on productivity is 0.17. This tends to show that overall public R&D is very valuable to the economy. The effect of public R&D on productivity is also larger in countries where the share of universities (as opposed to government laboratories) in public research is higher.

Public funding of business R&D investments: the first policy instrument aiming at stimulating business R&D is direct financial support of research performed by the business sector. These subsidies are targeted to specific goals chosen by the funder, *e.g.* “generic technologies”, “pre-competitive research”, health, defence. Government-funded R&D has a positive and significant effect on business R&D as the long term elasticity is 0.08. An alternative specification of the equation allows to approximate the average optimal subsidisation rate of business R&D. The results suggests that the effectiveness of government funding increases up to a particular threshold and decreases after that, which can be represented by an inverted U-shaped curve.

Fiscal incentives and private R&D investments: Government can also help firms through tax breaks. Most OECD countries allow for a full write-off of current R&D expenditures, which implies that depreciation allowances are deducted from taxable income. The long term elasticity of business R&D with respect to tax breaks is negative (-0.33). The estimates also suggest that the effect of tax breaks is quicker than the effect of government funding, as business spending reacts immediately to a change in taxes.

Public research and business R&D investments: government and university research have both a negative and significant impact on business funded R&D. Long term elasticities are respectively –0.08 and –0.05. Moreover, this negative impact is spread over several years (although there is no contemporaneous impact). The crowding-out effect – which is due either to an induced increase in the cost of R&D or to direct displacement – appears to dominate the stimulating effect. Public laboratories are supposed to meet public goals, however, not those of business; spillovers may occur but they are not instantaneous and are not the primary goal. The negative impact of university research on business funded R&D may also point to the difficulties in transferring basic knowledge to firms.

Interaction between the various policy tools: the estimates show that government funding of business R&D is substitute to fiscal incentives, complementary to university research, and does not interact with government research. In other words, increasing the direct funding (tax incentives) of business research reduces the stimulating effect of tax incentives (direct government funding). In addition, increased government funding of business research appears to reduce the negative effect of university research on business funding, possibly because government funding helps firms to absorb knowledge from universities that may be poorly used.

Defence oriented public support to business R&D: Defence technologies are less likely to be characterised by spillovers, as they are often specific, with little emphasis on cost but primarily on extreme performance in extreme conditions. Secrecy constraints may also imply that the results will only diffuse slowly to civilian applications. Furthermore, because defence contracting is attractive - it generates high rewards at low risk - firms might allocate resources that would otherwise have been used for civilian research. The estimates show that the higher the share of defence, the lower the positive effect of government funding on business R&D.

3.3.3.2. Macro-Economic Evaluation of the Effects of Community Structural Funds (CSF) Interventions with QUEST II (Röeger, 1996)

- **Operational steps for method implementation**

Goal variable(s): This model analyses the short, medium and long run macroeconomic effects both on the demand and the supply sides of CSF on key macroeconomic variables such as GDP and its components, employment, real wages as well as government deficit and public debt. Since prices, interest rates, exchange rates and wages are allowed to respond to the CSF induced public investment, the simulation results allow to shed some light on the question to what extent public investment adds to total GDP, rather than at least partly displacing investment activities of the private sector. The model can therefore be used to look at the major determinants of displacement effects on a macroeconomic level.
Econometric model: The paper traces the macroeconomic impact by using the respective country modules of DG IIs macro econometric model QUEST II. The current version of QUEST bases behavioural relationships on principles of dynamic optimising behaviour of households and firms. Since the model has a supply block based on a neoclassical production function, it is possible to model explicitly the supply side effects of infrastructure and human capital investments. The model is also closed with respect to stock flow interactions. Those stock variables which are identifiable on a macroeconomic level, such as physical capital, net foreign assets, money and the government debt, are endogenously determined and wealth effects are allowed to influence savings, production and investment decisions of households, firms and governments. Moreover trade and financial linkages of each country to the rest of the world are explicitly modelled as well, which allows for an endogenous determination of interest and exchange rates. The labour market is based on a standard bargaining model.
Data base: The analysis is based on payments data for Greece, Ireland, Portugal and Spain over the period 1989-93 as well as data on planned CSF spending for the same countries over the period 1994-99.
Econometric method: The model solution method, which solves a forward looking model with rational expectations is based on a linearisation of the model around the steady state and applies closed form solution algorithms to the linearised model [33].

3.3.3.3. Dynamic Input-Output Model to Evaluate the Economic Impacts of the CSF (Beutel, 1996)

- **Policy instruments/interventions to be evaluated**

This input-output model has been developed for the Directorate-General for Regional Policies and Cohesion to evaluate the economic impacts of the CSF.

- **Operational steps for method implementation**

Goal variable(s): Evaluation of the economic impacts of the CSF interventions on economic growth, structural change, foreign trade and employment and induced changes in technology, imports, labour and capital use.
Econometric model: In previous studies for the periods 1989-1993 and 1994-1999 the main issue was to identify the short-term supply and demand effects of the Community Support Frameworks for the objective 1 regions. The impact analysis system was designed as a comparative static input-output model to assess the

[33] The model is estimated using the TROLL software. See Roeger and Veld (1997) for more details as well as the simulations methods.

quantitative impacts of the Structural Funds on economic growth, structural change, foreign trade and employment. Based of these former studies, a dynamic input-output model was developed which is capable to evaluate the long-term supply and demand effects of the Community structural policies. Expenditures of the Structural Funds will affect the structure and level of final demand but will also induce changes in technology, imports, labour and capital use. In particular the long-term effects on capital and labour, output and productivity are in the focus of interest and are covered by the dynamic input-output approach.

Data base: A set of harmonised input-output tables with labour and capital stock data is used which has been established by Eurostat in cooperation with the author. The projected input-output tables are based on harmonised National Accounts of Eurostat and the latest official economic forecasts of the Directorate General for Economic and Financial Affairs.

Econometric method: The dynamic input-output model is designed in line with the macroeconomic multiplier-accelerator theory. According to this theory it is expected that new capacities are required if final demand components are growing. Therefore, induced investment is estimated which can be related to the activities of the Structural Funds. The first part of the model estimates how an increase of gross fixed capital formation will affect the economy which is financed by the Structural Funds to improve the infrastructure of public and private institutions. The second part analyses how the contributions of Community interventions affect value added. A dynamic version of the input-output model is used with the third element (induced investment) of the impact analysis system to evaluate the long-term supply effects of the CSF.

- **Results of policy evaluation**

In previous studies, the impact of Structural Funds expenditure was analysed for individual years assuming that the Funds were still active in the previous year. The short-term impact of the Structural Funds activities revealed that the growth potential of the economy would be substantially reduced in individual years if the Structural Funds were not in existence. The dynamic version of the model estimates the impact for a sequence of years and consequently the supply effects are more profound. The results of the dynamic input-output model reflect a different growth path of the economy which would be realised in the absence of the Structural Funds.

So far, the dynamic impact analysis was conducted for four countries EUR4 (Greece, Spain, Ireland, Portugal) on a national scale and for two countries on a regional basis (East Germany, Mezzogiorno). For 1994-1999 it is expected that in EUR4 the gross domestic product will grow at an average annual rate of 3.2 %. Without interventions of the CSF the annual GDP growth rate would be reduced to 2.5 %. The labour force is projected to grow at an annual rate of 1.4 % versus 0.6 % without Community interventions. The capital stock is expected to grow in the same period at an average annual rate of 3.7 % versus 2.7% without Community interventions. It is estimated that in 1999 approximately 35% of Community interventions is leaking into the rest of the Union and another 10% into the rest of the world.

3.3.3.4. ‘An empirical model for endogenous technology in the Netherlands’ (den Butter and Wollmer 1996, the Netherlands)

- **Policy instruments/interventions to be evaluated**

The authors develop an empirical simulation model for the Dutch production sector which is inspired by modern endogenous growth theory. The model is used to simulate various technological impulses. The parameters of the model are determined by considering empirical results from the literature as well own estimates and by calibrating the model over the reference period 1972-1987.

- **Operational steps for method implementation**

Goal variable(s): Economic growth, long term industrial output.
Econometric model: The core of the model consists of a production block of nested CES functions, where investments in technology capital and in human capital play a major role. The external effects of R&D are modelled in such a way that R&D investments not only lead to more technology capital, but also have a positive impact on human capital through learning by doing and learning by designing. Technology capital enters into the production block in two related way: firms accumulate knowledge by either undertaking R&D or importing knowledge. Technology capital is then assumed to augment the human capital of workers as they work with new technology. Human capital is assumed to be a substitute for raw labour. Raw capital and technology capital, considered as complementary are combined to produce efficient units of physical capital. The demand for domestic and imported R&D is positively related to output. The production block is extended by adding output demand and monetary equations.
Data base: Macroeconomic aggregates.

- **Results of policy evaluation**

The simulations show the importance of incorporating elements of new growth theory into macroeconomic policy models. An impulse in R&D investments leads to higher labour productivity and consequently increase the long term demand for all inputs, except labour, and increases final output. In order to avoid negative employment effects, any public policy of enhancing economic growth through impulses to private R&D should be accompanied by a strong appeal to the social partners in order to avoid that the rise in labour productivity is fully absorbed in wage demands.

3.3.3.5. 'Endogenising technological progress: The MESEMET Model' (van Bergeijk, van Hagen et al. 1997, The Netherlands)

- **Policy instruments/interventions to be evaluated**

The model aims at illustrating the macroeconomic consequences of technology policies in different institutional settings. Technological progress and knowledge formation are endogenised in the model. Results are presented for three simulations: public R&D expenditures, public education expenditures and tax allowances for private R&D. The paper puts forward that endogenising technological progress in an empirically relevant context turns out to be important.

- **Operational steps for method implementation**

Goal variable(s): Macroeconomic aggregates.
Econometric model: The model is a macroeconomic semi equilibrium one that attempts to bridge the gap between applied general equilibrium models with a microeconomic foundation and the macroeconometric models typically applied in Dutch policy-making. The production structure contains various nested constant elasticity of substitution functions representing the relationships between the inputs of a representative firm and the corresponding output. Human capital is considered to be a public good so that individual firms as well as households have no incentive to invest in it. Nevertheless investment in physical capital exerts a positive external

learning-by-doing effect on the stock of total capital. Both private and public R&D expenditures are considered to have a similar effect. Conversely, the taxation is expected to have a negative effect on the stock of total capital. Private R&D is assumed to be a continuous variable input. Both the relative stocks of human capital and technology capital exert a positive effect on exports.
Data base: The model is parameterised for the Dutch economy in 1992.

- **Results of policy evaluation**

Tax allowances for private R&D expenditures and public expenditures on both education and R&D are effective instruments to stimulate economic growth through the accumulation of knowledge. Technology policies have a positive impact on overall employment. The spillover effects from R&D on human capital seem to be crucial for the economic consequences of public R&D and tax-free allowances on private R&D. Furthermore, the degree of complementarity between physical and technology capital on the one hand and human capital on the other hand is important for the degree in which public expenditures crowd out private investments in physical and technology capital.

3.3.3.6. The Hainaut economic-lead-in model (DULBEA-CERT 1998 and 2001, Belgium)

- **Policy instruments/interventions to be evaluated**

In the framework of Objective 1 programmes, the Belgian Hainaut has benefited from the Community intervention over the period 1994-1999. For the programming period 2000-2006, it will continue to benefit from structural interventions. In order to evaluate the impact of the policy intervention, a small-scale macroeconometric model has been implemented. In 1993, some part of the model was initially used to evaluate the expected impact of public funds (ex ante evaluation exercise). The results of the model have also been used to appreciate the expected impact over the period 2000-2006.

- **Operational steps for method implementation**

Goal variable(s): Regional added values, employment, investment, labour productivity, R&D.
Econometric model: The structure of the model is based on the Kaldorian theory according to which the manufacturing industry is the leading sector of economic growth. The dynamics of technical change in the manufacturing industry is a main factor of the development process thanks to the productivity gains that can be accumulated. Hainaut being an old industrial region, the approach is well-suited to its economic situation. The province being a small open economy, these productivity gains explain to a large extent its competitiveness. The trade balance of the region is mainly composed with manufacturing products. It is the increases in competitiveness that determine the growth potential inside the region thanks to direct or indirect impulses to other economic activities.
The equations of the model dealing with the public support to R&D are specified as follows:

Business R&D

$$\log RDPR(t) = \alpha + \beta * \log SRDPU(t-4) + \gamma * \log VMAN(t-1) + \delta * \log RDPR(t-1) - \phi * \Delta\{1/AIDRD(t)\}$$

Manufacturing employment

$$\Delta \log EMAN(t) = \alpha + \beta * \Delta \log SFORM(t) + \gamma * \Delta \log SRDPR(t-3) - \delta * \Delta \log CTRA(t) + \phi * \Delta \log VEX(t) + \varepsilon * \log IAIDR(t) + \eta * TX(t)$$

Manufacturing production

$$\log PROD(t) = \alpha + \beta * \log SRDPR(t-1) + \gamma * \log INTCAP(t) + \delta * \Delta \log DUC(t) + \phi * \log SADM(t-1) + \varepsilon * \log TX(t)$$

External demand

$$\log LIVEXT(t) = \alpha + \beta * \log PROD(t) + \gamma * \log SMAN(t) + \delta * \log \{IRD(t-3)/IRDE(t-3)\} + \phi * \log ECU(t-1) + \varepsilon * \log DEFLEN(t)$$

where: RDPR = private R&D expenditures, SRDPR = private R&D stock, SRDPU = public R&D stock, VMAN = manufacturing value added, AIDRD = R&D subsidies, EMAN = manufacturing employment, SFORM = human capital stock, CTRA = labour cost, VEX = mining value added, IAIDR = investment subsidies, TX= investment subsidies/subsidised investment, PROD = labour productivity, INTCAP = capital intensity per manufacturing worker, DUC = degree of use of Belgian production capacity, SADM = public physical capital stock, LIVEXT = exports, SMAN = manufacturing physical capital stock, IRD = Hainaut R&D intensity, IRDE = European R&D intensity, ECU = value of the Belgian franc in Euros, DEFLEN = deflator of the energy value added.

Data base: Macrosectoral annual data were collected for the period 1964-1993. The six sectors covered by the model are respectively: agriculture, manufacturing, energy, construction, market services and non-market services. Four groups of variables were considered: total investment, employment, production and demand. Use of the concept of total capital is made, which leads to make a distinction between physical capital (with a distinction between agriculture, tourism, government and other private expenditures) human capital (based upon the “weighted” number of persons following a formation cycle in the higher education system as well as in the on-the-job training system) and knowledge capital (based on a distinction between private and public R&D). The model is composed with 50 variables, 28 econometric equations and 12 definition equations.

- **Results of policy evaluation**

The simulation results put forward the impact of interventions will be lagged (first significant impacts can only be expected in 1999) and it cannot be expected that the intervention will be enough to reverse the divergence trajectory of the province. Regarding the impact of public support to R&D on the growth process, it has a high long term effect on output, employment and physical investment. Nevertheless, investment in human capital appears to be crucial for the recovery process of the region, they are characterised by indirect important positive effects on physical investment and employment. Conversely to investment in R&D and human capital which have long term effects on the growth process, physical investment subsidies have only a temporary effect on economic growth.

3.3.4. Conditions for Application

To apply satisfactorily the macroeconometric methodology, it is worth keeping in mind that this approach is mainly adapted to measure the global expected economic and social impacts of a programme or a set of programmes. If its main advantage is to give an evaluation of direct, indirect and induced effects of a policy in a structured and systematic way based on the economy theory, it is not suited to appreciate the effects of small scale programmes, to estimate the benefit of programmes whose expected economic outcomes are only very marginal or to select projects. As this approach allows taking into account a large variety of effects, it requires:

- the availability of a large scale socio-economic dataset;
- a high degree of expertise;
- enough time to build the model prior to any evaluation exercise;
- the implementation of policies which do not lead to a high dilution of economic effects.

Regarding the first constraint, the data requirement bears not only on data linked to the public policy but also on the quantitative information necessary to model with a sufficient degree of reliability the functioning of the socio-economic system. It is a little bit trivial to say that the power and the potential of a large scale econometric model will be higher than what can be evaluated from a small scale econometric one. Given the high complexity of socio-economic systems, not any econometric model can be used to evaluate the impacts of a public policy. Generally, econometric models are calibrated to assist decision makers to deal with specific economic questions as monetary policy, employment policy, public finance… or to formalise economic mechanisms restricted along some time periods (short terms, medium term and long term). Yet, a large number of current macro models, at least their structural basis, could be adapted to measure the impact of S&T policy instruments.

The design of macroeconometric models implies a high level of expertise, not only in the field of economic theory but also in the field of applied economics (understanding of hypotheses underlying the model, translation of theoretical concepts into empirical terms, identification of appropriate variables…), econometrics (estimation procedures, calibration of the model, simulation…), as well as in the building of economic databases (sources of information, collection, transformation and limits of data, data analysis...). Once estimated and well calibrated for the needs of evaluation, a macro model allows one to simulate a scenario with the public intervention and a scenario without the public intervention. The difference between both scenarios gives an estimation of the impact of the public intervention on the modelled macroeconomic variables.

The building of macro models is time and resources expensive. The elaboration of a small scale macro model takes a minimum of one year and its usefulness is often limited to the study of a preliminary defined economic problem (for example, employment policy or fiscal policy). The construction of large scale macro models often require a team composed of four to five persons for several years. Once the model built, a permanent team is necessary to ensure its utilisation and its periodical updating. This explains why macro models are rarely built for one-shot evaluations. It is generally made use of existing macro models to which some adaptations or extensions are brought to satisfy the needs of evaluation. In the present state-of-the-art, macro models should be adapted to improve the relationships between S&T indicators and macroeconomic aggregates.

A last word of caution bears on the use of macro models for the evaluation of S&T policy. Indeed, the use of a model makes only sense if the intervention (the programme or the set of programmes) has a sufficient critical mass compared to the weight of the macroeconomic aggregates, or at the least, by comparison with the main economic aggregates on which the policy is calibrated.

3.3.5. Steps for Implementation

The construction of a macroeconometric model is composed with nine operational steps, which are in fact strongly interconnected. Nevertheless, these different steps can be schematised as follows:

1. Defining the objectives of the model and the feasibility of the experiment: what are the policy instruments to evaluate and do they interact with macroeconomic variables? Is the intervention directed towards the economy as a whole or is it limited to some agents or sectors? What do we want to measure? Is the macroeconometric approach suited to highlight policy makers given the questions asked?
2. Investigating data availability: what are the data requirements? Are all the data necessary available? How to solve the problem of missing or deficient data? Is it possible to use proxies?
3. Specification of the model: what is the economic canvas to formalise (causality links, conceptualisation of the different blocs of the model, theoretical foundations of the model, empirical background)? Can we adapt an existing model and if yes what are the modifications to make?
4. Collecting, analysing and transforming the data: besides its economic foundations, the quality of an econometric model depends to a large extent on the reliability of the dataset. It is only rarely that raw data can be used. Some transformations are often necessary as it is the case to obtain deflated data or capital stocks. A good knowledge of data (their content, their limits, their sources, the influence of exogenous shocks…) is indispensable before estimating the equations of the model.
5. Econometric estimations of the equations of the model: it is the most exciting as well as frustrating step of modelling. Exciting because it is at this level that the theoretical canvas takes its empirical content and that economic interdependencies take shape. It is also frustrating because the economic reality is more complex than the economic theory and leads to some adaptations to the methodological canvas.
6. Testing and calibrating the model: despite good estimates, a model can give a poor performance due to exogenous shocks, the omission of some phenomena or the limited quality of some data. This implies to go back to the estimation procedure or to improve the quality of data in order to reproduce with a sufficient degree of reliability the economic dynamics. It is the fine tuning of the model. A macroeconometric model can be considered to be fine tuned when it is able to reproduce correctly the economic dynamics over the past ten or fifteen years.
7. Simulating the reference base scenario and sensitivity analysis: to measure the impact of public policy it is necessary to appreciate how should have performed the economy in the absence of the policy intervention. This step of the modelling will give the reference (or central) situation from which will be estimated the impact of the public policy. This situation is called the counterfactual situation

because it will not be observed if effectively the policy instruments are implemented. The sensitivity analysis reveals which parameters are crucial and to what extent alternative parameter values (or alternative specifications) influence substantially the results of the model, it is also a cornerstone of calibration econometrics.

8. Simulating the policy options: in this simulation exercise, exogenous quantitative values of instruments (the extent of the intervention) are injected into the model to measure how much they influence the aggregates.
9. Interpreting the results: the comparison of results obtained from simulation exercises (the reference scenario against the policy scenarios) allows one to give estimations of the global impact of the policy upon the economy as well as upon main macroeconomic aggregates.

Once the model is operational, it can be used to evaluate ex ante the expected impact of a policy. It can also be implemented to evaluate in itinere the effects of a policy and so to give some guidelines for a revision of the policy. Ex post, the model allows to measure to what extent the expected effects have been achieved, to appreciate how efficient has been the policy and to suggest improvements for the new policy to implement.

3.3.6. Data Requirements

The studies surveyed in section 4 use macro and meso level aggregated data in general for several industrialised countries and long time periods (more than 20 years). Among these variables we can distinguish between socio-economic variables, e.g. GDP, GDP per capita, multifactor productivity growth, employment, real wages, government deficit, public debt, prices, interest rates, exchange rates,… and S&T variables, e.g. public and private R&D expenditures and stocks, human capital stock, ICT stock, knowledge spillover stocks, … Data referring to the policy instruments whose impact is to be estimated concerns among other multiannual RTD programmes budgets, (planned) CSF spending, R&D tax breaks, R&D subsidies… More and more macro evaluation studies use panel data sets, which as compared to purely cross section or time series data, provide more information, variability, efficiency and less collinearity across aggregates and finally a higher flexibility in the modelling of behavioural differences across individual units, e.g. countries or regions. Despite the development of new and improved indicators, more detailed indicators regarding S&T activities, e.g. absorptive capacity, channels of knowledge exchanges, spatial and technological proximities, quality embodied deflators are needed as well as data at the regional level.

3.3.7. Scope and Limits

The evaluation toolbox consists of a large variety of complementary rather than substitutable methods (Capron, 1992a, 1992b; Capron and van Pottelsberghe de la Potterie, 1997). Globally, a distinction can be made between qualitative, e.g. peer review, semi-quantitative, e.g. matrix and systemic approaches, and quantitative methods, e.g. cost/benefit and econometric analysis. There is no perfect (or complete) assessment method: each method has its own advantages and drawbacks. The choice of a method depends of the issues that are to be investigated, the data availability and the level of analysis, e.g. the macro-level. Answers and highlights obtained from a single method will always be partial and will often give rise to new questions. Taken individually, each of them is able to provide relevant additional piece of information in the evaluation process.

To increase the credibility of evaluation results, alternative methods should be ideally used to consolidate the foundations of policy recommendations. Despite their drawbacks and the apparent mistrust from which they suffer, and given the absence of any firmly established substitutes at the time being, econometric methods appear to be the most appropriate to quantitatively assess the socio-economic impacts of RTD programmes. Yet, the lack of relevant and sufficiently detailed S&T indicators is certainly a main bottleneck preventing a more intensive use of such quantitative methods.

The impact of S&T policies shows itself at different levels of aggregation of economic activities. For instance, evaluations at the micro level give a good insight into the direct impact of a policy but do not generally provide any information on the indirect ones. Only inter-industry studies (at the meso level) and studies at the macro-level can provide information on the global impact, i.e. direct and indirect effects, of S&T policies. Yet, the global impact is a net effect since the indirect effects may either counterbalance or reinforce the effects of the direct ones. Hence, it is not easy to disentangle, at the macro-level, between the variety of direct, indirect and induced effects that contribute to specific outcomes.

The evaluation at the macro-level measures the global socio-economic impact in the long-term (global impact of a system of programmes or benefits to the society as a whole) rather than the short-term partial impact (benefits to the participants of any specific RTD project or programme). The timing itself needs to be addressed since the socio-economic impacts are not immediate.

Macro-economic effects are in general more difficult to assess than micro-economic mechanisms. Given the range of S&T public interventions and of public policies in general, it is not easy to isolate the impact of a particular RTD programme, which accounts for only a small part of total public intervention on macro-aggregates such as GDP growth, exports, and quality of life. More generally, socio-economic performance is conditioned by many other factors that influence the broader economic, institutional and social context of innovation. Despite these issues, it seems useful before any evaluation, to identify the different types and nature of the socio-economic impacts of RTD programmes.

3.3.8. Bibliography

Commented Bibliography

Economic Modelling, Volume 15, Issue 3, July 1998

Special issue on empirical models and policy making which gives a good idea of interactions between models and policy making : G. Zalm, The relevance of economic modelling for policy decisions, pp. 309-316; Olav Bjerkholt, Interaction between model builders and policy makers in the Norwegian tradition, pp. 317-339; Jan Donders, Johan Graafland, CPB models and employment policy in the Netherlands, pp. 341-356; Pierre Duguay, David Longworth, Macroeconomic models and policy making at the bank of Canada, pp. 357-375; David G. Mayes, W.A. Razzak, Transparency and accountability: Empirical models and policy making at the reserve bank of New Zealand, pp. 377-394; Jacques J. Polak, The IMF monetary model at 40, pp. 395-410; Hali J. Edison, Jaime Marquez, US monetary policy and econometric modeling: tales from the FOMC transcripts 1984--1991, pp. 411-428; Ron Smith, Emergent policy-making with macroeconometric models, pp. 429-442; Frank A.G. den Butter, Mary S. Morgan, What makes the models--policy interaction successful?, pp. 443-475.

Ginsburg V. and Keyzer M. (1997), The Structure of Applied General Equilibrium Models, MIT Press, 555 pages

The book bridges the gap between the theoretical foundations of the general equilibrium analysis and the applied models. Chapter 1 of the book provides an overview of the competitive model, and Chapter 2 reviews the standard theory of producer and consumer behaviour. Chapter 3 defines the basic formats used in the following chapters to represent and analyse the models. It discusses the main steps used to construct a numerical application and gives a full GAMS application for the simplest model. The remaining nine chapters present specific models.

Granger C. (1999), Empirical Modelling in Economics, Specification and Evaluation, Cambridge University Press, Cambridge, 108 pages
This books consists of three essays that explain the process of constructing and evaluating an empirical model with comments on the evaluation of econometric models and of forecasts.

Klein L., A. Welfe and W. Welfe, (1999), Principles of Macroeconometric Modelling, Advanced Textbooks in Econometrics, C. Bugg and M. Intriligator (eds), North-Holland, 366 pages
State of the art of macroeconometric modelling : what is a macroeconomic model, demand and supply determined models, disequilibrium macro models, use of macroeconometric models, production functions, price and wage equations, financial flows, some LINK applications, simulation, dynamic analysis, VAR models, stationarity and non-stationarity, equilibrium relationships and cointegration analysis, rational expectations, multiplier and policy analysis, forecasting, ex ante prediction, forecast preparation.

Research Policy, Volume 29, Issue 4-5, 01-April-2000
Special issue on the evaluation of S&T policies : Stephen Martin, John T. Scott, The nature of innovation market failure and the design of public support for private innovation, pp. 437-447 ; Bronwyn Hall, John Van Reenen, How effective are fiscal incentives for R&D? A review of the evidence, pp. 449-469 ; Tor Jakob Klette, Jarle Møen, Zvi Griliches, Do subsidies to commercial R&D reduce market failures? Microeconometric evaluation studies, pp. 471-495 ; Paul A. David, Bronwyn H. Hall, Andrew A. Toole, Is public R&D a complement or substitute for private R&D? A review of the econometric evidence, pp. 497-529 ; F.M. Scherer, Dietmar Harhoff, Technology policy for a world of skew-distributed outcomes, pp. 559-566 ; John Hagedoorn, Albert N. Link, Nicholas S. Vonortas, Research partnerships, pp. 567-586 ; P.A. Geroski, Models of technology diffusion, pp. 603-625 ; Barry Bozeman, Technology transfer and public policy: a review of research and theory, pp. 627-655 ; Luke Georghiou, David Roessner, Evaluating technology programs: tools and methods, pp. 657-678.

Romer D. (2000), *Advanced Macroeconomics*, 2nd edition, Irwin-McGraw-Hill, 672 pages.
Topics covered: the Solow growth model, infinite-horizon and overlapping-generations models, new growth theory, real-business-cycle theory, traditional Keynesian theories of fluctuations, microeconomic foundations of incomplete nominal adjustment, consumption, investment, unemployment and the labour market, inflation and monetary policy, budget deficits and fiscal policy.

References

Bayoumi, T., Coe D. et al. (1999), "R&D Spillovers and Global Growth", Journal of International Economics, 47(2), 399-428.

Beutel J. (1996), "Dynamic Input-Output Model to Evaluate the Economic Impacts of the Structural Funds", European Commission DGII, Brussels.

Broer P., de Mooij R. et al. (1998), "CPB models and their uses", *CPB Report*, n°3, 17-21.

Bradley, J. and Whelan K. (1992), "Econometric Modelling and Technological Change a Review of Issues" in Quantitative Evaluation of the Impact of R&D Programmes , European Commission, Brussels.

Capron H. (1992a), Economic Quantitative Methods for the Evaluation of the Impact of R&D Programmes. A State of the Art, Monitor-Spear Series, European Commission, Brussels.

Capron H. (1992b), "A State of the Art of Quantitative Methods for the Assessment of R&D Programmes", in M.T. Khalil and B. Bayraktar (eds.), *Management of Technology III,* Industrial Engineering and Management Press, Georgia, 1195-1204.

Capron, H. and van Pottelsberghe B. (1997), "Public Support to Business R&D: A Survey and some quantitative evidence" in Policy Evaluation in Innovation and Technology: Towards Best Practices, Paris, OECD.

David, P. and Hall B. (2000), "Heart of Darkness: Modelling Public-Private Funding Interactions inside the R&D Black Box." NBER Working Paper No. 7538, Cambridge.

David, P., Hall B. et al. (2000), "Is Public R&D a Complement or Substitute for Private R&D? A Review of the Econometric Evidence." *Research Policy* (29), 497-529.

den Butter F. and Wollmer F. (1996), "An empirical model for endogenous technology in the Netherlands", *Economic Modelling*, 13(1), 15-40.

DULBEA-CERT (1997), "Evaluation des impacts du DOCUP" in SEMA GROUP, Evaluation intermédiaire du DOCUP Objectif 1 Hainaut, Rapport au Ministère de la Région wallonne, Namur.

DULBEA-CERT (2001), Evolution de la situation économique hennuyère, Tableau de bord et Modèle économétrique HELM, Rapport au Ministère de la Région wallonne, Namur.

Eaton and Kortum (1997) "Engines of Growth: Domestic and Foreign Sources of Innovation", *Japan and the World Economy* 9(2), 235-259.

European Commission (1999), *Options and Limits for Assessing the Socio-Economic Impact of European RTD Programmes*, ETAN, Brussels.

Giorno, C., Richardson P. et al. (1995), "Technical Progress, Factor Productivity and Macroeconomic Performance in the Medium Term," *OECD Economics Department Working Paper* No. 157, Paris .

Goolsbee, A. (1998), "Does Government R&D Policy Mainly Benefit Scientists and Engineers?", *American Economic Review* 88(2), 298-302.

Guellec D. and van Pottelsberghe B. (2000), The Impact of Public R&D Expenditure on Business R&D; STI Working Paper 2000/4, Paris.

Guellec D. and van Pottelsberghe B. (1999), Does government support stimulates private R&D ?, *OECD Economic Studies*, 29, 1997/II, 298-122.

Guellec D. and van Pottelsberghe B. (2000), The effectiveness of public R&D policies in OECD countries, STI Working Papers, DSTI/DOC(2000)4, OECD, Paris.

Guellec D. and van Pottelsberghe B. (2001), R&D and productivity growth – A panel data analysis of 16 OECD countries, STI Working Papers, OECD, Paris.

Hall S. (1995), "Macroeconomics and a Bit More Reality", *Economic Journal*, 105(431), 974-988.

Jones, C. and Williams J. (1999), "Measuring the Social Returns to R&D." *Quarterly Journal of Economics* 113, 1119-1135.

McKibbin W. and Wilcoxen P. (1999), "The theoretical and empirical structure of the G-Cubed model", *Economic Modelling*, 16(1), 123-149.

OECD (1998), Technology, Productivity and Job Creation – Best Policy Practices, OECD, Paris.

Richards G. (2000), "Endogenous technological advance in an econometric model: implications for productivity and potential output in the United States", *Economic Modelling*, 17(1), 13-34.

Roeger W. (1996), "Macro-Economic Evaluation of the Effects of Community Structural Funds Interventions with QUEST II", European Commission DGII, Brussels.

Roeger W. and Veld J., (1997), "QUEST II. A Multi-Country Business Cycle and Growth Model", Economic Papers, No. 123. European Commission, Brussels.

Smith R., (1998), "Emergent policy-making with macroeconometric models", *Policy modelling,* 15(3), 429-442.

Tancioni M. and Simoetti R. (forthcoming) "A macroeconometric model for the analysis of the impact of technological change and trade on employment", *Journal of Interdisciplinary Economics*.

van Bergeijk., van Hagen G. et al. (1997), "Endogenizing technological progress: The MESEMET Model" *Economic Modelling*, 14(3), pp.341-68.

3.4. ECONOMETRIC MODELS: MICROECONOMETRIC MODELS

Authors: **Spyros Arvanitis** (ETHZ) and **Max Keilbach** (ZIW)

3.4.1. Introduction

Microeconometric Evaluation Models aim to quantify the impact of a public policy scheme on the level of the subsidised (supported) units which would be firms in the case of technology transfer programmes or individuals in the case of labour market programmes. To quantify this impact correctly, it would be necessary to compare the level of the goal variable of an individual after participation (e.g. the innovation output of a firm) with the level of this goal variable of that individual if it had not participated. Since this latter – counterfactual – state is not observable, a number of econometric methods have been suggested that aim to estimate this state.

In principle, these methods differ in that they estimate the counterfactual state using either (a) data on previous states of the participating individuals (henceforth firms) or (b) contemporaneous data on states of other firms. However, in either approach it is necessary to correct for biases that would occur from a „naive" use of corresponding observations. (a) If an estimation uses data on previous states of the same firms, the approach must correct for autonomous evolution of the goal variable. Otherwise, also those parts of the evolution of this variable that would have occurred even without participation in a policy measure would be accounted for as result of it. (b) If an estimation uses data on states of other firms, the approach must be designed such as to avoid or correct biases that would occur from comparing structurally different firms. Often, firms that participate in a public policy measure differ systematically from non-participants. Even more so, entry conditions of a programme might be designed such as to favour the participation of firms that are "weaker" with respect to the goal variable.

In the following section we provide a survey on microeconometric evaluation methods which are by now well-established. They will differ along the line drawn above. For all these methods to obtain statistically robust results, it is necessary to include a large number of individuals in the evaluation such as to ensure that the analysis is representative, whereby the definition of "large" depends on the specific programme to be evaluated.

3.4.1.1. Methods that Use Data on Previous States of Participating Firms

These methods aim to quantify the impact of a public policy measure by comparing the state of the goal variable of participating firms before and after the implementation of (hence participation in) the programme.

The subscript t denotes a point in time after implementation of the programme and t-1 a point in time before implementation. Further, Y denotes the goal variable and P a variable indicating programme participation. That is Pi > 0, if firm i participates and 0 otherwise. A "naive" measurement of the implications of the programme would refer merely to the difference Yi,t - Yi,t-1. However, this difference would not allow to account for the autonomous evolution of Yi, i.e. for changes in this variable that are not due to the programme. Rather this "naive" metric would assign also these changes to the policy outcomes. To avoid this problem we can introduce correction variables that are able to describe this autonomous evolution of Yi. Then, the impact of the programme could be estimated with the help of the following simple econometric model

$$Y_{i,t} = \alpha_i + \beta_1 Y_{i,t-1} + \beta_2 P_i + \text{control variables} + e_{i,t} \quad (1)$$

where *control variables* is the set of variables that account for the autonomous evolution. In this setting, β_2 gives an estimation of the impact of the public policy measure. Which variables are to be included in this set, depends on the programme under investigation (see section IV). In a, say, programme of R&D subsidising, it would be e.g. firms' own R&D expenditure. Of course, this very example illustrates possible difficulties of this approach. Data on firms' own R&D expenditure are possibly difficult to obtain either due to methodological problems (firms might have difficulties in discriminating "own" expenditure from subsidies) or strategic considerations (firms might under- or overstate own expenditures). Therefore, any application of this evaluation method must find out which variables have to be included in the set of control variables.

3.4.1.2. Methods that Use Contemporaneous Data on States of Other Firms

Methods that fall within this group choose a different approach. The *counterfactual* – i.e. the state in which participating firms would have found themselves in, if they had not participated – is estimated using observations on non-participating firms, both measured at the same point in time. As mentioned above, participating firms very often differ systematically from non-participants. There are two possible ways to account for this effect, called in the econometric literature *sample selection bias*: one can utilise information on non-participants which have to be chosen carefully in order to assure maximum comparability of participating and non-participating firms, or the effect has to be corrected for explicitly in the econometric estimation. The following subsections present one well-established method for either case:

(i) Matching approaches

The underlying idea of matching methods is to imitate an experiment, in which a set of pairs of two identical individuals are chosen, only one is treated, and the members of each pair are compared with each other thereafter. In the social context of EU member states, such experiments are legally unfeasible. However, such a setting can be in a way simulated, if we are able to identify for each firm that participates at a public policy measure a "twin" firm that did not. It has been shown in the econometric literature that this approach is possible, if for each participating firm, a "twin" firm can be found with an identical set of variables X that can explain the goal variable Y (the *conditional independence assumption*, Rubin, 1977). Variables that enter the matrix X should be those that are responsible for the selection of the firms into the programme, which depends of course on the specific nature of the programme under investigation. In a R&D subsidy programme e.g. for small firms, possible candidates would be size, age, R&D intensity, etc.

Depending on the size of X and the scaling of its constituting vectors it may turn out to be very difficult or even impossible to identify perfect "twins" to participating firms. Therefore, a number of generalisations of the matching methods have been suggested in the literature (Rosenbaum and Rubin, 1983) . A very intuitive one is the estimation of a *propensity score* that expresses the probability (the propensity) of a firm i to participate in a public policy measure as a function of its realisation of X_i. Once this value is determined for each participating firm and a (preferably much larger) set of non-participating control firms, the "twin" pairs can be determined by matching to each participant *the* non-participant whose propensity score deviates at least.

In many cases an evaluator would like to make sure that certain – though not all – variables of *X* are exact "twins". This cannot be assured by the above approach. However, this method can be modified accordingly by requiring that a subset of *X* should be identical for both groups of firms. The resulting matrix is called *balancing score* and contains a number of variables which is smaller than in the case of the untransformed matrix *X* but larger than in the case of the propensity score.
In any of these three cases, to obtain meaningful results it is necessary to have a sufficient large number of (non-participating) potential matching partners available. With increasing number of potential matching partners, the probability of identifying a comparable "twin" firm also increases. The larger the dimension of *X* (or of the balancing score), the more difficult it is to identify viable matching partners; as a consequence, the larger should be the set of potential matching partners the evaluator can choose from. Therefore, matching approaches can be very data-demanding; they are appropriate when large datasets are available. It should be mentioned however that a number of statistical procedures (such as bootstrapping) are available for handling problems with small data sets.

(ii) Selection correction approaches

A second approach is to explicitly model the firms' decision to participate in a public policy measure or not. Thus, this approach amounts to correcting for potential sample selection biases. To this end, a participation equation is set up that regresses a set of variables *Z* against the participation vector *P* (see equation 1), hence:

$$P_i = \gamma_0 + \gamma_1 Z_i + u_i \quad (2)$$

By taking account of equation (2) and after somer transformations the regression equation of the evaluation becomes:

$$Y_{i,t} = \alpha_i + \beta_1 X_{i,t} + \beta_2 P_i + \gamma_1 Z_i + e_{i,t} \quad (3)$$

where *X* is a set of variables that can explain the level of the goal variable *Y*. Thus equation (3) is a reduced form estimator that includes variables that explain *Y* and others that account for the firms' decision to participate or not.
An alternative approach is to specify the selection equation separately. This allows to specify recursively the probability of participation as a function of the goal variable. This simultaneous equation model has the form

$$A_i = \alpha + \beta_1 X_i + \beta_2 P_i + \text{control variables} + e_i \quad (4a)$$

$$\mathrm{P}_i = \delta + \gamma_1 Z_i + \gamma_2 A_i + u_i \quad (4b)$$

A particular advantage of this approach is that important features of the political economy of firm participation to government programmes can be explicitly taken into consideration in the evaluation procedure.
This selection correction approach is less data-demanding than the matching approach since it does not need a large set of non-participants to choose from.

3.4.1.3. Panel Data Regressions

Matching and selection correction approaches are probably the most established procedures in the current evaluation literature. Approaches based on panel data combine the principles of these approaches using observations on participating and non-participating firms and for each of these from more than one point in time. Panel data regressions – fixed-effects or random-effects estimators – are appropriate when the selection process – i.e. the process that determines why firms participate in a public policy measure or not – cannot be modelled on the basis of observable variables *Z*, as has been implicitly assumed in the previous section. The panel data method specifies an underlying latent model that can account for firm-idiosyncratic unobservable influences of the selection process. Such influences could be the quality of management, firm strategies, etc.

3.4.2. Application to Policy Instruments and Interventions

As it follows from section II, these methods are appropriate when the implications of a public policy are to be scrutinised at the level of participating individual agents (firms, organisations, etc.). Hence, all policy interventions that target firms are possible evaluation candidates, e.g. programmes that aim to foster R&D activities (tax policy, subsidies, etc.), R&D collaborations among firms, the adoption of new technologies, start-up assistance for new high-tech-firms, etc. (for a detailed discussion of an important category of policy instruments which can be evaluated by microeconometric methods see the chapter written by Patel referring to a series of policy initiatives under the heading of ‘technology transfer’).

3.4.3. Good Practice Examples

Seven studies are selected which cover clear-defined government programmes in four European countries (two large und two small ones, three of them members of the European Union), the United States (represented by two studies because of a longer tradition in evaluation practice based on econometric models in this country) and Japan. The government programmes under evaluation are related either to R&D subsidies (Germany, Spain, Japan) or to the promotion of computer-controlled manufacturing technologies (USA, Austria, Switzerland) and took place in the nineties (with the exception of the Japanese project). The seven reviewed studies cover practically all the econometric methods described in section above.

3.4.3.1. “The Effects of Public R&D Subsidies on Firms’ Innovation Activities in a Transition Economy: The Case of Eastern Germany” (Almus/Czarnitzki 2001, Germany)

- **Policy instruments/interventions to be evaluated**

German firms receive R&D subsidies from several government sources. Public R&D schemes are intended to stimulate the privately unprofitable R&D projects which would not carried out without public funding. But this public support may bear a problem. Firms that receive funds may simply substitute public for private investment, thus leading to an inefficient resource allocation. This potential crowding-out effect has to be taken into consideration when public authorities decide on the level of their engagement in public R&D programmes. The evaluation analysis reported in this study tried to find whether firms in Eastern German which received public R&D funds in 1994, 1996 and 1998 have had a higher R&D intensity on average compared to firms that did not receive public support in the respective year.

- **Operational steps for method implementation**

Goal variable(s) : R&D intensity (ratio of R&D expenditure to sales)

Econometric method : Using a matching approach, the R&D intensity of the group of firms receiving R&D subsidies was compared to that of a matched control group of non-subsidised firms. To define a suitable control group a non-parametric matching approach has been used. This approach assigns to each firm of the treatment group (subsidised firms) a firm from the potential control group (non-subsidised firms) which is as close as possible with respect to a number of characteristics ("perfect twin"). These characteristics included firm size (number of employees), firm age, export ratio, import share (at industry level), market share, capital intensity, sellers concentration, affiliation to a foreign mother company and the 2-digit industry a firm belongs to.

Database: The sample contained 966 manufacturing firms located in Eastern Germany from which 653 participated in public R&D schemes. The available data referred to the years 1994, 1996 and 1998 and are taken from the Mannheim Innovation Panel (MIP) conducted by the Centre of European Economic Research (ZEW).

- **Results of policy evaluation**

Firms that received public funding showed a higher R&D intensity on average than firms included in the control group. The estimated difference amounted to about 2.6 percent points on average. This means that a subsidised firm with a turnover of 100'000 marks would have invested 2'600 marks less in R&D on average if it did not participate in public R&D schemes. Moreover, evidence was found that this difference decreases with firm size. This means that a crowding-out effect, thus policy ineffectiveness, is more likely to occur if larger firms are involved.

3.4.3.2. "Evaluation of the ITF Programme FlexCIM" (Geyer et. al. 2000, Austria)

- **Policy instruments/interventions to be evaluated**

The Austrian government started in 1991 a programme to promote the use of "Computer Integrated Manufacturing Technologies" (CIMT) (FlexCIM programme) among manufacturing firms. This programme, which ended in 1996, offered the firms information and training services as well as subsidies for consultancy and development projects. The programme evaluation was conducted on behalf of a government agency.

- **Operational steps for method implementation**

Goal variable(s): Intermediate goal: several adoption variables (change of the CIMT intensity, measured by the number of the CIMT elements used in the period 1992-

1998; CIMT intensity in 1998; number of a firm's functional areas in which CIMT was used for the first time in the period 1992-1998; number of a firm's functional areas in which CIMT was used 1998, etc.).

Final goal: impact of the use of CIMT and programme participation on firm performance and firm organisation (changes in the production techniques, changes in the workplace organisation, firm competitiveness, increase of employment, of sales, of productivity, of exports and of R&D expenses).

Econometric model: Two equations were specified, an adoption equation which, besides the main explanatory variables as proposed by the economic theory of technology diffusion, included policy variables which discriminated between government-supported from non-supported firms, and a policy equation which contained, among other things, also the endogenous variable of the adoption equation.
The model of technology adoption used for the specification of the adoption equation contained several groups of explanatory variables. A first group of variables included a set of anticipated benefits of new technology; a second category of factors referred to anticipated barriers to the adoption of new technology. A third one was related to the (product) market conditions under which the firms were operating (competitive pressure, intensity of quality competition). A fourth one contained measures for physical and human capital intensity. A further group of variables served to characterise a firm's type of products (degree of product differentiation) and production technique ("length of production run"). The firm's ability to absorb knowledge from external sources is important not only for innovation but also for technology adoption; for this reason a proxy for knowledge absorptive capacity was included in the adoption model. Finally, it was controlled for firm size, experience with new technology prior to programme launching, and type of industry a firm was operating in.
The policy equation contained a series of possible factors determining the probability of a firm participating in a government programme. First, a measure of R&D intensity was inserted in the policy equation reflecting a firm's capacity to claim successfully policy support as well as to absorb and apply new knowledge. Further variables included in the policy equation were a profitability measure, a measure of capital intensity, a wage variable, a variable for firm age and dummies for firm size, type of firm organisation (parent vs. affiliate firm) and ownership structure (foreign vs. domestic firm).
The results of the econometric estimation were cross-checked by conducting a second evaluation based on the "propensity score" matching method. This method is based on direct comparisons of participating and non-participating firms. In this case the matching method was applied using the same vector of explaining factors as in the econometric model described above (adoption equation).
Finally, simple probit models were used to investigate the influence of CIMT use (for participating firms) and programme participation on firm performance.

Database: The analysis is based on data for 301 firms (84 participants of government programme, 217 non-participants) collected 1999 by means of a survey which was conducted for this specific purpose.

Econometric method: The technology adoption equation and the policy equation (which also included the endogenous variable of the adoption equation as an additional regressor) were estimated simultaneously. For this special version of a simultaneous probit model the estimation method was based on a "mean- and covariance-structure" model and was implemented in the software programme MECOSA 3 (Arminger et al. 1996). "Propensity score" matching was applied in the standard version.

- **Results of policy evaluation**

The evidence suggests that the policy goal to foster the diffusion of CIMT was attained in the case of firms characterised by a low intensity of AMT use and for small firms with less than 200 employees. The use of CIMT showed a considerably stronger positive impact on employment and sales growth for programme participants than for non-participants.

3.4.3.3. "An Empirical Evaluation of the Effects of R&D Subsidies" (Busom 2000, Spain)

- **Policy instruments/interventions to be evaluated**

The impact of a R&D subsidy programme on the R&D activities of the supported firms at the year 1988 was empirically investigated. Public funding for R&D projects of the participating corporations came from the "Centro para el Desarrollo Tecnologico e Industrial (CDTI)", an agency of the Spanish Ministry of Industry.

- **Operational steps for method implementation**

Goal variable(s): Absolute and relative effort in research and development (R&D expenditure, R&D personnel; ratio of R&D expenditure to sales, ratio of R&D personnel to total employment).

Econometric model: A two-equation framework was developed, the first one explaining the probability of participating in the R&D programme and the second one the level of firms' absolute and relative R&D effort respectively as a function, among other things, of programme participation. The participation equation contained following explanatory variables: firm size (employment), firm age, exports share of sales, type of ownership (public, foreign), pricing behaviour (regulated prices, monopoly, etc.), type of R&D strategy (increase of a firm's own R&D in response to a rival's increase), number of patents obtained in the last 10 years and industry dummies. The equations of R&D effort included, besides most of the independent variables of the participation equation, a dummy for participation in R&D programmes of the European Union.

Database: The data came from a sample of 154 Spanish firms that were conducting R&D activities in 1988, 75 of which received public funding for their R&D projects through CDTI.

Econometric method: In a first step the participation equation was estimated in form of a simple univariate probit model. In a second step the estimation of R&D effort equations was conducted. In a third step the relationship between R&D effort and programme participation was econometrically implemented. Each of these equations was estimated by four procedures. First, by OLS (Ordinary Least Squares) using the entire sample of firms and including a binary variable for participation. The second procedure consisted of splitting the sample to participants and non-participants and estimate OLS separately for each group of firms, thus removing the restriction of equal regression coefficients for both groups of firms. Third, a two step Heckman procedure was used to correct for sample selection. Finally, both equations were jointly estimated by maximum likelihood. The last two procedures allow to take consideration of the endogeneity of programme participation.

- **Results of policy evaluation**

The results of the study show that smaller firms have a higher chance of participating in the programme than larger ones. The same holds also for foreign firms in comparison to domestic ones. On the whole, public funding had a positive impact on R&D efforts of the programme participants, but for a sizeable portion (about 30%) of supported firms complete crowding-out effects between public and private investment cannot be ruled out.

3.4.3.4. "The Effectiveness of Government Promotion of Advanced Manufacturing Technologies (AMT): An Economic Analysis Based on Swiss Micro Data" (Arvanitis et al. 2001, Switzerland)

- **Policy instruments/interventions to be evaluated**

The Swiss government launched in 1990 a programme to promote the use of "Advanced Manufacturing Technologies" (AMT) among manufacturing firms. This programme, which ended in 1996, offered the firms information and training services as well as subsidies for consultancy and development projects; the latter were mostly based on joint ventures between firms or between firms and research institutions embedded in regional networks. The programme evaluation was conducted on behalf of a government agency.

- **Operational steps for method implementation**

Goal variable(s): Change of the AMT intensity between 1990 and 1996 (adoption variable).

Econometric model: Two equations were specified, an adoption equation which, besides the main explanatory variables as proposed by the economic theory of technology diffusion, included policy variables which discriminated between government-supported from non-supported firms, and a policy equation which contained, among other things, also the endogenous variable of the adoption equation. The model of technology adoption used for the specification of the adoption equation contained six groups of explanatory variables. A first group of variables included a set of anticipated benefits of new technology; a second category of factors referred to anticipated barriers to the adoption of new technology. A third one was related to the (product) market conditions under which the firms were operating (competitive pressure, market structure). A further group of variables served to characterise a firm's type of products (degree of product differentiation) and production technique

("length of production run"). The firm's ability to absorb knowledge from external sources is important not only for innovation but also for technology adoption; for this reason some proxies of knowledge absorptive capacity were included in the adoption model. Finally, it was controlled for firm size, experience with new technology prior to programme launching, and type of industry a firm was operating in.
The policy equation contained a series of possible factors determining the probability of a firm participating in a government programme. First, some policy-related variables such as proxies for firms' experience with other government programmes with similar goals and measures of financial barriers to investment and innovation were used as regressors. Secondly, the age of the firm, the ownership structure (foreign vs. domestic; parent vs. affiliate vs. independent firm), the firm size and industry dummies were included as control variables in order to capture further unspecified effects.

Database: The analysis is based on data for 463 firms (96 supported and 367 non-supported firms) collected in the course of the Swiss Innovation Survey 1996 as a supplement to the standard questionnaire, thus allowing the combination of technology-specific information with basic data on innovation and technology use.

Econometric method: The technology adoption equation and the policy equation (which also included the endogenous variable of the adoption equation as additional regressor) were estimated simultaneously. For this special version of a simultaneous probit model the estimation method was based on a "mean- and covariance-structure" model and was implemented in the software programme MECOSA 3 (Arminger et al. 1996).

- **Results of policy evaluation**

The evidence suggests that the policy goal to foster the diffusion of AMT was attained in the case of firms adopting AMT for the first time or characterised by a low intensity of AMT use, and for small firms with less than 200 employees (with some overlapping between these two categories).

3.4.3.5. "Manufacturing Extension and Productivity Dynamics" (Jarmin 1998, USA)

- **Policy instruments/interventions to be evaluated**

In recent years a consortium of state, local and federal agencies have created a nationwide network of manufacturing extension centres designed to help small and medium manufacturers improve productivity and become more competitive. Typical services provided by centres include changes in plant layout, process redesign, software selection, preparing plants for ISO-9000 certification and marketing assistance. The evaluation study investigated first whether there is a performance improvement (in terms of productivity) of the client plants in comparison to non-client plants and second whether this improvement can be traced back to plants' participation to a programme of manufacturing extension.

- **Operational steps for method implementation**

Goal variable(s): Labour productivity, total factor productivity

Econometric model: Two models were used to estimate the impact of extension services on plant performance: (a) a labour productivity model containing labour, the ratio of capital to labour, dummies for industry, year and state and a dummy variable for participation in an extension programme as explanatory variables and (b) a total factor productivity model with dummies for industry, year and state as well as a dummy for using services of an extension programme as regressors.

Database: For this evaluation a (balanced) panel data set was constructed with annual data for 726 client and 5818 nonclient plants from 1987 to 1993. The data used in this evaluation were from two sources. First, plant level production data were taken from the Census Bureau's Longitudinal Research Database (LRD). The LRD was constructed by linking plant level data from the Censuses and Annual Surveys of Manufactures. Second, manufacturing extension client data came from nine manufacturing extension centres located in three states.

Econometric method: Unbiased estimates of the programme impact parameter (coefficient of the programme impact dummy in the productivity equations) were obtained by using fixed-effect estimators for the OLS regressions, thus assuming that the error term in the regressions has both permanent and transitory components. These fixed-effect models were estimated for within, difference and growth rate regressions.

- **Results of policy evaluation**

The main conclusions from the evaluation analysis are the following: a) there is clear evidence for a positive impact of the programmes of manufacturing extension on productivity; b) the estimated impact on total factor productivity is smaller than that for labour productivity and is statistically significant only in a part of regressions; c) the range of estimates for programme impact in this analysis is similar to those obtained in previous studies; d) finally the results provide evidence that the timing of performance improvements is consistent with positive impacts of participating in manufacturing extension.

3.4.3.6. "Does manufacturing Extension matter? An Evaluation of the Industrial Technology Extension Service in New York" (Oldsman 1996, USA)

- **Policy instruments/interventions to be evaluated**

The Industrial Technology Extension Service (ITES) of New York State was established in 1990 to help small and medium-sized manufacturers upgrade production technologies and management practices. The evaluation study investigated the programme influence on several types of outcomes such as specific actions or changes in behaviour within the firm, intermediate effects (e.g. manufacturing lead time reduction), business outcomes (e.g. cost savings) and desired policy outcomes (e.g. increased employment). The evaluation was conducted partially on behalf of a government agency.

- **Operational steps for method implementation**

Goal variable(s): The model used for the evaluation analysis called attention to the chain of events leading from extension services to desired policy outcomes. The provided services were intended to lead to specific actions or changes in behaviour within the firm (goal variable: "action", e.g. "change of plant layout" or "adoption of total quality management"). Actions taken by the firm were intended to resolve identified problems, resulting in improvements along a number of dimensions (goal variable: "intermediate impacts", e.g. manufacturing lead time reduction). By resolving problems, it was expected that the firm would increase sales, reduce costs,, or otherwise generate higher profits (goal variable: "business outcomes"). Finally, these favourable business outcomes were expected to lead to desired policy outcomes, e.g. increased employment (goal variable: "policy outcomes").

Econometric model: An econometric model was specified for every type of goal variable defined above ("actions", "intermediate impacts", "business outcomes" and "policy outcomes"). The equation for "actions" contained besides policy variables (variable for the combined total of meetings and telephone calls that the agency had with a firm; dummy variable for receiving direct assistance) firm size (number of employees), industry dummies, dummies for the company being publicly held and for the firm having more than one plant. The equation for "intermediate impacts" included, besides policy variables, firm size and industry dummies, a series of possible "actions" leading to the intermediate impacts (improving plant layout to facilitate work flow, using CNC tools, adopting SPC, employing total quality management, etc.). The equation for "business outcomes" included industry dummies in addition to the policy variables already described. Finally, the equation for "policy outcomes" which used the employment after receiving programme support as dependent variable contained, besides the policy variable (number of months in which activity was reported for the firm in 1992 by the agency), firm size, regional and industry dummies as well as employment prior to programme support.

Database: A mail survey of clients was used to obtain information on the characteristics of companies participating in the ITES programme, nature and magnitude of participation, satisfaction with services and perceived benefits. The 12-page questionnaire was sent out to all companies that had been active in the programme between July 1990 and March 1993. A total of 222 complete surveys were returned, representing an effective response rate of 20%. Further, the New York State Department of Labour provided quarterly employment data for the period between January 1991 and March 1993.

Econometric method: A maximum-likelihood logit model was used to estimate the probability of a firm taking a specific action and attributing it to services received from the extension programme (equation for "actions"). A maximum-likelihood ordered logit model was used to estimate the relationship between benefits reported by companies, actions taken as results of services received and firm characteristics (equation for "intermediate impacts"). Median regression was applied as estimation method for the equation for "business outcomes" (cost savings). Finally, OLS (Ordinary Least Squares) was used to estimate the equation for policy outcomes" (employment increase).

- **Results of policy evaluation**

The analysis suggests that the duration and intensity of received services matter. Companies having more contact with the supporting agency are likely to report higher benefits. Moreover, companies are able to reap greater cost savings if direct assistance from the supporting agency is provided. It also appears that manufacturing extension programmes can have a beneficial impact in terms of increased productivity. This dampens demand of labour in the short-term. However, participating firm may, in time, be able to secure a larger share of the market for the products, thus adding jobs in the long-term if sales grow faster than productivity.

3.4.3.7. "Evaluation of Government-Sponsored R&D Consortia in Japan (Sakakibara 1997, Japan)

- **Policy instruments/interventions to be evaluated**

Between 1983 and 1989 a large number of government-sponsored co-operative R&D projects (including R&D contracts, R&D consortia and research joint ventures) were set-up in Japan. These projects were organised by MITI. To the extent that the MITI officials could observe it, it is quite likely that MITI officials picked firms with higher "research quality" for participation in the sponsored R&D consortia. The evaluation was conducted in close co-operation with government agencies.

- **Operational steps for method implementation**

Goal variable(s): R&D spending per year, patents generated per year ("research productivity"), extent of knowledge spillovers among co-operating firms.

Econometric model: Three types of econometric equations were specified. First, a R&D expenditure equation was developed which contained the log of firms'capital stock to control for firm size, industry dummies as additional control variables and the number of consortia in which a firm was involved in year t as participation (policy) variable. This equation was formulated in order to test the hypothesis that increases in the intensity of participation were associated with increases in R&D investment. Second, a patent production function was specified which included, besides industry dummies, the log of a firm's own R&D investment (as control variable) and the policy variable (the number of consortia in which a firm was involved in year t) as an additional independent variable. A positive coefficient of policy variable demonstrates a positive relationship between participation and patenting; this effect can be interpreted as empirical evidence for an increase of "research productivity" (measured as patents generated per year, controlling for R&D spending, among other things) of participating firms in comparison to non-participating ones. For a part of the estimations also a participation equation was specified; it contained primarily lagged policy variables. Finally, the patent production function was extended to include proxies for knowledge spillover effects. The sample was divided into non-participants/infrequent participants and frequent participants and the parameter of the knowledge spillover variable was allowed to vary across the two sub-samples. In this way two parameters, one for every sub-sample, could be estimated. This last equation was used to estimate (indirectly) the impact of consortia on knowledge spillovers. If the coefficient of the spillover variable for the group of participating firms was larger than that for the non-participating ones, there was an augmentation of knowledge spillovers due to the (frequent) participation in R&D consortia.

Database: Data were collected on 226 firms' R&D spending, sales, capital stock, labour and materials usage and patenting in the United States as well as in Japan, for the period 1983-1989. In addition, data on participation in consortia were available. Of the sample firms, 141 participated in at least one consortium.

Econometric method: The R&D equation was estimated with ordinary least squares (OLS) as a fixed-effects as well as a random-effects model. To estimate the patent production function two approaches were taken. The first was the fixed-effects approach. The second was the standard two-stage least squares (2SLS) approach. In the first stage exogenous and lagged endogenous variables were used to predict the number of research consortia a given firm would be involved in during a given year (participation equation); in a second step, the participation measure was instrumented using these predicted values. Alternatively to OLS estimates the patent production function was also estimated as a Poisson model and as a Negative Binomial fixed-effects model, in order to take account of the specific character of patents as a count variable.

- **Results of policy evaluation**

Participation in R&D consortia tends to be associated with higher levels of R&D spending of participating firms. Participation in R&D consortia also seems to raise the research productivity of participating firms. Finally, the results suggest that at least one channel through which consortia have these positive effects may be through effectively augmenting knowledge spillovers among participating firms.

3.4.4. Data Requirements

Econometric estimations require data for a large number of individual agents (firms, organisations) in order to obtain statistically adequate results. The number of observations needed for these estimations depends heavily on the *magnitude of the project* to be evaluated and on the *evaluation method* applied. For a programme which covers some hundred firms one would require information for all participants, when many thousands of supported firms exist, one has to take a sample of the population of the participants which is sufficient large for econometric estimations. The reviewed studies used at least data of 150 to 200 firms, these numbers being some kind of absolute lower bound for a statistically satisfactory econometric estimation. One needs data not only for programme participants but also for *non-participants,* especially when a matching method is used, a procedure which is especially data-demanding (see also section II).
Which type of information is needed, depends on the informational requirements of the underlying economic model(s). Models with *many explanatory variables* models need more information per available individual (firm, organisation) than models which contain only few variables. When *several goals* are pursued (intermediate, final goals), then one needs also several goal variables, several estimation equations and so on, and the data requirements are high.

Ideally, the evaluator would dispose of *time series* of data from before to after the policy support programme; often only cross-section information is available, a fact restricts the number of questions that can be adequately treated in the evaluation.

Which kind of indicators is to be applied, depends on the type of economic relations to be considered (model), the data availability or, alternatively, the possibility to collect new data. The evaluator would try to use such indicators that are already well-known in the economic literature and allow the comparison with similar studies. In most cases, the well-established indicators can be used. When new indicators have to be constructed, one has to pay attention that the new measures are clear-cut and economically plausible. This also increases the chances of these new indicators being widely accepted.

3.4.5. Conditions for Application

In sum, following conditions should be fulfilled in order to be able to apply satisfactorily the econometric methodology:

- The implications of the applied policy measures are scrutinised at the level of participating individual agents (firms, organisations, etc.).
- The goals of the support programme can be formulated in such a way, that their degree of attainment can be measured by some indicator or a set of indicators, at best with indicators which are already well-known and therefore widely accepted.
- The relationship between the policy goals and the used policy instruments can be modelled based on some piece(s) of available economic theory.
- Data for a large number of firms are available or is planned to be collected specifically for the evaluation study.
- Data are needed not only for programme participants, but also for non-participants.

3.4.6. Steps for Implementation

We distinguish following operational steps:

- *Definition of goal variables (intermediate, final goals, etc.)*: depending on the nature of the policy measures to be evaluated, an "evaluation model" with intermediate, primary and secondary final goals, etc. can be postulated.
- *Construction of an econometric model*: given the "evaluation model", state of the art of economic theory and economic empirics in the relevant field and some (preliminary) considerations with respect to data requirements, an econometric model can be constructed.
- *Search for / Collection of appropriate data*: given the econometric model final data requirements can be stated; the required data may be already available or have to be collected specifically for the evaluation task.
- Given the model and the available data and, mostly, after some adjustment of the model specification in accordance with the final data availability, an appropriate *econometric method* has to be chosen.
- Run of econometric estimations.
- *Interpretation of the results* of the econometric estimations

3.4.7. Scope and Limits

A first advantage of econometric approaches, especially those applying an explicit modelling of goal and policy variables, relative to alternative methods is that the evaluation analysis is based on the explicit formulation of theory-based causal relationships between the goal variables of a policy support programme and the factors influencing these goal variables. Second, factors related to the political economy of the selection (or participation) of firms with respect to the support programme in question are explicitly taken into consideration in the evaluation procedure.

These methods quantify implications of public policy measures on the firm level and can give a precise assessment of its impact at this level. Thus, microeconometric methods can capture direct effects only for participating firms. More general (indirect) effects which may also affect non-participants such as knowledge spillovers within one industry or the economy as a whole cannot be assessed by these methods. Hence, if a public policy measure aims at generating, in addition to direct, also indirect effects (for non-participating agents), microeconometric methods should not be used exclusively. Rather, they should be combined with methods that investigate a measure’s impact on the appropriate – i.e. industry or economy-wide – level (see also chapter on macroeconometric models).

The most serious shortcomings of the use of econometric methods in policy evaluation are related to data limitations. Mostly, the evaluators do not dispose of time series of data from before to after the policy support programme or only data on a few variables not allowing an adequate modelling of the underlying relationships. The problems related to correctly specifying the empirical model and jointly testing the assumed causal relationships - to be confronted with in practically every econometric study - are the main sources of difficulties.

A further practical problem for the evaluator refers to the effort needed in order to persuade a large number of participants and non-participants to disclose sensitive firm data. The collection of such data in combination with a national (e.g. R&D survey or investment survey) or European survey (e.g. CIS survey) conducted by well-known and trustworthy institutions could help to reduce this type of difficulties.

In general, part of a good evaluation strategy should be the parallel use of several evaluation methods in order to check up the results and correct for the shortcomings of the single methods. For example, properly-designed case studies based on a number of “characteristic” firms (e.g. supported vs. non-supported; large vs. small; belonging to a high-tech sector vs. belonging to a traditional sector, etc.) could be used not only to control for the plausibility of the results of the econometric analysis but also to help focus to the really relevant factors in the evaluation equations. The analysis of subjective assessments of the firms of the impact of a policy measure can also be utilised to check the consistency and plausibility of econometric analysis. Of course there is also a price to pay for such a multi-dimensional strategy, which is the high cost of conducting evaluation work along several lines.

In spite of the shortcomings of the microeconometric evaluation approaches sketched above there are still many advantages on their side which force us to consider them a core element of policy evaluation. Improvements of the database would significantly increase its reliability because most of the weaknesses of the approach lie at the empirical level. Therefore, it is crucial that the preconditions for an evaluation should be secured from the very beginning of policy implementation, i.e. at the preparatory stage of a promotion programme. This means, among other things, that the institutions responsible for policy implementation should be obliged to collect the necessary data and have the authority to enforce the participating firms to deliver the required information.

3.4.8. Bibliography

General Surveys:

The special issue of *Research Policy*: "The Economics of Technology Policy", Guest Editors A.N. Link and J.D. Rossner (Vol. 29, 2000) contains a collection of articles on evaluation methods which cover the most relevant aspects on this topic.

Georghiou and Roessner (2000) give a critical survey of evaluation methodologies in general. They discuss advantages and limitations of different (not only microeconometric) methods. They also review a large number of different public policy programmes within different contexts.

The Special Issue of the *IPTS Report* of December 1999 gives a survey of RTD programmes of the European Union which have been recently evaluated.

OECD (1997) reports the results of an international conference on techniques of technology evaluation reflecting the state of the art in this field.

Surveys of Evaluations Based on Econometric Methods:

David et al. (2000) investigate the question whether public R&D is a complement or a substitute for private R&D by doing a large survey on the econometric evidence. They make a point that 11 out of 33 studies find public R&D spending to be a substitute while the rest of the studies do not. It is noticeable that the results depends heavily on the level of aggregation: Only 2 studies out of 14 on the industry level assess public R&D as substitute for private R&D, while this is the case for 9 out of 19 studies on the firm level.

Hall and Van Reenen (2000) review in a recent article the empirical evidence on the effectiveness of fiscal incentives for R&D and also discuss econometric methodology.

Klette et al. (2000) give a comprehensive survey of the empirical studies dealing with R&D subsidies including a good presentation of the methodological problems related to such microeconometric investigations (comparable to that in section II 2). They give special attention to the question of knowledge spillovers and its implications for R&D subsidies. They argue in favour of a combined application of methods.

Salter and Martin (2001) give a survey on a large body of evaluation literature on publicly-funded basic research. They argue that the benefits of basic research differ as a function of scientific field, type of technology and industry under consideration. They conclude that an evaluation has to take account of these facts.

Surveys of Econometric Methodology:

Heckman et al. (1999) give a very comprehensive, although technical survey on microeconometric evaluation methods from the point of view of labour market programmes.

Literature:

Almus, M. and D. Czarnitzki (2001): 'The Effects of Public R&D Subsidies on Firms' Innovation Activities in a Transition Economy: The Case of Eastern Germany', *ZEW Discussion Paper No. 01-,* Mannheim.

Arminger. G., Wittenberg, J. and A. Schepers (1996): *Mean- and Covariance-Structure Analysis (MECOSA)*, User Guide, Friedrichsdorf.

Arvanitis, S., Hollenstein, H. and S. Lenz (2001): 'The Effectiveness of Government Promotion of Advanced Manufacturing Technologies (AMT): An Economic Analysis Based on Swiss Micro Data', *Small Business Economics* (forthcoming).

Busom, I. (2000): 'An Empirical Evaluation of the Effects of R&D Subsidies', *Economics of Innovation and New Technology*, 9, 111-148.

David, P., Hall, B. and A. Toole (2000): 'Is Public R&D a Complement for Private R&D?' *Research Policy*, 29, 497-529.

Geier, A., Polt, W., Pointner, W., Rammer, C., Arvanitis, S., Donzé, L. und H. Hollenstein (2000): *Evaluierung des österreichischen ITF-Programms FlexCIM*, Schlussbericht, Wien.

Georghiou, L. and D. Roessner (2000): 'Evaluating Technology Programmes: Tools and Methods', *Research Policy*, 29, 657-678.

Hall, B. and J. Van Reenen (2000): 'How Effective are Fiscal Incentives for R&D?', *Research Policy*, 29, 449-469.

Heckman, J., LaLonde, R.J. and J.A. Smith (1999). 'The Economics and Econometrics of Active Labour Market Programmes', in: O. Ashenfelter and D. Card (eds.), *Handbook of Labour Economics, vol. 3A*, Elsevier, Amsterdam.

IPTS Report (1999). Evaluation of R&D Activities: Special Issue, European Commission, Seville.

Jarmin, R.S. (1998). 'Manufacturing Extension and Productivity Dynamics', *CES Working Paper No. 98-8*, U.S. Bureau of the Census, Washington, D.C.

Klette, T.J., Moen, J. and Z. Griliches (2000): 'Do Subsidies to Commercial R&D Reduce Market Failures? Microeconometric Evaluation Studies', *Research Policy*, 29, 471-495.

Link, A. N and J.D. Rossner (eds) (2000) 'The Economics of Technology Policy: Special Issue', *Research Policy*, Vol. 29.

OECD (ed.) (1997): Policy Evaluation in Innovation and Technology - Towards Best Practices, OECD, Paris.

Oldsman, E. (1996): Does Manufacturing Extension Matter? An Evaluation of the Industrial Technology Extension Service in New York, *Research Policy*, 25, 215-232.

Rosenbaum, P. and D.B. Rubin (1983): The Central Role of the Propensity Score in Observational Studies for Causal Effects, *Biometrica*, 70, 41-55.

Rubin, D.B. (1974): 'Estimating Causal Effects for Treatment in Randomised and Nonrandomised Studies', *Journal of Educational Psychology*, 66, 688-701.

Sakakibara, M. (1997): 'Evaluation of Government-Sponsored R&D Consortia in Japan', in: OECD (ed.), *Policy Evaluation in Innovation and Technology – Towards Best Practices*, Paris.

Salter, A. and Martin, B. (2001): 'The Economic Benefits of Publicly Funded Basic Research', *Research Policy*, 3, 509-532.

3.5. ECONOMETRIC MODELS: PRODUCTIVITY MEASUREMENT

Author: **Jonathan Eaton** (Boston University)

3.5.1. Introduction

Productivity refers to the amount of output that a given set of inputs can produce. The ultimate purpose of RTD activity is usually to raise productivity. Hence productivity measurement is central to the *ex post* evaluation of the success or failure of such policies. The *ex ante* evaluation of a proposed RTD activity in most cases would hinge ultimately on its effect on productivity.

In measuring productivity the unit of observation can be an individual worker, plant, or firm (at the micro level), a sector (the meso level), or region or country (macro level). A distinction is between single-factor productivity (where, outside agriculture, the factor is usually labor, while agricultural productivity is occasionally reported as per unit of land) and multifactor productivity (MFP), where several factors (capital, labor, occasionally adjusted for education and experience, land, etc.) are taken into account. While measuring MFP allows the effects of variation in nonlabor inputs to be distinguished from changes in productivity, measuring MFP requires substantially more data and assumptions about technology.

A manifestation of technical progress is increasing productivity over time. A more or less consensus view since Solow (1957) is that technical progress accounts for at least half of economic growth (with the accumulation of capital explaining the remainder) at the macro level. Somewhat more controversial is the view that productivity differences across countries are a major reason for cross-country income differences. Mankiw, Romer, and Weil (1992) argue that technology is inherently mobile, so that productivity differences across countries are negligible. They attribute cross-country income differences entirely to differences in capital stocks per worker and educational attainment. But Parente and Prescott (1994), Eaton and Kortum (1996, 1999) and Hall and Jones (1999) report evidence of substantial cross-country differences in technology.

At the meso level, a major debate has concerned the sources of productivity growth over the last decade. A claim is that advances in the computer sector have been the major source of technical progress. An issue is how to allocate the contributions of better computers between the computer sector and the industries that achieve higher productivity through the use of better computers. Another issue has been the measurement of productivity in the service sector, where obtaining a quality-adjusted physical output measure is particularly problematic.

At the firm and plant level an issue has been the role of entry by more productive units and exit by less productive units, rather than improved productivity at surviving units.

3.5.2. Application to Policy Instruments and Interventions

Since raising overall productivity across the economy is usually the ultimate goal of virtually all RTD policy, the effect on productivity at the national level provides an ultimate test of the efficacy of any instrument or type of intervention. Other policies might have a regional or sectoral focus so that productivity outcomes at that level are key.

Connecting the use of a particular instrument or intervention with the ultimate effect on macro or meso level productivity may often prove extremely difficult. Too many other factors affecting productivity vary to allow for appropriate controls. Hence the effect of many interventions may be better observed at the micro level, especially if an experiment is undertaken involving a control group of production units. At the micro level any intervention that affects some producing units more than others can in principle be evaluated in terms of its effect on the measured productivity of those units. The difference in the effect on units could be by design (e.g., if a control group of units is excluded from the intervention) or by accident (e.g., if the intervention happens to affect some units more than others, maybe because of its limitation to a particular region). In the second case adequate control variables may be required to account for other differences between different units.

Several studies have examined the effect of trade liberalisation on productivity at the producer unit. Tybout (2001) provides a survey. Other studies have used micro level productivity measures to examine technological spillovers across firms (e.g., Currie and Harrison, 1997).

3.5.3. Good Practice Examples

At the national level productivity measurements are provided according to a number of sources. The most comprehensive cross-national source is the International Monetary Fund data on Gross Domestic Product (GDP) per worker. The methodology for creating these data is described by Summers and Heston (1991). An objection to the Heston-Summers data as measures of productivity is that they are primarily intended as indicators of welfare. The measure attempts to capture the real resources that a worker could acquire with her output. While differences in labor productivity probably account for most of the variation in the measure both across countries and over time, these measures are also affected by changes in the terms of trade.

Van Ark (1996) provides cross-country national measures that seek to isolate productivity directly. Van Ark also pursues some sectoral disaggregation and investigates the effect of variations in hours worked. A limitation of the van Ark measures is that they are available only for Europe, the United States, and Japan.

An early attempt to measure aggregate multifactor productivity is by Solow (1957).

The U.S. Bureau of Labor Statistics measures aggregate multifactor productivity and MFP in various sectors of the U.S. economy on an annual (and in some cases quarterly) basis. The bureau provides a description of their methodology, as well as the data themselves, on their at website, http://www.bls.gov/mfp/home.htm.

While there are various efforts to compile cross-country measures of productivity differences at the macro and meso levels, comprehensive micro-level data are available only for production units within national boundaries. At this point there are no datasets that combine producer units across a broad range of countries. Various national statistical offices maintain micro datasets for individual producers. Examples are those datasets maintained by the U.S. Census of Manufactures (primarily for U.S. manufacturing plants) and by the French INSEE (for French firms in all sectors).

A wide range of studies have measured and analysed productivity at the producer level. Studies that limit themselves to labor-productivity measures include Roberts and Tybout (1996), Clerides, Lach, and Tybout (1998), Aw, Chung, and Roberts (1998), and Bernard and Jensen (1999). Studies providing measures of multifactor productivity include Olley and Pakes (1996), for the U.S. telecommunications equipment industry, Levinsohn and Petrin (2000), Pavcnik (2000), both for Chilean manufacturing firms, and Frazer (2001), for Ghanaian firms.

3.5.4. Data Requirements

Productivity measures require data on outputs and inputs over some period of time (usually a year). In the case of meso and macro level measures output is always measured in terms of the value of production. At the micro level output can be measured in terms of value or occasionally in physical terms. If it is a value measure the output measure at the micro level is usually value added: the value of shipments (for a plant) or sales (for a firm) less the value of material inputs.
Labor inputs are typically measured as the physical number of workers, which may or may not be adjusted for hours worked. Some MFP measures attempt to distinguish among different types of workers. Frazer (2001) adjusts labor input by schooling at the firm level. Bils and Klenow (2000) discuss schooling adjustments at the macro level. Macro level data on educational attainment, the Barro-Lee (1996) dataset, are available at the World Bank website:
http://www.worldbank.org/research/growth/ddbarlee.htm.
Capital inputs are usually measured as the value of the capital stock. Capital stock data are frequently constructed from data on investment using the perpetual inventory method. Jorgenson (1995) provides a discussion.

3.5.5. Steps for Implementation

Behind the notion of productivity is the production function, which relates output (or value added) to the inputs (or factors of production) used to produce it. The production function could apply to the output or value added of a firm, plant, sector, region, or country. Denoting unit i's output in period t as Y_{it}, and the vector of physical resources (labor, capital, land, etc.) used to produce this output as X_{it} we can write:

$$Y_{it} = G_{it}(X_{it}). \qquad (1)$$

where G_{it} is the relationship between these inputs and output at unit i at time t.

Higher productivity is a change in the function F_{it} that generates as much or more output for any set of inputs, and strictly more output for at least some set of inputs. In general we cannot identify each unit's production function at each unit of time without imposing some restrictions on the form of the production function. A common assumption is that differences in the production function across units at any given time or over time at any given unit arise only in a multiplicative factor A_{it}.

Hence we can write:

$$Y_{it} = G_{it}(X_{it}) = A_{it}F(X_{it}). \tag{2}$$

which imposes the condition that F is invariant across units or over time. Another common assumption is that the function F is characterised by constant returns to scale.[34]
The term A_{it} is often referred to as total factor productivity (TFP) and provides a scalar indicator of productivity. As discussed in the next section, most standard productivity comparisons are scalar in nature so require these assumption, although some practitioners have erroneously attempted to perform such comparisons in contexts in which the assumption was clearly violated.
An additional restriction that is typically imposed is that the function F is of the Cobb-Douglas form and is characterised by constant returns to scale. In the case of two inputs, labor L_{it} and capital K_{it},

$$F(X_{it}^{P}) = L_{it}^{1-\alpha}K_{it}^{\alpha},$$

where $\alpha \in [0,1]$ is the capital share. The classic (1957) analysis of Solow's assumed a Cobb-Douglas technology. A subsequent paper by Arrow, Chenery, Minhas, and Solow (1961) introduced a more flexible (CES) functional form imposing only a constant elasticity of substitution across factors. Cobb-Douglas emerges as the special case in which the elasticity of substitution is one. An advantage of the Cobb-Douglas specification is that, if factors are paid their marginal products, then factor shares in payments correspond to their shares in the production function. Jorgenson (1995), in some cases with various co-authors, provides productivity measures using a translog functional form that is not nested within nor nests the CES form. While the translog is very flexible, a disadvantage is that its dual forms has no closed form. Unless noted otherwise, the discussion here will assume Cobb-Douglas technologies.

3.5.5.1. Growth Accounting

Under the assumption that factors are paid their marginal products, α corresponds to the capital share in the units total expenditure, a number that is typically estimated between ¼ and 1/3.
Under these assumptions, along with data on Y, L, and K, the TFP term A_{it} can be calculated linearly from the logarithmic expression:

$$\ln A_{it} = \ln Y_{it} - (1-\alpha)\ln L_{it} - \alpha \ln K_{it}. \tag{3}$$

The absolute level of A_{it} is of no independent interest since it depends on the units in which inputs and outputs are measured. But ones these units are fixed comparisons of A_{it} across time or across units are meaningful.
Note that calculating A_{it} requires knowing the full set of inputs and their share in production. In the absence of any knowledge of the capital stock practitioners often measure labor productivity as output per worker $y_{it} = Y/L_{it}$. Unless the capital share is 0 an obvious problem with this measure is that it fails to control for potential variation in the capital stock. That is, $y_{it} = A_{it}k_{it}^{\alpha}$, where $k_{it} = K_{it}/L_{it}$, the capital stock per worker.

34 In fact, with nonconstant returns to scale productivity can only be defined conditional on a scale of production. Moreover, disentangling scale effects from productivity differences in actual data is highly problematic.

However, if it is the case that the interest rate r is the same for each observation, and the marginal product of capital equals the interest rate, then for each observation:

$$r = \alpha A_{it} k_{it}^{\alpha-1}.$$

Solving for k_{it} and substituting into the expression for y_{it} yields:

$$y_{it} = A_{it}^{1/(1-\alpha)} \left(\frac{\alpha}{r} \right)^{\alpha/(1-\alpha)}.$$

Hence labor productivity does reflect differences in underlying TFP monotonically. Note that under these assumptions differences in y_{it} exaggerate differences in A_{it} by a factor of $1/(1-\alpha)$. The reason is that more capital is used where it is more productive.

3.5.5.2. Production Function Econometrics

Growth accounting requires independent knowledge of factor shares in production (which, as discussed above, can, under an assumption of perfect competition in factor markets, be inferred from payment shares). In many cases the researcher does not have this information independently, in which case econometric estimation of the production function and its parameters has been undertaken. The basic approach is to use equation (3) above as the basis for an estimation equation using data on output and factor inputs. The approach also requires identifying factors that might account systematically for differences in productivity, such as the passage of time (denoted by t) or some policy intervention P_{it}.

One usually estimates the equation:

$$\ln Y_{it} = f(t) + \beta P_{it} + \gamma_L \ln L_{it} + \gamma_K \ln K_{it} + u_{it} \qquad (4)$$

where $f(t)$ is some function of time and u_{it} an error term. A standard approach is to assume $f(t) = \lambda t$, where the coefficient λ is an estimate of the "rate of technical progress." A more general approach, if panel data are available, is to include a dummy variable D_t for each period. In this case all units of observation are assumed to be subject to a common "technology shock" associated with the period of observation. The coefficient β provides an estimate of the efficacy of the policy intervention. The coefficient γ_i denotes the contribution of factor i to output. The assumption of constant returns to scale implies the restriction that $\gamma_L + \gamma_K = 1$. The term u_{it} captures variation in TFP that cannot be systematically related to observable parameters.

Note that this methodology provides a direct means of assessing the contribution of policy intervention on productivity, assuming that quantitative measures of the intervention are available. The hypothesis that instrument P_{it} has no effect can be tested against the alternative that it has a positive effect, for example.

Equation (4) can be estimated using data on factor inputs, output, and the policy intervention. Applying standard ordinary least squares is highly problematic, however. Griliches and Mairesse (1998) provide an excellent survey of the econometric problems. A particular criticism, going back to Marschak and Andrews (1944), is that the choice of inputs is unlikely to be independent of TFP. In particular, more factors are likely to be employed where productivity is higher. An implication is that the parameter estimates of γ_i are biased upward. If constant returns to scale are not imposed then estimation will tend to exaggerate the role of increasing returns, leading to downward bias in TFP. Random or policy induced variation in A_{it} will consequently be understated. Intuitively, if a policy intervention that leads to higher TFP causes more factors to be employed, the regression will mistakenly attribute too much of an increase in output to more factor employment and too little to random or policy induced changes in TFP.

A potential solution is to impose some assumption about returns to scale, such as constant returns to scale, on *a priori* grounds. However, as Marschak and Andrews point out, a problem remains if not all factors are equally variable during the period of observation. At the micro level employment L_{it} might vary with u_{it} more than capital K_{it}. At the macro level the opposite is more likely as capital appears to be much more mobile internationally than labor (as suggested by the greater cross-country similarity in interest rates as opposed to wages).

Following Griliches and Mairesse's (1998) useful formalization of the problem, we rewrite equation (4) as:

$$\ln Y_{it} = f(t) + \beta P_{it} + \gamma_L \ln L_{it} + \gamma_K \ln K_{it} + a_{it} + e_{it} + \varepsilon_{it} \,, \qquad (5)$$

decomposing the TFP term u_{it} into three components: a_{it} is observed by the unit in time to make its current decision about L_{it} but not K_{it}, which was predetermined by last period's investment decision; e_{it} is a shock to TFP observed by the unit but too late to affect any factor employment decisions; ε_{it} is an error in measurement. It has mean zero and is serially uncorrelated (otherwise it would not be a total shock). The first two errors are true components of the production function while the third is simply noise introduced by measurement problems.

Under these assumptions the firm will hire labor in order to maximise its expected profit knowing a_{it} (and its previously chosen K_{it}) but not knowing the realisation of e_{it}. (It doesn't care about ε_{it}; that is only the econometrician's problem.) Facing a wage w_t to maximize expected profit it will hire:

$$L_{it} = \left(e^{f(t) + \beta P_{it} + a_{it}} K_{it}^{\gamma_K} / w_t \right)^{1/(1-\gamma_L)} \qquad (6)$$

workers. Substituting (6) into (5) one sees that a_{it} appears both in the error term and in the labor choice, inducing a positive correlation between L_{it} and the error in the equation. As a consequence, even if returns to scale are properly constrained, γ_L will overstate the contribution of labor and γ_K understate the contribution of capital. Since L_{it} is varying more than K_{it} in response to TFP, and since labor's contribution to production is overstated relative to that of capital, policy-induced or random changes in TFP are underestimated.

One approach is to hope that all of the variation in a_{it} occurs across units rather than over time, so that it is a constant, unit specific effect a_i. Using panel data one can then estimate:

$$\ln Y_{it} = \beta P_{it} + \gamma_L \ln L_{it} + \gamma_K \ln K_{it} + \sum_{i=1}^{N} \delta^j D_i^j + \sum_{\tau=1}^{T} \delta^\tau D_t^\tau + u_{it} \qquad (7)$$

where D_i^j is a variable that equals 1 if $i = j$ and 0 otherwise while D_t^τ is a dummy variable that equals 1 if $t = \tau$ and 0 otherwise. The unit fixed effects control for the a_i and will lead to unbiased estimates of γ_i.

There are a number of problems, however. One is simply that a_{it} is likely to change over time. But even if it is constant, parameter estimates may be much more weakly identified, since variation over time tends to be much less than variation across units. Moreover, if the policy intervention does not have a great deal of variation both across time and across units, its importance will be harder to assess.

The standard remedy for simultaneity bias is the use of instrumental variables. But a valid instrument for factor use must be correlated with factor use but not with output. Such an instrument is difficult to imagine, at least at the micro level.

Olley and Pakes (1996) propose an alternative simultaneity correction which they apply to estimate productivity dynamics in the telecommunications equipment industry. (Griliches and Mairesse provide a particularly clear description of their methodology which is drawn upon here.)

Olley and Pakes assume that the a_{it} follow a first-order Markov process. The simplest case is a random walk, which is assumed here for purposes of illustration:

$$a_{it+1} = a_{it} + \mu_{it+1}^a$$

where μ_{it+1}^a is a serially uncorrelated error term. They also assume that the investment decision I_{it} in period t, like the employment decision, is made with knowledge of a_{it}. The capital stock evolves according to:

$$K_{it+1} = (1-\delta)K_{it} + I_{it}.$$

Because there is persistence in the a_{it} (a higher value today means that the value is likely to be higher next period), optimal investment will rise with a_{it} even though it is too late to affect the stock available today. Hence the investment decision, which is observable, reveals information about a_{it}.

A further issue, which Olley and Pakes do not address, is how to examine the effect of a policy intervention P_{it}. An issue is whether it affects employment and investment decisions as well. The answer depends on the nature of the intervention, e.g., is it permanent or temporary? When did producers learn about it? The approach described below assumes that the intervention has two components. One, P_{it}^{temp}, is transitory, and like e_{it}, raises productivity too late to affect the unit's current employment decision, and since it is serially uncorrelated, has no effect on I_{it}. A second P_{it}^{temp}, analogous to the productivity shock a_{it}, follows a random walk:

$$P_{it}^{perm} = P_{it}^{perm} + \mu_{it+1}^{perm}$$

and occurs in time to affect the producer's choice of L_{it} and I_{it}. Here μ_{it}^{perm} is a serially uncorrelated error term.

Optimal investment will be a function of a_{it}, K_{it}, and P_{it}^{perm}:

$$I_{it} = \iota(a_{it}, K_{it}, P_{it}^{perm}).$$

Since the function ι is monotonically increasing in a_{it}, following the procedure in Olley and Pakes the function can be inverted to yield:

$$a_{it} = h(I_{it}, K_{it}, P_{it}^{perm}) \tag{8}$$

Substituting back into (5) gives:

$$\ln Y_{it} = f(t) + \beta^{temp} P_{it}^{temp} + \beta^{perm} P_{it}^{perm} + \gamma_L \ln L_{it} + \gamma_K \ln K_{it} + h(I_{it}, K_{it}, P_{it}^{perm}) + e_{it} + \varepsilon_{it} \tag{9}$$

Note that this formulation allows for the permanent and transitory components of intervention to have different effects, denoted β^{temp} and β^{perm} respectively. Note furthermore that the anticipated, permanent component of the policy can affect output both directly and indirectly through its effect on the producer's investment choice.

Estimation proceeds in two steps:

1. Define the function:

$$\phi(I_{it}, K_{it}, P_{it}^{perm}) = \gamma_K \ln K_{it} + h(I_{it}, K_{it}, P_{it}^{perm}).$$

The function $\phi(I_{it}, K_{it}, P_{it}^{perm})$ is then approximated by third or fourth order polynomials in I_{it}, K_{it}, and P_{it}^{perm}, denoted $\tilde{\phi}(I_{it}, K_{it}, P_{it}^{perm})$. Estimate the equation:

$$\ln Y_{it} = f(t) + \beta^{temp} P_{it}^{temp} + \gamma_L \ln L_{it} + \tilde{\phi}(I_{it}, K_{it}, P_{it}^{perm}) + e_{it} + \varepsilon_{it} \tag{10}$$

to obtain consistent estimates of γ_L, β^{temp}, and the parameters of $f(t)$, since both remaining error terms are now independent of any of the right-hand-side variables.

2. With the estimates of γ_L, β^{temp}, and the parameters of $f(t)$ in hand, one can define and construct the variable:

$$V_{it} = \ln Y_{it} - [f(t) + \beta^{temp} P_{it}^{temp} + \gamma_L \ln L_{it}] - \tilde{\phi}(I_{it-1}, K_{it-1}, P_{it-1}^{perm}).$$

The final step is to estimate the equation:

$$V_{it} = \gamma_K (K_{it} - K_{it-1}) + \beta^{perm} (P_{it}^{perm} - P_{it-1}^{perm}) + \mu_{it}^{a} + e_{it} + \varepsilon_{it},$$

which, under the assumptions, yields consistent estimates of γ_K and β^{perm}.

Another issue that arises in trying to estimate the effects of RTD policy on productivity is the exit of producers. As pointed out by Griliches and Regev (1995), exit is likely to remove firms that have both low u_{it} and K_{it}, generating negative correlation between the two variables among surviving firms. The effect again is to bias the estimate of γ_K downward. Olley and Pakes deal with the selectivity issue by estimating a probability of survival function where exit is also a function of the realised a_{it}.

Griliches and Mairesse (1998) provide a list of criticisms of the Olley-Pakes procedures, but it remains the state of the art procedure in production-function estimation. Pavcnik (2000) provides a recent application to Chilean firms, analysing the effect of trade liberalisation on productivity and exit. A study by Levinsohn and Petrin (2000) introduce a variant of the Olley-Pakes technique. They propose using a unit's purchases of intermediate inputs, rather than investment, as a proxy for the unobserved TFP term a_{it}. An advantage is that intermediate purchases, which can typically be observed, reveal information about a_{it} under much broader assumptions about its distribution over time. Levinsohn and Petrin apply their methodology to a panel data set of Chilean firms.

3.5.6. Scope and Limits

Ultimately, the success of RTD policies depends on their ability to raise productivity. Hence productivity measurement is inevitable, despite the problems that it generates. But the user must be aware of the severe limitations involved. So far the analysis has been agnostic about whether the output variable y_{it} is expressed in physical or value terms. When it is measured as value added, as is often the case, it is necessarily in value terms.

If output is measured in physical terms the production function analysis discussed above can be carried out without any assumption about the market structure in which the output is sold. However, if the unit is producing a range of products, or various qualities of the same product, no single measure can summarise output. Moreover, if the composition of outputs varies across units or over time, output measures are not comparable.

The standard response is to use value rather than physical output measures. The problem here is that now the output measure involves both quantities and prices. Under most reasonable assumptions about market structure, prices as well as physical quantities respond to TFP. In fact, in the case of perfect competition all increases in TFP are passed on to the buyer in the form of lower prices. If the production function estimation fails to correct for price changes it will fail to recognise TFP gains at the producer unit.

The point can be made simply. Rewrite the production function (2):

$$Q_{it} = A_{it} F(X_{it}), \tag{11}$$

where Q_{it} replaces Y_{it} to emphasise that it applies to physical quantity produced Q_{it}. Invoke the zero-profit condition:

$$Y_{it} = P_{it} Q_{it} = W_t X_{it}$$

where P_{it} is output price and W_t a vector of factor costs. If one tries to measure productivity as value of output over input one gets:

$$\frac{Y_{it}}{F(X_{it})} = \frac{W_t X_{it}}{F(X_{it})}. \tag{12}$$

TFP, the term A_{it}, has disappeared!

To the extent that productivity is rising across all sources of employment for the factors in question, the overall gains in productivity in the entire factor market will be incorporated in higher factor rewards W_t, and hence reflected in each unit's productivity measure. Hence macro level comparisons of productivity where the unit of observation can reasonably be assumed to span factor markets make sense. For example, a countries with high average productivity across its individual producers will see the benefits through higher W_{it}. But different producing units within the factor market with relatively higher TFP will have offsetting lower output prices.

In fact, productivity measures based on value measures of output vary enormously across producer units that can reasonably be interpreted as tapping the same factor markets. An explanation is that competition is not perfect and units are charging different markups. But in this case what the econometrician is estimating is not any technological parameter of the production unit but an indicator of its market power. If a policy intervention is seen to have made units more monopolistic rather than more efficient, the conclusion may be very different.

However, Bernard, Eaton, Jensen, and Kortum (2000) show that under reasonable assumptions about technology and market structure markups may reveal more about underlying efficiency that is initially evident. Say that a particular good i could be produced by any of a number of potential units whose TFP drawn from some distribution $H(A)$. In either a Bertrand equilibrium (in which only the lowest cost supplier actually produces the good) or in a Cournot equilibrium with several active producers, lower TFP translates on average into a higher markup. The higher a producer's TFP draw the higher price it can charge and still keep a large market share.

In the case of Bertrand competition, the producer with the highest realisation of A, denoted A_1, in order to keep its competition out of business, will set a price just below its closest rival's cost of production. If this rival has TFP A_2 the markup is A_1/A_2. Under plausible assumptions about the distribution H (for example, that it is Frechet) the markup is increasing in the realisation of A_1: units that are more efficient tend to be further ahead of their rivals, so can charge higher markups.

An observation in many micro data sets is that firms that appear to be more productive also tend to be larger. Bernard, Eaton, Jensen, and Kortum can explain this outcome as well: Units that are more efficient, while they tend to be further ahead of their rivals (so charge higher markups) at the same time tend to have rivals who are also further ahead. Hence more efficient firms charge lower prices. (In other words, their greater efficiency is shared between themselves in the form of a higher markup and their customers in the form of a lower price.) With elastic demand these units sell more, so are larger.

The overall point, however, is that microlevel productivity measurement needs to be more alert to the nature of competition in the output markets.

3.5.7. Conditions for Application

Productivity measurement is relevant whenever an RTD policy has been targeted toward increasing the efficiency of production. Measurement requires that relevant input and output variables be available both for the treated and nontreated periods or production units.

An important issue is whether increased efficiency at the aggregate level is achieved through an increase in efficiency of individual producers or through a reallocation of production from less to more efficient producers, sometimes called "rationalisation." Average productivity in a sector or the economy increases whenever less productive units exit or more productive units enter, or when workers move from less to more productive producers.

Estimates from longitudinal studies about the extent to which aggregate productivity growth results from rationalisation as opposed to increases in productivity at existing producers vary enormously. Foster, Haltiwanger, and Kazin (1998) provide a very thorough survey. Levinsohn and Petrin provide recent evidence from Chile. The evidence suggests that during economic upswings a larger share of productivity growth results from increased efficiency at existing producers, while during economic downturns productivity growth is more the consequence of the exit or shrinkage of less efficient producers.

Micro and meso level productivity measures do not need to distinguish between within producer efficiency gains and reallocation of production across units with different

efficiencies as sources of productivity growth. As discussed above, however, assessing the impact of an RTD policy on aggregate or sectoral level productivity is problematic: controlling for other factors is difficult. For this reason
Assessing the impact of RTD policies has been most successful by looking at what happens to individual producers. It is important to take into account, however, not only how policies might change productivity among individual producers, but how it also affects the reallocation of production across producers.

3.5.8. Bibliography

Arrow, K, Hollis B. Chenery, B.S. Minhas, and Robert M. Solow (1961), “Capital-Labor Substitution and Economic Efficiency,” *Review of Economics and Statistics*, 43: 225-250. This article introduced the constant-elasticity-of substitution production function and applied it to measuring productivity and factor intensity differences across countries.

Aw, Bee Yan, Sukkyun Chung, and Mark J. Roberts (1998), “Productivity and the Decision to Export,” *NBER Working Paper* No. 6558. One of several papers analysing the connection between measured productivity and export behavior at the level of individual producers, in this case for Taiwan.

Barro, Robert and J.W. Lee (1996), "International Measures of Schooling Years and Schooling Quality, *American Economic Review, Papers and Proceedings,* 86: 218-223. A paper that introduced widely-used measures of educational attainment across a broad set of countries over a number of years.

Bernard, Andrew , Jonathan Eaton, J. Bradford Jensen, and Samuel Kortum, “Plants and Productivity in International Trade” *NBER Working Paper No. 7688,* May 2000. A paper that provides a theoretical explanation for the observed links between measured productivity, size, and export status, with an application to U.S. manufacturing plants.

Bernard, Andrew and J. Bradford Jensen (1999), “Exceptional Export Performance: Cause, Effect, or Both?,” *Journal of International Economics*, 47: 1-25. A paper that demonstrates, using longitudinal U.S. plant-level data, that the observed relationship between productivity and export status emerges from a selection of more productive plants into exporting, rather than productivity improvements derived from exporting.

Bils, Mark and Peter J. Klenow (2000), “Does Schooling Cause Growth?,” *American Economic Review*, 90: 1160-1183. A paper that provides an assessment of the contribution of education toward worker productivity.

Clerides, Sofronis, Saul Lach, and James R. Tybout (1998), ``Is Learning by Exporting Important? Microdynamic Evidence from Colombia, Mexico, and Morocco,” *Quarterly Journal of Economics*, 113: 903-947. A paper that demonstrates, using longitudinal data from a set of developing countries, that the observed relationship between productivity and export status emerges from a selection of more productive plants into exporting, rather than productivity improvements derived from exporting.

Currie, Janet and Ann E. Harrison (1997) “Sharing the Costs: the Impact of Trade Reform on Capital and Labor in Morocco,” *Journal of Labor Economics*, 15. This paper is one of several that uses productivity measures to infer technological spillovers across firms.

Eaton, Jonathan and Samuel Kortum (1996), “Trade in Ideas: Patenting and Productivity in the OECD,” *Journal of International Economics*, 40: 251-278. A paper that measures the role of cross-country technology diffusion in income differences. It makes use of bilateral cross-country data on patent applications to infer the magnitude of technology flows.

Eaton, Jonathan and Samuel Kortum (1999), “International Technology Diffusion: Theory and Measurement,” *International Economic Review*, 40: 537-570.This paper develops a model of innovation and cross-country technology diffusion to identify the sources of growth in the five major research economies.

Foster, Lucia, John Haltiwanger, and C. J. Krizan (1998), “Aggregate Productivity Growth: Lessons from Microeconomic Evidence,” *NBER Working Paper* No. 6803. This paper provides an exhaustive survey of the evidence from longitudinal data on the contribution of the reallocation of production across producers to aggregate productivity growth.

Frazer, Garth (2001), “Linking Firms to Workers: Heterogeneous Labor and Returns to Education,” mimeo, Yale University. This is a very recent paper that considers the role of worker heterogeneity in measured productivity differences across firms in Africa.

Griliches, Zvi and H. Regev (1995), "Firm Productivity in Israeli Industry, 1979-1988" *Journal of Econometrics*, 65: 175-204. A study longitudinal study of productivity dynamics that introduced a method to take into account the role of firm exit.

Griliches, Zvi and Jacques Mairesse (1995), "Production Functions: The Search for Identification," *NBER Working Paper* No. 5067. A comprehensive review and critique of production function estimation techniques.

Hall, Robert E. and Charles I. Jones (1999), "Why do Some Countries Produce More Output per Worker than Others?", *Quarterly Journal of Economics*, 114: 83-116. A well-known analysis of the determinants of productivity differentials across countries.

Jorgenson, Dale (1995), *Productivity*, Cambridge: MIT Press. A comprehensive tome on productivity analysis that provides a detailed explanation of the author's lifetime contribution to the topic, in particular the exploitation of the translog production function.

Levinsohn, James and Amil Petrin (1999), "When Industries Become More Productive, Do Firms?" *NBER Working paper* No, 6893. The authors decompose productivity growth in Chile between productivity gains within producers and rationalisation of production among producers.

Levinsohn, James and Amil Petrin (2000), "Estimating Production Functions Using Inputs to Control for Unobservables" *NBER Working paper* No, 7819. The authors develop a variant of the Olley-Pakes methodology for production function estimation using intermediate purchases. They apply the methodology to Chilean manufacturing firms.

Mankiw, N. Gregory, David Romer, and David N. Weil (1992), "A Contribution to the Empirics of Economic Growth," *Quarterly Journal of Economics*, 107: 407-437. An influential paper interpreting the growth of a broad range of countries in terms of the Solow growth model.

Marschak, Jacob and W. Andrews (1944), "Random Simultaneous Equations and the Theory of Production," *Econometrica*, 12: 142-205. A paper that pointed out how standard estimation of production functions were open to sever measurement bias and underestimation of the contribution of TFP gains.

OECD (2001), *Measuring Productivity: OECD Manual*, Paris. This manual provides a detailed overview of applied productivity measures, implementation methods, empirical applications and interpretation of evidence.

Olley, Steven G. and Ariel Pakes (1996), "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, 64: 1263-1297. This paper introduced a new method of production function estimation that attempts to overcome the Marschak-Andrews critique, applying it to longitudinal data in the U.S. telecommunications.

Parente, Stephen L. and Edward C. Prescott (1994), "Barriers to Technology Adoption and Development," *Journal of Political Economy*, 102: 298-321. A paper that provides an explanation for income differences in terms of barriers to technology diffusion.

Pavnik, Nina (2000), "Trade Liberalisation, Exit, and Productivitiy Improvements: Evidence from Chilean Plants," *NBER Working Paper No. 7852.* This paper develops a methodoloy for assessing the contribution of trade liberalisation to productivity improvements in Chile.

Roberts, Mark and James R. Tybout (1997), "The Decision to Export in Colombia: An Empirical Model of Entry with Sunk Costs," *American Economic Review*, 87: 545-564. A paper that examines the dynamics of productivity and exporting behavior using longitudinal data from Colombia.

Solow, Robert M. (1957), "Technical Change and the Aggregate Production Function," *Review of Economics and Statistics*, 39: 312-320. The classic analysis of aggregate productivity growth.

Summers, Robert and Alan Heston (1991), "The Penn World Tables (Mark 5): An Expanded Set of International Comparisons," *Quarterly Journal of Economics*, 106: 327-368. This article presents the now widely-used data on cross-country trends in productivity.

Tybout, James R. (2001), "Plant and Firm Level Evidence on 'New' Trade Theories," *NBER Working Paper No. 8418*. This paper surveys micro analysis on export behavior and productivity.

Van Ark, Bart (1996), "Productivity and Competitiveness in Manufacturing: A Comparison of Europe, Japan, and the United States," in K. Wagner and B. van Ark (eds) *International Productivity Differences, Measurements and Explanations: Contributions to Economics Analysis, Volume 233.* Amsterdam: North Holland. This paper presents estimates of sectoral productivity differences across three major economies.

3.6. CONTROL GROUP APPROACHES

Author: **Eamonn Kinsella** (The Circa Group)

3.6.1. Introduction

The evaluation of an instrument may cover *inter alia* the efficiency and effectiveness of the instrument's management and operation; attainment of objectives; quality of the projects supported: amount of deadweight: profile of the participants supported: attention to recommendations contained in evaluations of preceding instruments: adequacy of the budget: benefits and costs to the participants and impacts on them and on the economy at large.

Arguably, the last is the most important. No methodology can exist which can answer this question, as an organisation's growth and success always depend on many unobservable influences in addition to public support, such as changes in management, the business environment, improved marketing or changes in the political scene. This approach attempts to exclude the effects of variables external to the sphere of influence of the instrument being evaluated.

It is not possible to prescribe a procedure, using this approach or any other, which can be rigorously followed in evaluating a programme or instrument. Pragmatic considerations must always prevail and restrictions such as the evaluation budget and data quality must be observed in selecting the level of sophistication of the analysis. This is particularly true for a small population.

The use of control groups in programme and instrument evaluation is attractive. To be rigorous, it requires statistically significant sample sizes, stratified exactly the same as the test population. It may be that such samples do not exist for a particular application, either in terms of numbers or exact comparability of objectives. The use of control groups in this context must not be regarded as having the accuracy or precision of laboratory experiments nor of clinical trials in the pharmaceutical and healthcare industries.

A framework for an evaluation typically involves indicators for the following:

- The objectives of the activity and targets set
- The resources available for the activity
- The methods used to implement the activity
- The outputs and achievements of the activity
- The impacts of the activity

Only projects which have been completed, or nearly so, may be considered. In every case adequate data on input, output and impact indicators is required. It is also assumed that the instrument's objectives and targets are clearly stated. This section does not deal in detail with survey techniques for beneficiaries, or questionnaire design, which should be as simple, clear and amenable to processing as possible.

Three groups of actors may be defined:

- **Group 1** consists of beneficiaries of the instrument
- **Group 2** consists of those who applied unsuccessfully for funding, but who completed the project, using their own funds or other support instruments such as venture capital or a loan. Ideally, the only difference between Group 1 and Group 2 is that Group 2 did not receive funding under the instrument being evaluated, but on

average Group 1 are likely to be more experienced in submitting proposals and in carrying out R&D. The original Group 2 project may have been slightly modified or executed slower, and in fact the different funding mechanism may have changed aims and objectives as well.
- **Group 3** consists of those who did not seek funding under the instrument, but executed a comparable project using their own funds or other means of support. They will clearly be most useful in gaining insights into the management and promotion of the instrument, and its relevance to the needs of the target clientele.

Groups 2 and 3 are control groups. There are few literature references to the use of Group 3 data.
These will also give useful information on the management of the instrument: publicising the scheme, bureaucracy, commercial sensitivity, user friendliness, speed of response and easier or preferred ways of getting funding. Convergence of opinions and views normally happens quite quickly.

There are four principal ways of carrying out an evaluation:

- Evaluation at an aggregated level of the instrument against external criteria, typically financial, laid down by a national Ministry of Finance. This is a Value-for-Money review, or Cost/Benefit Analysis. Indicators are limited to the external criteria. No control groups are needed.
- Benchmarking the instrument against other public instruments designed to achieve the same results in the same sector. The instrument and its benchmark may be in different countries, but the contexts must be sufficiently similar to allow conclusions from the benchmark to be applied to the instrument being evaluated. No control groups are needed.
- Comparing the performance of an instrument with its own internal aims, objectives, and targets. No control groups are needed.

The control group approach, comparing the performance of the instrument in supporting Group1 beneficiaries with the performance of samples taken from the control Groups 2 and 3 above. The comparison is based on rigorous analysis of the same indicator values and measures for all three groups. Some qualitative issues, such as the reason for non-participation in the case of Group 3 firms, obviously vary from group to group.
In all four cases the data for input, output and impact indicators are assumed to be available for Group 1. The control group approach is of primary interest here. In the case of public support for RTDI, which is the focus of EPUB, the control group approach is the only one which can evaluate a public instrument against a free market support mechanism.
The previous chapters described quantitative non-experimental approaches to the micro-economic evaluation of RTD policies, that is evaluations starting at the participant level. These models use control variables and sophisticated statistical methodology to answer the difficulty of constructing true control groups. Some of the ideas, which are described in this section in a more pragmatic and intuitive manner, are formulated algebraically. Both approaches require similar input data at participant level, and both face many of the same dilemmas in dealing with variables external to the instrument, such as long-term spillover within an industry (which is an important justification for publicly-funded R&D), which is related to the elapsed time between the instrument's use and the evaluation, or changes in companies' business strategies.
There are however some clear differences in the two approaches:

The quantitative models described before do not allow the legal feasibility of experimental groups in the EU context: the approach described here probably identifies them as Group 3, a "voluntary experimental group".

- The statistical models specify control groups by probability of participation in a programme, whereas the approach described here regards these indicators as the profile of the participants and conformity with eligibility criteria.
- This section specifies control groups by their profile, not their eligibility.
- The indicators of interest in the approach described in this section are the outputs and impacts on the target group having the same, as far as possible, profile as the control groups.
- The differences between the outputs of, and impact on, Groups 1 and 2 are used to evaluate the effect of the instrument, where as the corresponding differences between Groups 1 and 3 reflect principally the effectiveness of the operation and management of the instrument.

The algebraic approaches described in the previous chapters could be applied to the more pragmatic approach described in this section with advantage where precise and accurate data are available. Great care must be exercised in sifting these data, and even more in interpreting the results.

3.6.2. Application to Policy Instruments and Interventions

EPUB is concerned with the evaluation of publicly funded RTD, both at the national level and at the EU level. Apart from the size of budgets, aims and objectives, closeness to market of projects funded and other features, there are significant structural differences between national and EU instruments:
Most of the EU instruments of relevance here require a consortium of at least two organisations in separate states, led by the private sector partner. Consortia of six or even ten partners are common.

- National instruments are often available to single applicants.
- Specific programmes of the Framework Programme are normally over-subscribed by a factor of three or even four. Funding is highly competitive.
- The funding success rate for national programmes may be much higher, and they are less competitive.

These differences give rise to a number of important considerations in evaluating national as opposed to EU instruments:
Difficult as it may be to find a match for a firm, it is much more difficult to match a large consortium. The control group approach will work far better for evaluating national as opposed to EU instruments.
Given the over-subscription to EU instruments, the likelihood of funding being awarded is related to the characteristics of the project rather than of the applicants, though of course there will be a bias towards more experienced R&D performers. Some of the descriptors of the applicants may be criteria for eligibility, e.g. size for SME programmes such as CRAFT, rather than indicators of likelihood of participation.
This control group approach will not easily work for EU or national instruments which act for public good, such as infrastructural support, or which address a market failure, as there are unlikely to be any control groups. Any of the other three approaches may be used.

Nor will it be feasible to use the control group approach to evaluate wide-ranging instruments such as the Framework Programme, or the ERA. The drawback is not in the scale of the instruments, but in the fact that they cover such a wide range of activity that no one control group will suffice. It is however possible to use this approach to evaluate specific programmes of the Framework Programme, for which distinct control groups can be found. In practice, many of the specific programmes of the Framework Programme each cover such a wide range of topics and productive sub-sectors as to make dis-aggregation necessary. Control groups can then be found for each sub-sector, or smallest homogenous activity. The control group approach cannot be used for a large number of projects unless they are very homogenous.

The control approach would be meaningless in the context of a programme for a large infrastructural development, such as a large item of scientific equipment. This is because there are very unlikely to be any controls.

It follows that this approach will be suited to evaluating specific national or EU instruments primarily directed at private sector economic activity, for which control groups can be found. These instruments include:

- Grants for RTDI in existing companies
- Fiscal measures, such as tax concessions
- Venture capital for high potential start up firms
- Loans for established firms
- Support for employee education and training

There may be differences between government research support grants and government research contracts, especially if the latter may lead to production contracts. Equally, government or EU support measures for industry which are non-competitive will affect industry differently, and cognisance must be taken of these variations in selecting control groups, as the participants are self-selecting.

3.6.3. Good Practice Examples

Detailed evaluations of specific programmes of the Framework Programmes FP4 and FP5 have been carried out. However, they were generally carried out against internal criteria, and did not involve the use of control groups. These evaluations indicated how well the specific programmes were carried out, in relation to their own stated objectives, but not in comparison to alternatives or in terms of value for money. For reasons of cost and practicality, the control group approach is not often used for large programmes, as mentioned above. Large single projects are generally of an infrastructural nature, carried out for the public good. There are few, if any, private sector counterparts which could be used as controls.

An evaluation of a Japanese government and an American government sponsored initiatives to promote collaborative industrial research has been described by Sakikabira (OECD, 1997).

This study compared the R&D performance of firms which did not participate in government sponsored collaborative R&D, did so occasionally, or frequently. A questionnaire was sent to 226 firms in the US and Japan, asking the change in scale and pace of projects that would have been carried out without the government sponsored collaboration and the likely BERD. The outputs used as indicators were research intensity (R&D/Sales), numbers of patent and sales. There is no discussion on the selection or definition of the controls. Matched pairs were not used and the statistical analyses were carried out at a macro level.

The control group approach has been used recently in evaluating national RTDI instruments in Ireland. Examples of this approach were the evaluations of Ireland's participation in the European Space Agency, and of the Research, Technology and Innovation initiative, RTI, which is nationally funded. Both instruments support RTDI in industry by contributions to the ESA and through grants to firms respectively. There were relatively few firms involved in the ESA programmes (20) but a much larger number in the RTI initiative (414). Neither evaluation has been published.

The evaluation of the ESA participation was effectively at an industry level, as all of the 20 firms involved were active in space sector work. The evaluation did not employ Group 2 controls, as such do not exist. Three Group 3 controls, all of which had received national funding, were used. This sample size was determined by the evaluation budget, and by the experience that it is more informative to conduct a small number of in-depth interviews face to face when seeking qualitative information, such as non-quantifiable benefits or reasons for not participating. Data of a superior quality is obtained this way. The same quantitative indicators were used as for the participants, who also were interviewed personally. However, this rationale suffers from the drawback that it is not meaningful to carry out statistical analysis on a small sample, especially if it is stratified.

In addition to the control group approach, an attempt was made to evaluate ESA participation by benchmarking against five other national RTDI instruments and the Framework Programme, all of which had previously been evaluated. However the objectives of the benchmarks and of ESA were found to be so different as to make this impractical.

The RTI initiative also was evaluated using a control group approach. As RTI funding was, and is, available to firms in very many sectors of industry this evaluation related more to the firm or sub-sector level. Sixty-five firms were selected at random from the total of 414 involved, comprising 55 participants, 10 Group 2 controls and 4 Group 3 controls selected as being companies that could have benefited from RTI funding. Personal interviews were conducted with all of them, using questionnaires specifically designed for each group. The sizes of the three samples were dictated by the evaluation budget, which had to support other activities as well as firm interviews.

3.6.4. Data Requirements

The Terms of Reference for the evaluation, the budget and the time available all determine the data, which must and can be collected for an evaluation. Further limits are imposed by the timing of the evaluation in relation to the maturity of the instrument or programme: there will be few outputs and fewer impacts early in the life of the programme, while after a significant time lapse, records of indicators may be missing or defective, particularly where there have been staff changes. In addition, it may be difficult to separate impacts due to the programme from those arising simultaneously from other unobservable influences and spill-over from the same or other sectors of the economy. The control group approach using a Group 2 sample goes some way to measure additionality and overcome this problem.

Using Group 3 controls also gives information about aspects of the management and operation of the programme, which may have a bearing on why no application was made. Reasons commonly given are: excessive bureaucracy; lack of confidentiality; lack of awareness of the programme; lack of flexibility in changing project orientation; failure of the programme to address bottlenecks, such as shortage of skilled labour or

space; alternative, easier ways of securing funding or cash flow and possible requests by the funding agency for equity in the company.
Attention tends to concentrate on indicators for which there are quantifiable measures. However, many indicators, particularly related to impacts, are non-quantifiable but nonetheless important. These include enhancement of reputation among peers, building RTI capability and networks, raising management levels by exposure to international best practice, increasing international credibility, and employee career prospect improvement. Some of the indicators available include:

- *Objectives and targets:* numbers of first time R&D performers; increased commitment to continuing R&D performance; encouraging the formation of consortia to carry out large scale projects; encouraging international linkages and staff mobility; increased innovation in products and processes
- *Input indicators*: available budget for disbursement; budget for overhead charges; staff time allocation
- *Management and operation:* effectiveness of promotion and guidance literature; ease of the application process; speed of the selection process and decision making; promptness in dealing with contract negotiation, administration and payments; amount of deadweight; confidentiality in relation to applicant's sensitive information, especially in knowledge based industry; measures to publicise and exploit results; budget drawdown; overtime worked; overheads as a fraction of total budget; improvements in the programme over its predecessors, if any; adequacy of the budget
- *Output indicators*: project completion rate; refereed papers, invited papers at conferences, books; theses, patents, degrees, other qualifications and prizes awarded; licenses awarded or obtained; other technology transfer; increased or maintained employment; return on investment and time required to realise the return
- *Impact indicators*: changes in level of employee education; establishment or growth of a research facility; increases in sales, profitability, entry into new markets; enhancement of reputation and credibility; improved management, especially innovation management, capability; improved research capability; socio-economic impacts at national or EU level, such as contribution to the intellectual capital and foreign direct investment, quality of life; knock-on effects on non-participants

3.6.5. Steps for Implementation

The control group approach is not in principle particularly complex. The target group, Group 1, is or should be, contractually obliged to co-operate. Members of Group 2 can be identified by the funding agency and they may be prepared to co-operate subject to confidentiality assurances. However, their views may contain bias against the funding agency, and in addition there may be other sources of bias as discussed later. Members of Group 3 can be identified in various ways, such as by discussion with Research Associations, professional bodies, deans of research in universities, or Chambers of Commerce. Some Group3 members could be selected from those who participated in earlier rounds of the funding programme, who are known to the funding agency, but who have participated in this round.
The extent to which stratified samples are possible or realistic clearly depends on the degree of stratification, and on the size of the universe. In the case of EU programmes, stratification may be usefully done by:

Type of organisation (3, Industry, University, Research Institute)

Geographic location (15, State or region)

- Sector of the economy (6, Broadly defined)
- Number of employees (5, Size range)
- Ownership (2, EU or non-EU)
- R&D history (2, New entrant or established performer)

This gives a total of 5,400 cells, such that a meaningful statistical analysis would require at least 100,000 participants. This number is clearly impractical and therefore the entire Group 1 sample should be used. In practice, most EU instruments are available only to networks or consortia of one sort or another, each having as many as ten partners. The consortium must therefore be described in some unique global manner summed over all partners, or the description confined to the lead partner. Strictly speaking, the instrument should be evaluated from the perspective of each partner, but common sense must prevail, and global impacts could be recorded. National instruments demand smaller, if any, consortia, making for easier evaluations.

In the case of EU instruments, Group 2 will be three or four times bigger than Group 1, but even that size will not be enough, and the entire Group 2 should be surveyed. Furthermore, while the Group 1 responses will be complete, those of Group 2 are very unlikely to be so. A final response rate from Group 2 equal to that of Group 1 would be the most to be expected, but it is unlikely that Group 3 will be as big as Group 1 or Group 2. However, the contribution desired from Group 3 will be heavily oriented towards reasons for non-participation, and these converge fairly rapidly.

The evaluation budget will determine how much effort can be expended in seeking Group 2 responses. Costs are also influenced by the data collection method selected; for example postal surveys, face to face interviews, or focus group discussions, as there are different associated costs. Most of the above information will be on file for Group 1 and Group 2.

This approach using all of Group 1 and a self-selecting sample of Group 2 (assuming the Group 2 response is incomplete) will not allow a deliberate stratification of the respondents. It will provide a statistical comparison.

It is possible to select applicant consortia from Groups 1 and 2, by examination of the file data and the questionnaire returns, which can be matched in pairs, giving a one on one comparison. An algebraic approach to minimising the balancing score between the members of the pair is described in Section XXX: it could be advantageous to weight the consortia indicators in order of their importance to the particular evaluation. The matched pairs would be closer to having a true control group, the only difference being that the Groups 2 and 3 members did not receive funding, but they will be few in number and of lesser statistical significance.

Nonetheless, deeper insights could be obtained by carrying out a limited number of case studies on closely matched pairs of Group 1 and Group 2, and Group 1 and Group 3 respondents. These could very usefully be done by face to face structured interviews. Face to face or telephone interviews with the control groups may give deeper insights into alternative support mechanisms and reasons for non-participation. Interviewees are often prepared to express views which they are not prepared to put in writing, and furthermore a particular point may be easier explored verbally.

The number of interviews will be limited by the evaluation budget.

The same indicators for input, output and impact should be used for all three groups, but not so for objectives and funding methods. These have been discussed in more detail in

Section V above. The techniques in questionnaire design and data collection for all three groups are essentially the same, but vary with the method of data collection, whether postal survey, face to face or telephone interview. In every case, it is essential to keep questionnaires as short and simple as possible.

A questionnaire for the Group 2 and Group 3 organisations covering two sides of an A4 sheet should give enough data for most surveys. Questions should be designed to be answered by ticking boxes or scoring a numeric answer. This serves to reduce the workload on the interviewee, and also makes it possible for junior staff to input results into a database and carry out unambiguous statistical analyses. Such a compact questionnaire may be sent to the interviewee in advance of an interview without the risk of causing rejection.

The control group approach is useful in examining a number of facets of the effectiveness of an instrument:

Additionality, results that would not have been realised without the instrument.

- Impacts, such as cluster or network formation, which is encouraged by collaborative research, especially EU funded, more commonly than non-publicly funded R&D.
- The effectiveness of the promotion and administration of the instrument and its relevance to the needs of the research community.
- Deadweight, the volume of work supported which did not contribute to the objectives of the instrument.

It is very difficult to determine the effects of investment in RTDI on economic development, as the model is so complex and susceptible to interference from other changing environmental factors. Given the difficulty, if not the impossibility, of establishing clinical control groups, this approach will not solve the problem, and results must be interpreted very cautiously.

3.6.6. Scope and Limits

This approach is the best suited to estimating additionality in public funding awarded to redress market failures at national or EU level. The results and the level of confidence in them will depend on the completeness and reliability of the data collected for each indicator, which in turn will depend on the evaluation budget. The sophistication of the statistical analysis carried out should be related to the size and quality of these data sets, and to the needs of decision-makers. All evaluation techniques must be accurate, but precision may vary.

Caution must be exercised when selecting Group 2 members in the situation where the support instrument is over-subscribed, as is normally the case with the specific programmes of the Framework Programme. It is unusual to have even the top quartile of applications funded. Many projects, especially in the second quartile, not funded, proceed with alternative funding supports, both because of the effort involved in preparing a project and because they are good projects, even if not the best. There may therefore be inherent bias, giving a non-representative sample. Reasons for proceeding must be carefully determined in the case of Group 2 controls.

The fact that some rejected proposals proceeded contributes indirectly to the success of the instrument, though it would be difficult to quantify this.

A paradox arises in the case of selecting Group 3 firms. Any spill over from the programme, which is desirable, will act to narrow the gap between Group 1 and Group 3, thereby reducing the apparent effectiveness of the programme. On the other hand, a

bigger gap in performance between Group 1 and Group 3 indicates greater effectiveness of the programme, but less spill over. Spill over effects may be reduced by picking Group 3 firms that are technically close to Group 1: Group 2 are close anyway. This problem is addressed in detail in the paper by Klette et al (Research Policy, 2000).
Data collection at the firm level can be difficult, and expensive, even when members of Group 1 are contractually obliged to co-operate. It can be much more difficult to obtain data for Group 2 and Group 3. Group 2 are known to the funding agency. Group 3 will not be known and it may be hard to find them. There are also problems of lack of motivation in co-operation, and perhaps antagonism to the funding agency, as well as "questionnaire fatigue". Confidentiality also raises problems for Group 2 and Group 3.

3.6.7. Conditions for Application

A true control will be identical to the test in every respect but one: the only difference allowed between Group 1 and Group 2 is that Group 2 were refused funding. Because changing the funding mechanism as between Group 1 and Group 2 will almost inevitably change some procedures, aims, objectives and speed of execution, it will make it very difficult to find a true control, especially in a small population.
Because it would be necessary for a rigorous evaluation to find one on one sets, the cost will be essentially double that of an uncontrolled evaluation.
It is therefore necessary to be pragmatic and to accept some position less than the ideal one on one, and to interpret results in the light of judgement and experience.
Projects executed by Group 3 may have aims very different from those of Group 1 or the programme itself. Surveying them may yield no more than usable information on perceptions of the management and operation of the programme.
The control group approach is one approach used in judging additionality. It may be combined with other approaches, especially as the data collected may be used in other evaluations as well, such as value for money.

3.6.8. Bibliography

The literature on evaluation methodologies and problems is very extensive, but the amount of information available on the definition and use of control groups is surprisingly sparse. Two publications are relatively useful sources of further information on the use of control groups:

A Norwegian programme to support the IT industry was evaluated using control groups, and found no significant difference between those funded and not. The programme was judged a failure.

Aravanatis, S: The Swiss AMT programme: p325 et seq. This review of the effectiveness of the Swiss AMT programme for industry did use control groups, which gave a higher response rate than the beneficiaries. A total sample of 667 firms, a representative sample, was surveyed, but the proportion of controls was not stated, nor their profile or exactly how they were used.

Capron, H. and B. Van Pottelsberghe de la Potterie: Policy implications in seven countries of R&D subsidies: p171. The paper concludes that publicly funded R&D has a lesser direct effect on productivity than privately funded R&D, but there are scant details of the definition or selection of control groups.

David, P.A., Hall, B.H., and Toole, A.W: Is public R&D a complement or a substitute for private R&D? pp *497-529.* This paper refers to one particular survey covering data on firms that received an award, those that applied but failed, and those that did not apply.

Fayle, G. and Durieux, L: The scheme used to evaluate EU programmes: p373 et seq. This combines continuing annual monitoring and a five-year review of each specific programme, but interestingly, no control groups are mentioned.

Georghiou, L. and D. Roessner: Evaluating technology programmes: Tools and methods: pp 657-678. This article reviews the analytical tools, methods and designs being used to evaluate public programmes intended to stimulate technical advance. The use of control groups in a number of evaluations of national programmes is discussed. Some of the problems inherent in selecting control groups are addressed.

Hall, B.H., and Van Reenen, J: How effective are fiscal incentives for R&D? pp*449-469.* A brief mention is made of control groups, and of the problems in selection. An Australian evaluation is described, as are other evaluation methods which are alternatives to using control groups.

Klette, T.J., Moen, J. and Z. Grilliches: Do subsidies to commercial R&D reduce market failures?: pp471-495. This paper deals with microeconometric studies and in detail with the problems of selecting control firms, as they do not constitute random samples and may be affected by spill over and other systematic biases. The evaluation of the SEMATECH programme in the US, which was a research consortium established for the semiconductor industry is discussed in detail. An evaluation of another US programme, the Small Business Innovation Research, also used control groups and is described.

Luukkonen, T: Evaluation of mission-oriented research in Finland: p347 et seq. The evaluations have been moving away from peer review to professional evaluations. However, better control groups are needed, but there is no detailed discussion.

OECD Conference: In 1997 the OECD organised a conference on “Policy evaluation in Innovation and Technology”. This was organised in the context of the OECD work on “Towards best practices in Innovation and Technology Policy”. While a small number of papers mentioned the use of control groups, there was little or no detail of

the definition or selection procedure for them. Papers which referred to control groups were:

Papaconstantinou, G. and Polt, W: Conference overview: p9 et seq. Control groups were beginning to be used, but with difficulty. Several approaches are needed to any evaluation, both qualitative and econometric, which latter often gives qualitative answers. Evaluations should not go too far, as the final decision is always political.

Research Policy, North Holland, Amsterdam: A recent special issue of *Research Policy (Vol. 29, 2000)* was devoted to evaluation methodologies. This is one of the most up to date reviews of the subject. It includes four papers of relevance to the use of control groups:

Sakakibara: The effectiveness of government sponsored industrial collaborative research in Japan and in the United States respectively: p225 et seq.

See also:

Airaghi et al *(1999):* Options and limits for assessing the socio-economic impact of European RTD programmes, *ETAN Expert working group:* EUR 18884, 1999

IPTS Report (December, 1999) Special Issue: Evaluation of R&D activities, European Commission, Sevilla. The difficulty of measuring additionality is discussed. A possible tool is to construct a model of what industry would have done in the absence of the programme being evaluated, but this is judged to be almost impossible. Control groups per se are not discussed.

Salter, A., and Martin, B. (2001): The economic benefits of publicly funded R&D: *Research Policy, Vol. 30, 509-532*

3.7. COST BENEFIT ANALYSIS

Authors: **Wolfgang Polt** and **Birgit Woitech** (Joanneum Research)

3.7.1. Introduction

Especially since governments re-assess their roles, try to cut their expenses, and have an eye on the leverage effects of their policy instruments, the role of CBA in evaluating all kinds of future and past investment projects increased. As an analytical tool cost-benefit analysis (CBA) has been widely used in various contexts to estimate the economic costs and benefits of investment projects, e.g. in health and safety issues, to assess the external effects of energy projects and for the valuation of environmental projects and policies. Contrary to a private investment decision problem, the (social) CBA considers internal as well as external costs and benefits.

The specific feature of CBA rests on its effort to appraise all economic and social consequences which are entailed by a project (see for an overview of these effects Table 12). It is therefore a tool for determining whether a project or a program is economically efficient. This is in fact the major rationale to carry out CBAs for public policy programmes: In standard economic theory it provides the ultimate justification for government intervention if it can be shown that social returns are high compared to private ones and if the costs of intervention and investments are outweighed by the benefits. It even carries the promise to be able to choose between different projects, programmes or policies based on their respective relation of costs to benefits.
As illustration of the possible components of CBA from different perspectives is provided in the following table.

Table 13: Types of Private and Social Costs and Benefits

	Individual Partners	Programme Sponsor (Government)	Society
Benefits	In Net Earning Profit	Tax Revenues	Increase in National Income (direct effect)
	Additional Benefits from Transfers	Decrease in other Subsidies	Spill-over Effects (indirect effect)
	Non-Economic Benefits		
Costs	Opportunity Cost of Participation	Tax Costs	Opportunity Costs (Cross Earnings from Other Potential Programmes)
	Direct Participation Cost	Project & Administration Costs	Programme Costs
	Loss of Subsidies from Other Programmes		

CBA does not consider the distribution of the cost and benefits. It seeks to aggregate individual preferences by taking the marketplace as the prime context in which those preferences are expressed. CBA thus is a stringent framework which demands that all (or at least most) costs and benefits have to be expressed in monetary terms. As every project or programme comprises a various number of costs and benefit, it is useful to first divide them in two main categories: *direct and indirect effects*. Direct effects are those which are directly produced by the project and are related to the objectives of the project. Indirect effects are benefits and costs associated with responses to the project in other markets that are not fully captured by the willingness to pay and opportunity costs of the direct benefits/costs of the project. Both categories may include *tangible and/or intangible components*. By definition, tangibles are quantifiable and can be appraised in monetary units, while for intangibles market prices cannot be applied. Very often such effects are spillovers and/or external effects (Mishan, 1994, p. 203) which are difficult to measure and hence quite challenging for evaluators. Examples for intangibles are a higher standard of living, health, safety and quality of life gains, noise pollution etc.. Several valuation techniques (see shadow prices) have been developed, many of them remains very controversial.

There are now several ways to "monetize" costs and benefits:[35]

First, benefits can be measured directly; this case requires discounting.

Second, if sufficiently perfect markets exist market prices can be used to determine the figures in a CBA. In the absence of a market failure, the price of an input or output reflects then its opportunity costs. Whenever markets are incomplete, market prices do not reflect the true marginal cost or benefits. Therefore *shadow prices* have to be calculated. Typically the measurement of intangibles is essentially a shadow pricing issue. The term has long been used in mathematical programming and can be defined as the increase in social welfare resulting from any marginal change in output or input factors. The three main concepts for the calculation of shadow prices are the application of the Lagrange multiplier, the use of the Ramsey rule and the application of producer prices as an approximation[36].

Furthermore, also in the presence of external effects there exist considerable problems to appraise them in monetary units since they are often intangibles (Fritz et al., 2001). Externalities exists when the activities or the outcomes of one agent (or project participant) positively or negatively affects someone else (consumer, firm, industry etc.). Even where markets exist it cannot be automatically assumed that an externality will be internalised (e. g. Coase Theorem).

Third, in case there is high quality-data, like panels and time series, econometric estimations could be performed.

Fourth, where no markets exist, also concepts like "willingness to pay" and "willingness to accept" can be introduced to measure benefits. Benefits are based on the consumer’s willingness to pay for the project. Costs are then defined as the amount losers are willing to accept as a compensation.

Due to the fact that costs and benefits usually occur at different point in time, each "monetized" component of CBA has to be discounted, otherwise different projects cannot be compared adequately. The discount rate chosen generally reflects several factors like inflation adjustment, opportunity costs, use of capital, etc.,[37] although ideally it should represent the society’s time preference. It will be the smaller the further into the future the benefit is received. Since the choice of the discount rate is a major difficulty, sensitivity analysis is often conducted. Thereby a range of discount rates is chosen to see how the results are effected by altering the rate.

As the general model points, the aim of CBA is to maximise the difference between benefits (B) and costs (C), which is equivalent to the efficiency of a project. Three measures are frequently calculated:

[35] see Polt, W. (1998) and Fritz, O. et al. (2001).

[36] Brent R. J. (1996), pg. 79-106 gives a detailed explanation of the concepts

[37] see Polt, W. (1998)

10. The net present value (NPV) of the investment is equal to the difference between discounted benefits and the discounted costs. If the NPV is positive, the investments can be regarded profitable and therefore justified.

$$NPV = \sum_{t=0}^{n} \frac{B_t - C_t}{(1+r)^t}$$

B_t = benefits occurring in year t,
C_t = cost occurring in year t,
r = discount rate and
n = duration of the project (in years)

11. The internal rate of return (IRR) is derived by setting the NPV zero and calculating the discount rate r. The IRR can be interpreted as some sort of a "ROI-benchmark", which means that an investment project is justified if the IRR exceeds the expected real interest rate.[38] Thus,

$$IRR = \sum_{t=0}^{n} \frac{B_t - C_t}{(1+r)^t} = 0$$

B_t = benefits occurring in year t,
Ct = cost occurring in year t,
r = discount rate and
n = duration of the project (in years)

12. Additionally, the Benefit-Cost-Ratio (BCR) can be used by expressing benefits and costs as a ratio to provide the relative efficiency of a project:

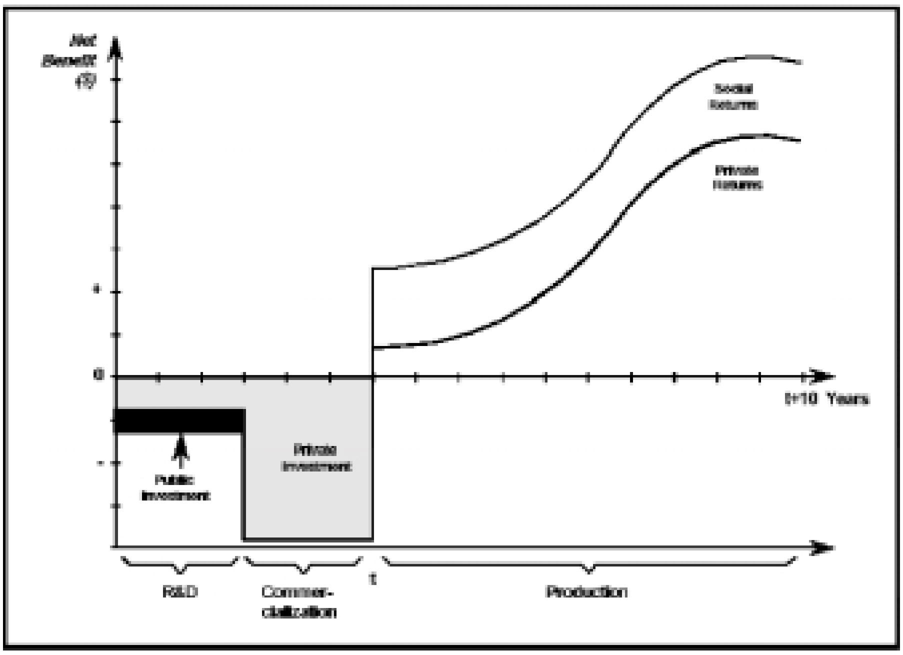
$$\frac{B}{C} = \frac{\sum_{t=0}^{n} \frac{B_t}{(1+r)^t}}{\sum_{t=0}^{n} \frac{C_t}{(1+r)^t}}$$

B_t = benefits occurring in year t,
C_t = cost occurring in year t,
r = discount rate and
n = duration of the project (in years)

The construction of an appropriate time horizon, i.e. setting the assumptions about the timing of the costs and benefits with regard to a special research project, is a crucial question for discounting. Costs and benefits of a technology project being successfully introduced onto the market are unevenly spread over time. Investments in for example new technologies often do not result in benefits to society and the private company for a number of years. We can therefore distinguish three different phases which usually generate typical cost and benefit patterns: (a) the *R&D phase, (b) the commercialisation phase, and (c) the production phase*. Thus, modelling and evaluating the social and private returns of R&D projects may require the calculation of costs and benefits for every stage. The connection between them and the net benefit are presented beneath in the figure below.

[38] see Smith, K. (1997), pg. 267

Figure 5: The Timing of Costs and Benefits in the different stages of an RTD Project



Source: Martin, Sheila A. et al. (1998), pg. 2-3

In the *R&D phase* companies pursuing a research project extend their R&D efforts to increase the probability of technical success (Sheila, Al et al., pg 2-3 ff). Obviously, no revenue is generated during this phase. In the following *commercialisation phase* the firm invests in sales, marketing, and manufacturing infrastructure. These activities should bring the results of the R&D to the market. The distinction between the R&D phase and the commercialisation phase is that uncertainty relates to technical success in the former, while it relates to market success in the latter. The *production phase* comprises all activities involved in producing the product or service that embodies the new developed technology. At this stage the company starts to accrue revenues, thus, both the private and social returns may become positive. This phase continues until the company closes the production. Determining the length of this production phase is difficult because it requires to forecast the emergence of a new product that replaces the technology in question. This is just one of the issues of considerable uncertainty in using cost-benefit approaches.

Uncertainty is one of the most technically difficult parts of CBA. The most common mistake in trying to deal with the uncertainty of costs and benefits is to argue that one should use a higher discount rate. This might lead to absurd results, since a higher discount rate would reduce the present value of the costs today, and a project, although highly uncertain, looks more and not less attractive (Stiglitz J. 1999, p. 289). In the literature there is a distinction is made between risk and uncertainty: Risk is there defined as measurable uncertainty. This means that under risk one knows the probabilities, while under uncertainty these probabilities are completely unknown. For simplifying the analysis situations of uncertainty were converted into risky situations by assessing probabilities either objectively or subjectively. This type of approach is called “risk analysis”. Techniques for deciding among uncertain outcomes are the concepts of expected value or expected utility. To evaluate the risk economists have introduced the concepts of certainty equivalents (Brent, R. J. 1996, Stiglitz, J. 1999).

3.7.2. Application to Policy Instruments and Interventions

CBA has recently been used especially to evaluate large scale projects like aeronautics and space projects, energy projects and projects with environmental impacts. In the field of the evaluation of RTD programmes, there have been but a few applications of this approach – mainly for some practical and conceptual reasons. Some types of RTD programmes and projects are more amenable to CBA than others. In the US, CBA has been applied to the evaluation of projects funded by technology programmes (like the ATP) or institutions (like the NIST) in several studies.

3.7.3. Good Practice Examples

Though not standard practice, there have been promising/interesting examples of CBAs in the context of the US Advanced Technology Program (ATP) and the National Institute for Standards and Technology (NIST).
NIST is a non-regulatory federal agency within the Technology Administration of the U.S. Department of Commerce. Founded in 1901 NIST is a key element of the technology-based infrastructure supporting the U.S. economy. Its primary mission is to leverage private-sector economic activities and enhance economic growth by working with the U.S. industry and by developing and applying measurements, standards, and technology.[39]. NIST is divided into four co-operative programs: The NIST laboratories, the Baldrige National Quality Program, the Manufacturing Extension Partnership and the Advanced Technology Program (ATP). Since 1992 economic impact studies have been conducted on a regular basis. In general, economic impact measures such as the benefit-cost ratio (BCR) and social rate of return (SRR) were used to measure the economic consequences associated with NIST research and services. Once of the recent economic impact assessments will be illustrated below.
The ATP has been initialised in the USA in 1990 in order to facilitate the development of market-orientated innovations and to ease the communication process between scientists and industry, thus, boosting national wealth through co-operation with the private sector. The main rationale of the ATP's decisions which specific projects to foster is based on its social returns. Thus programme management had to develop tools and methods to assess social benefits (consisting of private profits and consumer benefits) generated by the different projects.[40] One of the CBA studies recently completed for the ATP is presented here.

3.7.3.1. A CBA of Medical Technologies

The Economic Assessment Office (EAO) of the Advanced Technology Program (ATP) commissioned the Center for Economics Research at Research Triangle Institute (RTI) to estimate the social success of certain medical research programs financially fostered with ATP means.[41.] In addition to the assessment of the *potential benefits of an inclusive portfolio group of ATP projects* [42], the RTI was requested to perform three supplementary tasks: 1) Seven case studies had to be performed using a consistent methodology. 2) An evaluation framework, adjustable to a wide range of different technology programmes, had to be established. 3) The ATP's Focused Programme in

[39] see http://www.nist.gov/public_affairs/nist_mission.htm
[40] see www.atp.nist.gov/atp/overview.htm
[41] see Martin, Sheila A. et al. (1998), pg. 3-1
[42] see Martin, Sheila A. et al (1998), pg. iii

tissue engineering [43] shall be informed of the potential socio-economic benefits which can be reaped in this technology area.
From the following seven projects, four were analysed in detailed case studies. The rest was discarded for reasons of data availability:

Table 14: Medical research programs-selected case studies

ATP Project Title[44]	Duration	Funding Level
In-Depth Case Studies		
Stem Cell Expansion	2 years	$ 1,220,000
Biopolymers for Tissue Repair	3 years	$ 1,999,000
Living Implantable Microreactors	3 years	$ 4,263,000
Proliferated Human Islets	3 years	$ 2,000,000
Brief Case Studies		
Biomaterials for Clinical Prostheses	3 years	$ 1,999,000
Gene Therapy Applications	3 years	$ 1,996,000
Universal Donor Organs	3 years	$ 1,999,000

Source: Martin, Sheila A. et al. (1998), pg. 1-13

The information required for the in depth studies were gathered by interviews with company representatives and physicians, using ATP proposals and the regular progress reports of the projects as well as medical databases and journals and by exploiting publicly available company and industry information.

3.7.3.1.1. Approach and Methods

One of the main objectives of the survey we focus on was to develop a methodology to assess the social return on ATP investments in research projects with medical applications. The approach is based on the methodology recommended by Mansfield (1996). In this model private returns also include spillovers i. e. revenues to other companies, which are therefore not part of the social returns.
Due to time, budget, and data constraints which quite often are restrictions to the scope of CBA-studies it was necessary to reduce the complexity of the task by limiting the scope of the study in several ways:45 First, just one application of each project was investigated (the most likely one), even if the technology might lead to a wide range of possible medical treatments. Second, the authors of the survey limited the time horizon for each project, thus, they supposed that the technology will be substituted by a newer one after a ten year production phase; consequently all costs and benefits lie within an overall time horizon of 20 years. Thus the estimates of both costs and benefits were on the conservative side.
The authors identified several channels by which ATP funding can influence the medical research programmes of private companies[46]:

[43] "Tissue engineering integrates discoveries from biochemistry, cell and molecular biology, genetics, material science and biomedical engineering to produce innovative three dimensional composites having structure/function properties that can be used either to replace or correct poorly functioning components in humans or animals or to introduce better functioning components into these living systems."; see http://www.atp.nist.gov/www/press/97-07ts.htm

[44] Just the abbreviated title of the projects are listed below, for the full titles see Martin, Sheila A. et al. (1998), pg. 1-13

[45] see Martin, Sheila A. et al. (1998), pg. 1-3

[46] see Martin, Sheila A. et al. (1998), pg. 1-4

- Accelerating the technology's benefits
- ATP funding can speed up the R&D phase and therefore facilitate an early introduction of the new technology, which benefits the private sector, the patients and the society in general. This result enables the company to introduce its development onto the market sooner than originally expected, thus, the period in which social benefits are accrued is – ceteris paribus - lengthened.
- Increase the likelihood of success
- ATP funding might increase the overall R&D efforts and therefore the likelihood of its success. This possible impact arises from the decreasing R&D costs through ATP funding and thus encouraging additional R&D effort
- Widen the technology's applications
- ATP funding might broaden the scope of the R&D projects, i.e. due to more funds invested in the development of a possible product the company might be capable to apply its results to more patients or diseases than originally expected.

The approach chosen to estimate the social return on public investment was to compare two alternative scenarios; one considers the situation with ATP funding and the other constructs a world without it. The difference between these to states of the world varies with the impact the above mentioned three factors.
Furthermore, in order to assess the social economic returns it was attempted to specify the **costs incurred by various diseases**. Three different types of costs were identified[47]:

- *direct medical costs* like the costs of medical treatment
- *indirect costs* like the costs of decreased productivity and the unpaid care
- *intangible costs* like the pain and the suffering of patients

In the next step the medical benefits to patients were calculated:

Assessment of medical benefits to patients

Total health benefits can be calculated by multiplying the number of patients that will be treated with the benefit per patient. To derive an estimate of the value of change in health outcome attributable to the new medical technology (= benefit per patient) a step-by-step approach was adopted[48]:

- Step 1: Setting up a model in order to assess the technology's impact on health outcomes
- Step 2: Changes in health outcomes, induced by the new technology, have to be quantified in terms of patient well-being
- Step 3:Changes in the patient's well-being have to be expressed in terms of money.

Ad Step 1.: After identifying the currently applied method of treatment (defender technology) and therefore the alternative to the newly developed medical technology two models to estimate the outcome of each of the treatments were introduced: The *Chronic Disease Model* and the *Acute Illness and Injury Model.* The chronic disease model quantifies the impact of a new technology on the progression of a chronic disease over time. It is a multiple-step model being repeated each year and includes a transition probability matrix which expresses the likelihood that patients switch to another health

[47] see Martin, Sheila A. et al. (1998), pg. vi
[48] see Martin, Sheila A. et al. (1998), pg. 2-13

state with different benefits. The so called quality-adjusted life years (QALY) are based on a Markov process where patients transition from one health state to the next over time. Treatment costs are assigned to each of these health states. A new medical technology is likely to cause a change in the dispersion of persons to different health states and therefore alter the total annual QALYs for all patients and the total treatment costs. The acute illness an injury model is a single-period case of the former model. Expected benefits and cost for each technology are calculated by multiplying the probability of each health outcome by the associated QALYs and treatment cost.

Ad Step 2.:The attribution of utility to changes in the patients well-being is measured via the concept of quality-adjusted life years (QALY). It weights each year with a factor within the range of 0 and 1, where 0 is associated with death and 1 with perfect health. Consequently a year of life is worth more to a person than a year experiencing a painful disease. QALYs quantify this difference in well-being and therefore capture the effects of pain. For this study QALY values from other empirical studies were used.

Ad Step 3.: Finally, derived from the literature, an estimation of $5 million for the value of avoiding premature death at an average age of 40 was entered. In assuming a life expectancy of 76 years the annual QALY value (V) was derived by getting the 36-year annuity value (– which is equivalent to the expected life loss at the age of 40 -) of this amount of $5 million:

$$V = \frac{5{,}000{,}000}{\sum_{t=1}^{36}\left(\frac{1}{1+d}\right)^{t}}$$

d = discount rate and
t = index of the year

Assuming for example a 3 % discount rate the QALY Value (V) amount to $229,019.[49].

After the derivation of the benefit per patient the **number of beneficiaries** was determined. For this reason a specific diffusion model (the Bass Model) was applied. In general, a diffusion model tries to assess the number of applications occurring during the "life time" of a newly introduced technology since typically firms and/or customers do not adopt them simultaneously. The Bass Model contains two parameters to characterise the diffusion process: *p* is the coefficient of innovation and reflects the "external influence" i.e. the adoption due to influence from external activity (publications etc.). q is the coefficient of imitation, representing the "internal influence". Thus, the model shows the following structure[50]:

$$\underbrace{a(t+1)}_{\text{Number of new adoptions}} = \left[p + q\underbrace{\frac{A(t)}{M(t)}}_{\text{market penetration proportion}}\right]\underbrace{\left[M(t) - A(t)\right]}_{\text{Number of potential adopters remaining}},$$

with: A(t) = cumulative number of adopters in year t,
M(t) = total market potential in year t

[49] see Martin, Sheila A. et al. (1998), pg. 2-27
[50] see Martin, Sheila A. et al. (1998), pg. 2-29

The number of new adoptions is proportional to the difference between market potential (M(t)) and the number of previous adopters (A(t)). After the collection of data for M(t) and A(t), the parameters of the model can be estimated and be used for forecasting the number of adoptions of the new technology.

3.7.3.1.2. Estimation of Private Returns

The expected private return is influenced by the probability of technical success, the expected investments and cost for R&D, commercialisation, and production as well as the expected revenues. Therefore for both the with ATP and without ATP scenario the following information is required: R&D investment for each year of the R&D phase, investment in commercialisation for each year and phase, annual expenditures on the fixed and variable costs of production, annual revenues, and the probability of technical success.

The probability of technical success was derived by using the companies' own estimation adjusted to the expected completion date of the examined project:[51]

$$P_r = \frac{TP}{PF}$$

where TP stands for the percentage of progress the company has made towards demonstrating technical feasibility. It is calculated as the midpoint of the range of the reported progress. PF denominates the percentage of the projects' calendar time already passed at the date the project was estimated.

Private R&D investments was calculated by deducting the ATP contribution from the total project budget.

Due to the fact that *costs for commercialisation and production* are hard to estimate while a technology still is in its development phase, the costs for commercialisation were derived by examining a composite balance sheet of the biotechnology industry, which unveils that selling, general, and administrative expenses match approximately 37 % of total revenues. These costs were divided in fixed and variable ones, whereas the former occurs during the commercialisation phase (γ) and the latter during the production phase. The authors derived therefore following equation for the fixed costs (CC_F):

$$CC_F = \gamma \left[0.37 \sum_{t=1}^{n} TR_t \right]$$

with: TR_t = total revenues in year t
γ = share of cost incurred during the commercialisation phase
n = number of years of production

Thus, a (1 - γ) share incurs during the production phase. Estimates of production costs were also developed from balance sheets. The composite balance sheet for the biotechnology industry shows that the production costs of the industry match about 42 % of total revenues, which can be implemented in a straightforward way.

The *expected revenues* were derived on the basis of company statements.

[51] see Martin, Sheila A. et al. (1998), pg. 2-35

3.7.3.1.3. Estimation of Measures of Economic Return

After gathering the data, the economic return on ATP-funded investments was calculated whereby three different perspectives were considered: the social return on public investment, the social return on public and private investments, and the private return on private investments. To develop the measures of social and private returns first a time-profile for benefits and cost for each scenario has been constructed. Afterwards NPV and IRR were chosen as the appropriate measures for the economic return. Sensitivity analysis was conducted in order to check the outcomes under different discounting rates.
In order to obtain the social return on public investment the difference between the expected net benefit to society in the with-ATP case (ENB_t^w) and in the without-ATP case (ENB_t^{wo}) for each year were computed. Thus the incremental net benefit resulting from ATP-funding is calculated by:

$$IENB_t = ENB_t^w - ENB_t^{wo}$$

with: $IENB_t$ = incremental net benefit in year t
ENB_t^w = expected net benefits to society with ATP
ENB_t^{wo} = expected net benefits to society without ATI

Afterwards the annual values of the IENB's were used to calculate the social return on public investment.

3.7.3.1.4. Results

As shown in the following table, the social outcome of the different projects (in terms of net present value and the internal rate of return) are widespread. The figures range from NPV of \$ 47 million to \$ 17.7 billion and from a IRR of 21 % to 148 % respectively. The overall NPV and IRR derived from the individual results add up to \$ 34.3 billion and accordingly to 116 %. Thus, the ATP funding invested in these projects results in net benefits of over \$ 34 billion.
The table also provides an insight into the proportion of the total social return contributed by ATP investments. The different projects exhibit a wide range to which public investment influences the final social outcome; the figures show that there is a 100 % contribution in the case of "Biopolymers for Tissue Repair" but just an 23 % share at the "Living Implantable Microreactors". Nevertheless, the composite figures in the last row unveil that 31 % of the social benefit is induced by ATP funding.

Table 15: Expected Social Returns

ATP Project Title	Project Time Horizon	Expected Social Return on Investment		Expected Social Return on Public Investment	
		NPV (1996$ millions)	IRR (%)	NPV (1996$ millions)	IRR (%)
Stem Cell Expansion	1992 to 2009	$134	20%	$47	21%
Biopolymers for Tissue Repair[52]	1994 to 2009	$98	51%	$98	51%
Living Implantable Microreactors	1994 to 2009	$74,518	149%	$17,750	148%
Proliferated Human Islets	1995 to 2008	$2,252	36%	$1,297	34%
Biomaterials for Clinical Prostheses	1993 to 2010	$32,855	118%	$15,058	128%
Gene Therapy Applications	1995 to 2011	$2,411	106%	$945	111%
Universal Donor Organs	1992 to 2011	$2,838	91%	$783	92%
Composite	1992 to 2011	$109,229	115%	$34,258	116%

The composite measure of return is based on a sum of expected benefits and costs in each year across all projects. The time period for the composite measure includes all years from all the individual project periods.
The composite IRR is not an average of the individual project IRRs because IRR is not additive.
The composite NPV is not a simple sum of individual NP because the time periods are different.
Source: Martin, Sheila A. et al. (1998), pg. 1-22

There are several reasons why social returns to these projects differ: e.g. the number of patients treated, the value of the health benefits of the new development, and the impact on health costs influence the outcome as well as the probability of technical success. The main channels by which these benefits materialise are depicted in Table 16.

Table 16: Impact of ATP on technology development

ATP Project Title	Project Acceleration (years)	Increase in the Probability of Success (percent)	Widening of Technology Applications
Stem Cell Expansion	1 to 2	9%	None reported
Biopolymers for Tissue Repair[53]	At least 10	171%	Significant but not quantified
Living Implantable Microreactors	2	11%	None reported
Proliferated Human Islets	3 to 5	2%	None reported
Biomaterials for Clinical Prostheses	2	1%	None reported
Gene Therapy Applications	2	20%	Some effect reported but not quantified
Universal Donor Organs	1 to 2	16%	None reported

Source: Martin, Sheila A. et al. (1998), pg. 1-23

If we take the project *"Biopolymers for Tissue Repair"*, it is clear, why 100 % of the social returns are attributed to the ATP promotion. Public contribution boosted the research effort in this field resulting in an acceleration of the materialisation of social benefits by at least ten years, and an increase in the probability of success of 171 %. The company representatives reported that without public funding the project either would not have been pursued or that it might have been developed to tardily to make sufficiently use of the market opportunity. In general, the table indicates that at least two out of three channels improve the social return of the regarded research projects [54]

Table 17 clearly shows the importance of ATP funding: Due to the fact that the private returns and hence the IRR on private investments are substantially smaller than the

[52] (see Martin, Sheila A. et al. (1998), pg. 1-22)

[53] For Biopolymers, the two sets of figures are identical because all of the social return can be attributed to ATP investments

[54] see Martin, Sheila A. et al. (1998), pg. 1-23 – 1-24

overall benefit the society accrues, just relying on market mechanism may lead to under investments in R&D projects, which were from the society's point of view less optimal.

Table 17: Private returns, Additionality of ATP

Composite Private Returns[55]	NPV (1996$ million)	IRR (%)
Project returns (composite)	$1,564	12%
Increment attributable to ATP	$914	13%

Source: Martin, Sheila A. et al. (1998), pg. 1-23

This study was able to demonstrate some of the strengths of the approach, namely in that it incorporates all relevant and available information and varies all of the parameters simultaneously. On the other hand, the ATP cautioned that the results are – of course - sensitive to the underlying assumptions and should be used with care, though they generally support for the rationale of the program put forward by the ATP.

3.7.3.2. Measuring the Economic Impact of Role-based Access Control

In the early 1990's NIST (National Institute of Standards and Technology) started working on role-based access control, the first comprehensive RBAC model was published in 1992. Access control systems are used within computer networks to control the actions, functions, applications, and operations of legitimate users [56]. Role-based access maps the control system to the organisational-specific structures in order to reduce direct and indirect administrative costs and improve security. NIST's RBAC activities include the development of generic technologies that provide the technology base for RBAC market applications, and the development of infratechnologies that support implementation and interoperability across different systems.
The objective of the impact assessment study, conducted by the Research Triangle Park, was to measure the benefits of RBAC relative to alternative access control systems and the economic return from NIST/Information Technology Laboratory (ITL).

3.7.3.2.1. Approach and Methods

Telephone interviews and Internet surveys with software developers and companies and organisations (end users) that integrate RBAC products into their business operations were used to estimate the benefits of RBAC and the technical and economic metrics for RBAC'S impact on end-users. In addition a case study with a multi-product insurance company was conducted in order to capture an insight on benefits and costs associated with the implementation of RBAC on different users (company's employees and extranet users). Based on an impact and cost metrics a theoretical model was developed to quantify the incremental benefits of RBAC relative to other access control methods. Time series were constructed that compare the net benefits of RBAC with and without NIST contributions. The difference, i.e. the change in net benefits attributable to RBAC, between them is the economic impact of the NIST/RBAC project.
For the measurement of the counterfactual time-series of costs and benefits had to be developed for three key segments of the RBAC supply chain:

[55] see notes *18*, *19*, *20* and *21*
[56] see Gallaher, Michael P., O'Connor, Alan C., and Kropp, Brian (2002), pg. 1-1

- NIST's expenditures
- Software developers' R&D expenditures with and without NIST
- Industry users' benefits and adoption cost with and without NIST

The expenditure time-series on NIST were collected straightforward from the NIST database. The average R&D costs for software developers were developed from the responses of the telephone interviews and to a limited amount from the Internet survey responses. Firm-level benefits and costs of establishing RBAC were collected via an Internet Survey among subscribers of Information Security Magazine, a leading publication for information and systems security administrators. Prior to the survey a series of impact hypotheses and cost metrics were developed which had to be investigated during the interviews with the industry experts. The economic impact is therefore separated into two main categories. Each of them is then assigned to specific impact hypothesis.

- The benefits of RBAC (compared to the alternative technologies)
- Reducing the administrative processing time,
- increasing productivity, and
- reducing the likelihood of security violations
- NIST's impact on development and adoption of RBAC products and services
- Lowering the private costs of R&D
- Lowering the private costs of implementing RBAC systems
- Accelerating the adoption of RBAC products

Technical and economic impact metrics where then used to quantify each specific hypothesis[57] within the scope of the interviews and surveys. Software developers were asked to reflect on R&D expenditures for RBAC, future developments and the impact of NIST; the Internet survey among software end-users was focused on the benefits of RBAC. Based on these impact and cost metrics firm-level costs and per-employee RBAC benefits were then calculated.
Costs attributed to RBAC consists of R&D expenditures and the costs of implementation and customisation per employee. R&D expenditures are expressed as average software developer R&D expenditures (R&Dsd) and average R&D costs for a typical user (R&Dih, in-house costs):

$$R\&DExp = R\&Dsd * Nsd_t + R\&Dih * Nih_t$$

with: R&Dsd = total R&D costs for a typical software development company,
R&Dih = total R&D costs for a typical user
Nsd_t = number of software developers in year t that developed an RBAC product
Nih_t = number of users that developed in-house RBAC products in year t

Furthermore the end-users' customisation and implementation costs (IC_t) in all industries are given by the sum of the implementation costs per employee (IC_{it}) in industry i and year t times the number of employees in industry i managed by RBAC systems in year t (Emp_{it}):

[57] see Gallaher, Michael P., O'Connor, Alan C., and Kropp, Brian (2002), pg. 4-8 – 4-10

$$IC_t = \sum^{i} IC_{it} * Emp_{it}$$

Accordingly the sum of benefits per employee of RBAC in all industries can be defined as:

$$OB_t = \sum^{i} ((\underbrace{AC_{it} + PB_{it} + SB_{it})}_{OB_{it}} * Emp_{it})$$

with: OB_{it} = operating benefits per employee
AC_{it} = reduction in administrative costs per employee
PB_{it} = productivity benefits per employee
SB_{it} = security benefits per employee
Emp_{it} = number of employees in industry i being managed by RBAC

The time-series net benefits (NBt) from RBAC can then be expressed by summing benefits and costs to software developers and users in all industries:

$$NB_t = R \& Dsd * Nsd_t + R \& Dih * Nih_t + \sum^{i} (OB_{it} + IC_{it}) * EMP_{it}$$

Using the above-described equations a baseline time series of benefits and costs of RBAC (i.e. "with NIST scenario") has been calculated in order to measure the economic return of the NIST/ITL RBAC. It is based on the statements of industry experts who indicated that many industries would implement some form of RBAC-enabled system by 2005. Hence, benefits and costs associated with RBAC are therefore modelled through 2006, one year after the predicted penetration. Then this baseline time series was shifted as to represent a counterfactual world without NIST. The key parameters [58] for the determination of this scenario are based on the interviews with software developers.
The economic impact of NIST was then calculated by the change in net benefits (ΔNB) of RBAC with and without the contributions of NIST. Rewriting the above equation the changes resulting from NIST's contribution can be expressed as[59]:

$$\Delta NB_t = \underbrace{\Delta R \& Dsd * Nsd_t}_{\text{change in R\&D costs for software development companies}} + \underbrace{\Delta R \& Dih * Nih_t}_{\text{change in total R\&D costs for end users costs}} + \sum^{i} [(OB_{it} + \underbrace{\Delta IC_{it}}_{\text{change in implementation}}) * \underbrace{\Delta EMP_{it}}_{\text{change in number of employees}}]$$

As described above the impact of NIST may result from the change in R&D and implementations cost as well as in the number of employees being managed by RBAC (change in adoption rate).

[58] i.e. the change in R&D costs, in the acceleration of R&D activities, and in the enhanced diffusion due to the contribution of NIST
[59] see Gallaher, Michael P., O'Connor, Alan C., and Kropp, Brian (2002), pg. 8-5

The Net Present Value (NPV) was computed by calculating the difference between the time-series of the net benefits associated with NIST's contribution (ΔNB_t) and the NIST/ITL RBAC project expenditures: Employee-level impacts are therefore adjusted for inflation and discounted using a 7% social discount rate.

3.7.3.2.2. Results

Based on the interviews with software developers and end-users it was estimated that the activities of the NIST RBAC project have accelerated the development and adoption of RBAC by one year and lowered the costs of R&D by approximately 6%[60]. Using this information the time series for net benefits have been calculated. The results are shown in Table 18:

Table 18: Net benefits of RBAC with and without NIST's contribution (million$)

Year	Baseline (with NIST	Counterfactual (without NIST)	Total Change in Net Benefit (ΔNB_t)
1992	-	-	-
1993	-	-	-
1994	-	-	-
1995	-	-	-
1996	-5.05	-	-5.05
1997	-5.05	-5.50	0.45
1998	-5.05	-5.50	0.45
1999	-5.05	-5.50	0.45
2000	-18.08	-5.50	-12.58
2001	-0.36	-16.90	16.54
2002	19.84	-0.33	20.17
2003	60.26	18.54	41.72
2004	207.08	56.32	150.76
2005	308.51	193.53	114.97
2006	337.85	288.33	49.52

Source: Gallaher, Michael P. et al. (2002), pg. 8-7

The net present value of NIST's impact on the benefits of RBAC sums up to $ 295 million. NIST's expenditures were then used to calculate the NPV of the NIST/RBAC project. Results are shown in Table 19:

Table 19: Economic return on the NIST/ITL RBAC project (million$)

	High	Medium	Low
NPV change in net benefits	427.42	294.77	185.71
NPV NIST expenditure	2.70	2.70	2.70
NPV of the NIST/ITL RBAC project	425	292	183
Benefit-Cost ratio	158	0109	69
Internal rate of return	90%	62%	39%

Source: Gallaher, Michael P. et al. (2002), pg. ES-7

The economic impact was measured with subject to three different diffusion scenarios, reflecting the uncertainty concerning the future rate of penetration. But even under the

60 see Gallaher, Michael P. et al. (2002), pg. 8-6

low scenario the NPV of NIST's impact still amount to $ 186 million. The net present value of the NIST/RBAC project ranges from $ 425 million to $ 183 million.
The net benefits quantified in this study include only the administrative and productivity benefits (reduced downtime for new employees). Data on organisational productivity, upper management costs, and systems security and integrity have not been collected due to insufficient information gathered. This once more indicated the difficulty of data collection and availability, which is in fact not only a problem in CBA but for other methods, too. Even by conducting a survey not all necessary information can be obtained. Furthermore one has to consider that the sample of RBAC end-users – the basis for calculating the benefits of RBAC - is rather small: 9,530 e-mail messages were sent to the subscribers of the Information Security Magazine, 92 individual companies responded.

3.7.4. Data Requirements

CBA, if done accurately, poses high demands on data requirements. While private costs associated with the project can be identified without too great a burden of data collection (and should be already incorporated in the standard reporting and monitoring scheme of the programme), already the costs to society are hard to capture. These would include the opportunity costs of funding this programme against another one, or against leaving the monies which had to be raised by taxes beforehand in the hands of the private sector.
In order to track social returns, normally one has to rely on the estimates of existing literature (as was the case in our example). Where such estimates are not available, cost of data collection for the estimation of social costs will be high, especially when it is a generic technology with very widespread effects. Here, one would have to revert to macroeconomic studies of economic impacts.

Also, its applicability hinges crucially on the possibility to monetise as much as possible of these impacts. Where market prices would not be available or distorted, a CBA analysis would have to embark on other ways to estimate benefits by consumer interviews (e.g. to measure willingness-to-pay and willingness-to accept) and the like. Again, these will be costly.
Thus, CBA seems best suited when a limited number of cases are analysed with effects large enough to justify the efforts to identify social returns.

3.7.5. Steps for Implementation

Step 1: Identification and separations of relevant costs and benefits

As mentioned in Section I costs and benefit can either be direct or indirect, tangible or intangible, internal or external. Problems can arise from costs or benefits that are not assigned to a certain project or program.

Step 2: Monetary valuation of costs and benefits

All relevant costs and benefits should be quantified in monetary terms. Where sufficiently perfect markets exist market prices can be used to determine the impacts. Difficulties arise whenever the market prices do not reflect the true marginal cost or benefits. This is often the case for intangibles and external effect. As described in section I shadow prices should be calculated to value these effects.

Step 3: Discounting

Since costs and benefits occurring at different times they should be discounted at an appropriate rate to make them comparable. Choosing the right discount rate is an important task. Therefore sensitivity analyses are recommended in order to systematically vary uncertain parameters to see what effect they have on the outcomes.

Step 4: Aggregating

After discounting the various effects of a project or program net present values, benefit-cost ratios or internal rate of return can be computed.

3.7.6. Scope and Limits and Conditions for Application

CBA mainly builds on the concepts provided by standard economic theory. While useful in some respects, these concepts have some shortcomings, especially when it come to the analysis of innovation and technical change.

First there are the often demonstrated difficulties to ex-ante assess most of the benefits from technological innovation which are inherently uncertain or even unknown as evolutionary approaches would stress (see Metcalfe 1995). Therefore, CBA has limitations as a tool for ex-ante evaluation and selection of programmes and projects.

Second, with respect to ex-post evaluation, the approach seems more promising, but problems remain regarding difficulties in tracing back costs and benefits and attributing them. These difficulties increase as more time has elapsed, which on the other hand is a pre-condition for economic outcomes to materialise.

Third, there are limits of CBA especially when it comes to taking into account 'intangible effects'. These effects could be essential programme and project outcomes even if there are no (other) tangible results in terms of organisational innovation, technological learning, increase of adoption capacities and various forms of 'behavioural additionality'. Thus, CBA would bias evaluation results against programmes predominantly aiming at such effects.

And finally, there are practical limitations to CBA. While the approach seems feasible for a small number of large programmes and projects (such as energy, space and aeronautics and the like), it is hardly suited for a large number of small projects in which case the costs of data collection and benefit estimation would be very high.

On the other hand, as the practical examples from the ATP have shown, CBA can have a value insofar as it leads programme and project managers to think about the types of costs and benefits of a programme/project in a more structured and systematic way, and to take into account a great variety of effects.

Thus we would conclude that in some circumstances CBA is a tool that might be applied to RTD programmes and projects, if one keeps in mind the limits of the approach, uses additional approaches for the evaluation of economic impacts and does not end up with short-circuited conclusions.

3.7.7. Bibliography

Basic literature

The literature on cost-benefit analysis is very extensive, since it is widely used as an analytical tool in assessing the costs and benefits of various investment projects. The following publications provide useful insights into the basic principles of CBA – theory as well as application –, strengths and limits of the concept.

Brent, R. J. (1996): Applied Cost-Benefit Analysis, Edward Elgar. Cheltenham. The book starts with the basic principles of CBA, addressing the fundamental parts and provides useful and detailed illustrations and case studies of actual applications.

Georghiou, L.; Roessner, D. (2000): Evaluating technology programs: tools and methods. In: Research Policy 29, 657 – 678. Elsevier. Amsterdam. The article discusses methods and designs used in evaluating public research and development programmes. It critically addresses advantages and limits of different evaluation methodologies.

House, E. R. (2000): The Limits of Cost Benefit Evaluation. In: Evaluation. Volume 6, Number 1, January 2000. The article critically examines the scopes and limits of CBA, with a special focus on fairness and justice considerations.

Mishan, E. J. (1994): Cost-Benefit Analysis, Routledge. London and New York

Stiglitz, J. (1999), Economics of the Public Sector. 3rd edition. Norton New York

References

The following articles are references within the literature on special issues of impact assessment of public R&D programmes.

Fritz, O.; Grasenick, K.; Gruber, M.; Hartmann, C.; Novakovic, M.; Polt, W.; Sturn, D.; Woitech, B.; Zinöcker, K. (2001): EvinA – Evaluierung on innovativen Aktionen in der Technologie-, Struktur- und Arbeitsmarktpolitik: Entwicklung von Methoden, Indikatioren und “Good Practice”. Projetkbericht gefördert durch das Bundesministerium für Wissenschaft und Verkehr. Wien – Graz. The report gives a detailed description of evaluation practice in the Austrian Technology-, Structural-, and Labour Market Policy. It provides methods, indicators and examples of good practice.

Hall, B. H. (1995): The private and social returns to research and development: what have we learned? Paper presented at AEI-Brookings Conference on The Contributions of Research to the Economy and Society, Washington D.C.

Jaffe, A. B. (1996): Economic Analysis of Research Spillovers – Implications for the Advanced Technology Program. Paper prepared for the Advanced Technology Program http://www.atp.nist.gov/eao/gcr708.htm. The paper in general describes sources and pathways of research spillovers and gives a review of economic literature on this topic. It relates the concept of spillover to the ongoing impact analysis at ATP and develops a framework for explicit evaluation of spillovers in ATP project proposals.

Link, A. N.; Scott, J. T.; (1998): Evaluating Public Sector R&D: A Retrospective Examination of U.S. Experiences. Paper prepared for the APEC Symposium on Evaluation of S&T Programmes among APEC Member Economies, Wellington

Mansfield, E. (1996): Estimating Social and Private Returns from Innovations Based on the Advanced Technology Program: Problems and Opportunities. Paper prepared for the Advanced Technology Program, National Institute of Standards and Technology, U.S. Department of Commerce

Metcalfe, S. (1995): The Economic Foundations o Technology Policy: Equilibrium and Evolutionary Perspectives, in: Handbook of the Economics of Innovation and Technological Change, Oxford

Polt, W (1998): Evaluation and Impact Assessment" – Advanced Training Course in Science & Technology Policies and Programmes. Paper prepared for presentation at Universidad Carlos III. Madrid

Smith, K. (1997): Can we measure the economic impact of public research and development? In: Callon, M. et al: The Strategic Management of Research and Technology. Paris

Tassey, G. (1998): Lessons Learned About The Methodology of Economic Impact Studies: The NIST Experience. www.nist.gov/director/planning/epp-paper.htm. The paper describes the economic impact assessment activities at the National Institute of Standards and Technology (NIST). It gives an overview of frequently used impact variables and briefly discusses the most common impact measures.

Reports

Reports and articles cited below provide good examples for the use of cost-benefit analysis in evaluating the impacts of R&D programmes or projects.

Doremus, P. N. (1998): Performance Evaluation at the NIST Measurement and Standards Laboratories. Paper prepared for Workshop on Methods to Evaluate Research Outcomes. Office of Strategic Planning and Economic Analysis. http://www.nist.gov/director/planning/perfeval.pdf. The presentation gives a basic outline of the NIST Planning and Performance Evaluation System, briefly discusses quantitative metrics and impact assessments and concludes with

Foundation for Research, Science and Technology (1997): The Benefits of Meat Research in New Zealand: A Pilot R&D Outcome Review. Project report. The report is a good example for using CBA to estimate the benefits of public research in New Zealand. The report concludes with a brief rationale for the chosen CBA-approach and also discusses limits and advantages of cost-benefit analysis in general.

Gallaher, Michael P., O'Connor, Alan C., and Kropp, Brian (2002): The Economic Impact of Role-Based Access Control. Final report. Research Triangle Institute, Center for Economics Research. The evaluation of NIST'S RBAC-programme is based on a cost-benefit approach. By collecting the required data via interviews and surveys the report presents the methodology used to calculate the impact of NIST's funding.

Hertzfeld, H. (1992): Measuring the returns to space research and development. In: Hertzfeld, H.; Greenberg, J. Eds,. Space Economics. American Institute of Astronautic, Washington D.C.

Martin, S. A.; Winfield, D. L.; Kenyon, A. E.; Farris, J. R.; Bala, M. V.; Bingham, T. H. (1998): A Framework for Estimating the National Economic Benefits of ATP Funding of Medical Technologies – Preliminary Applications to Tissue Engineering Projects Funded from 1990 to 1996. Final report. Research Triangle Institute, Center for Economics Research. The evaluation report contains a detailed description of the methodology used in evaluating the benefits of seven ATP-funded tissue-engineering projects and also provides a case-by-case analysis of each of the projects. It is a good example for using CBA

Tassey, G. (1999): Assessing the Economic Impacts of Government R&D Programs.Presentation to Technology Transfer Society, May 20, 1999. http://www.nist.gov/director/planning/methodologytalk.pdf

The Advanced Technology Program. webpage: www.atp.nist.gov/atp/overview.htm. ATP Homepage, gives good information on the Programme, its projects, and evaluation processes.

3.8. EXPERT PANELS AND PEER REVIEW

Author: **John Rigby** (PREST)

3.8.1. Introduction

This section addresses the use of Expert Panels and Peer Review as mechanisms for the investigation and evaluation of RTD programme impacts. Expert / Peer Review is one of the most fundamental of decision making processes. It lies at the heart of many programme evaluation activities and it is moreover, the basis for the governance and progress of scientific enquiry itself. Recently, there has been concern that peer review, which is based on the same principle as panel review, has failed to provide the quality of judgement which is expected of it. Despite this, peer review remains popular because of its flexibility, wide scope of application and apparent fairness.

The terms Panel Review and Peer Review, while often regarded as synonymous are in fact distinct forms of review process with significant differences between them. These two methods though, comprise two of the most important forms of review based upon the principle of using a group which combines the most relevant and expert knowledge and experience available – Expert Review - to make a judgement of worth or value. In the modern age, the original use of experts to give judgement about the worth of an idea or proposal is normally regarded to date from the early eighteenth century when learned societies such as the Royal Society of London and the Royal Society of Edinburgh convened groups of experts to pass scientific judgement on articles submitted to them for publication. Bozeman and Melkers (1997) though, put the beginnings of peer review slightly early, arguing that it began with the foundation of the first publication of the "Philosophical Transactions" of the Royal Society in 1665.
The use of the term "Peer Review" by the European Commission for project evaluation by groups of experts assumes an equivalence with the review process by which scientific papers are assessed for quality; however, the groups which carry out the evaluations of EU projects are not composed solely of "peers" but of experts often with a broader range wide range of skills and experience than would be available in a true group of "peers". Despite this, the term Peer Review within the context of the evaluation of European Union projects has been retained to describe what is strictly Panel Review. Within the general category of Expert Review, there are a number of sub-types. Table 21 identifies those types of Expert Review which are currently used.

Table 20: Sub Types of Expert Review

Sub-Type of Expert Review	Level of Specialisation	Level of Professionalization
Traditional Peer-Review (Canonical Academic Review)	Generally Increasing	Generally Increasing
Direct Peer Review		
Modified Direct Peer Review		
Pre-Emptive Peer Review		
Indirect Peer Review		
Merit Review (extended form of Peer Review)		
Ancillary Peer-Review		
Expert Panels/Peer Review		
Panel Review		
Professional Evaluators		

As table 21 shown above indicates, peer review is the most specialist type of expert review. Experts who are normally within the academic context are not paid for their work but give their time freely. Moving down the table generally indicates a greater level of professionalisation – i.e. a greater use of professional evaluators and a greater involvement of a wider range of experts from broader disciplinary backgrounds.

Pre-emptive peer review is the form of review in which the whole decision-making process is delegated to the group of experts. The experts therefore have the sole right of decision-making, and consequently, this form of review, leaving no discretion to those appointing the group, is not used widely, although the National Institutes of Health (NIH) in the United States have operated evaluation in this form occasionally (Bozeman and Melkers, 1993). Ancillary Peer Review uses peer review along with other methods to decide a question of value or impact. Modified Direct Peer Review uses traditional peer review but peers are asked to look for broader impacts than at the narrow definition of the quality of publication outputs.

While Expert Review is often regarded as a specific research evaluation methodology, a Panel or Expert Review team can itself commission further studies which employ other research methodologies, such as surveys, questionnaires, benchmarking and case studies. Expert Review, whatever its precise form therefore, is more than simply a specific method, such as benchmarking or econometric analysis, being a process of decision making in which judgement is, to an extent delegated or parcelled out. The judgements of expert panels is dependent upon the expert’s own understanding and a range of evidence which the experts are best able to interpret.

Finally, Panel Review or Expert Review has a strong political aspect. Wherever programmes involve different stakeholders, such as countries, firms, funding bodies and research institutes, the use of experts with high reputations may help to ensure that the findings of the evaluation report will meet with acceptability at the political level (Massimo, 1997), increasing the likelihood that recommendations for change will be implemented.

3.8.2. Application to Policy Instruments and Interventions

Application of Panel and Peer Review methods is widespread and takes place both prospectively, i.e. ex ante, and retrospectively, i.e. ex post. The typical use of ex ante peer review is for the allocation of funding to research grant applications. The most typical form of the Panel Review process is the ex-post evaluation of RTD programmes. Panel Review is suitable where evaluators and policy makers require a general picture and assessment of the quality of the research carried out and the contribution which it has made to socio-economic development and where no other means of generating data or opinions have been developed. Panel Review is also appropriate when evaluators wish to look at programmes from entirely new perspectives. As panels bring "fresh-eyes" to the evaluation process, they can be a source of a great many new ideas.

Panels may also contribute to evaluation structure and systems by recommending to the programme managers how the Programme might be evaluated in the future. In the case of the EUREKA Programme, a major contribution to the development of the Programme was the introduction of the continuous and systematic evaluation methods which were subsequently adopted for use in the evaluation of the Programme. The wider use of peer review is well described in Guston (2000) who gives a wider range of options for peer review, including the use of the technique to assess changes proposed to regulatory structures, general policy and the evaluation of courtroom evidence.

3.8.3. Good Practice Examples

Critically examining the expert review process is commonplace within the academic context, particularly the scientific and medical publishing field where large numbers of studies (see for example Godlee and Dickersin, 1999) have analysed the process in an attempt to make improvements, especially in the reduction of bias. In the area of application of panel review to the impacts and outcomes of programmes seeking to fund research leading to the development of technology, the number of studies published is smaller, although it is still significant enough to make some tentative conclusions. The process of comparison and benchmarking though is not easy, for two main reasons:
Firstly, identifying principles or even examples of good practice from a limited number of cases of very different character provides the first difficulty. Secondly, as panel reviews often have the multiple aims, single examples which contain instances of good practice may at the same time contain practices which are less desirable. Attempting to single out one particular evaluation as the definitive good practice example may lead to a failure to draw attention to limitations which were also present. Finally, because evaluation units, rather than panels are responsible for the use of evaluations, much of the responsibility for the success of evaluations carried out by expert panels lies with those who commission the research, rather than the panel itself. Good practice in evaluation of RTD programmes by expert panels therefore should focus on good *practices* rather than simply on good *examples of evaluation*.
The following is a list of principles of good practice based on a wide expert review literature:

- Ensure that experts declare their interest – perhaps even publishing a bias statement and potential conflict of interest – to ensure that the panel's reputation for fairness is upheld (Bozeman and Melkers, 1993). The UK Research Assessment Exercise employs the following rules to ensure that potential and apparent conflicts of interest are avoided, see box 2. Concern that statements of financial interest should be made are rising, even for authors. The publication *Nature* took the step in Mid 2001 of asking authors of papers to declare their financial interests as there is increasing evidence that failures to do so are bringing the process of peer review into dispute.
- Restricting either the number of evaluations on which panel members serve or the number of years in which they are active, following the UK's Research Assessment Exercise guidelines (RAE2/98), is likely to reduce bias and complacency. Reducing the term of service to a panel beyond a certain level though can also have adverse effects, see note 6 below.
- Broadening the panel as much as possible without introducing those who do not have relevant skills is a means to ensure effective debate and discussion and the generation of new ideas.

- The number of panel members is a problematic area, with a major research effort in social and developmental psychology given to the effective functioning of groups of different sizes. No studies are reported in the literature of the relationship between group size and decisions made specifically within the context of panel review.
- Steps can be taken to publicise the area of expertise of a particular panel member thereby protecting the credibility of the expert process itself and the assessments made by the panel, although openness about the suitability of particular individuals' expertises could also be used to question or even to undermine the judgements which the panel makes.
- It is suggested that those within any panel who have more specialist knowledge of an area should not be allowed to make decisions as to quality and value of proposals but to submit their views to their peers within the panel itself for a joint decision.
- The appointment of panel chairs should or could be from the membership of previous panels to ensure continuity.
- The panel chair and other experts should all have high reputations in the areas in which they are required to make judgements so as to instil confidence in those affected by the evaluation.
- Occasionally it may be necessary to seek experts or expert witnesses from outside the geographical area where the programme is being carried out.
- In terms of increasing the efficiency of peer review, consideration should be given to the use of a wider variety of techniques such as the use of "remote reviewer participation" (Kostoff et al, 2001) thereby removing the need for panel members to be present at all meetings.

Box 2: Declaration of Interests in the UK Research Assessment Exercise 2001

All panel members, subpanel members, external advisers including –UK based advisers and panel secretaries will be asked to make a **declaration of their interests**. There are two categories of interest, material and minor. A **material interest** will prevent the panel member for participating in the assessment of submission from the institution concerned. A **minor interest** will not have this effect but may mean that there are parts of a submission on which the member should not comment.

…

A **material interest** will exist in respect of:

a The institution(s) at which the member is employed in 2001;

b Any institution at which the member has been employed during the assessment period;

c Any institution(s) at which the member's partner and/or an immediate family (parent, sibling or child) are employed during the assessment phase. This applies regardless of the nature or subject of the employment; it does not apply where a family member is a student.

d Any institution(s) which the member has advised, whether paid or unpaid, on the development of research strategies and activities or the preparation of RAE submissions.

…

A **minor interest** will exist in respect of any institution(s) at which the member has been engaged in substantial teaching or research collaboration during the assessment period. Examples would include being a co-holder of a research grant, principal investigator on a joint project or Director of a joint teaching programme.

3.8.4. Conditions for Application

While peer review is one of the most flexible methods for determining value, capable of application to a wide number of fields, in order to apply it, a number of essential pre-conditions must be met. Firstly, peers or experts with knowledge of a particular area must be available and be willing to participate. Because it can be difficult for government officials to identify the relevant peers as they not usually part of the social or professional networks of scientific peers, it is important for programme evaluators and responsible bodies to maintain access to such networks.
Secondly, the panel of peers or experts cannot be expected to answer questions which are beyond the scope of the available knowledge. Terms of reference need therefore to be set with some sense of what it is possible for the experts themselves to know or to infer and to judge collectively from their specialist knowledge.
Thirdly, the panel should only be asked to come to a judgement on a single area of knowledge or expertise rather than more than one as peer review is know to be weak where comparative judgements between different fields of expertise have to be made, see Weinberg (1966).
Fourthly, while the costs of peer review are low, sufficient resources should be made available to facilitate the work of the panel. Some panel reviews are often supported by a secretariat.

3.8.5. Steps for Implementation

Panel Review is often thought of as a relatively simple evaluation method which requires the client – the evaluation unit or the programme sponsor – to write a terms of reference, select a chairman and then set back and wait for the group of independent outsiders to be picked to carry out an investigation and submit a report. The operational steps required to establish and maintain a Panel Review Evaluation are however far more complicated and involved; and while panel reviews may devolve some important decision making activities, there are many unseen costs involved in panel reviews, some of which are born by the expert reviewers and their organisations. The steps involved in implementing a panel review are as described below.
Deciding the precise terms of reference which the panel is to be given to work to [Terms of Reference] is the first stage which precedes the operational steps and indeed the work of the panel itself. Deciding the terms of reference is, a task of considerable importance as all that follows is dependent upon it. The terms of reference define and clarify the brief of the panel and should cover issues of programme strategy, its scope, impact, appropriateness, specific issues and the form and range of its recommendations. Timing the peer review is also important. An evaluation should be completed in time for it to have a useful impact on policy. The timing of the evaluation should also ensure that there are likely to be some significant observable effects for the panel to review, in the case of an ex post evaluation.
The second most important step in the implementation of the Panel Review is the appointment of a suitable person to chair the Panel [Appointment of the Panel Chair] and the appointment of the panel members themselves. Checks on the relevance of an individual's experience for the work and any possible conflicts of interest which might arise are vital [Appointment of Panel Members]. This process should begin as soon as possible because, as Massimo (1997) suggests, appointing the panel members can be lengthy, in some cases the search for suitable panel taking up to sixteen months, although nearer nine months is the average. When potential appointees are found to be ineligible for membership for whatever reason, it may be necessary to look more widely

than the names on the original short-list so as to ensure the resulting panel remains as balanced as possible.

Once the panel is constituted, the Panel itself becomes responsible for the conduct of the review and must address the question of the procedures by which the panel will operate [Operating Procedure]. The operating procedure is also of signal importance as the rules by which the panel works have been shown in various studies to exercise a strong influence on the conclusions which panels arrive at. Once the rules for the conduct of the review are agreed, the Panel may address the issues of deciding the schedule to which the panel will work and report so as to allow time for it to deliberate and make its report [The Schedule].

Ensuring that the panel has the time and resources and necessary studies commissioned upon which it may wish to base its judgement is another major task for the client [Operational Support]. This operational support for the panel is another major requirement for successful panel review. Successful Panel and Expert Review will depend not only upon external support but on healthy and effective links between the Panel and the Programme Management itself [Links from Panel to Programme/Client]. Without good links, the evaluators may be unable to gain access to parts of the programme and to data upon which their work depends.

Panels also require support in the form of a Secretary whose principal function is to record the proceedings of the group, to write up minutes and to liaise with those who are responsible for the programme [Appointment of Secretary]. In addition to this, the Secretary of the panel is often the person who is best able to carry out the essential task of maintaining the “corporate memory” of evaluation activities and good practice which will inform future programmes and also future evaluations. If the Secretary is unable to carry out these other functions [Maintaining the Corporate Memory] and [Providing Knowledge of Evaluation Methodologies], then the panel itself should contain someone who can. This will ensure that the Programme evaluation employed the most appropriate methods and that the findings are available for future programmes and their evaluations. As evaluation techniques and methods are subject to continual change and development, it is essential that Panel members are kept up-to-date in their knowledge of methodology. Only if the scope and limitations of particular techniques and methods used by panels are known will the evaluation reports be reliable.

The relationship between the panel and any outside contractors which is also carrying out evaluation work either to assist the panel or independently can be a source of tension. As Bobe and Viala (1997) make clear, panels may feel that their independence is challenged and their expertise questioned when other studies are carried out by consultants.

Table 21: Establishing A Peer Review – Key Steps

Main Steps	Specific Functions and Actions	Location of Key Responsibilities
Setting the Terms of Reference	Often the sole responsibility of the Client (Programme Managers)	Client
Overall Time Available	Set often within the terms of reference, the deadline for the programme	Client
Appointment of Panel Chair	Carried out by the Client	Client
Appointment of Panel Members	Carried out jointly by Client and Panel Chair – often a very lengthy process	Client – Panel Chair
Appointment of the Panel Secretary or Scribe	Secretary can, in addition to making notes and communicating with the client and other stakeholders, provide the a) - corporate memory for the client and b) – bring some knowledge of evaluation methodologies to the work of the panel	Client – Panel Chair
Operating Procedure	Rules of how to operate discussion, use of evidence, areas of responsibility, functions of meetings, etc.	Panel
Schedule of Work of the Panel	Panel decides how to achieve the deadline given in the terms of reference	Panel
Links from Panel to Programme/Client and other Sub-contractors	Managing contact between the panel and the client and between the panel and other contractors	Panel - Client
Identifying the Requirement for External Support	External support in the form of resources and further data/ indicators / access to individuals may be required from the client	Panel
Interim Reporting	Giving the client the opportunity to see the report and make comments	Panel
Final Reporting	Final report delivered in the light of comments received from the client	Client
Dissemination	Client takes on the responsibility to use the report effectively through dissemination to relevant and interested parties	Client

3.8.6. Data Requirements

Panel Reviews may use a very wide variety of data and indicators to allow them to apply their expert knowledge of the field of RTD. The key issues which Panels should address are:

a) Will the data be available in the form in which the panel members can use it – have they had an opportunity to specify the format of the data which they are required to interpret? (Relevance)
b) Data sets which panel reviewers are required to use should also be comparable and congruent with each other so that they can be used together to make assessment of consistency – a feature which is common to other multiple methods approaches (Triangulation of Methods and Data)
c) Will the data available when the panel require it? (Timing)

3.8.7. Scope and Limits

Panel and Peer Review are flexible and effective tools for both ex ante and ex post evaluations. They are relatively simple to operate and relatively inexpensive although the support costs can be significant. Panel Review also depends upon the availability of those with sufficient expert knowledge of the area. In small-scale scientific fields, where peer review attempts to agree an ex ante judgement of quality, just as in large scale ex post socio-economic evaluations, a supply of suitable experts is essential for panel review to work. The availability of experts therefore defines the scope of application for panel review. In the absence of suitable experts with a general overview, the Programme Managers or evaluation unit may need to engage specialists to examine particular aspects of the RTD programmes. While the range of activities to which peer review assessment is being directed increases, a range of concerns about the suitability of peer and panel review remains.

Table 22: Strengths and Limits of the Panel Review Process

Strengths	Limits
Perceived Independence	Scope of application is defined by availability of experts
Panels can employ methods which their specialist expertise suggests might be relevant, for example Scientometrics techniques such as bibliometrics, or surveys	Risk of geographical bias, self-interest
Broad evaluation of quality approach which specialists are unable cover	Programme Officers could manipulate evaluators
Inexpensive	Peer Review is subjective – it is best supplemented by technical – metric based measures
Can be continuous	Costs can be significant, although they may be born by their panellists and their organisations rather than the client
Generally flexible - can work off-line	Peer Review and variants shown to be conservative and leading to "institutionalised orthodoxy" (US Department of Health and Human Services, 1989, quoted in Bozeman and Melkers, 1993)
	Not suitable for comparisons across fields – single domain area is best
	Identification of suitable peers is sometimes difficult for government staff, who are not researchers themselves
	Where research is likely to be non-public domain, the peers might be competitors. In this case peer review is not helpful.
	Group Dynamics are important – (although there is very little research on this in the context of R&D evaluation (Porter and Rossini, 1985)
	There may be a reluctance on the part of panellists to state their views on paper as peer review is normally non-anonymous (Grigson & Stokes, 1993)

3.8.8. Bibliography

Evaluation Reports

The following Eureka Evaluation Reports provide good evidence of the type of work carried out by panels. The procedures of the Higher Education Funding Council for England are well documented on the HEFCE website. The panel discussions are however, confidential, with outputs taking the form of ratings. An example of a scientific report which judges expressions of interest in major for integrated projects in the life sciences is the Panel Report for the QoL integrated projects.

Eureka Evaluation Reports (1995) (1996) (1997) (1998) (1999) (2000)

Report of the panel of experts for the evaluation of topics suitable for QoL integrated projects (2001) Expressions of Interest for Topics concerning Integrated Projects in Functional Genomics Relating to Human Health

UK Higher Education Funding Council – (www.hefce.ac.uk) web-site giving details of the refereeing procedures for the Research Assessment Exercise

References

The following are references within the literature to work carried out on the form and role of peer review. Works cited here are more theoretical than in the next section and give descriptions of the essential principles of peer review. A number of critiques of the process are presented.

Guston, D. (2000) “Expanding the Role of Peer Review Processes in the United States”, paper presented to the US-European Policy Workshop on Learning from Science and Technology Policy Evaluation, Bad Herrenalb, Germany

Grigson, D. & Stokes, T. (1993) “Use of peer review to evaluate research outcomes” Research Evaluation, Vol. 3, No. 3, pages 173-177

Campbell, P. (2001) “Opinion: Introducing a new policy for authors of research papers in Nature and Nature journals” Nature 412, 751 (2001)

Godlee, F. & Jefferson, T. (1999) Peer Review in the Health Sciences, BMJ Books: London

Roy R (1984) Alternatives To Review By Peers - A Contribution To The Theory Of Scientific Choice Minerva 22 (3-4): 316-328

Roy R (1985) Funding Science - The Real Defects Of Peer-Review And An Alternative To It Science Technology & Human Values (52): 73-81

Weinberg, A. (1966) Reflections on Big Science, MIT Press, Cambridge

Operational Guidelines to Evaluation

Works cited below outline the strengths and weaknesses of peer and panel evaluation, with greater emphasis upon operational requirements.

Barker, K. (1994) “R&D Evaluation – Strengthening the impact of R&D evaluation on policy making: methodological and organisation considerations”, Research Evaluation, Vol. 21, No. 6, pages 405-413

Bobe, B. & Viala, H. (1997) “Panels of experts of the European Commission” in Callon, M., Laredo, P. & Mustar, P. (eds.) The Strategic Management of Research and Technology – Evaluation of Programmes”, Economica: Paris

Bozeman, B. & Melkers, J. (eds.) (1993) Evaluating R&D Impacts: Methods and Practice, Kluwer: Boston

Campbell, P. (2001) “Opinion Introducing a new policy for authors of research papers in Nature and Nature journals” Nature 412, 751 (2001)

Massimo, L. (1997) “Lessons from the experience of the European Commission” in Callon, M., Laredo, P. & Mustar (eds.) The Strategic Management of Research and Technology – Evaluation of Programmes”, Economica: Paris

Porter, A. & Rossini, F. (1985) Peer Review of Interdisciplinary Proposals” Science, Technology and Human Values, Vol. 10, No. 3, pages 34-42

The Economist “Truth or Consequences”, September 15 2001

3.9. FIELD STUDIES AND CASE STUDIES

Author: **Elliot Stern** (Tavistock)

3.9.1. Introduction

The distinguishing feature of field studies is that they involve the *direct observation of naturally occurring events*. This implies that, instead of investigating behaviour under artificial conditions, such as in the laboratory or under rigorously controlled condition; and instead of telescoping the process of enquiry into a 'snapshot' of time and space, such as in an interview or questionnaire, field studies involve the prolonged and relatively uninterrupted investigation of behaviours in their indigenous social setting.

As will become clear, field studies are made up of many potential methods and techniques. A common format within which these methods and techniques are assembled is that of a 'case study'. Such studies generally involve some amount of genuine *social interaction* in the field with the subjects of the study. However, this is not to suggest that such social interaction necessarily involves direct communication between evaluators and subjects.

The field and case study approach (Yin 1994, Bickman and Rog 1998) calls for the application of multiple methods of data gathering and analysis, in which a range of respondent types are involved within a number of different applications settings. It entails the collection of both quantitative and qualitative data, including surveys, content analysis, statistical analysis of secondary data and observation, and the interpretative synthesis of these different data sources to provide an overall interpretation of each case, using the generic assessment criteria illustrated in Table 22.
Within this context, the broad demarcation between field study methods (Zelditch, Morris 1962) is based on the extent to which the methods adopted are *'obtrusive'*. 'Obtrusive' field studies involve the simultaneous participation in the process of enquiry of both investigators and subjects of investigation. The main technique involved is overt *participant observation*, but within the broad category of obtrusive field studies can be included *multi-method case studies*, including interviews together with shorter time scale obtrusive field study techniques such as *protocol analysis* and *critical incidents analysis*.
Unobtrusive field studies are mainly focused on *covert observation*, where the subjects engaging in the behaviour under study are unaware they are being observed, and on the analysis of *archival and physical trace material*.

3.9.2. Application to Policy Instruments and Interventions

There is no one to one match between field studies and case studies and particular policy instruments. They could be applied to financial incentives to firms, encouraging networks among research and development actors, dissemination initiatives etc. The key determinant of appropriateness are the characteristics of the policies and programmes being evaluated and the kinds of evaluation questions being asked.
Thus field studies are characteristically associated with exploratory and descriptive modes of investigation. (See introductory chapter). The techniques are not generally conducive to the testing of causal relationships, except under some circumstances. For example, they can contribute to causal explanations where previous experimentally-determined outcomes are investigated within a 'natural' setting. It is also becoming increasingly common to undertake comparative case studies that seek to locate examples in a particular setting or look for associations between different attributes of the case study concerned.

3.9.3. Good Practice Examples

The examples in this section are drawn from a study undertaken for DG 13 which sought to assess the impact of telematic based innovation on the health care sector. Impact was variously construed to include a number of different socio-economic impacts, such as:

- Impact on national healthcare structure
- Organisational impacts (work organisation, training needs)
- Access to healthcare
- Quality and costs of healthcare

The study was constructed around a number of case studies of technological innovations in healthcare. A first issue that had to be addressed was how to select individual cases for the assessment exercise. This is a frequent issue that has to be addressed. The authors are explicit about their selection criteria:
"Case study analysis was chosen as the appropriate methodology to gain insight into this complexity because it allows the use of different data capture methods and a more interpretative data analysis approach. Thus, whilst each case was analysed in relation to a set of 'generic' criteria reflecting the main research questions addressed by TELMED, the analysis allows for contextualisation of cases to enable differences in scale - for example between an application involving single users at home and the implementation of a national infrastructure programme - to be explored.
It should be emphasised that 'assessment' in the context of case study analysis is not exclusively confined to 'impacts' assessment, but denotes a broader evaluation of the concept or world view a particular case reflects; how far this concept is consistent with what is happening in the domain, and what is likely to happen in the future, and what are the effects on the socio-cultural environment precipitated by the introduction of the innovation. The overall assessment criteria that were used to provide a comparative integration of the different cases are illustrated in table 23 below:

Table 23: Criteria for interview questionnaires

Criteria	Indicative questions
Representativeness	How far does the application represent a particular typology (size, user groups , partnership, funding arrangements, policy initiative)?
Goodness of fit	Does the application suggest new types of scenario not yet covered in the literature? Does it represent new organisational, funding or trans-national arrangements?
Policy impacts	What is the likely impact of the application on regional/national/EC healthcare policy (i.e. does it vindicate policy)?
Structural impacts	Effects on distribution of healthcare services nationally and regionally Integration of services? Displacement effects (does it shift the burden of care outside health services?) Additionality effects (does it introduce more service functions that were not there before? Does it generate a need for other support services like training, new administration ?) Redundancy (does it eliminate previous functions like paperwork, support services?)
Organisational impacts	Effects on distribution of healthcare services within and between organisations Integration of inter-departmental services? Displacement effects (does it shift the burden of care outside the organisation?) Additionality effects (does it introduce more service functions that were not there before? Does it generate a need for other support services like training, new administration ?) Redundancy (does it eliminate previous functions like paperwork, support services?)
Service provision effectiveness	Does it create improvements in service provision (e.g. reduced in-patient times; higher patient turnover; more users covered; reduced time/staff resources?) Does it improve cost-effectiveness?
User acceptance	Are users satisfied with how the applications performs? Does it create any problems for users? Does it enhance user quality of life?
Technical effectiveness	Is the technology used reliable? What problems do users have using the technology? Are there problems of interoperability?
as above	Why is the application more or less effective/acceptable? What factors contribute to effectiveness/lack of effectiveness?
Scenario development	What organisational changes need to be made to make the application more effective and/or acceptable? What technological improvements are necessary? Is the application transferable to other healthcare domains? What adaptations would be necessary? What is the likely future demand for this application?
Policy relevance	What can we learn from this application that can inform future policy?
Conceptual coherence	Is the scenario used consistent with state of the art? Is it outdated or innovative? Is the technological configuration chosen consistent with the aims and objectives of the application? Does it meet user needs?
Symbolic complexity	Is the application consonant or dissonant with the environment in which users live and work (is it appropriately contextualised?)
Action contingency	Are users engaged properly in how the application is developed and implemented? Is there appropriate interactivity between different types of actors (e.g. providers/users)?

The format illustrated in the Table above was applied to the cases in order to summarise the main features of each, and to facilitate comparison.

Specific methods included both content analysis of documents, the observation of the applications in use, and interviews with users. Various analyses were undertaken across the twenty-nine cases that were eventually selected, in terms for example of the application environments (diagnosis, medical imaging, decision support systems etc); in terms of settings (regional GP networks, single-site and multi-site) and in terms of financial arrangements (e.g. funding sources, partnership arrangements, the role of public finance etc.). On the basis of this analysis a number of scenarios were identified based on clusters of cases that represented, different ways in which healthcare telematics could be arranged. A common set of questions were then applied to these scenarios, this is represented below in Table 24:

Table 24: Questionnaire dimensions

Dimension	Criteria	Question
IMPACT ON NATIONAL HEALTHCARE STRUCTURES	Representativeness: how far does it represent 'state of the art'?	How far does the application represent a particular typology (size, user groups , partnership, funding arrangements, policy initiative)?
	Goodness of fit: how important a contribution will it make?	Does the application suggest new types of scenario not yet covered in the literature?
		Does it represent new organisational, funding or trans-national arrangements?
	Policy impacts	What is the likely impact of the application on regional/national/EC healthcare policy (i.e. does it vindicate policy)?
	Structural impacts	Effects on distribution of healthcare services nationally and regionally Integration of services? Displacement effects (does it shift the burden of care outside health services?) Additionality effects (does it introduce more service functions that were not there before?
ORGANISATIONAL IMPACTS	Organisational impacts	Effects on distribution of healthcare services within and between organisations Integration of inter-departmental services? Displacement effects (does it shift the burden of care outside the organisation?) Additionality effects (does it introduce more service functions that were not there before? Does it generate a need for other support services like training, new administration ?) Redundancy (does it eliminate previous functions like paperwork, support services?)
ACCESS	Service provision effectiveness	Does it create improvements in service provision (e.g. reduced in-patient times; higher patient turnover; more users covered; reduced time/staff resources?)
QUALITY AND COSTS	User acceptance	Are users satisfied with how the applications performs? Does it create any problems for users? Does it enhance user quality of life?
	Technical effectiveness	Is the technology used reliable? What problems do users have using the technology? Are there problems of interoperability?
	Cost-effectiveness	What are the development and operating costs? How are they linked to other benefits?
	as above	Why is the application more or less effective/acceptable? What factors contribute to effectiveness/lack of effectiveness?

Table 24 (cont.): Questionnaire dimensions

Dimension	Criteria	Question
FUTURE PROSPECTS	Scenario development	What organisational changes need to be made to make the application more effective and/or acceptable? What technological improvements are necessary? Is the application transferable to other healthcare domains? What adaptations would be necessary? What is the likely future demand for this application?
	Policy relevance	What can we learn from this application that can inform future policy?
	Conceptual coherence	Is the scenario used consistent with state of the art? Is it outdated or innovative? Is the technological configuration chosen consistent with the aims and objectives of the application? Does it meet user needs?
	Symbolic complexity	Is the application consonant or dissonant with the environment in which users live and work (is it appropriately contextualised?)
	Action contingency	Are users engaged properly in how the application is developed and implemented? Is there appropriate interactivity between different types of actors (e.g. providers/users)?

On this basis an assessment was undertaken of the impact of these innovations at a macro level. This was summarised as in Table 25 under four headings: technological innovation; institutional innovation; economic innovation; service integration.

Table 25: Technological Innovation; Institutional Innovation; Economic Innovation; Service Integration

Dimension	No. high impact projects	Case study examples	Key impact areas
Technological innovation	7	VSAT Insurrect Prescence	Satellite, ATM, Internet, SuperJanet, Videophone
Institutional innovation	7	SAVIOUR COPA	New practitioner partnerships Structure for EDI
Economic innovation	10	VITALE FSH/GP Links Medibridge	Reimbursement Subscription GP Shareholders
Service Integration	6	Medisystem Loginat/SIPR MEDSERVE TEAM	Integration of platforms (e.g. EDI/HIS/EPR) Regional service integration

3.9.4. Conditions for Application

Conditions for the application for this methodology fall under a number of headings. These variously relate to the stage at which an RTD instrument is applied; the problematics being addressed; and the skills and time available. With regard to stage of application, field studies are best suited to the exploratory analysis of innovative programmes which have not previously been analysed in depth. They are also best suited to those kinds of interventions where socio-economic impacts are understood within a complex system, rather than in terms of simple proxy measures. The skills available include a range of data capture, observational and interview skills. These will vary depending on the particular mix of field study activities undertaken. (See section 6 below). In general, field studies are time intensive at the point of data collection but less time intensive in terms of data analysis. One of the conditions therefore necessary for the application of these methods is the willingness of practitioners, technology providers, industrial partners and the like to provide access and make time available for planning - even though the data collection itself may well be unobtrusive.

3.9.5. Steps and Techniques for Implementation

A range of methods and techniques are brought together within the general heading of field studies and case studies (Miles and Huberman 1994, Bickman and Rog Op Cit.). These can include interview and survey data, but generally require observational techniques of various kinds. The main operational steps are therefore to select among the range of techniques that are appropriate for the particular evaluation task. We therefore list a number of such techniques in this section to enable choices to be made:

3.9.5.1. Unstructured Observation

Frequently labelled by its exponents as the 'naturalistic' technique par excellence and by its critics as the most unscientific and unrigorous, unstructured observation has been likened to the social science equivalent of zoological studies involving the unobtrusive recording of animal behaviour in natural settings. The technique typically involves the retrospective application of analysis to recorded events, for example the classification and frequency counting of meaningful incidents.

Recall of past events through questionnaires and interviews is often unreliable. As we have discussed earlier, such techniques are prone to problems of interviewer bias, respondent bias and respondent compliance. Direct observation is probably more reliable than recall measurements, because it places fewer demands on the respondent's category systems. Observation avoids placing the observer and respondent in the artificial context of the laboratory or experimental situation and provides opportunities to examine 'natural' behaviours.

The disadvantages of unstructured observation relate firstly to the great demands based on observer's skill and judgements. It is also inherently subjective in terms of the interpretations placed on events by the observer. Typically involves time scales of relatively long duration, with a limited number of subjects inhibiting the degree of generalisability of results to larger populations.

3.9.5.2. Structured Observation

In contrast to unstructured observation, which implies a 'see what happens' role for data collection, structured observation involves the selection of recording and classification of events according to some predetermined structure. The main categories of behaviour normally of interest are non-verbal, spatial, extra-linguistic and linguistic forms (Weich, A & Pope, L. 1988).
Linguistic measures used in structured observation techniques have focused on Bales' Interaction Process Analysis Scheme, a modified version of which has been used to study leadership roles, group performance characteristics, effects of group setting on performance. Another set of instruments, the Encounter Group Observation measures, provides a means of analysing group processes in terms of 'transitions', 'episodes' and 'interludes'. For a definitive list of structured observational instruments see Simon, A and Boyer E.G. (1974).
Extralinguistic measures cover stylistic aspects of linguistic communication including loudness, timbre, rates of speaking and rhythm. A typical indicator is Mahl, G. (1957) Speech Disturbance ratio, which identifies differences between social, situational and personality anxiety and could be applied to the identification of interviewing variables in the assessment of user response to learning technology application. Non-verbal observation measures have focused on identifying and interpreting body movements and facial expression. Hall, Edward T. (1963) shows how proxemic measures can be used to analyse spatial behaviours, such as the way in which the arrangement of desks, chairs and other artefacts within settings might influence learning behaviour within groups.
Structural observation provides greater opportunity to test pre-determined hypotheses than unstructured observation techniques. It is relatively precise in specifying the parameters of what is legitimate data to be recorded and what should be ignored, leading to improvements in efficiency in terms of data collection and resources expended. Structured observational techniques place fewer inferential demands on the observer, and similarly fewer demands on the subject, since subjects may have no symbolic language for communicating their behaviours to the observer.
The technique requires sensitive apparatus or highly trained observers. The techniques also tend to underplay the complexity of interaction, particularly within the context of cross-cultural differences in subjects where meanings attributed to one cultural group may be inaccurate when applied to another group. Traditionally concerned with sequential interaction whilst ignoring simultaneously occurring interaction.

3.9.5.3. Protocol and Critical Incidents Analysis

Over the past decade much attention has been directed to developing procedures that can be used to extract appropriate information for the development of artificial intelligence and expert systems. The 'knowledge acquisition' processes have utilised a range of 'conventional' data capture techniques including interviews, simple observation and multi-dimensional scaling, but one of the most commonly used and productive techniques has been protocol analysis.
This involves the recording and analysis of the behaviours engaged in by experts when performing tasks. In addition to recording what happens during task performance, for, example by using video, experts are prompted to 'think aloud' and describe the protocols they are using to solve problems.
Variations on the technique involve 'interruption analysis', where the observer stops the expert when a particularly complex or unintelligible protocol is identified, in order that it can be unpicked and analysed. Other elaborations on protocol analysis involve the use of 'recall trees', networks and repertory grids, where the experts are prompted to depict

diagrammatically the associations between the different objects and attributes of knowledge they use to solve problems.

Recent software-driven acquisition system incorporating protocol analysis include KRIMB Cox L.A. Jr, and Blumenthal, R.(1987), an intelligent expert prompt system which models the decision process; AQUINAS Boose, J.H. and Bradshaw J.M. (1986) which uses repertory grids; and KADS Anjewierden A, Wielemaker J. and Toussaint C. (1988) which incorporates a life-cycle model within the knowledge acquisition process.

Critical incidents are exceptional circumstances which may not be representative of habitual events and behaviours. They tend to encompass a wide spectrum of variations in substance and content and are thus difficult to measure against 'standard' yardsticks.

Protocols analysis is well suited where evaluators want to investigate the decisions underlying strategic design choices, and to investigate ways in which technologies are utilised, that is descriptive and explanatory research rationales although the techniques are used to test hypotheses.

Critical incidents analysis involves identifying and analysing the factors associated with the events or behaviours which have a significant effect on some outcome.

In the development of a technological application, for example, it is frequently possible to identify a pivotal point in the development process from which the ultimate success or failure of the application can be traced. Similarly, the success or failure of a trainee within a learning situation might be attributable to specific moments or incidents occurring within the duration of the learning process.

Critical incidents analysis isolates these pivotal moments and assesses the criteria explaining the incident and its ramifications through the perception of action involved. Observers may be external evaluators or involved actors and the analysis can be retrospectively applied or based on immediate recall.

Similar to unstructured observation in providing an opportunity to capture data or behaviour generated in the settings in which they naturally occur, with the added advantage of providing the language to describe such behaviours.

Protocol analysis may reflect the world view of particular experts or individual actors, and there is no guarantee that the apparent structure of response reflects the underlying structure. The techniques are also inadequate for some processes for which there is no natural verbalisation: for example perceptual motor tasks.

3.9.5.4. Physical Trace Analysis

Closely associated with archival record content analysis, this unobtrusive technique is broadly sub-divided into: erosion measures (signs of wearing) and accretion measures (signs of material deposits).

These techniques could conceivably be utilised to assess the relative frequency and nature of use of learning materials, since they are best suited to measuring incidence, frequency and attendance. They have possibly more utility within the context of the ergonomics of applications, for example, recording the rates of erosion and wear and tear on hardware and software.

As with content analysis, the main advantage of physical trace analysis is its resistance to reactive measurement effects.

Physical trace analysis is similar to content analysis in terms of difficulties in constructing analytical constructs interpreting outputs.

3.9.5.5. Participant Observation

The last two techniques discussed within the 'field studies' category of method are examples of multi-technique approaches.

In contrast to both structured and unstructured observation, which emphasises the need to minimise the active involvement of the observer in the events to be studied, there is a degree of involvement and interaction with the subjects of investigation over a relatively long time scale. This involvement at one extreme can be tantamount to action research. In his classic study of low income groups in Columbia, Liebow, Elliott (1967) not only documented the social life of the subjects of the research, but was involved in legal hearings, employment support and other aspects of his subjects' social life.
The research rationale of participant observation is described by Zelditch, Morris (Op. Cit.) in terms of three broad classes of information: incidents and histories; distributions and frequencies; generally known rules and status's. Typically, participant observation relies heavily on the use of *informants*, that is anyone who is knowledgeable about the participants to be observed. Informants can help to specify times, settings and actors that can provide richer information than might otherwise be available.
Associated with the use of informants is the adoption of *snowball sampling*, which involves obtaining suggestions for other participants from those already observed. Conversely, deviant sampling involves selecting subjects who do not fit patterns that have emerged through prior observations.
As with other forms of observation, the setting up of analytical categories and operationalisations, together with coding strategies, for interpretation, is not conducive to easy sets of rules. Most participant studies involve *dynamic hypothesis formulation*, which essentially means the iterative revision and reworking of different lines of reasoning exploration.
Observations can be statistically analysed in terms of frequency of occurrence and correlation of meaningful events, as with other forms of observation.
Participant observation is more flexible and adaptable to the exploration and reformulation of emerging theoretical preoccupations, particularly in areas which are characterised by uncertainty and innovation. This implies less commitment to an established 'world view' and existing mis-conceptualisations by the observer and a greater probability that 'rich' and innovative findings may emerge.

3.9.6. Data Requirements

As suggested in the previous sections, different kinds of data will be required for different types of methods that might be applied within a field study.

3.9.7. Scope and Limits

Field studies and case studies are appropriate when one is seeking to observe socio-economic impacts in defined settings under naturalistic conditions. It would nonetheless be possible to apply this range of methods to extensive areas of impact, provided a number of different field studies and case studies were undertaken. They are best applied to exploratory and descriptive means of investigation and are less useful for testing causal relationships. Rather, they should be seen as a means of understanding how particular contexts affect and shape impacts in different settings.
Finally, we would consider that many forms of field study would be suitable in innovative settings where pre-existing, well theorised understandings are not available.

3.9.8. Bibliography

The following references will help the interested reader go further with field studies and case studies. Those marked * are useful overviews that should be the first priority for the non-specialist reader.

A. Anjewierden, J. Wielemaker, and C. Toussaint. Shelley - computer aided knowledge engineering. Knowledge Acquisition Journal, 4(1), 1992. Special issue: "The KADS approach to knowledge engineering".

Bickman, Leonard & Rog, Debra J (1998) Handbook of applied social research methods. Thousand Oaks, CA: SAGE

Boose, J.H. and Bradshaw, J.M. (1987). Expertise transfer and complex problems: using AQUINAS as a knowledge acquisition workbench for knowledge-based systems. International Journal of Man-Machine Studies 26 3-28.

* Hall, Edward T. 1963. A System for the Notation of Proxemic Behavior, American Anthropologist 65, 1003-1026.

L.A. Cox, Jr. and R. Blumenthal, 1987. KRIMB: An Intelligent Knowledge Program for Interactive Model Building, Technical Report, U.S. West Advanced Technologies, Englewood, CO

Liebow, Elliott. 1967. Tally's Corner: A study of Negro street corner men. Boston: Little, Brown and Co.

Mahl, G. 1957. Disturbances and silences in the patient's speech in psychotherapy. Journal of Abnormal and Social Psychology, 42, 3-32

* Miles, Matthew B. and Huberman, Michael A. (1994) Qualitative data analysis: an expanded sourcebook (2nd ed). Thousand Oaks, CA: SAGE

* Simon, A. and Boyer, E.G. 1974. Mirrors for behavior III: An anthology of observation instruments. Philadelphia: Research for Better Schools.

Weich, A. & Pope, L. (1988, January). Knowing what's best: A new look at self determination. Social Casework, 69, 10-16

Yin, R. K. (1994) Case study research: Design and methods (2nd ed.) Thousand Oaks, CA: SAGE

* Zelditch, Morris (1962). Some methodological problems of field studies. American Journal of Sociology, Vol. 67, No. 5: 566-576.

3.10. NETWORK ANALYSIS

Author: **Susanne Bührer** (Fraunhofer ISI)

3.10.1. Introduction

In the area of research, technology and innovation policy, the promotion of "innovation *networks"*, "competence *centres"* "competence *networks*“ etc. is increasingly being discussed (see Boekholt et al. 1998). These initiatives have in common that not (only) a narrowly defined research or technological goal should be achieved, but also that an impetus towards a *structural and behavioural change* in the participating institutions should be provided. The following contribution shows which instruments evaluation research has at its disposal to cope with this complex policy measure and, based on one concrete example, discusses the strengths and weaknesses of the applied method.

In the social sciences a special research focus – social network research – has emerged, which serves to analyse the structure of (co-operation) relationships and the consequences for actors' decisions on actions. The premise of this approach is to be able to formulate explanations for the actions/activities of individuals by describing and analysing their social embeddedness, i.e. individual actions are neither attributed to the normative convictions of the actor nor to the mere membership of a certain category, such as e.g. age groups, but to the individually structured relationships between the actors (cf. Laumann 1979, Milardo 1988, Marsden 1992). Another approach is the „Actor-Network-Theory“, a concept to explain scientific and technological innovations developed by Callon and Latour (Callon/Latour 1992, Callon 1986, 1987). The central thesis is that science and technology development is the result of the connection of heterogeneous components into networks. This process is as more successful as more the components involved are able to act in a way taking the perspectives of the other components into account.

Social networks can be regarded from various perspectives. A central difference is the question whether the relationships among several units is being considered (*total network*) or a network is being examined from the *perspective of a single actor*: in this case, we are dealing with ego-centred or personal networks (cf. Fisher 1977, Craven/Wellman 1973, Kaufmann et al. 1989). A further differentiation can be made as to whether partial networks are being studied, i.e. those which are restricted to a certain type of relationship (e.g. relationships with friends), or total networks, which comprise all possible relationships within the total network (Pappi 1987) [61].

In social network research it is customary to differentiate between the structural or morphological criteria (of the total network) and content characteristics (the dyadic relationships). Although almost every author who has written an overview of the characteristics of social networks has utilised different systematics, the most significant network features can however be resumed as follows (cf. Hall/Wellman 1985, 28):

- *Characteristics of relationships*: [62] (1) strength of the bond, (2) frequency of interaction [63]; (3) multiplexity (number of resources which are exchanged by two

[61] Network studies do not only record the existing relationships, more interesting are often the potentially possible, factually non-existent relationships (‘structural holes’) (Knoke/Kulinski 1982: 12). Relational analyses must also be differentiated, in which the kind of transaction as well as density, cliques and clusters of the total network are investigated, and positional analyses, in which the relationship of the actors to each other is examined, for example, questions about the structural equivalence and alterations in stable relationships in formal organisations (Kaufmann et al. 1989: 15ff.; Pappi 1987: 18ff.).

[62] Larson/Bradney (1988, 109) differentiate in addition interactions (specific exchange at a certain point in time) and relationships (relationships existing beyond the moment).

[63] The measurement of the frequency of contact was repeatedly criticised, as it says nothing about the content or type of relationship (Allan 1979, 10; cf. Allan 1989, 93, Jackson/ Fischer/ McCalliser-Jones 1977, [47]; Grieco 1987, 42; Mitchell 1969, 28).

network members in the relationship) [64]; (4) duration of the relationship [65]; (5) symmetry and reciprocity of the exchange; (6) intimacy (emotional attachment).

- *Characteristics of networks*: (1) size or scope of the network; (2) density; (3) extent to which a network member is directly connected with others; (4) demarcation (share of all bonds of network members which are contained within the network); (5) availability (average number of attachments which are necessary to connect the network members as a couple); (6) homogeneity (extent to which the network members possess similar personal characteristics); (7) cliques (network areas in which all members are directly connected); (8) clusters (network areas with high density, but less stringently defined connection criteria than for cliques); (9) components (network areas with which all members are directly or indirectly connected).

3.10.2. Application to Policy Instruments and Interventions

In research and technology policy programmes, *co-operation and communication processes* are gaining ever increasing importance. The reason is that classical programme promotion is being abandoned in favour of comprehensive attempts to influence whole innovation systems by means of policy measures. An outstanding example in this context is the promotion of competence networks, respectively competence centres, which often emerge alongside certain technologies (biotechnology, medical technology, nanotechnology). The promotion of innovation networks to encourage business start-ups should also be mentioned in this context.

Why do policy measures target network dynamics? For several years, scientists have been emphasising the increasing importance of networks as a mechanism to improve knowledge and technology transfer (see Freeman 1991, Lundvall 1992, Metcalfe 1996). Due to the increased complexity and dynamics of the innovation processes in a globalised economy, the capability of institutions to absorb and transfer knowledge and their overall learning capacities are seen as crucial success factors (Foray/Lundvall 1996). Especially inter-organisational co-operation is becoming more and more important in order to cope with the challenges of modern innovation systems.

The following section describes an example of a network analysis in the context of clinical research centres in German university hospitals. This example is less typical in the sense of stimulating innovation processes, but delivers some useful insights into the advantages of such an approach to improve policy measures.

3.10.3. Good Practice Examples

Communication and Co-operation in the Interdisciplinary Centres for Clinical Research

Since 1996 the Federal Ministry for Education and Research (BMBF) has been supporting eight model centres for interdisciplinary clinical research at university hospitals (called IZKF in the following text). This promotional measure was monitored evaluatively from the beginning by the Fraunhofer Institute for Systems and Innovation

[64] In the literature on the subject the assumption is to be found that the degree of dependency increases with the *multiplexity*, i.e. a multiplicity of relational contents. Moreover, it is presumed that multiplex relationships are particularly intensive, trusting ties. Uniplex relationships on the other hand are used more for instrumental and material aid (Jackson/Fisher/McCalliser-Jones 1977, 41; cf. Boissevain 1978, 32).

[65] The duration serves often as a measurement for joint experience and is significant above all in the area of emotional support (Jackson/Fischer/McCalliser-Jones 1977, 46). According to Boissevain (1978, 34) it is also better suited to determine the investment of an actor in the relationship than the frequency of contact.

Research, Karlsruhe. The decision to examine the communication and co-operation structures within the eight centres in the programme in-depth, stems from the catalogue of objectives of the promotional measure itself, where the *encouragement of junior staff* and the establishment of efficient structures for clinical research on an *inter-disciplinary* plane, respectively the *intensification of interdisciplinarity* and scientific quality assume a prominent role[66]. As the encouragement of junior staff and the extent of interdisciplinary co-operation are reflected in daily work, an analysis of the co-operation and communication forms and mechanisms can contribute to answering the question, to what extent the programme participants already meet these objectives.
The communication and co-operation structures in the IZKF were determined and interpreted via a *survey of total networks*. Based on the possibility of a comparison among the eight centres, such an approach was promising, as taking into account the differing framework conditions of each centre, it was possible to identify influential factors for successful communication and co-operation. The following topics were examined in detail:

- How successful is the integration and socialisation of the junior scientists? Are the up-and-coming scientists sufficiently integrated in the centre structures?
- How are the measures implemented to encourage communication and co-operation judged? How are the premises, the existence of central labs etc. assessed?
- Which formal and informal communication mechanisms could be ascertained? Do the researchers from all levels of the hierarchy participate? Are there differences regarding social integration between the hierarchy levels?
- How does the interdisciplinary discourse take place in practice? How often does a co-operation with external persons (including industry) take place?

The co-operation and communication analysis was based on a *written survey* of all persons identified as members of the clinical centres (total survey). The questionnaire consisted of three groups of questions:
Group A was designed to ascertain from the questionees an assessment of the efficacy of different measures to promote communication and co-operation (colloquiums, seminars) and structures (central lab space, spatial proximity).
Group C inquired after personal data (highest scientific degree, professional position, age, research orientation, scientific discipline, self-definition as the up-and-coming scientific generation, length of activity in the IZKF, membership in committees) as well as the question of general job satisfaction in the IZKF.
Group B formed the core of the questionnaire. It consisted of the following parts:

- A list of all employees of the IZKF was supplied; this list could be added to by the questionees if required; the names of the employees were provided with a code for anonymity, which – with the corresponding "choice" of person - should be entered into the questionnaire.
- All in all, four *network generators* were used [67], in order to survey four different kinds of networks: the co-operation network, the network to pass on information, the network to receive information and the "sympathetic" network.

[66] A further goal refers to the increase of transparency and performance orientation in the financing of research projects (cf. Bührer et al. 2001).

[67] The wording of the question was: (1) "With whom do you cooperate in your project work in the Clinical Centre?"; (2) "Are there other persons to whom you pass on important professional information – also outside the Centre?"; (3) "Are there persons – besides the above named - from whom you receive important work-related information – also outside the Centre?"; (4) "Are there colleagues to whom your relations are primarily of a "social nature" (spending time together in the breaks from work or in leisure time) and who have not yet been named?"

- A number of *name interpretators* were added, i.e. questions which were designed to elicit information about the characteristics of the persons named in the first question. The questionnaire took on the form of a matrix. In detail, the question was posed how long the questionee had known the person named, which highest academic degree and professional status this person has, which research orientation, scientific discipline and faculty he/she belongs to and whether the person is regarded as influential in certain pre-determined areas.
- In addition, the questionee was asked to describe the type of relationship to the named partners in more detail; this should encompass the criteria origin of the contact, duration of the relationship, frequency of contact, subjects of discussion and perceived quality of the relationship.

A total of 713 persons were addressed, of which 270 responded. This corresponds to a response rate of 36% (cf. in the following Bührer/ Peter 1999). Information was given on a total of 2,390 network members, whereby the average number was nine persons. The largest share fell - not least because of the formulation in the questionnaire – to the co-operation network with 1,508 citations in all.

One of the results of the study was that the *satisfaction or identification* with the job in the centres is *high*: 83% of those questioned stated that they would recommend a job in the IZKF further. In-depth analyses have made clear that the *quality of social integration* (measured according to the indicators duration of the relationship, frequency of interaction, quality of the relationship, number of subjects of discussions) exercised hardly any influence on the identification. An important factor however consists in the promotion of *independece of the junior scientists*, i.e. their possibility to head independent projects. Moreover, the junior scientists asked (both post-docs and post-grads) displayed a higher degree of satisfaction than those with a habilitation thesis.

A core question of the survey was, to which extent the promotional measure led to *co-operation not only on an inter-disciplinary but also cross-hierarchy level*. Regarding this last point, it could be demonstrated that habilitated scientists tend to name "their own kind" as co-operation partners. Scientists with doctorates and graduates also cited habilitated scientists more frequently as members of their network than members of their own group, although the networks of these two groups were more widely diversified. The higher the level in the hierarchy, the greater the tendency to demarcate, i.e. the higher those questioned are on the hierarchy ladder, the more inclined they are to confine their "social circle" to the own group or to perceive them as possible co-operation partners.

The analysis of the cross-disciplinary co-operation produced results which must give cause for concern in the sense of the promotional programme. For example, *over two thirds* of the networks of the clinicians questioned were comprised of members of clinical research orientations. This applies for all centres, i.e. the two-thirds share represents the *minimum*. The networks of the natural scientists as well as the pre-clinicians were clearly inter-disciplinarily structured, even if here too the majority co-operated with members of the same research field.

The communication and co-operation analysis has two structural types – identified "old" and "new" networks, which are significant for the further evaluation. "Old" networks are characterised by actors who are older and have a higher professional status. This group of persons is characterised by less frequent interactions, spatial proximity is regarded as less important for initiating and cultivating communication and co-operation relationships. In contrast, "new" networks are characterised by a multiplicity of younger staff, who as a rule have graduated, received doctorates or have just recently habilitated. For the first two groups in particular frequent interactions and spatial proximity play an important role.

What does this mean for the success of the promotional measure, i.e. the creation of new, interdisciplinary research structures which are able to integrate the younger generation of scientists appropriately? Old networks, for which spatial proximity[68] is less important and which are often characterised by less frequent interactions, have far less need of measures to intensify co-operation and communication than new networks. The explanation is obvious: often, people in new networks do not yet have relevant own contacts and have to build these up first. The spatial proximity is the most obvious possibility for these persons to make contacts. Additionally, measures to intensify communication – also a first step towards possible co-operation - encounter a clearly higher need in these persons and respectively a greater willingness to use them. If there is no spatial proximity, then the measures promoting communication are of far greater importance. A centre with new networks must therefore attach considerable importance to the communication aspects.

Central results have also emerged regarding *interdisciplinarity*. That interdisciplinarity is a result of the quality of communications, respectively of the size and duration of exchange relationships, can be clearly confirmed by the results obtained. In centres with old networks it was seen that longstanding communicative relations can even, under certain circumstances, compensate for unfavourable research frame conditions, i.e. for example the lack of central "meeting places", in the form of central labs. For the new networks, spatial proximity is the outstanding criterion for establishing new networks and – where given – is encouraging. As the lack of sufficiently large, established and diversified networks, which are characteristic for younger people, also reflects on the interdisciplinarity, special efforts are necessary in order to anchor the targets of the promotional measure sufficiently: one way to do this is to integrate the younger generation of scientists better in the "established" networks of experienced colleagues.

3.10.4. Conditions for Application

As mentioned above, the conduct of a network analysis entails great effort on both sides – the evaluators and those to be evaluated. Before starting the survey, it is necessary to know precisely the universe, i.e. the members of the networks – not only the formal ones, but especially those who are involved in the daily network activities. In a second step, different pre-tests are necessary – not only to test the appropriateness of the questions, but also to enhance the acceptance for this kind of investigation. Finally, the results of the quantitative survey should be complemented by qualitative interviews with some selected network members. It is also possible to organise workshops with the members in order to validate the results and discuss ways to optimise the network structure.

To sum up: network analysis is not only a measure to find out certain parameters of the network characteristics like density, centrality etc., but it is also a tool to promote the communication within the networks. This is why network analysis should be used in the context of intermediate evaluations. Our experience in different contexts shows that using network analysis as a learning medium to improve network communication is highly appreciated even by those to be evaluated.

[68] "Spatial proximity" refers in the context of the IZKF primarily to the question whether the individual institutes and departments of a university hospital are mainly centrally located, or whether they are spread over large grounds or even over the whole town.

3.10.5. Steps for Implementation

Generally, in a network analysis we differentiate between three forms of social networks: (1) exchange networks, (2) information networks, (3) sympathy networks[69]. Under exchange networks is understood a multiplicity of interactions, which range from networks of friends over neighbourly aid up to recording professional relationships. For the field of science and technology policy evaluation, exchange and information networks are especially important.

In order to survey such networks, certain sector-oriented *network generators* are utilised (cf. Schenk/ Pfenning 1993). Network generators are a listing of persons/ institutions according to a selection criterion, for example the question about the persons/ institutions with whom information is most frequently exchanged. The second step comprises gathering information about the characteristics of the persons/ institutions named in the first question *(name interpretators).* Possible characteristics are age, profession, type of institution (basic research, applied research, development etc.). In a third step the type of relationship between the questionee and the persons/ institutions named by him is determined *(form and content of the dyad).* Examples of the questions to be put in this context are duration of the relationship and frequency of contact. The last step consists in analysing the *network structure*. For this, details of the size, the density or unity, the microstructure (existence of cliques etc.) as well as the homogeneity or heterogeneity of the networks is studied.

What does this mean for the evaluation of modern research and technology policy initiatives? First of all, a decision must be taken whether to examine total networks or ego-centred networks and what type of network is the object of the study (partial network or totality of the relationships). Also, it must be decided whether the approach should rather be formalistic (e.g. position analyses or examination of "structural holes", cliques etc.), or whether the primary question is which consequences the network characteristics have for the network members' options to act.

To conclude: network analysis requires detailed data about the actors belonging to the network (e.g. their institutional background, resources) and the kind of relation-ships between these actors. Usually, the "measurement" of the network takes place once, but it is also possible to repeat the survey in order to investigate changes within the network structure. Network analysis is a quantitative approach using highly formalised questions; the data gathered, however, should be complemented by some qualitative insights through personal interviews.

3.10.6. Data Requirements

Data gained by network analysis are process data which deliver insights in "objective" characteristics of the network like density, central actors, accessibility etc. They are able to measure „impacts“ if the stimulation of innovation networks is the aim of the policy measure. It should be noted, however, that the stimulation of innovation networks is usually (only) a way to improve science and technology transfer and not the final aim. Yet it can be assumed that networks which function well are more likely to end up with the desired results (more innovative products, services etc.) than networks which function less well.

[69] Knoke/Kulinski (1982, 14ff) differentiate between following networks: (1) exchange of control over goods or means, (2) communication relationships, (3) ‘boundary penetration relations’ (overlapping memberships), (4) instrumental relationships, (5) emotional relationships, (6) authority and power relationships, (7) relationships with relatives and descendents.

3.10.7. Scope and Limits

The study of total networks by means of written surveys produces comprehensive empirical material, which can be utilised in various ways for purposes of evaluation as well as for policy improvement:

- it can point to weak spots in the communication and co-operation of network members;
- the analysis of total networks provides a multiplicity of surprising results, even for the members, which can spark off intensive debates;
- because of the high standardisation of the procedure (open questions are the exception, most questions have (given) multiple choice answers), results of the separate participant clusters can be easily compared;
- by comparing different network types it is possible to derive specific success factors;
- especially in those cases in which several networks or centres are being promoted which have been explicitly granted a free hand in organisational matters, a comparison against the background of different context conditions offers the potential to identify "best practices" and so trigger learning and optimisation processes.

These positive results are also confronted with difficulties: the main problem consists in the time involved in such a written survey, especially for the questionees: in order to engender acceptance for such a survey and to attain a response rate necessary for empirically sound analyses, it is unavoidable to conduct not only pre-tests, but also to convince the main actors (as a rule the co-ordination or business office of the promoted institutions) in advance of this form of investigation.

3.10.8. Bibliography

Evaluation Reports

Boekholt, Patries; Clark, John; Sowden, Philip, in collaboration with Jörg Niehoff (1998): An International Comparative Study on Initiatives to Build, Develop and Support „Kompetenzzentren". Technopolis.

Braun, D., Hinze, S./ Hüsing, B./ Kuhlmann, S./ Menrad, K. /Peter, V. (1997): Interdisziplinäre Zentren für Klinische Forschung in der Frühphase der Bundesförderung. Erste Ergebnisse einer Prozeßevaluation. Fraunhofer IRB Verlag. Stuttgart.

Bührer, Susanne; Peter, Viola (1999): Kommunikation und Kooperation in den interdisziplinären Zentren für klinische Forschung. Stuttgart: Fraunhofer IRB Verlag.

Bührer, Susanne; Menrad, Klaus; Peter, Viola; Braun, Dietmar; Kuhlmann, Stefan (2001): Modelle der Finanzierung interdisziplinärer Forschung an Hochschulkliniken. Stuttgart: Fraunhofer IRB Verlag. mimeo.

Scientific references

Allan, Graham A. (1979): A Sociology of Friendship and Kinship. London: George Allen & Unwin.

Allan, Graham A. (1989): Friendship. Developing a Sociological Perspective. New York, London, Toronto, Sydney, Tokyo: Harvester Wheatsheaf.

Boissevan, J.F. (1978): Friends of Friends. Networks, Manipulators and Coalitions. Oxford: Basil Blackwell. First edition 1974.

Callon, M. (1986): Some Elements of a Sociology of Translation: Domestication of the Scallops and the Fishermen of St. Brieuc Bay, in: Law 1991: 196-232

Callon, M. (1987):Society in the Making: The Study of Technology as a Tool for Sociological Analysis, in.: W.E. Bijker et al. (eds.): The Social Construction of Technological Systems. New Directions in the Sociology and History of Technology, Cambridge, Mass.: MIT Press, 82-103

Callon, M.; Latour B. (1992): Don't Throw the Baby Out with the Bath School! A Reply to Collins and Yearley, in: Pickering 1992: 343-368

Craven, Paul / Wellman, Barry (1973): The Network City. Sociological Inquiry 43: 57-88.

Fisher, Claude S. (1977): Network Analysis and Urban Studies. In: Fisher, Claude S., Robert M. Jackson, Stueve, C.A., Gerson, Kathleen and Jones, Lynn McCallister, with Mark Baldassare (ed.): Networks and Places: Social Relations in the Urban Setting, pp. 19-37. New York: The Free Press.

Foray, D.; Lundvall. B.A. (1996): The Knowledge-based Economy: From the Economics of Knowledge to the Learning Economy. In: Employment and Growth in the Knowledge-based Economy. OECD, Paris, pp. 11-32.

Freeman, C. (1991): Networks of Innovators: A Synthesis of Research Issues. In: Research Policy 20, pp. 499-514.

Grieco, Margaret S.(1987): Keeping it in the Family. Social Networks and Employment Chance. London, New York: Tavistock Publications.

Hall, Alan, Wellman, Barry. (1985): Social Networks and Social Support. In: Sheldon Cohen and S. Leonard Syme (eds.): Social Support and Health, pp. 23-41. Orlando: Academic Press.

Jackson, Robert M.; Claude S. Fisher and Lynne McCallister Jones (1977): The Dimensions of Social Networks. In: Fisher, Claude S., Robert M. Jackson, Stueve, C.A., Gerson, Kathleen and Jones, Lynn McCallister, with Mark Baldassare (Hrsg.): Networks and Places: Social Relations in the Urban Setting, pp. 39-58. New York: The Free Press.

Kaufmann, Franz-Xaver; Engelbert, Angelika; Herlth, Alois; Meier, Brigitte; Strohmeier, Klaus Peter (1989): Networkbeziehungen von Familien. Bundesinstitut für Bevölkerungsforschung, Materialien zur Bevölkerungswissenschaft. Sonderheft 17. Wiesbaden.

Knoke, David & James H. Kulinski (1982): Network Analysis. Beverly Hills / London: Sage.

Larson, Reed W.; Bradney, Nancy (1988): Precious Moments with Family Members and Friends. In: Milardo, R.M. (ed.): Families and Social Networks, pp. 107-126. Newbury Park: Sage.

Laumann, E.O. (1979): Network Analysis in Large Social Systems: Some Theoretical and Methodological Problems. In: Paul W. Holland and Samuel Leinhard (eds.): Perspectives on Social Network Research, pp. 379-402. New York: Academic Press.

Lundvall, B.A. (1992) (ed.): National Systems of Innovation: Towards a Theory of Innovation and Interactive Learning. Pinter Publishers, London.

Marsden, Peter V. (1992): Social network theory. In: Edgard F Borgotta, Marie L. Borgotta (eds.): Encyclopedia of Sociology, Vol. 4, pp. 1887-1894. New York.

Metcalfe, S. (1995): The Economic Foundation of Technology Policy: Equilibrium and Evolutionary Perspectives. In: P. Stoneman (ed.): Handbook of the Economics of Innovation and Technological Change. Blackwell, Oxford.

Mitchell, J. Clyde (1969): The Concept and Use of Social Netzworks. In: Mitchell, J. Clyde (ed.): Social Networks in Urban Situations: Analysis of Personal Relationships in Central African Towns, pp. 1-50. Manchester: Manchester Univ. Press.

Milardo, R.M. (1988): Families and Social Networks: An Overview of Theory and Methodology. In: Milardo, R.M. (ed.): Families and Social Networks, pp. 13-47. Newbury Park: Sage.

Pappi, Franz Urban (1987): Die Netzwerkanalyse aus soziologischer Sicht. In: Franz Urban Pappi (ed.): Methoden der Netzwerkanalyse, pp. 11-37. München: Oldenbourg Verlag.

Schenk, Pfenning (1993): Ego-zentrierte Netzwerke in Massenumfragen. Mannheim.

3.11. FORESIGHT AND TECHNOLOGY ASSESSMENT AS COMPLEMENTING EVALUATION TOOLS

Author: **Stefan Kuhlmann** (Fraunhofer ISI)

3.11.1. Introduction

Traditionally, the evaluation of publicly funded research, technology and innovation programmes has been conceptualised as a hindsight exercise: what direct and indirect impacts were achieved? Were the policy targets attained? Was the target group reached? etc. This kind of ex post questioning, though, provides only to a limited extent answers to *strategic questions* like: what basic technical, scientific, economic or societal problems are calling for a policy intervention? On what functional assumptions would a policy programme's concept be based? Under which conditions would a programme be "strategically efficient"? Any attempt to answer such questions raises a whole series of methodical, conceptual and empirical problems, which must be solved in each case – one solution, nevertheless, becoming a promising model in recent years, is the *amplification and integration of evaluation procedures with foresight exercises and technology assessment*. Roughly, one can describe the basic concepts of foresight and of 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).
- 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).

3.11.1.1. Foresight[70]

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)

[70] This section is taken mainly from Kuhlmann 2001, based on Kuhlmann et al. 1999.

were based on the assumption that the future is pre-defined as a linear continuation of present trends (Helmer 1966; Flechtheim 1968; Linstone 1999). Albeit these approaches largely failed due to the in-build 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, strategic intelligence can be enforced (see EPUB "Tool Box", chapter 4). 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.

3.11.1.2. Technology Assessment[71]

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.

3.11.2. Application to Policy Instruments and Interventions

Basically, it is useful to amplify evaluation procedures by combinations with foresight exercises and technology assessment if the evaluation is put in a *strategic perspective*, including the analysis of *users' and market expectations and needs*, understood as a critical frame conditions of a policy measure's potential success.

[71] This section is taken mainly from Kuhlmann 2001, based on Kuhlmann et al. 1999.

The interest of policymakers in such combinations is increasing, but there are only few examples of implemented systematic exercises yet. Thus, the following assessment of the potential use of new combinations for *different research and innovation policy instruments* is based rather on plausible consideration than on broad practical experience:

- *Financing R&D:* foresight exercises and technology assessment can help with *priority-setting* under the condition of scarce public budgets and competition for funding – evaluation might either assess funded research and innovation activities *ex post* in the light of foresight and technology assessment results by using them as a kind of benchmark, or rank envisaged funding themes *ex ante*.
- *Provision of R&D infrastructure:* foresight exercises and technology assessment can help to evaluate the actual or envisaged *priorities of research institutes*, by using the exercises' results as a benchmark (see example of Fraunhofer evaluation, below).
- *Technology Transfer/Innovation Diffusion*: foresight exercises and technology assessment can help to identify the quality and extent of the present or future demand for research results and technological developments, i.e. for the *likelihood of successful innovation*.
- *Standards, Regulations, IPRs:* foresight exercises and technology assessment can help to *characterise the need* for technical standards, regulations, and for the appropriateness of IPR regimes, in the light of identified present or future technical, social or economic risks and potentials, thus enlightening the evaluation of related policy measures.

3.11.3. Good Practice Examples

3.11.3.1. Example 1

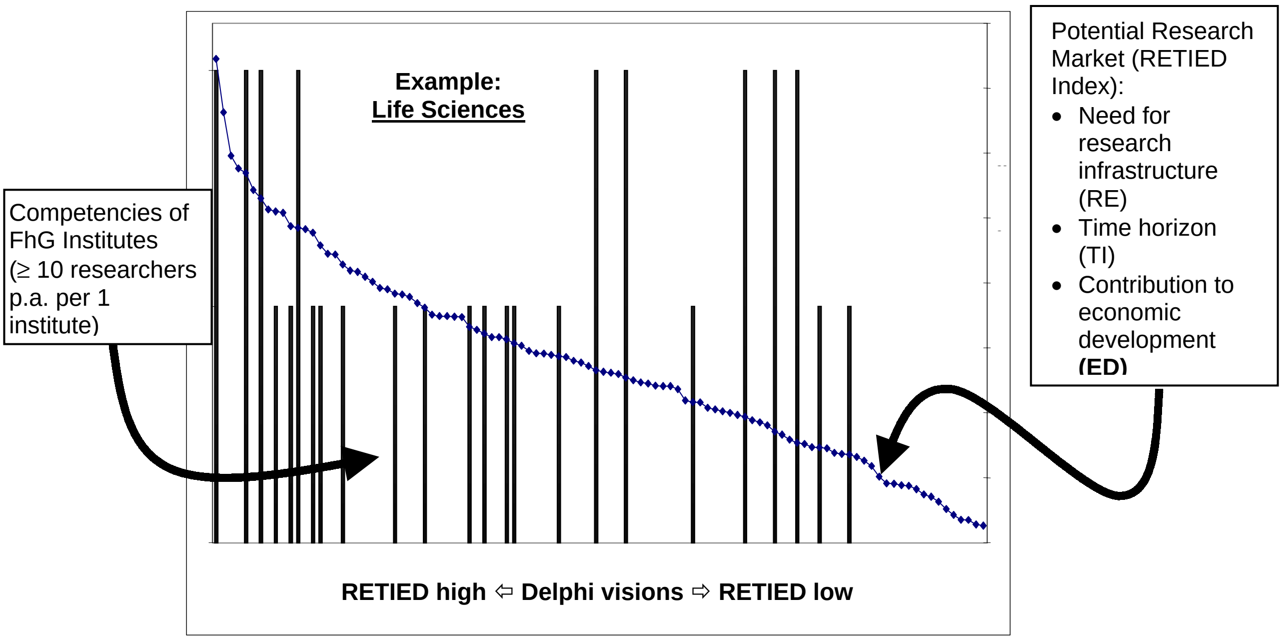
Using technology foresight results in order to evaluate a research institution enables evaluators to get an overview of the fit between perceived future developments in science and technology world-wide and the performance portfolio of a given publicly (co-) funded research organisation. By constructing 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. 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 et al. 2002) 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 6, 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 6, left hand).

Figure 6: Combining Foresight Results with Evaluation Purposes - Example: System Evaluation of the Fraunhofer Society (FhG)



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.

3.11.3.2. Example 2

Foresight methods might be further improved for the purpose of policy evaluation by combining it with technology assessment efforts. The German study *Technology at the Threshold of the 21st Century* (Grupp 1993), for example, was rather a foresight study, but indicated at the same time the relevance of extending foresight methods to technology assessment. The experts involved were assumed to have *some understanding of the potential – (non) desirable, (un-)intended – effects and impacts of new technology*. In other words, an informal technology assessment competence was required, profiting from exposure to foresight methods and experience.

3.11.4.Conditions for Application

Since the amplification of evaluation practices with foresight exercises and technology assessment is of an experimental character it is not advisable to fix a detailed set of conditions, operational steps, and data lay-outs for this methodological approach. There are, nevertheless *two basic conditions* of a useful implementation:
The envisaged evaluation procedure (programme or policy evaluation, institutional evaluation, *ex post* or *ex a*nte) must be embedded in *a strategic decision-making process*, calling for disposability of alternative perspectives.
Foresight exercises and technology assessments are costly. Their conduct just for the purpose of an evaluation would mean a considerable investment. Rather it is recommendable to *use the results of already available major foresight and assessment exercises in the context of evaluation*, the precondition of which is the willingness of policy authorities to initiate and support repeatedly the conduct of fresh exercises (e.g. regular exercises on a European scale).

3.11.5. Steps for Implementation

Addressed in the section on "Conditions for methodology application".

3.11.6. Data Requirements

Addressed in the section on "Conditions for methodology application".

3.11.7. Scope and Limits

The amplification of research and innovation policy evaluation practices with foresight exercises and technology assessment helps to broaden the scope of actual or potential, intended or non-intended impacts and effects of public interventions.
Foresight and technology assessment can jointly contribute to strategic 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 evaluation, 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.
Obviously, there are also *limits* of these methodologies. An important limitation of *foresight* is the well known fact 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 strengths and limitations 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.

Finally, since foresight and assessment are *complex combinations of methodologies in itself*, the scope of options and limitations of the hybrid combination with evaluation is inevitably huge. There are two basic limitations that should be mentioned explicitly:

The *cost of combining such complex efforts* (resources, time) are potentially high.

The *problem of the causal attribution* of potential scientific, technological, social or economic developments to a public policy measure under evaluative scrutiny – a basic problem of any policy evaluation – becomes even more an issue in this case.

3.11.8. Bibliography

On the combination of evaluation with foresight and technology assessment:

RTD Evaluation Toolbox, chapter 4: Distributed Techno-Economic Intelligence for Policymaking.

Grupp, H. (ed.) (1993): *Technologie am Beginn des 21.* Jahrhunderts, Heidelberg (Physica/Springer)

Kuhlmann, S. (2001) : *Management of Innovation Systems: The Role of Distributed Intelligence*. Apeldoorn/Antwerpen 2001 (Publisher Maklu; series "Nijmegen Lectures on Innovation Management"; edited by the Faculty of Policy Sciences, Nijmegen University)

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)

On the use and methodologies of foresight:

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)

Cuhls, K. (1998): *Technikvorausschau in Japan. Ein Rückblick auf 30 Jahre Delphi-Expertenbefragungen*, Physica Verlag, Heidelberg

Cuhls, K., Blind, K.,Grupp, H., (2002): *Innovations for our Future. Delphi '98: New Foresight on Science and Technology*, Heidelberg, New York (Physica, Springer) (series 'Technology, Innovation and Policy' by the Fraunhofer Institute for Systems and Innovation Research, ISI)

Cuhls, K., Kuwahara, T. (1994): *Outlook for Japanese and German Future Technology. Comparing Technology Forecast Surveys*, Heidelberg (Physica/Springer Pub-lishers, Series "Technology, Innovation and Policy“, Vol. 1)

Flechtheim, O. K. (1968): *Futurologie - Möglichkeiten und Grenzen*, Frankfurt/M./ Berlin

Forrester, J. W. (1971*): Der teuflische Regelkreis. Das Globalmodell der Menschheitskrise*, Stuttgart

Gordon, T. J., Helmer, O. (1964): *Report on a Long-Range Forecasting Study*, Rand Corporation, Santa Monica/ California

Grupp, H. (1998) (ed.):; Special Issue on Recent National Foresight Activities, *Technological Forecasting & Social Change,* Vol. 57

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

IPTS (2000): *The IPTS Futures Project. Synthesis Report*, Seville, European Commission : EUR 19038 EN.

Irvine, J., Martin, B. R. (1984): *Foresight in Science*, Picking the Winners, London/ Dover

Linstone, H. (1999): *Technological Forecasting and Social Change*, Vol. 62 (1999) p. 1-8

Loveridge, D. (guest ed.) (1996) 'Special Publication on Technology Assessment', *International Journal of Technology Management*; Vol. 11, Nos 5/6.

Martin, B. (1995): 'Foresight in Science and Technology', *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

Sanz, L., Cabello, C. (2000): 'Expectations and Learning as Principles of Shaping the Future'. In: N. Brown, B. Rappert and A. Webster (Eds): *Constested Futures: A Sociology of Prospective Science and Technology.* Aldershot: Ashgate

On the use and methodologies of technology assessment:

Rip, A., Misa, Th.J., Schot, J. (eds.) (1995): *Managing Technology in Society. The Approach of Constructive Technology Assessment*, London/New York (Pinter)

Schot, J., Rip, A. (1997): 'The Past and the Future of Constructive Technology Assessment', *Technological Forecasting and Social Change*, 54, 251-268

Smits R., Leyten, J., den Hertog, P. (1995): 'Technology Assessment and technology policy in Europe: new concepts, new goals, new infrastructures', *Policy Sciences*, 28, 272-299

Sundermann, K., Bröchler, S., Simonis, G. (eds.): *Handbuch Technikfolgenabschätzung*, Berlin 1999 (Sigma)

3.12. BENCHMARKING

Author: **Wolfgang Polt** (Joanneum Research)

3.12.1. Introduction

Benchmarking has been defined as a "continuous, systematic process for comparing performance of for example, organisations, functions, processes of economies, policies, or sectors of business against the 'best in the world', aiming … to exceed [these performances]"[72] or as "… a continuous process of consideration… The flows of work are monitored constantly and compared with those of the world-leading actors to gather information which helps to take steps towards the improvement of own flows of work".[73] The Cambridge Dictionary (2000) defines a benchmark as "a mark made on something like a post used as a point for measuring things by".

Common elements of most definitions include: a comparison with the best, using quantitative indicators and the conceptualisation of benchmarking as a learning process, which goes beyond mere comparisons, but aims at an understanding of the underlying processes that cause different performances. Kastrinos (2001) points to the fact that (a) benchmarks were used by craftsmen, having to apply their tacit knowledge and skilled ability to judge correctly and (b) ways of benchmarking were characteristic of particular communities who shared a concern with the development of a particular kind of mostly tacit knowledge.

While benchmarking is a practice stemming from the realm of companies, it has become a widespread practice also applied to other institutions and processes. Especially in recent years, this practice has been extended to cover also public institutions and – most recently – also policies.

Benchmarking can be carried out in a great variety of ways[74]. It can cover products and services, processes, methods, structures and whole institutions. It can be carried out internally or externally, and can be done in a co-operative or competitive setting (in a co-operative setting the exchange of primary information between the institutions comparing with each other is possible). These different objects and setting of course demand different choices of indicators and process steps of benchmarking. Despite its variety in each concrete case, benchmarking processes generally involves

10. a *planning phase* (including the identification of the object to be benchmarked, the formation of a benchmarking team, the definition of the performance measures, identification of objects to be compares against and the identification of information sources)
11. a *phase of analysis* (elaboration and interpretation of data, identification of performance gaps, analysis of the potential reasons underlying the performance gaps)
12. an *action phase* (reporting, adjustment of goals and strategies, elaboration of action plans / policies)
13. a *control and revision phase* (checking the implementation of action plans / policies, identification of deviations, feed back into next planning phase)

The design of a typical benchmarking process is provided in Figure 7.

O'Reagain and Keegan propose to apply the same benchmarking procedure to an analysis of performances of innovations systems and policies, i.e. to select areas of improvement, to identify best practices in these areas, to develop a set of indicators (benchmarks) in order to position a process analysed vis-à-vis best practices, to study the best practices processes in great detail, especially concerning the conditions under

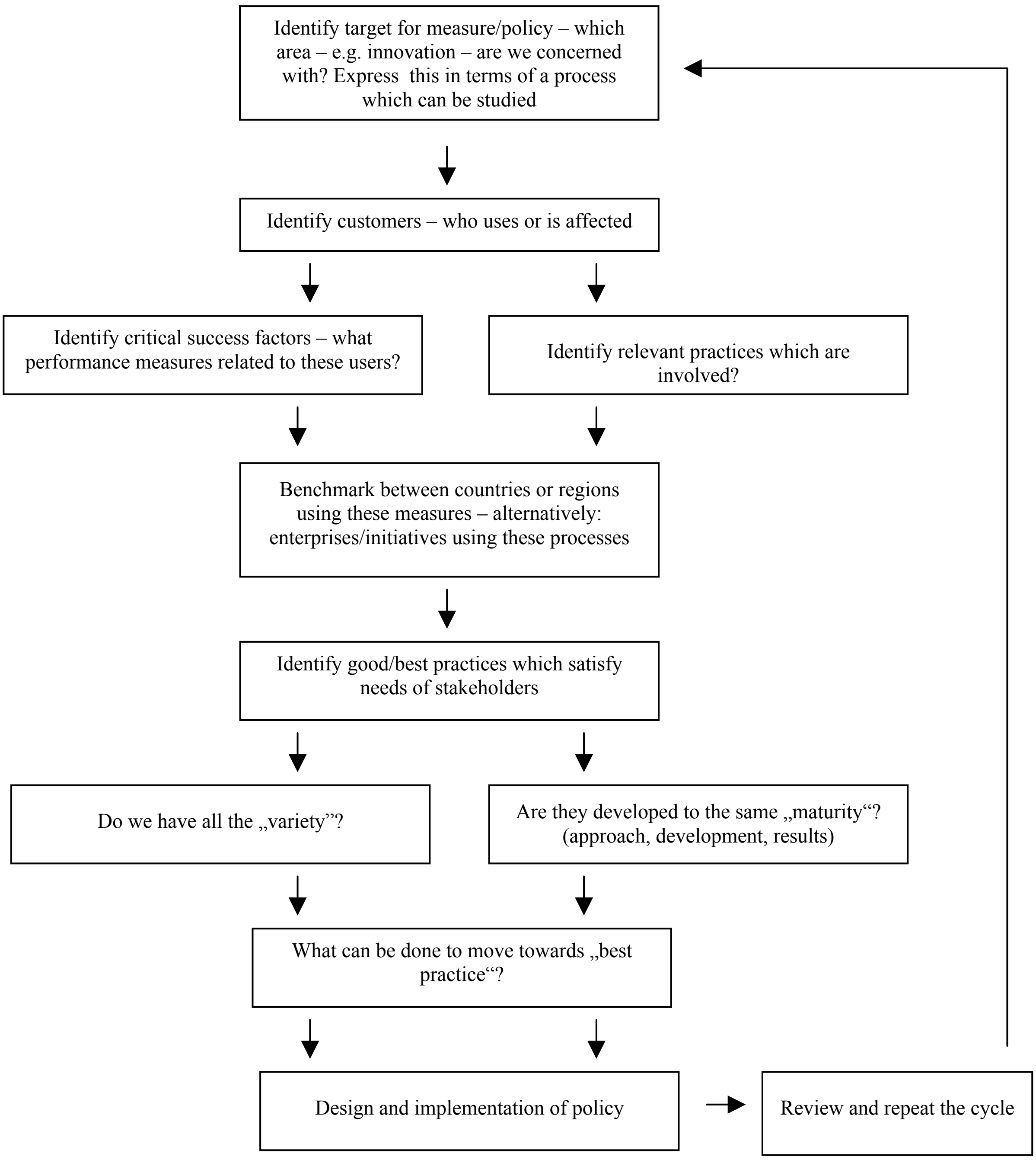
[72] European Commission, First Report by the High Level Group on Benchmarking. 1999, http://www.benchmarking-in-europe.com/rpt1hlg.pdf

[73] International Benchmarking Clearinghouse (IBC) Design Steering Committee

[74] Krech, J. , 'Benchmarking'. In: WISU 1/2001, pp 53-54

which best practice is achieved, and to derive with recommendations how to adjust framework conditions to the best practice case. These recommendations should then be used as inputs in a dialogue with concerned actors.

Figure 7: Typical steps of a benchmarking process



3.12.2. Application to Policy Instruments and Interventions

The practice of benchmarking stems from the enterprise sector. By and by it has been extended to other institutions, e.g. public authorities in the context of 'new public management approaches' (see e.g. the activities of the OECD committee on Public Management - PUMA). From the beginnings of benchmarking more narrowly defined performance characteristics of public bodies (e.g. number of clients, customer satisfaction and the like) it has been widened to cover whole institutions and – al last – policies. At present, the number of policy benchmarking initiatives is proliferating at the EU level (see especially www.benchmarking-in-europe.com for an extensive coverage of the various initiatives in all parts of the Commission as well as on the level of member states).

In science and technology policy, this development is a more recent one: at the Lisbon European Council on 23-24 March 2000 the Heads of State or Government called for the development of a 'new open method of co-ordination of RTD policies', which included establishing, "where appropriate, quantitative and qualitative indicators and benchmarks against the best in the world and tailored to the needs of different Member States and sectors as a means of comparing best practices".[75]

While international comparisons of performances in individual fields of science and technology have been undertaken already in the past, this was the first time that a benchmarking exercise in the field of RTD policies had been launched at EU level. Subsequently, the Commission services, together with the Member States have embarked on the development of indicators and a methodology that aims at establishing a benchmarking exercise on a regular basis addressing the needs of the stakeholders in RTD policies (decision makers, industry, academia, interest groups etc.).

In the first phase of this process, the applicability of the benchmarking approach to science and technology policy was a hotly debated issue in all the involved expert groups and for a of stakeholders (see below the section on the limits and scope of benchmarking). And in fact, there are limits to the application of benchmarking to policy in general and to science and technology policy in particular. These "fundamental reasons why international benchmarking related to policy design is sometimes dubious and always difficult" include:

- Policy has to take into account a number of objectives. No single, aggregate indicator could be employed to cover this diversity in goals.
- The relation between the policy intervention and the change in outcomes is not easily to be established nor understood. This is especially true in science and technology policy - a fact that has been underlined throughout the descriptions of the scope and limits of the various methods in this volume.
- The effects of policy – where they can reasonably be established – have a high degree of context dependency: there is normally not only one way to design a successful policy, but rather a few. Also, what works in the context of one innovation system, must not necessarily work in another one. Examples for such pitfalls can be derived from the lessons of other policy fields (e.g. in the attempts to emulate the US-type of labour-market flexibility in some European countries).

Compared to benchmarking on the level of whole policies, benchmarking on more disaggregate levels seem more suitable. E.g. international comparison of the performance of scientific disciplines (publications, scientific awards etc), of scientific

75 Presidency Conclusions of the European Council, 2000

and technological infrastructures (equipment at universities, large-research facilities, science parks) have been performed frequently and though not without difficulties of their own, were able to produce meaningful recommendations (see next section).

3.12.3. Good Practice Examples

As stated above, a large number of benchmarking activities of policies have been started recently in the EU. For most of the initiatives, in the realm of science and technology policy, it is too early to judge which of the exercises could reasonable be labelled as constituting good practice examples. The main reason for this being that the feed-back loop (see Figure 6) for this initiatives has yet to be run through and only benchmarking processes successfully implementing the results of the analyses can be called successful. While no such judgement seem feasible for the processes of benchmarking science and technology policies as a whole, in terms of conceptualising and setting-up the benchmarking process as well as in terms of the substantive outcomes of the analytical phase, some examples could already be given.
A recent *benchmarking of industry-science relations* carried out for DG Enterprise (see Polt et al. 2001) had the following features:

a. It combined policy makers from the member states, the European Commission, lead experts and experts from the various countries representing partly communities of concerned stakeholders from the beginning, that is already in the stage of definition of targets of the exercise and involved national experts throughout the process.
b. The process – though involving an extensive use of data, which were gathered both on the level of internationally comparable data as well as on the level – was not indicator-driven. Rather, it gathered a lot of information also on the level of qualitative information about processes and policies in the participating countries and advocated the use of
c. The diversity of the respective systems was acknowledged. No policy prescription of the one-fit-for-all –type were presented, but rather ‘good practice’ examples were identified. Such examples could also be found in countries with no overall good performance indicators for industry-science relations.

Another (emerging) good practice example is the benchmarking of research infrastructure (see Gheorgiou et al 2001). In this case, a methodology developed by the US National Science Foundation was adapted to the UK and recently to Ireland to benchmark the technical capabilities of eqiupment in selected fields. This pilot study could be transformed into an exercise on European scale, and be complemented with data from equipment suppliers to provide a first-time benchmark for the equipment of science in Europe.

The following table summarises the options available for benchmarking the provision of research infrastructures

Table 26: Approaches to International Benchmarking of research infrastructures[76]

Approach		Method	Possible Advantages	Possible disadvantages
1.	Opinion-based	Ask scientists to indicate whether they are better or worse off than other countries of foresight Delphi questions or reputational assessment	Simple and quick Suitable for Large Scale Infrastructures	Not clear who is being compared Presumes accurate knowledge of provision in other countries Not suitable for small infrastructures
2.	Case-studies	Identify comparable research groups in different countries and review inventories and equipment situation	Only true base for comparison is at research group level Suitable for Medium size and small infrastructures	Need to identify comparable teams at leading edge Leading teams may be unrepresentative of national situation Respondents may exaggerate situation
3.	Surveys to a common format	Census or large sample survey of state and condition of individual equipment items and situation of departments	Only means to understand overall national situation Suitable to Universities and other institutions managing small scale infrastructures	Relatively expensive and time-consuming Need careful definitions of equipment, costs and fields Have to factor in differences in institutional structure Need to consider additional costs such as facilities and maintenance

3.12.4. Conditions for Application

The use of benchmarking is well-established in industry, mainly as a management tool for improving productivity through comparing industrial processes and learning from those enterprises with the best performance. When it is transferred to science and technology policy with its complex relationships between inputs, processes, and outputs, many issues arise. These make it imperative for anyone, interested in using the results to improve areas such as research processes, management habits, or funding modes, to become aware of the opportunities and the manifold pitfalls and shortcomings of an indicator-based approach.

An example in case is the US Government Performance and Results Act (GPRA) from 1993, which requires all federal agencies quantitatively evaluate and report on the results of their activities annually. In a recent report, screening the implementation of the GPRA, the Committee maintains that for many institutions under scrutiny, "for research activities in science and engineering, .., especially those involving basic research, it is difficult or impossible to know the practical outcomes of activities in advance or to measure their progress annually with quantifiable metrics or milestones" (COSEPUP 2001, p.11)

When trying to derive conclusions one has to be clear about the scope and limits of comparative analysis of innovation systems and related policies (to which benchmarking is one approach).

- First, there is no single country that could be taken as a benchmark for performance. As performance can and should be measured taking into account different dimensions (and hence indicators), a number of countries appear to perform well in different indicators - though the respective profile is again different for each country. These countries may be used as a 'control group' for comparing the other countries.
- Second, even in countries where performance measures would indicate a low

[76] Based on Georghiou L, Halfpenny P and Flanagan K, Benchmarking the provision of scientific equipment, Science and Public Policy August 2001 pp303-311

performance level, there could be found good practice examples (as was the outcome of a benchmarking exercise looking into the framework conditions for Industry-Science Relations see Polt et al 2001). Thus, one should not restrict the comparison to well performing countries. Rather, one can learn a lot when looking into individual policy measures across all countries.
- Thirdly, though most of the framework conditions and policy measures address generic problems of innovations systems, what works in one country might not work in another. Good practice is specific to certain market and institutional environments, and addresses market failures and barriers stemming from these environments. To learn from good practice means first of all to learn to carefully identify these market failures and barriers, and then to select a proper mechanism to tackle them.

The learned (and learning) policy maker will not use the results of such a benchmarking exercise as a toolbox to be applied mechanically to the perceived problems in his/her country. Rather, these can be used to guide policy learning, and as an input in discussions among the actors concerned towards the establishment of a shared vision as a basis for future policy actions. Thus policy makers are well advised to improve innovation systems by taking up good practice examples, putting them into the context of their respective national innovation system, and integrating them into the broader policy context of improving the overall system.

3.12.5. Data Requirements

The role of indicators in the benchmarking process is an important and delicate one at the same time. On the one hand, it is evident that any benchmarking requires the definition of a "benchmark" and that indicators represent an important tool and basis for measuring and comparing performance. Thus, benchmarking processes have to be founded on timely, internationally comparable, and policy-relevant data.
Yet, the definition, selection and construction of indicators to be used in a benchmarking exercise is not at all a straightforward task:

- As has been stressed above, policies and public authorities – unlike enterprises - often pursue multiple goals. Therefore, a whole array of indicators rather that a single one has to be used to take into account these different goal dimensions. especially when there are trade-offs between different goals (e.g. raising competitiveness, fostering indigenous industries, environmental and societal goals), a policy excelling in the attainment of one is not necessarily the best combination of all goals.
- Innovation systems, be it on the national, regional or sectoral level differ in structure and composition. Therefore, polices addressing these systems not only have many, but also different targets. Again, this diversity would not be captured if only measured by a single, synthetic, highly aggregated performance indicator. Instead, comparisons should be undertaken at the appropriate level of aggregation, which could be on the level of specific industries, scientific disciplines, networks of enterprises, scientific team etc.
- But even if indicators are selected carefully not to 'compare apples with oranges' there are the problems of finding internationally comparable data. E.g. if one compares scientific and technological output by very traditional and well-established indicators like publications and patents, one has to be aware that definitions, concepts, ways of data collection differ largely between the countries (notably between the US, Japan and Europa). This can be exemplified by looking at

patents, which are sometimes counted as applications, as patents granted and with very different coverage (see ETAN Group on benchmarking scientific and technological productivity, 2002).

- The problem of international comparability is aggravated by the fact that for most *processes* underlying the development of performance indicators internationally comparable statistics are not available. One has to rely here very much on national expertise, ad-hoc studies and subjective judgement of the respective national community in assessing these processes. this again points to the fact that benchmarking involves a good deal of tacit knowledge and communication within a community of practitioners.

3.12.6. Scope and Limits

Benchmarking has been introduced as a management tool for comparing industrial processes and learning from those enterprises with the best performance. One has to be cautious when applying this method to policies or countries as a whole: "Countries are characterised by systemic differences and therefore what is best practice in one country or region will not be best practice in another. Therefore the more modest aim to develop 'good' and 'better' practices through 'learning by comparing' is more adequate".[77]
"Intelligent benchmarking" – as opposed to "naïve benchmarking" in which direct policy conclusions are directly drawn from the comparison of indicators - focuses on the development of "a common understanding and shared objectives which make it more meaningful to benchmark some specific aspects of the innovation system"[78] rather than comparing a set of quantitative indicators.
While meaningful measures of performance can reasonably be selected for individual processes (e.g. measures of technical efficiency), this selection gets more complicated once one arrives at higher levels of aggregation. In systems of a higher order of complexity, it becomes difficult to relate performance (however measured) to the great variety of processes and interactions between these processes that lie underneath the observed performance. This is a prominent feature of benchmarking national innovation systems and their scientific and technological productivity.
When trying to adopt benchmarking practices from the field of business to countries and policies one has to consider some requirements. A minimum requirement is that benchmarking must aim at deeper insights into the processes behind performance. Benchmarking in this sense must not stop at the quantitative comparison of different indicators. Rather, such a comparison is only the starting point for further analysis. Another requirement, especially if benchmarking is to play its role as an 'open method for policy co-ordination', is that it has to be carried out in a 'co-operative' manner. In co-operative benchmarking, the involved parties exchange first-hand information with the aim of mutually beneficial learning. In competitive benchmarking, one is often restricted to secondary sources of information and statistics. Countries often hesitate to enter benchmarking exercises if they fear to be ranked in 'league tables'. Therefore, the approach of the Commission is sensible, as "the exercise is not aimed at making rankings among Member States. Rather it will facilitate the identification and diffusion

[77] Lundvall, B.-A., Tomlinson, M., Learning-by-Comparing: Reflections on the Use and Abuse of International Benchmarking. In: Sweeny, G.: Technology, Innovation, and the Quality of Life. 2001, pp. 122

[78] Lundvall, B.-A., Tomlinson, M.,. 2001, pp. 131

of best policy and practice across the Union, while taking into account its adaptability in different national contexts."[79]

As benchmarking is intended to be a continuous learning process, the exercise should be repeated regularly. The main value of a benchmarking exercise thus arises from the dialogue with and among a broad range of actors concerned. Continuous learning processes will only be triggered by their sustained involvement.

3.12.7. Bibliography

Barré, Rémi (2001): Sense and nonsense of S&T productivity indicators, in: Science and Public Policy, Vol. 28, No. 4, , pp. 259-266, Guilford: Beech Tree Publishing. Questions the meaningfulness of frequently used indicators of scientific and technological output and the possibility to draw direct policy conclusions from the observation of differences in indicators.

Bessant, J.; Rush H. (1999): Benchmarking Framework Conditions. Paper prepared for the Benchmarking Co-ordination Office. Brighton November 1999. A methodology for benchmarking framework conditions for industry is outlined and recommendations for the design of such a benchmarking process are developed.

Committee on Science, Engineering, and Public Policy [COSEPUP] (2001): Implementing the Government Performance and Results Act for Research; A Status Report, Washington: National Academy Press. Reviewing the experiences and status-quo of the implementation of the GPRA, which demands public authorities to set quantifiable performance targets. Points to the difficulties in establishing such targets especially for institutions dealing with science and technological development

ETAN Expert Group on 'Benchmarking Scientific and Technological Productivity'. Final Report. Brussels. June 2002

European Commission: Benchmarking the Competitiveness of European Industry. COM 96 463 Final dated 9 October 1996. Brussels

European Commission: Development of an open method of co-ordination for benchmarking national research policies – Objectives, methodology and indicators. Working document from the Commission services SEC (2000) 1842. Brussels. Basic document outlining the process and objectives of the current benchmarking exercise in science and technology polices

European Commission, Benchmarking Enterprise Policy. First results from the Scoreboard. Working document from the Commission services SEC (2000) 1841. Brussels

Georghiou L, Halfpenny P and Flanagan K, Benchmarking the provision of scientific equipment, Science and Public Policy August 2001 pp303-311

Kastrinos, Nikos (2001): Introduction: Contribution of socio-economic research to the benchmarking of RTD policies in Europe, in: Science and Public Policy, Vol. 28, No. 4, pp. 238-246, Guilford: Beech Tree Publishing. Introduction and overview of a symposium on benchmarking RTD policies in Europe.

Lundvall, B.-Å., and M. Tomlinson (2001), "Policy learning through benchmarking national systems of competence building and innovation – learning by comparing. Report for the 'Advanced benchmarking Concepts (ABC) project. March 2001. Summarizes the main results from 10 different TSER projects for European benchmarking exercises from a systemic point of view.

Lundvall, B.-Å., and M. Tomlinson (2000), "Learning-by-Comparing: Reflections on the Use and Abuse of International Benchmarking", in G. Sweeny (ed.), *Innovation, Economic Progress and the Quality of Life,* Celtenham: Edward Elgar, 120-136. Critisise "naive benchmarking" which directly translates a practice from to enterprise to the public sector and to policy. Outline requirements for an "intelligent benchmarking approach" as a process of communication aiming at policy learning from comparing good policy practices involving all stakeholders in the process.

O'Reagain, Sean (2000): Benchmarking Public Policies for Competitiveness, in: Benchmarking in Europe – Working Together to Build Competitiveness', PSI Group, UK with the support of the European Commission, DG Enterprise. Overview of the proliferating approaches and initiatives to benchmarking in the EU, especially with respect to policies addressing the enterprise sector.

[79] Commission Services, Development of an open method of co-ordination for benchmarking national research policies - objectives, methodologies and indicators. Working document. SEC (2000)1842, p.4

O'Reagain, Sean; Keegan, Richard (2000): Benchmarking Explained – What is Benchmarking?, in: Benchmarking in Europe – Working Together to Build Competitiveness' Publication 2000, a magazine on benchmarking published by PSI Group, UK (with the support of the European Commission, Enterprise Directorate). Giving basic definitions of the concept, the objectives and the process of benchmarking

Polt W. et al (2001) Benchmarking Industry-Science Relations – the Role of Framework Conditions. EC-DG Enterprise, Federal Ministry of Economy and Labour, Austria, Vienna/Mannheim. Example of a recent benchmarking initiative trying to avoid the pitfalls of naïve benchmarking.Arriving at a multi-facetted picture of industry-science relatons and related policies. Find good practice examples also in countries with poor performance in aggregate indicators

Smith, Keith (2001): Comparing economic performance in the presence of diversity, in: Science and Public Policy, Vol. 28, No. 4, pp.277-284, Guilford: Beech Tree Publishing.. Points to the fact that innovation systems are divers in structure and policies are divers in objectives. Argues that there is a need to take this diversity into account in the construction and selection of appropriate benchmarking indicators.

UNIDO: Policy Benchmarking in the Developing Countries and the Economies in Transition: Principles and Practice. Vienna 1998. Gives an overview of recent policy benchmarking initiatives in developing countries. Indicates the value of co-operative benchmarking between countries. Points to the limits of benchmarking policies in the presence of divers institutions and policy goals.

4. DISTRIBUTED TECHNO-ECONOMIC INTELLIGENCE FOR POLICYMAKING

Author: **Stefan Kuhlmann** (Fraunhofer ISI)

4.1. INTRODUCTION

Who is in need of evaluations of public research, technology and innovation policies? And why, for which purpose? Meta-evaluations of evaluation practices give evidence of an increasing production of evaluative information for public policymaking in the area of research and innovation (e.g. Georghiou 1995; Kuhlmann/Holland 1995; Kuhlmann/Heinze 2001). At the same time, experience shows that both the theory and practice of evaluation has undergone important developments over the past decade. In particular, in countries where evaluation has taken root fairly early, the following *trends* can be observed:

- The major rationale for evaluations has shifted and evolved from an attempt to *legitimate past actions* and demonstrate accountability, to the need to improve understanding and *inform future initiatives.*
- Correspondingly, the issue focus of evaluations has broadened away from a narrow focus on quality, economy, efficiency and effectiveness of a programme, and towards *a more all-embracing concern* with additional issues, such as the appropriateness of a policy tool and a concern with performance improvement and strategy development.
- Approaches to evaluation have evolved 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 and mediators in learning exercises *involving all relevant stakeholders*, providing advice and recommendations as well as independent analysis.
- This has led to more *flexible and experimental concepts* of policy portfolios, and to even greater demands for well specified systems of monitoring, evaluation and benchmarking to aid analyses and feed back into strategy development.

Many evaluation exercises thus reflect an increasing concern with the link between evaluation and strategy, with an varying 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 *evaluation can inform strategy* – and quite often also in combination with benchmarking studies, technology foresight exercises, technology assessment efforts and other analytical tools. The combined use of such tools has been hallmarked *"strategic intelligence"*[80].

The present chapter discusses evaluation and neighbouring exercises as intelligence tools for research and innovation policies, in the context of 'systems of innovation' and related – often contested – arenas for policymaking. Some illustrative examples of strategic intelligence use will be sketched. Critically, also the need for a system of *"distributed intelligence"* is examined which could provide public and corporate

[80] See Kuhlmann et al., 1999: 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 developed an outline of 'Distributed Strategic Intelligence', providing a conceptual basis for the present chapter. The present text draws also on a related, more recent publication (Kuhlmann 2001a).

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 is explored. Then, the need for an effective European system of distributed intelligence is contemplated. The final section summarises the guiding principles and related requirements.

4.2. Complex Innovation Systems and the Need for Improved Strategic Intelligence

The likelihood of innovation in modern society's science, technology and industry has been shaped by national, regional or sectoral "systems of innovation": innovation systems were discovered by the social scientists (first of all by economists[81]), 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.

4.3. Policymaking in 'Systems of Innovation'

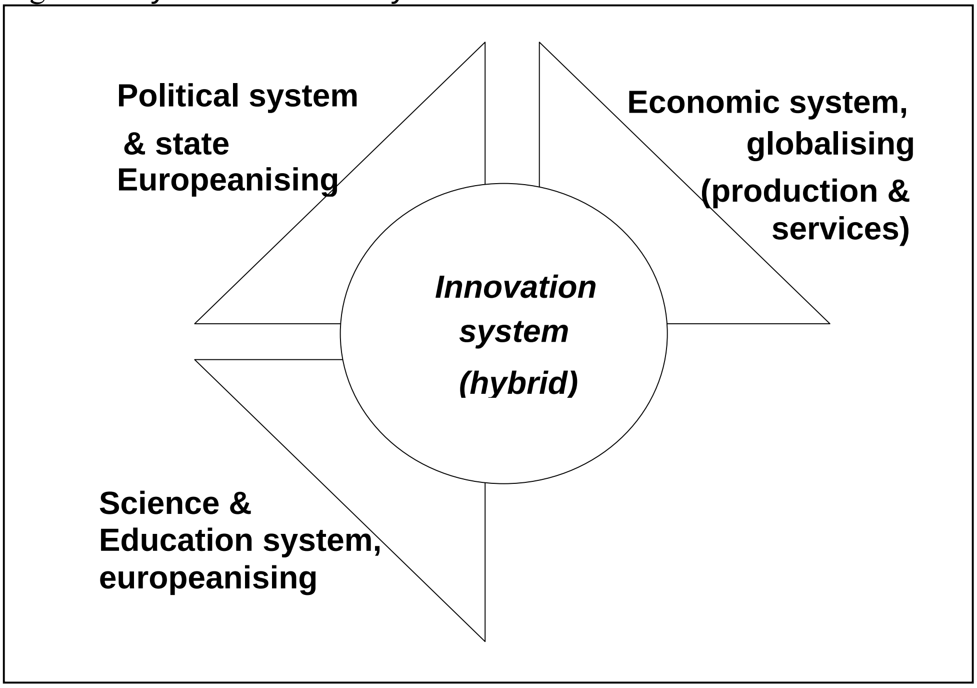
Innovation systems are encompassing, according to a meanwhile widely accepted understanding, the "biotopes" of all those institutions which are engaged in scientific research, the accumulation and diffusion of knowledge, which educate and train the working population, develop technology, produce innovative 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 and informal networks of the actors of these institutions. As "hybrid systems"[82] (see Figure 8) 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).

---

[81] 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.

[82] 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 8: Hybrid Innovation System



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. 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):

- *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.
- *The "soft side of innovation" is of growing importance* (den Hertog et al. 1997; Coombs 1999; Smits 2001). Non-technical factors such as design, human resource management, 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).
- 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.
- 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.
- 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 2001b).
- 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 "optimising production machinery" to "globally learning corporations" (Meyer-Krahmer/Reger 1999). Also, innovation managers in large multinational corporations run their strategies vis-à-vis 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 (changing) 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" research and 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 5th 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 (Georghiou/Rigby 2002; Kuhlmann/Meyer-Krahmer 2001, Airaghi et al. 1999).

Over the last two decades, considerable efforts have been made to improve the design and conduct 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).

Also, 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[83]. 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.

4.4. THE INTELLIGENCE POTENTIAL OF EVALUATION AND OTHER TOOLS

The sketched changes of the functional conditions for research and innovation have led to a growing interest in evaluation since the 1990s and have provided impetus for the application of relevant procedures (for an overview, see e.g. Shapira/Kuhlmann 2002; OECD 1997). The expectations of evaluation processes are divided between two functional poles: evaluation can in the first instance serve to *measure performance* and thus provide the legitimisation for promotional measures afterwards (*summative function*), or it can be utilised as a *"learning medium"*, in which findings about cause and effect linkings of running or completed measures can be utilised as intelligent information for current or future initiatives (*formative function*).
The *summative pole* is nurtured above all by the evaluation practice of Anglo-American countries: here in the framework of efforts to reform and cut costs in the public sector ("New Public Management") performance measurement procedures also gained great influence in research and innovation policy (Osborne/ Gaebler 1993, Shapira et al. 1997; Feller 2002). The US government and a majority of the states are increasingly implementing "performance-based management and budgeting systems" – not least in research and innovation promotion (Cozzens 2002; Cozzens/ Melkers 1997).
Promoter and promoted are under growing pressure to prove the advantages of their activities, financed as they are by tax-payers' money. This is not just because of relevant new legal requirements – such as the "Government Performance and Results Act (GPRA)" – or tight public budgets, but also because of an intensive public debate about justification, direction and advantages of public investments in research and innovation. An example of this is the "Advanced Technology Program (ATP)" – a hotly debated government programme in support of co-operative research and innovation projects between science and industry in risky high-tech areas, which aims in the long term at far-reaching diffusion effects; but the programme is always being confronted with expectations of short-term, measurable (quantifiable) impacts. Ruegg (1998) observed,

[83] See e.g. the OECD's Science, Technology and Industry Scoreboard; the European research policy benchmarking (EU Commission 2001); the "European Innovation Trendchart"at http://trendchart.cordis.lu; or the "European Benchmarking Inititatives" at http://www.benchmarking-in-europe.com.

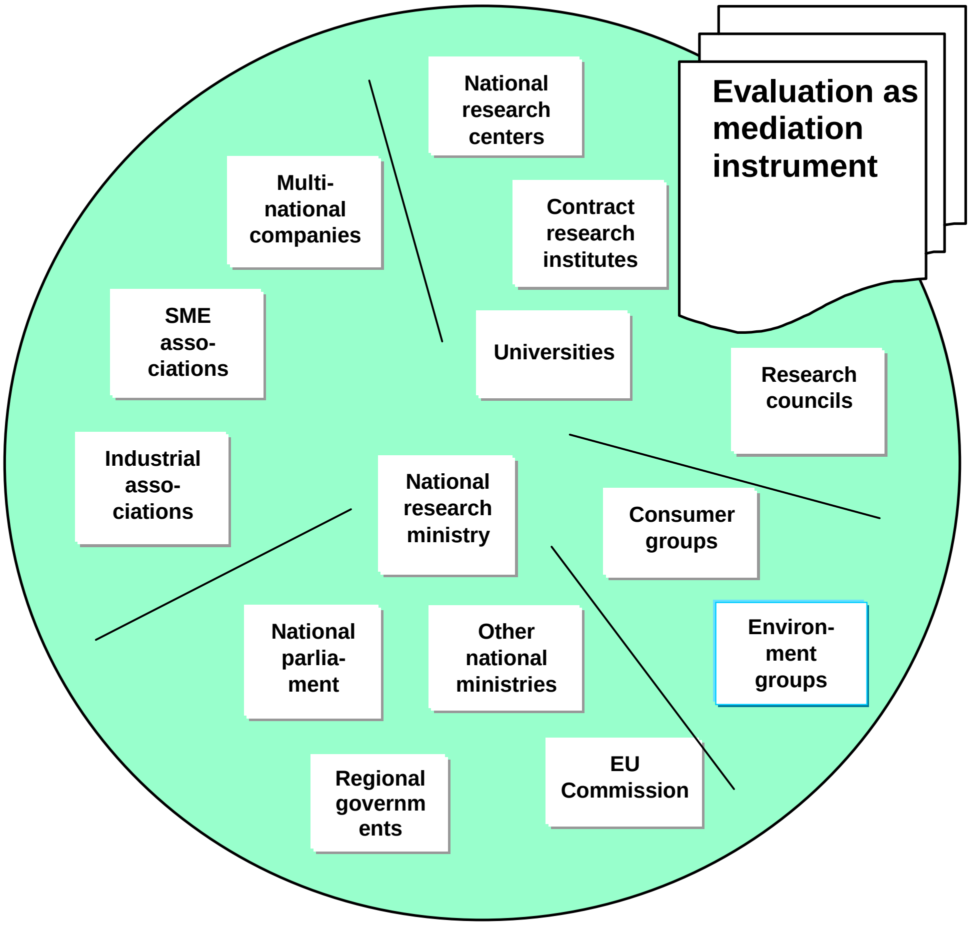
that the programme "has met nearly continuous demand for measures of impact of the program since the day it was established".

As the complexity of research and innovation policy programmes and also the tasks of institutions have grown rather than diminished, summative performance measurements soon reach their limits. *Formative, learning-oriented evaluation* approaches were therefore – partly in competition, partly as a complement to the summative – also further developed and applied. Proceeding from the lasting experience that evaluation results produce often only small impacts on political decision-making processes, and support only a few of the expectations and interests represented in a policy arena, evaluation experts (and increasingly also policymakers) tried to relax the boundaries between evaluation and decision-making processes, indeed even to partly integrate the two spheres. Guba and Lincoln (1989) have sketched the problems concisely. They see a basic problem of previous evaluation concepts in their naive science-theoretical statement, that investigation results inform "how things really are and function", as if they could not have been influenced by the evaluator, his commissioner and others.

The key concept of the new, wide understanding of evaluation is "negotiation" among the participating actors. The result of evaluations correspondingly designed, is, in contrast to conventional methodology, no longer "a set of conclusions, recommendations, or value judgements, but rather an *agenda for negotiation* of those claims, concerns, and issues that have not been resolved in the hermeneutic dialectic exchanges" (Guba/ Lincoln 1989: 13): decisions are made rather as a continuous process, in which competing actors achieve consensus interactively, or also not. Evaluation results are one piece of information among many. Here therefore the evaluation *process*, or more exactly the communicating of the participating actors in its course, takes centre stage; the process is consciously designed to be "participative" ("Participatory Evaluation"; see Patton 1997: 100; Worthen et al. 1997: 153-170). The *media* aspect of the evaluation procedure takes centre stage. Especially the following characteristics of the participatory approach can be further developed for use in research and innovation policy discussions (see also Figure 9):

- Evaluation is designed as a process of empirically, analytically prepared, *structured* presentation and confrontation of (partly conflicting) *actors' perspectives*; the whole spectrum of evaluation methods (see section 0) can be applied in this exercise.
- The evaluator acts as a "*facilitator"*, he supports the *mediation of the conflicts* in the negotiation system through actors of the policy-administrative system.
- The target of the evaluation is not only the assessment of the facts from the perspective of an individual actor (e.g. the policy-administrative system), or the "objective" examination of the appropriateness of a policy, but the stimulation of *learning processes* by breaking down rigid actor orientations.

Figure 9: Evaluation as mediation instrument in the research and innovation policy arena



These evaluation concepts aim primarily to facilitate a "re-framing" (Schön/ Rein 1994) of the orientations of corporatistic and policy-administration actors. In the context of the research and innovation system they can be regarded as an "intelligent" provider of strategies for negotiation and management, not only for the responsible political actors but also the interested public (Kuhlmann 1997). *"Intelligent" policy development processes* in this sense can moreover be enriched by combinations with (cf. Kuhlmann et al. 1999):

- "Foresight processes" ("Technology Foresight"; cf. Cuhls et al. 2002; IPTS 2000; Grupp 1998b; Cuhls/ Kuwahara 1994), with the intention of delivering debatable "visions" of more or less desirable future developments. "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).
- 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 Smits et al. 1995; Schot/Rip 1997; Loveridge 1996).

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; see also table 28):

- Though quite *some examples of integration* between the three bodies of experiences could be found in several countries, there is little 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.
- *Industry* has an *older tradition* of combining approaches when defining strategies to assess uncertain (technological) developments with potentially wide impacts, both commercial and societal.
- 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.
- 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.

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*.

Table 27: Combinations of evaluation, foresight and technology assessment to enhance decision-making

Input from: Main exercise:	**Evaluation**	**Technology Foresight**	**Technology Assessment**
Evaluation		• benchmarking to identify potential S&T developments • increase strategic dimension ex-ante evaluation • contribute to appropriateness issues • set the right context for evaluations	• knowledge on pro-cesses to assess externalities and effects on a wider set of stakeholders • make value issues more explicit (bio-technology) in evaluation
Technology Foresight	• benchmarking present capabilities with future developments for SWOT analysis • evaluation of foresight exercises to improve future use of foresight • Bring foresight closer to policy clients		• increase awareness of social issues in prospective outlooks • anticipating social barriers • articulation of public values • avoid tunnel vision (widen the technological context)
Technology Assessment	• expose problems in S&T programme due the lack of technology assessment at start • evaluation methodologies (as process and techniques causalities) can improve the technology assessment analysis on effects	• widen the technological context (avoid tunnel vision) • increase future outlook • expose strengths and weaknesses S&T infrastructure • expose user needs of various stakeholders	

source: Kuhlmann et al., 1999

4.5. 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: policy evaluation, foresight, 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 *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 emerging 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; see e.g. Bongers et al. 2000), 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:

- *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.
- *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.
- *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.
- *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, three examples of strategic intelligence for innovation policy will be discussed in the following.

4.5.1. 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 et al. 2002) 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:

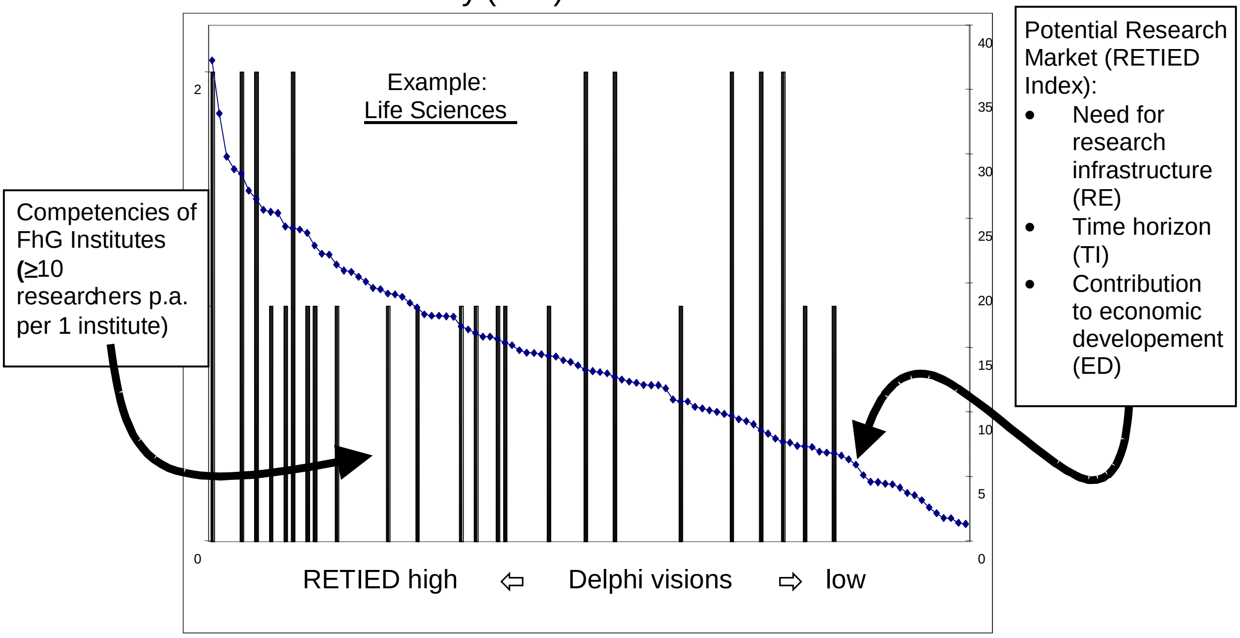
- necessity of an improvement of the research infrastructure (RE),
- time horizon of the realisation of a technological innovation (TI),
- contribution of an innovation to the economic development (ED).

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 10, 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 10). 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 10: Combining Foresight Results with Evaluation Purposes - Example: System Evaluation of the Fraunhofer Society (FhG)



- *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.
- *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.
- *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 10). 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.
- *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.

•

4.5.2. Evaluation and Foresight at ADEME[84]

The French environment and energy agency ADEME was established in 1992 as a merger of three former agencies: for waste management (ANRED), for air pollution (AQA) and for (non-nuclear) energy (AFME). ADEME's mission is broad. It reaches from the support of research (e.g. through a PhD scheme), to demonstration, education and local development projects, in all areas relating to environment and energy (except those related to water), and in virtually all economic sectors. ADEME is managed by three ministries, i.e. environment, research and industry. It receives funding directly from these ministries and through special levies and taxes, for instance on landfill.
ADEME does not conduct research itself, it supports other organisations to do so. Already in the mid-1980s with its predecessor AFME, this led to the problem how to legitimise the agency's activities, towards the public authorities. Since research does not immediately lead to tangible or measurable results, ways had to be found to account for the activities of ADEME, and to evaluate whether these were satisfactory.
Therefore, in the second half of the 1980s a collaboration was established with the Centre de Sociologie de l'Innovation (CSI). Together with AFME, CSI developed the *approach of techno-economic networks* to manage and evaluate the technological research programmes. The approach was based upon the formalisation of the action of ADEME's programme managers. It describes the agency's field of intervention in terms of actors and intermediaries, around three main poles (Science, Technology and Market) and two "transfer" poles (ST, TM). Hence the agency gave itself the task to stimulate the emergence and development of networks around energy technologies, and promote the interaction and exchange between users, developers, engineers and scientists in such networks. The evaluation of programmes became the *evaluation of the agency's capability to co-construct networks*.
The methods and concepts were taken over by the new agency in 1992. A new situation occurred: the agency moved into environmental technologies. However, the advice of outside specialists was no longer needed: the agency had made itself so familiar with the method that it easily extended it to include a new pole (Regulation) and corresponding transfer poles. The use of evaluation and its tools became routine business for the agency. Evaluation practice had been extended to other areas than those strictly linked to R&D. Also, the network approach has been adapted for use within strategy formulation.
Finally, other research organisations in France have started to adopt the techniques for their own strategy formulation.
The following lessons can be learned from this experience (see also de Laat 1996):

- *The principle of participation* was realised as an interplay between research policy practitioners and social scientists, and by the subsequent integration of the tools in the daily practice of policymaking.
- *Principle of "objectivisation"*: the tools developed have a strong theoretical basis (from innovation studies), but they have been continuously tested on real life situations, proposed to ADEME's people, nourished by new theoretical insights, transformed to account for new problems the agency is faced with.
- *Principle of mediation and alignment:* the interaction between various actors –and not a linear development whereby the policy advisors develop tools and policymakers subsequently adopt them – may explain that, after more than ten years, the approach

[84] This example is taken from Kuhlmann et al. 1999.

was fully absorbed by the agency.

- *Principle of decision support*: the strategic intelligence approach has become almost tacit knowledge in ADEME's daily practice. Today it serves as a general but nevertheless practical heuristic tool, no longer only for evaluation per se, but, far more broadly, to organise and evaluate the agency's action internally. It also helps to organise the dialogue with the relevant ministries, the beneficiaries of the agency's support and other parties involved in this field of policy.
-

4.5.3. Regional Technology Plans as Inter-Cultural Exchange of Policy Planning[85]

In 1993 the European Commission's Directorate General XVI (Regional Policy and Cohesion) launched a pilot initiative called Regional Technology Plans (RTP) which was to initiate the development of a regional policy strategy. The projects in this initiative were to be undertaken in so-called "less favoured regions" which had an Objective 1 and 2 status. 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 traditionally been the domain of national policymakers. Particular concern related to the top-down approach in regional technology policy initiatives either from centralist national authorities or inexperienced regional authorities.

What the Commission offered was a policy-planning model which included both an indication of the contents and a structure for the 18 months long 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. Seven regions entered the experimental action and went through what was to become an ongoing S&T policy-planning process. The Commission played a "mentor role" in the background, the regions themselves were responsible for running the RTP projects.

One of these regions was Limburg in the Netherlands. Prior to the RTP, research and innovation policy did not have high priority in the Province and was dealt with as a side line of mainstream economic policy. There was no explicit strategy and subsidies went haphazardly to the main innovation support agencies who put forward project proposals. Several lessons could be learned from this case:

- *Principle of participation*: the RTP kicked off a policy-planning process which involved people from industry, intermediaries, research centres, the regional development agency and the provincial government.
- *Principle of "objectivisation"*: the aim was to generate a broad base of support among all those involved in developing an innovation strategy. Outside experts were involved to conduct analyses. After this process, which took two years, the Province had a policy strategy for Limburg, consisting of ten priority areas and a number of pilot projects. It also put in place an agreed "support selection mechanism" which assessed whether new programmes and projects fit the issues set by the RTP.
- *Principle of mediation and alignment:* the RTP helped to create a more open policy "culture" where policymakers involve stakeholders in discussing and defining

85 This example is taken from Kuhlmann et al. 1999.

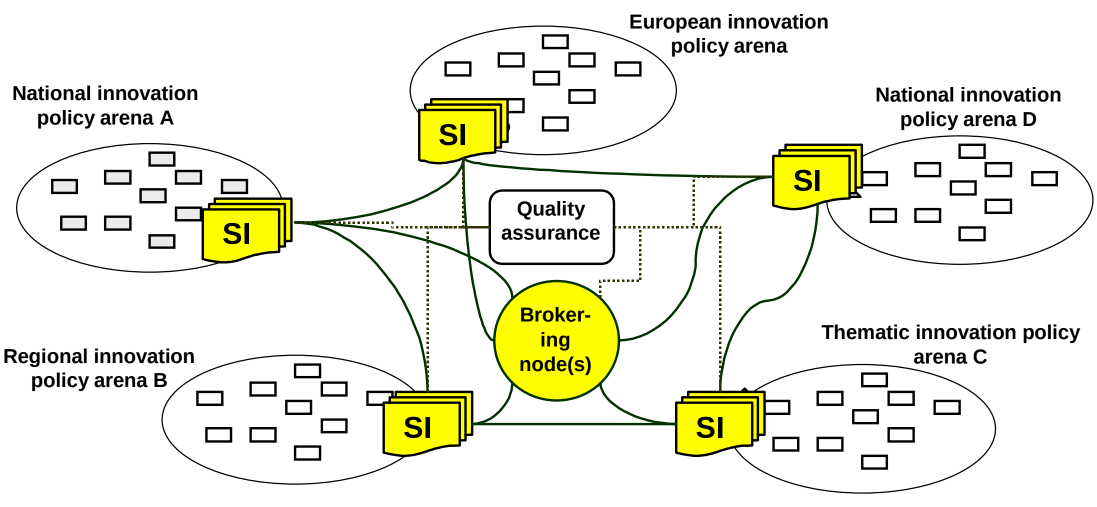
demand-oriented policies, through discussion platforms, steering committees, seminars, company visits and so on. Another result was that being part of an international network of RTP regions, exchange of experiences with other regions resulted in longer term international collaborations.

- *Principle of decision support*: an evaluation of the RTP Initiative showed that the general policy-planning model, first defined by the Commission 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. It is now continued under the name Regional Innovations Strategy (RIS) and Regional Innovation and Technology Transfer Strategy (RITTS) of DG XVI and DG XIII of the Commission, with more than 60 European regions going through the policy-planning process.

4.6. 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 11: Architecture of Distributed Strategic Intelligence



Consequently, an architecture and infrastructures of distributed intelligence (see Figure 11) 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:

- *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.
- *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. They would take care of various "reservoirs" of strategic intelligence as depicted in *Box 1*.
- *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.
- *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.
- *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 policymaking? Three major avenues of quality assurance can be followed: (a) bottom-up processes of institutionalisation amongst the providers of strategic intelligence may 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.

"Reservoirs" of Strategic Intelligence for Research and Innovation Policy

1. Types of public *policy interventions* in science, research and innovation: e.g. RTD funding in all variations ranging from institutional funding through classical project funding to funding of research networks or tax incentives towards companies; IPR policies; standardisation policies; VC policies; structural reform policies (like university reform; reshaping of publicly supported research institutes; reshaping of research council-like bodies); discursive policies like "public understanding of science", foresight initiatives; "new instruments" like ERA-oriented networks and "integrated projects" ...
2. *Theories* with respect to public intervention in science, research and innovation covering e.g. (a) discipline-oriented theories like evolutionary economics; neo-institutionalism; sociological systems theories; sociological actor theories; politological (neo-)corporatist approaches; innovation management concepts ... (b) interdisciplinary approaches like (national) "systems of innovation" ...
3. Analytical tools and *methodologies*: surveys of all kind; quantitative analysis and benchmarking of S/T indicators; policy analysis; policy evaluation; foresight; technology assessment; risk assessment ...
4. *Teaching* modules: e.g. for university curricula basic theory blocks (incl. competing approaches), methodology modules (e.g. "econometrics of innovation"; "innovation network analysis"; ...), interdisciplinary exercises, case study modules etc.; summer schools; application oriented continuing education courses like "R&D evaluation" ...
5. Core *competencies of defined members*: (a) research competencies; (b) teaching competencies, as a basis for staff exchange, for joint research efforts and policy advisory functions, for the compilation of EU-wide offered joint teaching activities,
6. Continuously updated agenda of *Forum* activities, including the amendment of the reservoirs and directories; the definition of several major lines of joint research; the improvement of theories, research methodologies and teaching concepts. The Forum agenda should allow for a variety of sub-agendas followed by Forum sub-groups.

4.7. ENHANCING DISTRIBUTED INTELLIGENCE FOR INNOVATION POLICYMAKING AT 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 research and innovation policymaking system* (see e.g. EPUB section on 'New Policy Instruments'; Kuhlmann/Edler 2002; 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 Observatory (ESTO), are attempting to work as "brokerage nodes" between the various strategic intelligence providers and users across Europe.

4.8. SUMMARY

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.
Four basic principles for effective strategic intelligence were figured out in this paper:

- *Principle of participation*: evaluation, foresight, or technology assessment exercises take care of the diversity of perspectives of actors avoiding maintaining one unequivocal "truth" about a given innovation policy theme.
- *Principle of "objectivisation"*: strategic intelligence facilitates a more "objective" formulation of diverging perceptions by offering appropriate indicators, analyses and information-processing mechanisms.
- *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.
- *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:

- *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.
- *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.
- *Transparent access requirement*: clear rules concerning the access to the infrastructure of distributed intelligence are needed.
- *Public support requirement*: distributed intelligence infrastructure is in need of a regular and reliable support by public funding sources.
- *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.

4.9. BIBLIOGRAPHY

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

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)

Cozzens, S. (2002): Frameworks for Evaluating S&T Policy in the United States. In: Shapira, Ph., Kuhlmann, S. (eds.): Learning from Science and Technology Policy Evaluation: Experiences from the United States and Europe, Cheltenham (E. Elgar) (forthcoming)

Cuhls, K. (1998): Technikvorausschau in Japan. Ein Rückblick auf 30 Jahre Delphi-Expertenbefragungen, Physica Verlag, Heidelberg

Cuhls, K., Blind, K., Grupp, H., (2002): Innovations for our Future. Delphi '98: New Foresight on Science and Technology, Heidelberg, New York (Physica, Springer) (series 'Technology, Innovation and Policy' by the Fraunhofer Institute for Systems and Innovation Research, ISI)

Cuhls, K., Kuwahara, T. (1994): Outlook for Japanese and German Future Technology. Comparing Technology Forecast Surveys, Heidelberg (Physica/Springer Publishers, Series "Technology, Innovation and Policy“, Vol. 1)

Cozzens, Susan, Melkers, Julia (1997): Use and Usefulness of Performance Measurement in State Science and Technology Programs. In: Policy Studies Journal, 25, 2

de Laat, B. (1996): Scripts for the future. Technology foresight, strategic evaluation and socio-technical networks: the confrontation of script-based scenarios, Amsterdam (PhD thesis)

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)

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

Feller, I. (2002): The Academic Policy Analyst as Reporter: The Who, What, and How of Evaluating Science and Technology Programs. In: Shapira, Ph., Kuhlmann, S. (eds.): Learning from Science and Technology Policy Evaluation: Experiences from the United States and Europe, Cheltenham (E. Elgar) (forthcoming)

Freeman, C. (1987): Technology Policy and Economic Performance: Lessons from Japan, London (Pinter)

Georghiou L. (1995): Assessing the Framework Programmes - a meta-evaluation, Evaluation Vol. 1, No.2, October 1995 pp.171-188

Georghiou, L., Rigby, J. (eds.): Assessing the Socio-economic Impacts of the Framework Programme (report to the EU Commission)

Gibbons, M., Limoges, C., Nowottny, 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)

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 a. (Springer)

Guba, Egon G./Lincoln, Yvonne S. (1989) Fourth Generation Evaluation, Newbury Park et al.: Sage.

Guzzetti, L., 1995, A Brief History of European Research Policy. Brussels/Luxembourg (Office for Official Publications of the European Communities)

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

IPTS (2000): The IPTS Futures Project. Synthesis Report, Seville (European Commission, EUR 19038 EN)

Kitschelt, H. (1991): Industrial Governance Structures, Innovation Strategies, and the Case of Japan: Sectoral or Cross-National Comparative Analysis? In: International Organisation, 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. (2001a): Management of Innovation Systems: The Role of Distributed Intelligence. Apeldoorn/Antwerpen 2001 (Publisher Maklu; series "Nijmegen Lectures on Innovation Management"; edited by the Faculty of Policy Sciences, Nijmegen University)

Kuhlmann, S. (2001b): 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, vol. 30, 6/2001, 953-976

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., Edler, J. (2002): Governance of Technology and Innovation Policies in Europe: Investigating Future Scenarios. In: "Technological Forecasting and Social Change", special issue "Innovation Systems and Policies", guest editors: P. Conceição, M. Heitor and G. Sirilli (forthcoming)

Kuhlmann, S., Heinze, Th. (2001): EU RTD Framework Programmes – a "Mini-Meta-Evaluation". In: Georghiou, L., Rigby, J. (eds.): Assessing the Socio-economic Impacts of the Framework Programme (report to the EU Commission)

Kuhlmann S., Holland D. (1995): Evaluation von Technologiepolitik in Deutschland. Konzepte, Anwendung, Perspektiven, Heidelberg: Springer

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)

Martin, B. (1995): Foresight in Science and Technology. In: Technology Analysis & Strategic Management vol. 7, no. 2, 140

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

Nelson, R. R. (Hg.) (1993): National Innovation Systems: A Comparative Analysis, Oxford/New York (Oxford University Press)

OECD (1999a): Managing National Innovation Systems, Paris (OECD)

OECD (1999b): Globalisation of Industrial R&D: Policy Issues, Paris (OECD)

Patton, Michael Q. (1997): Utilization-Focused Evaluation. The New Century Text, Thousand Oaks et al. (Sage)

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 Pharmazeutik, 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.

Ruegg, R. (1998): Symposium Overview. Symposium on Evaluating a Public Private Partnership: The Advanced Technology Program. In: The Journal of Technology Transfer, 23, 2, summer

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, Philip, Kingsley, Gordon, Youtie, Jan (1997): Manufacturing Partnerships: Evaluation in the Context of Government Reform. In: Evaluation and Program Planning, 2, 1, 103-112

Shapira, Ph., Kuhlmann, S. (eds.): Learning from Science and Technology Policy Evaluation: Experiences from the United States and Europe, Cheltenham (E. Elgar) (forthcoming)

Smits, R. (2001): Innovation studies in the 21st century. Questions from a user's perspective. In: Technological Forecasting and Social Change, 69 (2001), 1-23

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

Trischler, H., vom Bruch, R. (1999): Forschung für den Markt - Geschichte der Fraunhofer-Gesellschaft, München (Beck)

Worthen, Blaine R., Sanders, J.R., Fitzpatrick, J.L. (1997): Program Evaluation. Alternative Approaches and Practical Guidelines (2nd ed.), White Plains, N.Y. (Longman) New Policy Instruments

5. FUTURE POLICY INSTRUMENTS: EVALUATION OF THE SOCIO-ECONOMIC EFFECTS IN THE EUROPEAN RESEARCH AREA

Authors: **Luke Georghiou** (PREST) and **Stefan Kuhlmann** (Fraunhofer ISI)

5.1. INTRODUCTION

This chapter addresses the evaluation of new policy instruments. The development of evaluation practice has tended to mirror the evolution of technology and innovation policy, moving from an initial (and ongoing) focus on collaborative RTD programmes in the 1980s and gradually shifting towards measures intended to enhance the environment for innovation and technology transfer (Georghiou, 1998). Most recently there has been an increasing interest in policies designed to build research and innovation capacity, encompassing human capital and mobility enhancement, infrastructures and the building of networks. This has been accompanied by a shift in the rationale for innovation policy, or at least an extension from the market failure arguments developed in the 1960s and applied strongly in the 1980s. The structuralist-evolutionist approach now recognises that while information failures and lack of appropriability of returns may cause an under-investment in RTD, they do not necessarily guide the policymaker to the most appropriate actions. Such guidance may be obtained from the systems of innovation approach, which tends to highlight the absence of bridging institutions and the need to overcome firms resistance to adopt new technologies. A full review of developments in economic rationales and their implications for evaluation practice is available in the report of the ASIF Project (Georghiou, Rigby and Cameron (eds), 2002).

The focus of this chapter is upon the challenge to evaluation presented by the emergence of a series of innovative policy instruments under the European Research Area concept. At the same time consideration is given to policy measures emerging at national level.

Following a document published by the Commissioner for Research at the beginning of 2000, *Towards a European Research Area* (European Commission, 2000), the European Commission has been implementing a major change in its research policies. For the past fifteen years, its two principal instruments for research funding have been programmes of collaborative research and work in the Commission's own laboratories. The intention is to move away from these two instruments, which account for a small fraction (around 5%) of public funding, and to mobilise the entire research resource of Europe.

The principal aim of the ERA measures is to reinforce European competitiveness or to contribute to the solution of important societal problems through the mobilisation of a critical mass of research and technological development resources and skills existing in Europe. However, it also entails closer integration of research policies, which has implications both for policy implementation and subsequent evaluation.

In line with this aim, new policy instruments have been developed for application within and beyond the sixth Framework Programme. Key amongst these policy instruments are Networks of Excellence, Integrated Projects and provision for participation in Member States' programmes under Article 69.

5.2. NETWORKS OF EXCELLENCE

The objective of Networks of Excellence (NoE) is to reinforce European scientific and technological excellence by means of a progressive and lasting integration of research capacities existing across Europe.

Based on a particular scientific and technological theme and with a set of long term objectives, each Network aims to achieve a critical mass of competence and skills and to advance knowledge through the creation of a virtual centre of excellence. The networks will operate by fostering co-operation between existing centres of research excellence in universities, research centres, enterprises and technology organisations, expanding over time with the addition of new members.

The activities then undertaken by network members include not only joint programmes of research, but also associated collaborative activities, such as joint training activities, exchange of personnel, shared research infrastructures, equipment sharing and joint management of the knowledge produced. The minimum lifetime of these networks is five years, although it is anticipated that they will continue to function beyond the period of EC funding, the amount of which is determined in relation to the value and capacities of the resources to be integrated by participants.

5.3. INTEGRATED PROJECTS

Sharing the same rationale as Networks of Excellence, Integrated Projects (IPs) aim to reinforce European competitiveness or to contribute to the solution of important societal problems through the mobilisation of a critical mass of research and technological development resources and skills existing in Europe.

Integrated projects will have clearly defined scientific and technological objectives, and may include research, technological development and demonstration activities. They are expected to cover innovation and the dissemination, transfer and exploitation of knowledge. They will incorporate a degree of flexibility, allowing participants greater autonomy to modify the joint programme, to add new partners and to launch new activities in response to changing circumstances.

To reflect the more ambitious nature of the projects, they will be funded for up to five years, with a corresponding allocation of resources, possibly reaching tens of millions of Euros. Inter-related sub-projects may be integrated through a unified management structure.

5.4. ROLE OF THE COMMISSION ACCORDING TO THE TREATY

Article 169 of the EC treaty states: “In implementing the multinational framework programme the Community may make provision, in agreement with the Member States concerned, for participation in research and development programmes undertaken by several Member States, including participation in the structures created for the execution of those programmes.” There are no real experiences with this “géométrie variable” type of programmes yet. In big industries like aerospace there is a possibility that such initiatives will evolve. Also, a few member states, since the late 1990s, had begun to open some of their research funding programmes for international participation; e.g. a programme (PRO INNO) co-funding R&D co-operation of SMEs of the German Federal Ministry for Economic Affairs and Technology[86]. The level and extent of the member states' readiness to open their schemes and of the Community

[86] See www.bmwi.de/Homepage/Politikfelder/Technologiepolitik

participation will not at least depend on how difficult the related negotiation and decision procedures will be.

5.5. NEW AND SYSTEMIC INSTRUMENTS IN NATIONAL CONTEXTS

For about a decade now, we have been witnessing the emergence of new, systemically oriented RTD and innovation policy instruments in Europe, not at least reflecting the spreading of the "systems of innovation" approach of understanding the dynamics of research and innovation (OECD 1999). Such new, mostly rather experimental instruments may be characterised (see Smits/Kuhlmann 2002) as aiming at:

(1) *management of interfaces*; i.e. striving for the building of bridges and stimulating the debate, not limited to bilateral contacts but also focuses on chains and networks at system level;
(2) *building and organising (innovation-) systems* by the facilitation of new combinations (Neue Kombinationen) and deconstruction (creative destruction) of systems, initiation and organisation of discourse and alignment, consensus;
(3) providing *a platform for learning and experimenting*, such as: learning by doing, learning by using, learning by interacting and learning at system level;
(4) providing an *infrastructure for strategic intelligence*, i.e. identifying sources (Technology Assessment, Foresight, Evaluation, Bench Marking), building links between sources, improving accessibility for all relevant actors (Clearing House) and encouraging development of the ability to produce strategic information tailored to the needs of the actors involved;
(5) *stimulating demand articulation*, strategy and vision development, facilitating the search for possible applications of future science and technology, developing instruments that support discourse, vision and strategy development.

An *example* representing several of the above characteristics is the *German "Futur" initiative*, a new kind of foresight process. Futur is run on behalf of the Federal Ministry of Education and Research (BMBF) as a means of priority-setting for future innovation-oriented research policies. Futur is oriented towards the identification and inclusion of societal needs in future research agendas. "Leading visions" (*Leitvisionen*) are supposed to be the major outcomes of the process which shall be translated into funded research programmes or projects. The participation of a broader audience in various kind of workshops and the combination of different communication and analytical methodologies are characteristics of the process. Futur is intended to introduce "fresh ideas" into the research funding portfolio of the BMBF, by way of bypassing the traditional mechanisms for agenda-setting and prioritisation. The conventional process is characterised by a close and rather intransparent interaction between research institutions, industry, programme agencies (*Projektträger*) and ministerial bureaucrats in charge of research funding. The process started with *workshops* in early summer 2001 and is still running. Actors from industry, science, media and others were invited. These persons were identified because of their more broad, general knowledge. They were not supposed to be "specialists" in the narrow sense. From this first list, a kind of co-nomination process was conducted identifying about 1500 persons. Workshop participants were asked what they thought society might look like in the year 2020. As a next step, an "*open space conference*" was organised. The purpose of this conference was that "*focus groups*" should be founded and focus their themes. The groups had to produce "profiles" of their themes and a kind of competition was organised to write an interesting profile that meets a set of criteria given (new theme, societal need orientation, research link etc.). Out of the focus group's themes BMBF selected twelve areas to be more thoroughly debated in Futur. The Ministry organised an in-house workshop with department and division heads as well as the project management agencies' representatives asked to score the thematic areas according to their relevance. A similar process was organised in the internet: the persons already identified for the initial workshops were asked to give their votes. Finally, the Ministry asked the Futur participants to develop five of the twelve themes in detail, as a basis for the implementation of related funding measures. In the course of summer 2002 the Futur initiative will be evaluated by an international peer panel. Below we consider more broadly the evaluation of foresight as a policy instrument.

5.7. EVALUATION IN THE EUROPEAN RESEARCH AREA

Clearly, the precise form of specific future policy instruments at the time of implementation will determine the combination of evaluation methodologies that are relevant and their mode of implementation. At this stage it is more pertinent to consider the types of evaluation issues that the concept of the ERA and the instruments within it are likely to give rise to. This section considers these issues, as well as the specific evaluation issues associated with the two main future policy instruments. More general issues related to the ERA are also considered, such as the emphasis, within instruments, on personnel mobility, and on the relationship between benchmarking and evaluation.

5.7.1. Integration

The move towards greater integration of research activities, through both the networks of excellence and the integrated projects poses particular challenges for evaluation. By design, the whole greater is than the sum of parts, and so the ways in which the effects of integration can be assessed must be considered.

With the greater integration proposed in the ERA, evaluation at a European level must look beyond the evaluation of the Framework Programme, in terms of both scope and methodology. Closer integration of research policies requires mutual understanding of what has been achieved and how and of the distribution of benefits.

Developing greater linkages between national activities, must take into account the national variations that exist, such as in the institutional settings in which the type of work is carried out. For example, certain types of cancer research could be carried out in a university in one country, a branch of a national research organisation in another, a central laboratory in a third and by a non-governmental charitable foundation in a fourth. In each case, the core expertise has been acquired and developed through different sources, with different objectives, and with different modes of operation. An understanding of the dynamics of each national and institutional context maybe required for a thorough evaluation.

5.7.2. Excellence

The emphasis on research excellence requires consideration of the concept and measurement of “excellence” and efforts to move towards shared notions of both quality and excellence, and thus to the setting and adoption of standards. It could be argued that the development of a European Research Area requires a corresponding development of a “European Evaluation Area” in which there is a common methodological and procedural understanding that allows members to accept and validate each other’s findings. Thereby a proposal that a particular institution or centre meets a particular level of excellence in some dimension of its performance could be treated unequivocally.

5.7.3. Networks of Excellence

The long-term nature of Networks of Excellence and their mode of operation pose challenges for their evaluation. Research conducted by network participants should fall within the priority theme areas of the Framework Programme, or should respond to emerging policy needs, although the outputs and impacts of research may not be precisely specified at an early stage, to provide readily measurable and verifiable objectives.

Rather, it is anticipated that a long term evaluation perspective should be applied. The principal focus of evaluation of the operation of Networks of Excellence will centre on the added value they generate, bringing together the expertise of individual institutes into something larger than the sum of the parts, and its persistence beyond the initial period of funding.

The distinction between activities funded directly by the Commission and the other core activities and expertise of each of the network members is blurred, and thus a further challenge to the evaluator, as both must be taken into account. The true measure of the impact of the policy instrument will not be the productivity of a shared activity, as in a funded project, but in the enhancement to overall scientific productivity.

Any evaluation of the economic impacts of basic research must acknowledge the multiple dimensions of its effects on the economy, and, in this case, should pay particular attention to the production of human resources trained in the context both of scientific excellence and European added value.

Where concrete proposals for the evaluation of Networks of Excellence have been put forward these have centred on the arrangements for ex ante evaluation and monitoring. It is envisaged that the evaluation of project proposals will be based on a “peer review” approach, but may combine different methods depending on the characteristics of the areas under consideration, such as the use of independent expert panels, a referee system, and hearings of project representatives.

For the selection of the participants of these new instruments, it is indispensable to establish criteria for measuring excellence and quality. The long-term nature of Networks of Excellence requires a long-term evaluation perspective with a multi-dimensional assessment, focusing e.g. on the generated added value or the contribution to overall scientific productivity. Concrete proposals for Networks of Excellence can be evaluated in an "extended peer-review" approach, centring on eligibility criteria like excellence (number of publications or patents), managerial competence, Community added value, and the potential contribution to the integration of scientific efforts and the advancement of the scientific field; where necessary such extended information would have to be drawn from additional "background studies". Also, it is highly recommended that the peer panels undertake on-site visits of the "principal investigator", thus facilitating a critical debate between applicants and peers (see the Deutsche Forschungsgemeinschaft's (DFG) practice of appraising applications for multi-annual, multi-site integrated thematic research groups). An external scientific and technological monitoring council could monitor the progress of the network and assesses its work. The principal difficulty likely to be faced is that the size of individual grants limits the number which can be awarded. In these circumstances choices will have to be made between competing fields of research. The principles of peer review do not function in such circumstances, necessitating some higher strategic criteria for selection and a process to implement these. Ultimately this could be enhanced by information generated through foresight and other forms of strategic intelligence.

5.7.4. Integrated Projects

In terms of evaluation, the socio-economic dimension clearly has priority for this instrument. However, the approach is somewhat different to that for evaluation of smaller isolated projects. In the case of the Integrated Projects the project should achieve a critical scale whereby the strategic direction of a sub-sector of the economy is affected. This implies that the evaluation needs to engage with the socio-economic status of that sub-sector to understand the potential and actual impact of the IP (and the rationale for carrying out a project in this area). Effects could include such broader aspects as market structure. This can be described as a meso-level evaluation. Several of the approaches mentioned in earlier chapters of this book are applicable here. In terms of relevant past experience, the evaluation of projects such as JESSI (a successful joint EU/EUREKA programme in microelectronics) provide some relevant experience.

5.7.5. Mobility and Co-operation

More generally, the ERA objective of increasing mobility of researchers within and beyond the European Union, and in the context of different research policy instruments, raises broader issues for socio-economic evaluation. In particular, while the success of measures to promote researcher mobility may eventually be manifest in measurable impacts and outputs, studies to assess capability and potential are more relevant to study the state and dynamic of the knowledge economy, and the role of science, particularly a high quality scientific workforce, in sustaining and attracting economic activity.

5.7.6. Benchmarking

One dimension of assessment activity of the European Research Area, which is clearly relevant to evaluation and where progress is currently being made, is the idea of benchmarking policies for science and innovation. With its origins in the industrial domain[87], benchmarking aims to identify and spread best practice. To do so, it relies on the same types of data as evaluation, such as bibliometrics and peer assessments, and draws on a similar vocabulary, with terms such as “relevant performance indicators”, “qualitative understanding of best practice”, “monitoring mechanisms” in common use. However, for evaluators, the use of benchmarking data should be placed on the context of the systems that generate them. For example, in the domain of economic evaluation, indicators such as patents and the income from intellectual property are highly context-dependent and should be used with due caution.

5.8. Evaluation Issues and Methodologies

5.8.1. Monitoring, Analysing and Assessing of Changing National and Sectorial Innovation Systems at National, Regional and Sectoral levels

For decades now, we have been witnessing in Europe a co-evolution of regional, national and European research and innovation systems and related policy arenas, the latter meanwhile merging into towards a multi-level, multi-actor systems (Kuhlmann, 2001; Georghiou, 2002). Regional, national and transnational levels undergo a re-

[87] 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.

distribution of tasks, thereby experiencing new functional and informational linkages, vertically and horizontally. Thereby, the integration and redistribution is proceeding with different speed across Europe: initiatives of the “géométrie variable” type have been suggested repeatedly and will be implemented (§ 169, 6th Framework Programme).

In an extrapolation of this development we would see regional or national authorities concentrating their efforts on the competitiveness of “local” innovation systems, while the EU Commission – instead of running cumbersome own funding programmes – would “mediate” between the competitors and “moderate” their conflicts, and would take care of horizontal policy co-ordination. Public investment in, and regulation of RTD and innovation would originate mainly from regional or national initiatives and sources – but it would be concerted and matched with any parallel activities throughout Europe. Here, an important task of the EU Commission would be to carefully evaluate and facilitate the transferability of funding instruments, developed in heterogeneous national, regional or sectoral contexts, across Europe, thereby providing a *forum* to debate the degree of immediate imitation versus the need of "domestication" (Silverstone & Haddon, 1996); consider e.g. – as a thought experiment – the degree of transferability vs. requested domestication of the German "Futur Process" for innovation-oriented research funding priority-setting. Such debates would have to be grounded on the results of related policy evaluations as well as other sources of Strategic Intelligence (see section 5 of this volume). The EU Commission would have to facilitate the production of related intelligence inputs, such as:

- a *new breed of "national impact studies"*. Since the early 1990s, studies with a strong focus on the impact of EU FPs on Member States' science and technology policy and national actors in the science and economic sphere have been carried out. Laredo (1990) examined the role of public and academic research institutions in the FPs, Georghiou et al. (1993), Reger and Kuhlmann (1995), and later a full series of similar studies (e.g. Luukkonen and Hälikkä, 2000) drew conclusions about the impact on national academic institutions as well as national industry and research organizations, thereby shedding light on the interaction between European and national policies. These national impact studies, though certainly milestones in terms of their methodological stance towards impact measurement, were, however, not full evaluations of the FP as they focused largely on the effectiveness of the programme's impacts. Nonetheless, they remain the most detailed cross-Framework examinations ever done since the introduction of European RTD policies. As a result of this experience, the Commission made attempts at designing a common core methodology for subsequent impact analyses. A new breed of "impact studies" would in particular take care of the interplay between national, regional and EU RTD initiatives, combined with an analysis of the dynamics of sectoral innovation systems, not bothering about national borders.
- Consequently, in an integrating European research and innovation area the *study of sectoral innovation systems* will become an indispensable part of future RTD evaluation exercises.
- For the same reasons serious attempts would have to be made to extrapolate future research and innovation systems as well as policymaking mechanisms, as a basis for the design and ex ante evaluation of the next round of EU RTD policies ("7^{th} Framework Programme") (see Kuhlmann 2001; Georghiou 2002).

5.8.2. Example: Evaluating Foresight

As noted above the German Futur is being evaluated by an international peer panel supported by surveys of participants and users. The broader issue of evaluation of foresight is one major challenge in the coming decade. Several countries are coming into a second cycle of foresight activity and are keen to learn systematically from their own and others' experiences. So far, most evaluation of foresight has been of a fairly ad hoc nature. As with many of the new policies under discussion, a key problem is that the range of stakeholders has been broadened to include social as well as scientific and economic actors.

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 (Georghiou, 1996) or as martin and Johnston have put it, foresight is wiring-up the national innovation system (Martin and Johnston, 1999).

How then should foresight be evaluated? In a recent project in collaboration with Wiseguys Limited, PREST developed a framework for evaluation of the second UK programme. One of the key design considerations was to make allowance for the fact that the Programme 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 programme 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 programme. For this and other reasons a real-time evaluation was recommended.

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 Programme in terms of an indicator framework (see below) 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 which would facilitate interpretation of the data. In the event, this approach was superseded, at least in the short-term, by an internal high-level review prompted by the dissatisfaction of the responsible minister with the initial outputs of the Programme. The full evaluation is still pending.

Figure 12 Evaluation Framework for the UK Foresight Process

PANEL PROCESS

Suggested evidence:
Panel statistics, panel member satisfaction levels, stakeholder satisfaction levels

DIRECT DISSEMINATION

Suggested evidence: Programme and panel statistics, panel member satisfaction levels, takeholder satisfaction levels

DISSEMINATION VIA INTERMEDIARIES

Suggested evidence: Intermediary organisation statistics and documents, intermediary organisation satisfaction levels, stakeholder satisfaction levels

TARGET SECTORS FOR FORESIGHT IMPACTS

INDUSTRY & COMMERCE

Suggested evidence:

Use of Foresight outputs in business planning or technology strategy of case study firms

Evidence of use of Foresight outputs in business planning or technology strategy from Annual Reports, intermediary organisation documents, etc.

Use of Foresight methods in business planning or technology strategy of case study firms

Evidence of use of Foresight methods in business planning or technology strategy from Annual Reports, intermediary organisation documents, etc.

Formation and persistence of new networks within industry, and between industry, Government, the Science Base and other organisations (e.g. via intermediary organisation statistics,)

VOLUNTARY SECTOR

Suggested evidence:

Use of Foresight outputs in planning or technology strategy of case study organisations

Evidence of use of Foresight outputs in planning or technology strategy from Annual Reports, intermediary organisation documents, etc.

Use of Foresight methods in planning or technology strategy of case study organisations

Evidence of use of Foresight methods in business planning or technology strategy from Annual Reports, intermediary organisation documents, etc.

Formation and persistence of new networks within voluntary organisations, and between the sector, industry, Government and the Science Base (e.g. via intermediary organisation statistics,)

EDUCATION, TRAINING AND PUS

Suggested evidence:

Inclusion of Foresight approaches in business schools and professional training

Development and take-up of new scientific, professional or vocational training courses in line with Foresight recommendations

Use of Foresight by educational and training establishments

Increased numbers attending scientific, technical, engineering

SCIENCE BASE

Suggested evidence:

Alignment between Foresight and RC and other funding body objectives

Analysis of high quality proposals to RCs in priority areas

Increased funding for research in priority areas

Formation and persistence of new research networks in priority areas (e.g. via RC statistics, bibliometric analysis, intermediary organisation statistics etc.)

GOVERNMENT

Suggested evidence:

Use of Foresight outputs in planning or technology strategy of OGDs and agencies (via case studies, Annual Reports, Whitehall Foresight Audit, etc).

Importance of Foresight in co-ordination of policy (via OGD case studies and Whitehall Foresight Audit).

Effects on spend on S&T by Government departments (e.g. via OGD case studies, Forward Look, WFA) and on structure (eg WFG)

Formation and persistence of new networks with industry, Government, the Science Base and

COMPETITIVENESS

QUALITY OF LIFE

5.9. CONCLUSIONS

The flow of new policy instruments is likely to continue as policymakers attempt to adapt to changing circumstances and to innovate in the ways to stimulate innovation. Likely future directions include more effort to measure capacity (including the development of human capital where Bozeman and Gaughan (2000) are among those advancing the concepts for evaluation through their analyses of career developments using curricula vitae. Similar efforts are needed to assess the evolution of scientific infrastructure. Some preliminary ideas on how to benchmark these through equipment surveys are put forward by Georghiou, Halfpenny and Flanagan (2001).
A broader challenge for evaluation will be to assess the combined effects of different mixes of policies. This is a pressing issue as policymakers have come to realise that a portfolio of measures is likely to have greater effect than an imbalanced system which simply tries to reinforce one or two apparently successful measures.

5.10. BIBLIOGRAPHY

Selection of "national impact studies":

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

Larédo, P. (1990): L'impact de programmes communitaires sur le tissu scientifique et technique francais, Paris: La Documentation Francaise

Luukkonen, T., Hälikkä, S. (2000): Knowledge Creation and Knowledge Diffusion Networks. Impacts in Finland of the EU's Fourth Framework Programme for Research and Development. Publication of the Finnish Secretariat for EU R&D 1/2000. Painotalo Miktor, Helsinki

Reger G., Kuhlmann S. (1995): European Technology Policy in Germany. The Impact of European Community Policies upon Science and Technology in Germany, Heidelberg: Springer

Studies or papers on the future of EU RTD policies:

Georghiou, L. (2002): Evaluation of research and innovation policy in Europe – new policies, new frameworks? In: Shapira, Ph., Kuhlmann, S. (eds.): Learning from Science and Technology Policy Evaluation: Experiences from the United States and Europe, Cheltenham (E. Elgar) (forthcoming)

Kuhlmann, S. (2001a) '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, vol. 30, issue 6/2001, 953-976

Information on "New Instruments" Evaluation Approaches and Policy Rationales:

Bozeman, B and Gaughan, M (2000) Evaluating Scientific and technical Human Capital: An Event History Approach, in Shapira, P and Kuhlmann, S (eds), Proceedings of the 2000 US-European Workshop on Learning from Science and Technology Policy Evaluation, Bad Herrenalb, http://cherry.iac.gatech.edu/e-value/

Georghiou, L. (1998) Issues in the Evaluation of Innovation and Technology Policy, Evaluation Vol 4(1): 37-51

Georghiou, L, Halfpenny, P and Flanagan, K, Benchmarking the provision of scientific equipment, Science and Public Policy, August 2001, pp303-311

Georghiou, L., Rigby, AJ and Cameron, HM, (eds) (2002) *Assessment of the Socio-economic Impact of the Framework Programme* available from http://www.cordis.lu/

OECD (1999): Managing National Innovation Systems, Paris (OECD)

Seminar on the New Instruments of the Framework Programme 2002-2006: networks of excellence and integrated projects - 20.04.01 Discussion paper: Networks of excellence

Seminar on the New Instruments of the Framework Programme 2002-2006: networks of excellence and integrated projects - 20.04.01 Discussion paper: Integrated projects

Silverstone R. & L. Haddon (1996): ‘Design and the domestication of information and communication technologies: technical change and every day life’, in: Silverstone, R. & E. Hirsch (eds.). (1992) ‘Communciation by design. The politics of information and communication technologies’, Oxford: Oxford university Press.

Smits, R., Kuhlmann, S. (2002): Strengthening interfaces in innovation systems : rationale, concepts and (new) instruments. Paper for workshop "New challenges and new responses for S&T policies in Europe", Brussels, April 22-23, 2002 (EU STRATA)

6. AVENUES FOR RTD-EVALUATION IN THE FUTURE POLICY CONTEXT

Authors: **Gustavo Fahrenkrog** and **Alexander Tübke** (JRC-IPTS), **Wolfgang Polt** and **Jaime Rojo** (Joanneum Research)

This chapter presents the results of an expert-conference, which discussed the results of the work of the Epub[88] network against the new role of evaluation in the context of the European Research Area (ERA) and the 6th Framework Programme (6th FP). Special emphasis was set on the challenges for evaluation processes in the design of the new policy instruments, especially Networks of Excellence and Integrated Projects. The conference also showed how the "RTD evaluation toolbox" produced in the EPUB project could be applied in a real policy context.

During the conference, the 33 experts[89] from national governments, research institutions and the European Commission addressed these issues in a round table discussion. The following sections outline the main lines of thought expressed during the conference.

6.1. THE CHANGING POLICY ENVIRONMENT

Policymaking for Science and Technology (S&T) is currently confronted with new challenges. According to the experts, the increasing economic role of knowledge, the speed of technological change and the network effects in today's globalised world change the way S&T development can be influenced and produces new impact patterns. Policymaking processes therefore become more dynamic, which shows in the increasing role of science and governance issues and the demand for higher levels of participation in setting S&T policy agendas. In addition, public attitude makes greater demands on the transparency and accountability of the policymaking process. Taken together, these trends increase the complexity and uncertainties about the impact of S&T in society and raise the stake for decision-makers.

The establishment of the European Research Area (ERA) and the 6th Framework Programme (6th FP) adds further challenges because they focus on improving the impact of projects at the EU level (European Added Value) through the co-ordination of research efforts, which leads to larger projects and longer-term commitments. This increases the level of EU funding per project, but reduces the number of projects funded. That leads to a new profile of risk and impact assessment, which makes the policy-maker's task of establishing systemic policies more complex. It also increases the need to well-define the policy rationale.

The experts also outlined that the inertia of the system leads to continuity in the policy objectives and measures. How the new challenges are faced at the EU-level is therefore also a question of balancing what is desirable with what is feasible.

Within the ERA, the new policy objectives are implemented through Networks of Excellence and Integrated Projects. Networks of Excellence intend to progressively

[88] The "Socio-Economic Evaluation of Public RTD-Policies (Epub)" was a thematic network under the STRATA action (Strategic Analysis of Specific Political Issues) of the 5th EC RTD Framework Programme (see http://epub.jrc.es/). The Epub thematic network constituted a pilot action aiming at supporting European RTD policy making and improving efficiency in RTD policy actions. The network's main objective was to provide a better understanding of the dynamics of science and technology by helping to investigate the socio-economic impact of private/public RTD policies at a micro/macro level. A special emphasis was given to analyse and review the rationales, concepts, methodologies and indicators generally applied in RTD evaluation and to investigate the implications for evaluation of new S&T policy approaches. The network also facilitates the sharing and diffusion of knowledge, techniques and tools applied in RTD evaluation.

[89] See the Annex for the list of participants

integrate research capacity across Europe through bundling competence and skills in virtual centres of excellence for at least five years. The members of the network undertake a large variety of activities, share experience and infrastructure and jointly manage the network. Integrated Projects have a closer thematic focus with clearly defined objectives and run up to five years.

6.2. EVALUATION IN THE POLICY CONTEXT

The experts identified three main themes on which S&T-policy evaluation has focused in the past. The first evaluation studies concentrated on the scientific results, the second generation analysed the rather direct impact on technology and innovation and the third generation takes into account the wider societal fabric and stresses the structural impacts. By the stage of development of a policy measure, the evaluation process comprises ex-ante evaluation, selection, monitoring and ex-post (impact) assessment. Evaluation processes are applied both to projects and project lines (programmes).

For its present programme, the EU has a well-established evaluation system including all stages of the evaluation process. There is already a potential for matching data from the different databases by integrating information from different member states in order to gain a deeper insight on the connections between technology, productivity, employment and trade. With respect to the new policy instruments proposed in the ERA and the 6th FP, evaluation has also to address structural aspects. In the expert-opinion, evaluation would become much easier if the EU decided on the shape of the new policy instruments, so that the right indicators and success measures can be found (excellence, for example, could be measured by the change of research agendas or infrastructure creation).

Some of the strategic political objectives (e.g. scientific excellence or capacity building) are particularly difficult to measure. This requires the introduction of new indicators and methods, including benchmarking or detailed case studies. New indicators could be constructed based on the political reasoning that justifies intervention. However, there remain multiple problems of choosing the right measures and levels for impact assessment (especially with respect to European Added Value), difficulties in relating qualitative and quantitative measures and hurdles in the EU that inhibit the application of certain evaluation methods. There is also the need to find a balance between the EU, the Member States, and the research organisations. Methods that focus on behaviour, changes, and networks might provide new insights.

The new modes of governance also increase the need for having a well-written work-programme in order to be able to eliminate politically irrelevant projects. The experts pointed out that interesting examples can be found in the US, where the COSEPUB committee has defined the broad guidelines for criteria and objectives of the RTD performed in Federal government agencies (e.g. GPRA of 1993, 1999, 2002). The US Department of Energy and the National Institute of Standards and Technology (NIST) have introduced innovative methods for ex-ante evaluation, including real option approach, network analysis, capacity building and foresight scenario analysis. With respect to the type of policy instrument evaluated, the consistency and aggregation of project data into programme or policy instrument level data for comparison remains an unsolved issue.

The development of a self-organising arena around evaluation could provide an important stimulus for the policy-making process. At present, most evaluators in the EU are not experts, so external opinions are highly regarded. However, the present changes in the policy-making process lead to a new role of policy advice, which needs to find new forms of organisation.

As the new policy focus of policy intervention in the ERA and the 6th FP leads to the support of bigger projects, it also increases the risk per project. A question to address is what happens if the ex-post evaluation shows that the EU has gone wrong with such large projects like Networks of Excellence and Integrated Projects. This raises the question of who uses the results and recommendations of the evaluations, how they are used and what the consequences are. In this sense, the difference between strategy and evaluation should be respected. In theory, evaluation is just one input to strategy. In practice, however, strategic evaluation incorporates a political dimension that can have an influence in the shaping of policy.

In order to improve the policymaking process, currently separated activities like evaluation, technology assessment, forecasting and foresight could be integrated. In this way, specific societal objectives of policy can be better taken into account.

6.3. EXPERT FEEDBACK ON THE TOOLBOX

The toolbox is a collection of methods used in evaluation. It describes their basic structure, their strengths and limitations and provides examples of how they have been applied in evaluation. The toolbox provides a first collection of the tools considered in the evaluation cycle. According to the experts, it is adapted to the needs of evaluation practitioners and programme managers and provides a general overview on evaluation methodologies and policy instruments. It meets the aim of the STRATA programme of linking researchers to policy makers. Although labelled "toolbox of evaluation methods" it needs to be put into context and used together with other sources of "distributed intelligence" (e.g. foresight, assessment, benchmarking etc). Therefore, it was suggested to use the label "handbook" instead of "toolbox".

The toolbox cannot be applied in a standardised mechanistic way, but resembling projects might be evaluated similarly. The correct application of the methods in the toolbox requires a certain amount of tacit knowledge from the user. Policymakers can use the toolbox in order to become familiar with the principles of evaluation and thus better understand the results of evaluation.

6.4. EX-ANTE EVALUATION

In general, ex-ante evaluation is linked to the rationale behind the policy intervention. Within ex-ante evaluation, it has to be demonstrated what the strategic objectives are, what the framework of implementation is and whether the policy rationale can be fulfilled at the end. It is also connected to the previous policy cycle by learning from and using the results from previous evaluations. Ex-ante evaluation is thus part of the bargaining process. Although strategic evaluation is a key issue for policy making it will require a long maturation period. The experts pointed out that, especially with larger programmes, it might not be sufficient to evaluate infrastructure capacity or the participation in associations and networks. There is an increasing demand to incorporate legal, social and ethical considerations into ex-ante evaluation and adopting a more systemic perspective.

Some methodologies are particularly well adapted to ex-ante evaluation, especially modelling, simulation, cost-benefit analysis and foresight. There is currently a shift from an ex-ante evaluation that defines programme objectives and metrics towards an evaluation of the structural capacity of policies. For large programmes, there will most likely be a requirement to perform an ex-ante evaluation.

6.5. SELECTION AND EVALUATION OF PROJECTS

The selection criteria for projects reflect the expected (political) benefits, so finding the right evaluation process is extremely important in order to fund the "right" projects. The basic methodologies used for the selection and evaluation of projects are proposal evaluation, peer review and panel evaluation.

In the current selection process, proposals are evaluated on the basis of expected benefits. In the European Commission there is pressure for speeding up the evaluation process, which makes it difficult to perform the longer and more complex evaluations. The question was raised of what could be an alternative to the current approach of locking experts in a room to evaluate a series of proposals. One suggested alternative was to adopt a two tier approach. In this approach, the proposals would first be evaluated according to purely scientific and technical criteria. A strategic evaluation would then be applied only for those proposals that rank above a certain technical quality threshold. This procedure could also be applied using other criteria, e.g. assessing scientific quality first and socio-economic and technical criteria second. A negotiation phase could be introduced after the first cut-off.

The experts remarked that the selection and evaluation process is extremely difficult for large size projects in areas where there are very few candidates (e.g. when there are just one or two proposals in a particular area). The difficulty of assessing the policy relevance or the ethical dimension of projects by traditional peer review was outlined. An alternative to peer review could be expert panels as they are currently used in most of the large programmes of the European Commission. Another alternative could be reputation analysis. The importance of assessing the profile of rejected projects was underlined.

Selection criteria could be adopting past productivity as a predictor of future productivity, judgements from overseas experts, trustworthiness, practicality, the commercial potential of the project, or also social aspects. This raised the issue of quantitative vs. qualitative evaluation. The NSF uses the panel approach to incorporate social criteria in the evaluation of the proposals.

Another question that came up was how to improve the selection process in a way that enables all potentially interested candidates to participate in the process, which is currently expensive and requires specialised human resources and skills.

6.6. PROGRAMME MONITORING

Monitoring starts at the beginning of the project and covers its whole life span. It is a question whether the European Commission should be actively involved in project management. The experts emphasised that monitoring should fall within the responsibility of the project manager and centre on the expected use of evaluation. The results of monitoring should not go to the project manager himself, but to his hierarchical superior. Good monitoring and the existence of quality control measures are sign of good project management. Quality control also needs to take into account outside opinions and user perspectives. When a passive monitoring alone is applied, the possibility to influence the project's direction is lost and the control whether the project is carried out efficiently is reduced. An active monitoring provides more space for corrections, but also entails a certain responsibility for the project outcome.

The experts pointed out that the monitoring system should take into account that data between different projects are not always comparable. However, the access to a common knowledge base could improve monitoring.

6.7. Ex-Post Evaluation

One of the most important objectives of ex-post evaluation is to provide a legitimisation for the use of the taxpayers' money. Also for ex-post evaluation it is very difficult to make an impact assessment at the EU level. Both the point of view as well as the purpose of the evaluation need to be clarified. The objectives then determine the methodologies to be used. One methodology is generally not enough. The experts raised again the questions of which methodologies of the toolbox should be used and how to find the right mix of the methodologies. Econometric methodologies have the strength of providing measures for the impact of policy, especially at the microeconomic level. However, impact assessment should take into account other perspectives, e.g. from the social and political sciences.
With these considerations, themes related to the role of evaluation in the new policy environment came up again.

In their final statement, the experts outlined the necessity to follow the ongoing changes closely and intensify co-operation in order to further develop the role of evaluation in the new policy context.

7. CONCLUDING REMARKS

Authors: **Wolfgang Polt** and **Jaime Rojo** (Joanneum Research)

All modern economies adopt policies to support their science and technology system aiming to improve economic performance and social welfare. This increasing relevance of science and technology in shaping the forces driving our societies and economies is forcefully reflected in the Presidency's conclusions of the European Council held in Lisbon on March 2000 where the Heads of Government and State agreed on the new strategic goal of the EU to become the most competitive and dynamic knowledge-based economy in the world, capable of sustainable growth with more and better jobs and greater social cohesion. Fulfilling this challenging strategy requires a clear commitment on the part of governments to discern the socio-economic impacts of science and technology policies.

Evaluation is a key decision support tool providing policy makers with a better understanding of policy results, allowing learning from past experiences, providing elements for improving strategy definition, increasing the efficiency and efficacy of policy intervention, and demonstrating the effects of intervention. All these elements help to improve the transparency and accountability demanded on the policy making process. As a rapidly growing and expanding field, the evaluation discipline is closely connected to the policy decision making process. However, there is still a lot of confusion on what evaluation can realistically and effectively achieve.

The theory and practice of evaluation has co-evolved with the developments experienced in science and technology policy. Evaluation tools have expanded to provide not only a means for quantification of policy impacts, but to facilitate mutual learning from past experiences, supporting mediation, decision-making and policy strategy definition. Increase in the complexity and uncertainty present in policy decision making requires the emergence of strategic intelligence combining the synergies of capacities between evaluation, technology foresight and technology assessment, to produce objective, politically unbiased, independent information to support active decision-making.

The RTD evaluation toolbox combines the effort of a network of evaluation experts to summarise the knowledge available regarding the main instruments applied in science and technology policy, emphasising the methodologies employed to evaluate their socio-economic impacts. Rationales for policy implementation, applied evaluation methodologies, conditions for application, methodology scope and limitations, data requirements and good practices are all addressed in this toolbox. The toolbox contents are structured into five interrelated sections. The first section brings evaluation into context defining the user perspectives in evaluation, signalling the sometimes conflicting expectations arising between evaluators and policy makers. The second section is structured around a set of broadly defined policy instrument categories providing the rationales for public intervention, main evaluation techniques available for their evaluation, and specifying illustrative practical examples of how evaluation have been conducted in the past. The third section introduces a series of evaluation methodologies, providing a description, reviewing the requirements for their application and illustrating their use with good practice examples. The fourth section analyses the conceptual framework describing how to incorporate evaluation practices within a "distributed techno-economic intelligence system" to inform and assist the policy making process. The fifth section gives some indications on how to use the capacities accumulated in the evaluation of the new policy instruments envisaged for supporting and promoting RTD activities.

The main contents and messages that can be extracted from the toolbox can be summarised as follows. Current EU RTD evaluation practices (which comprise continuous monitoring, five year assessments and mid-term evaluation) are characterised by a strong focus on monitoring as compared to impact assessment, on projects and programmes rather than the broad context of policy and a strong reliance on expert panels rather than on studies using techniques specifically adapted to policy evaluation such as control groups or productivity measurement. Also, there is a constraint imposed by the limited time and monetary resources devoted to evaluation. With the increasing diversity of RTD policy instruments (e.g. funding of collaborative R&D, support to R&D infrastructure, measures for technology transfer and diffusion, standards and regulations, IPRs, networking…) relevant at the EU level, a sensible mix of available methodologies has to be applied in evaluations.

As reflected in the summary table below, there is hardly any policy instrument that would be best evaluated using any single method in isolation. Most evaluations will effectively benefit from a combined use of various methods providing complementary valuable information on policy effects. For instance, quantitative evaluation methods combined with performance indicators permit to capture the dynamics involved in science and technology providing good estimates of output and impact of public intervention. Policy makers could make use of these impact estimates as a means to legitimise and as supporting evidence of the rationale behind policy intervention. Qualitative evaluation methods provide policy makers with more detailed insights on the multiple effects of policy intervention, helping to improve and clarify the processes, instruments and behaviour induced by science and technology policy.

Evaluation Matrix: Matching policy instruments and methodologies

	Innovation Surveys	Econometric Models	Control Group Approaches	Cost Benefit Analysis	Expert Panels/ Peer Review	Field / Case Studies	Network Analysis	Foresight/ Technology Assessment	Benchmarking
Financing R&D	●●●	●●●	●●●	●		●●●		●	●
Provision of R&D infrastructure		●●		●●●	●●●	●●●	●●●	●●	●●●
Technology transfer and innovation diffusion	●●●	●●●	●●●	●●	●	●●	●●●	●●●	●●●
Legal frameworks (IPRs, standards and regulation)	●	●	●	●●●		●●●		●●	●●●
Integrated projects			●	●●●	●●●	●●●	●●●	●●	●●
Networks of excellence					●●●	●●	●●●	●●	●●

Legend: ●●● *Highly suitable* ●● *Suitable* ● *Less suitable*

The next two tables constitute an attempt to summarise the wealth of information included in the evaluation methodologies section. For a deeper appreciation on the methodology of interest it is advised to refer directly to the specific toolbox section. We find among others that the result of innovation surveys have to be better linked to the evaluation exercises on the macro level, the econometric impact assessments can be used on a wider scale as currently in use. This holds true not only for the macro-level but also – and especially – for micro-econometric tools like control group approaches.

In the future, also cost-benefit analysis might play a larger role (in the arsenal of evaluators) not least in the ex-ante and ex-post evaluation of large-scale projects. Even with peer review – probably the most widely used approach in European S&T evaluation – it is possible to outline a potential for improved applicability, based on refined ways of panel composition, task allocation and decisional power.

However, probably as important as the suggestions and recommendations with respect to individual tools and approaches is the perception that evaluation – to serve its purpose to empower policy learning – should follow some general good practice rules and ought to be embedded in a broader "system of distributed intelligence" which comprises other sorts of assessments of policy as well: benchmarking, foresight and technology assessment. It is in this setting that the use of evaluation results will yield most benefits to policy making.

Despite the availability of a number of evaluation methods, there is scope and need to look for further methodological improvements in evaluation. At present, consistent evaluations can be conducted at the project level, but undoubtedly more thorough evaluations at programme or policy level will require advances in knowledge both in the causal relations between inputs and outputs as well as to arrive at meaningful procedures for measuring and to aggregating these outputs.

Table 28: Overview of Evaluation Methodologies

Methodology	Type / Use	Data Requirements	Strengths	Limitations	Good Practices
Innovation Surveys	Semi-quantitative Quantitative *Monitoring* *Ex-post*	Micro data Expenditures Profits Patents, Innovation	Detect innovation trends and insights on the soft side of innovation. Findings from interviewed sample can be generalised to the population Permits to identify size and distribution of impacts Provides groups comparisons and changes over time	High cost and time consuming Processing and analysis of data requires large human resources Some types of information are difficult to obtain Long time series generally not available	Analysis of the innovation process using data on the EU Community Innovation Survey
Micro Methods	Quantitative qualitative categorical data *Monitoring* *Ex-post*	Micro data Expenditures Profits Patents	Results based on explicit formulation of theory based causal relationships R&D Additionality Control for different effects: firm size, expenditures, innovation capacity	Quality of data Persuade participant and non participant entities to disclose information Only private rate of return to R&D	Effects of public R&D subsidies on firms in West Germany Evaluation of the ITF Programme FlexCIM Effects of R&D subsidies in Spain Promotion of AMT technologies based on Swiss Micro data
Macro Methods	Quantitative modelling methodology *Ex-ante (simulation)* *Monitoring* *Ex-post*	R&D Expenditures R&D output Macroeconomic data	Social Rate of return to R&D Capture R&D Spillovers Estimate long term policy intervention impact Scenario simulations for policy supported geographical areas	Average returns Robustness of results Time lags for observation of the effects	Modelling approaches: OECD Interlink, IMF Multimod, EU Quest. R&D Spillover studies: Jaffe, Nadiri International spillovers: Eaton and Kortum, Mohnen, Evenson
Productivity Studies	Quantitative modelling methodology *Monitoring* *Ex-post*	Micro data Expenditures Profits R&D, Patents	Estimation of effect of R&D on productivity Estimate the rate of return to R&D	Quality of data Deflation of series Required assumptions for measurement of stock variables	Productivity studies (Van Ark) Growth accounting (Griliches, Jorgenson) Micro datasets: French INSEE and US Census of Manufacturers
Control group approaches	Quantitative *Ex-post*	Micro data Expenditures Profits Patents	Capture the impact of policy intervention on the programme participant entitity	Requires high technical capacity High Implementation Cost Data Demanding	Collaborative industrial Research between Japan and US Evaluation of RTDI instruments in Ireland Participation of Ireland in Eurpoean Space Agency
Cost Benefit Analysis	Quantitative (with qualitative elements) *Ex-ante (especially)* *Monitoring* *Ex-post*	Micro data Profit & cost estimates	Provides an estimate of socio-economic effect of intervention Good approach to assess the efficiency of an intervention Addresses by making them explicit all the economic assumptions of the impact of the intervention	Requires high technical capacity Some degree of judgement and subjectivity, depends on largely on assumptions made Not easily comparable across cases Careful interpretation of results when benefits are not easily quantifiable in monetary terms	US Advanced Technology Programme US National Institute of Standards Methodology
Expert Panels /Peer Review	Qualitative Semi-quantitative *Ex-ante* *Monitoring* *Ex-post*	Project programme data	Evaluation of scientific merits Flexibility Wide scope of application Fairness	Peers independence Economic benefits not captured	Evaluation of Large Infrastructures Evaluation of EU Programmes
Field /Case studies	Qualitative Semi-quantitative *Monitoring* *Ex-post*	Project programme data	Observation of the socio-economic impacts of intervention under naturalistic conditions Good as exploratory and descriptive means of investigation Good for understanding how contexts affect and shape impacts	Results not generalisable	Telematic innovation in the health care sector. Evaluation case studies reviewed in Georghiou and Roessner (2000)
Network Analysis	Qualitative Semi-quantitative *Ex-post*	Project programme data	Comprehensive empirical material Compilation for policy purposes Co-operation linkages	Time involved in collecting the survey information Persuasion requirements	RTO systems Interdisciplinary centers of medical research
Foresight/ Technology Assessment	Qualitative Semi-quantitative *Ex-ante* *Monitoring*	Qualitative data Scenario	Consensus building to reduce uncertainty under different scenarios Combination on public domain and private domain data Articulation and road mapping of development of new technologies	Impossibility to detect major RTD breakthroughs	Benchmarking of ISI/FhG capacities against Foresight results
Benchmark-ing	Semi-quantitative *Ex-post* *Monitoring*	Science and technology Indicators	Comparison method across different sectors Support to systemic evaluation of institutions and systems	Data detail requirements Non transferable	EU Benchmarking national policies Innovation Trend Chart Science-industry relationship

Table 29: Overview of Evaluation Methodologies cont.

Methodology	Data application level	Areas of application	Output	Outcome	Impact
Innovation Surveys	Firm Industry Economy-wide	Innovation IPRs Technology transfer Research collaboration	New products and processes Increase in sales Increase in value added Patent counts, IPRs	Creation of new jobs Innnovation capacity building	Enhanced Competitiveness Institutional and organisational efficiency, Faster diffusion of Innovation Employment
Micro Methods	Plant Firm Industry Economy-wide	Sectoral Returns to R&D	Output and value added (collect baseline info for before-after comparisons)	Sectoral productivity industry sectoral spillovers Additionality, Leverage effects	Firms competitiveness
Macro Methods	Firm Industry Economy-wide	Sectoral Regional Economy-wide	Output and value added	Change in R&D Capital, Human capital, Social capital International R&D Spillovers	Regional, country productivity Employment, Good governance Economic and social cohesion
Productivity Studies	Plant Firm Industry Regional Economy-wide	Sectoral Regional Economy-wide	Output and value added	knowledge, geographical and International R&D Spillovers	Regional, country productivity Employment Economic and social cohesion
Control Group Approaches	Firm Industry	Technology implementation Innovation	Output and value added (on supported and non supported firma)	Additionality Rate of return to R&D	Firm, industrial competitiveness
Cost Benefit Analysis	Firm Industry	Health Environment Energy Transport	Value added benefit-cost ratio IRR Consumer surplus	Health improvements Consumer protection Environmental sustainability	Quality of life Standard of living
Expert Panels/ Peer Review	Firm Industry Economy-wide	Scientific merit Technological capacity	Publication counts Technological output	Scientific and Technological capabilities	R&D performance
Field/ Case Studies	Firm Industry	Science-industry relationships	Detailed inputs and outputs	firms RTD capabilities on the job-training educational schemes	Industrial competitiveness Quality of life Organisational efficiency
Network Analysis	Firm Industry Regional	RJVs, co-operation science industry Clusters	Co-operation linkages	Co-operation in clusters Social embededness	Efficiency of institutional relationships
Foresight/ Technology Assessment	Institution Regional Economy-wide	Technology Trends	Identification of generic technologies Date of implementation	Technological capacities	Technological paradigms shifts
Benchmarking	Firm Industry Economy-wide	Efficiency of technology policy	S&T indicators	Technology capabilities	Industry competitiveness Good governance

Faced with a rapid and continuous evolution of science and technology policies in advanced economies, there is a constant need to devise and adapt existing methods to evaluate the efficiency and efficacy of those new policies. For instance, the European Research Area (ERA) concept requires to better understand the increasing interconnection and integration between science and technology organisations and the rest of the involved actors. This would require implementing methods that allow to evaluate institutional capacity within changing environments. Thus, evaluation practices are increasingly required to analyse the effect science and technology policy induces on actors' behaviors and on institutional change.

With respect to good practice rules in RTD policy evaluation, the following constitutes a synthesis extracted from the information contained in the toolbox as well as the discussions held in the context of this project:

Evaluation planning

- Provide an early and adequate scheme for the evaluation design and integrate it into the policy intervention design to ensure that intervention objectives are clearly defined and can be effectively evaluated.
- Base the public intervention on a demonstrated market or systemic failure, which the intervention should solve.
- Define requirements on data compilation and updating during the intervention design stage. Ex-post evaluation will critically depend on the quality of compiled data.
- Introduce new methods in ex-ante evaluation that favour diversity and the taking up of new risks and multidisciplinarity. Peer review is a significantly conservative approach in the evaluation of research proposals, risky projects are likely to get worse scores from peer review. Mainstream science is better positioned when adopting peer review methods.
- Operational and management issues
- Allocate sufficient time and monetary resources to evaluation. This is justified as the aim is to ensure that public money is efficiently and wisely spent.
- Promote independence to ensure credibility of results, for this purpose it might be relevant to use external evaluation experts (from other countries).
- Involve policy makers and project managers in the evaluation so that their perceptions and priorities are fed into the evaluation design and during the evaluation execution.
- Separate in evaluation the strategy function from the operational function. Evaluation as a demonstration of impact is only one input to strategy definition.
- Strengthen transparency by publishing the terms of reference, criteria's used in the evaluation and disseminating the produced evaluation results to a broad audience of interested bodies.

Evaluation priors

- Clarify the implicit policy rationale of the intervention when conducting an evaluation. Identify the objectives of the policy intervention being evaluated, discussing the intervention logical framework, including implicit assumptions and establishing the feasibility of evaluating them.
- Define the intervention jointly with concrete targets that will facilitate the evaluation of the instrument, e.g. "increase the publications in the field of genetic technology by 20 per cent or increase productivity by 10 per cent".
- Ensure the compilation of data before and after the intervention as well as on supported and non supported units to allow to control for the counterfactual.

Method implementation

- Adapt methodologies to deal with the particular evaluation requirements and to answer relevant questions. Evaluation should not be perceived as mechanical process. Definition of objectives determine the methodology selection.
- Combine different methodologies and different levels of data aggregation to improve the understanding of the multidimensional effects of the policy intervention.
- Incorporate systemic considerations into evaluation as science and technology is likely to modify institutions structure and behaviour.
- Separate when possible the evaluation of the scientific merit provided by traditional established methodologies such as peer review from the evaluation of the other socio-economic objectives provided using the support of expert panels.
- Evaluate the profile of supported and non-supported firms including those rejected and those who did not apply for support. Control group approaches are especially valuable in this context.
- Establish intended and unintended effects of the intervention. Analyse failure as well as success histories.

Strategic evaluation

- Integrate RTD evaluation practices with other sources of "distributed intelligence" such as technology assessment and foresight to support strategy policy definition.
- Develop criteria to evaluate the increasing strategic role of systems and institutions in science and technology. New applications of benchmarking, foresight and network approaches could be used to evaluate increasingly relevant topics such as institutional capacity, behavioural additionality and networking.

Dissemination of evaluation results and recommendations

- Broaden the use of evaluation results by incorporating the views of the potentially interested audience such as industry, target groups and social communities representatives.
- Introduce the requirement that programme managers report on implementation of recommendations made in the evaluations.
- Produce timely evaluation reports and in a clearly and understandable language to increase impact.

All in all, the effort of synthesis presented in this RTD evaluation toolbox should be interpreted as a move towards introducing good governance practices in science and technology policy.

GLOSSARY

Accountability: The responsibility for the exercise of decision making powers granted through an act of delegation. It generally refers to the responsibility of programme managers to provide evidence on the effective implementation of the intervention.

Additionality: The difference which public intervention makes on the policy intervention recipients.

Alternative hypothesis: The hypothesis that the restriction or set of restrictions to be tested in an econometric model does not hold (denoted by H_1).

Applied research: Research in which commercially useful methods are the subject of examination. Involves activities in which the objective can often be definitively mapped before hand and are of a definitely practical and commercial value.

Attrition: The action or process of gradually reducing the strength or effectiveness of someone through sustained pressure. This undesired effect could arise when individuals and firms receive too many questionnaires and surveys**.**

Audit: Management control function mainly concerned with verifying the legality and regularity of the implementation of resources in a programme and the verification of financial records (financial audit).

Basic research: Applies to research conducted with little or no regard to commercial applications.

Baseline with follow-up data surveys: Consist of information at two points in time: before and after the intervention. These surveys may or may not include data on non–program participants. If the evaluation is based on a simple before-and-after comparison of program participants (reflexive comparison), the results should be interpreted with caution.

Before and after estimator: Is a quasi-experimental design method which compares the performance of key variables during and after a program intervention with those prior to the program. The approach uses statistical methods to evaluate whether there is a significant change in some essential variables over time. This approach often gives biased results because it assumes that had it not been for the program, the performance indicators would have taken their pre intervention values. A particular weakness of this design is the possibility that something else besides the intervention accounts for all or part of the observed difference over time.

Benchmarks: Standards by which the performance of an intervention can be assessed in a non-arbitrary fashion. An obvious way of deriving benchmarks would be to examine the intervention's objectives as expressed by expected outputs, results and outcomes. Ideally, benchmarks should allow to compare the performance of an intervention with that of other policy instruments in the same field of action or in a related one.

Benefits: Net programme outcomes, usually translated into monetary terms. Broad sustainable changes looked for by a program.

Bibliometrics: Is the application of quantitative analysis of scientific literature. Impact factors are among the most widely used bibliometric indicators used to evaluate the quality of a research institution.

Case study designs: A class of evaluation designs in the descriptive rather than the causal approach. It is often the case that an evaluation design will be based on an in-depth study of one or more specific cases or situations. It involves the examination of a limited number of specific cases or projects which the evaluator anticipates will be revealing about the programme as a whole. Case studies tend to be appropriate where it is extremely difficult to choose a sample large enough to be statistically generalisable to the population as a whole; where generalisation is not important; where in-depth, usually descriptive data is required; and where the cases or projects to be studied are likely to be quite complex.

Certainty-Equivalent: A certain outcome that an individual values equally to an uncertain outcome. For a risk-averse individual, the certainty-equivalent for an uncertain set of benefits may be less than the mathematical expectation of the outcome.

Comparative change design: An example of a quasi-experimental design in which any known or recognisable difference between the programme and control groups is taken into account in the statistical analysis. The problems with this design are, firstly, that there may be some other factor which explains some or all of the variation in the intervention and in the observed effects, and, secondly, that there may be initial differences between the programme and control groups which have an influence on observed effects and which can therefore become confounded with the influence of the programme on these effects (selection bias).

Computable general equilibrium models (CGEs) Attempt to contrast outcomes in the observed and counterfactual situations through computer simulations. These models seek to trace the operation of the real economy and are generally based on detailed social accounting matrices collected from data on national accounts, household expenditure surveys, and other survey data. CGE models do produce outcomes for the counterfactual, though the strength of the model is entirely dependent on the validity of the assumptions. This can be problematic as databases are often incomplete and many of the parameters have not been estimated by formal econometric methods. CGE models are also very time consuming, cumbersome, and expensive to generate.

Consumer surplus: The maximum sum of money a consumer would be willing to pay to consume a given amount of a good, less the amount actually paid. In a graph representing the consumer's demand for the good as a function of its price, consumer surplus is represented by the area between the demand curve and the price curve.

Contingent valuation: The use of questionnaires about valuation to estimate the willingness of respondents to pay for public projects or programs.

Control group: A group of individuals in an evaluation who share the same characteristics as the treatment (or participant) group, but have not been exposed to the programme intervention. In this situation the systematic differences between the two groups may be attributed to the effects of the intervention.

Cost-Benefit analysis: A judgemental technique which compares all social and private costs and benefits of a programme with the aim to establish its economic efficiency and whether the benefits exceed the costs. The social costs and social benefits including side effects are generally computed in an indirect manner and converted into monetary terms to facilitate the comparison with the private costs and benefits of the programme.

Cost-Delivery analysis: Study of the cost incurred to deliver a specified set and quantity of goods and services (outputs) to a targeted population.

Cost-Effectiveness analysis: A judgemental technique which compares the costs of alternative means of achieving the same bundle of benefits. It uses the same principles of cost-benefit analysis although outcomes can be estimated in non-monetary quantitative units.

Counterfactual situation: The situation which would have arisen had the intervention not taken place. In order to derive the counterfactual situation we need an evaluation design. Except for the theoretical case of the ideal experimental design, we can never know the counterfactual situation with certainty. Real world evaluation designs tend to be based on an estimate of the counterfactual derived either from comparing subjects who were exposed to an intervention with a comparison group who were not exposed, or from examining subjects before and after exposure.

Co-word analysis: This method identifies keywords and relates the contents of papers with other scientific publications, grouping papers to show the structure and dynamics followed by science and technology. It is used for mapping scientific fields and for detecting new emerging fields.

Criterion-population design: An example of a quasi-experimental design, which attempts to improve on the comparative change design. In the latter, the programme and control groups are two distinct groups drawn from a hypothetical larger population. In the criterion population design, however, the hypothetical population is identified and used for the comparison group. In this case, the possibility of selection bias is confined to just one group - the programme group. This design is particularly appropriate where the evaluator cannot easily create a control group but does have access to information about the larger population from which the programme group is drawn.

Cross section estimator: Uses the outcomes of non-participants to proxy for what participants would have experienced, had they not participated in the programme.

Cross-section data: Contain information from program participants and non-participants at only one point in time, after the intervention. Evaluations using cross-section data usually cost less than studies using information from more than one point in time, but the results tend to be less reliable, except for experimental designs.

Crowding out: The effect that accounts for the fall in private R&D investment in the supported firm resulting from an increase in public R&D support.

Data: Observable facts which can be used as a basis for inference.

Data analysis: The main techniques used to interpret information about an intervention for use in an evaluation are statistical analysis, the use of models, non-statistical analysis and judgement techniques, such as cost-benefit analysis, cost effectiveness analysis and multi-criteria analysis.

Data collection: The main techniques used to gather information about an intervention for use in an evaluation are surveys, case studies, natural observations, expert opinion, reviews of programme documents and literature reviews.

Data dissaggregation: The division of the target population into groups of equal size to see how indicators vary across the population and to better grasp the impact of the programme.

Deadweight: Are the effects which would have arisen even if the intervention had not taken place. Deadweight usually arises as a result of inadequate delivery mechanisms which fail to target the intervention's intended beneficiaries sufficiently well. As a result, other individuals and groups who are not included in the target population end up as recipients of benefits produced by the intervention. Deadweight is really a special case of programme inefficiency.

Delivery mechanisms: The organisational arrangements which provide the goods and services funded by the intervention to its intended beneficiaries, i.e. its target population.

Demand curve: For a given good, the demand curve is a relation between each possible price of the good and the quantity that would be bought at market sale at that price.

Descriptive statistics: Is a commonly used data analysis technique used to summarise in a concise and revealing manner the information provided by the data.

Deterministic: Not random. A deterministic function or variable often means one that is not random, in the context of other variables available.

Difference in difference or double difference estimator: This estimator uses a comparison group of non-participants to remove common trends in outcomes. The application of the method compares a treatment and comparison group (first difference) before and after a program (second difference). Comparators should be dropped when propensity scores are used and if they have scores outside the range observed for the treatment group.

Discount factor: The factor that translates expected benefits or costs in any given future year into present value terms. The discount factor is equal to $1/(1+r)^t$ where *r* is the nominal interest rate and *t* is the number of years from the date of initiation for the program or policy until the given future year.

Discount rate: The interest rate, *r* used in calculating the present value of expected yearly benefits and costs. The discount rate used should reflect not only the likely returns of funds in their best relevant alternative use (i.e., the opportunity cost of capital or "investment rate of interest"), but also the marginal rate at which savers are willing to save in the country (i.e., the rate at which the value of consumption falls over time).

Displacement: Displacement and substitution are two closely related terms which are used to describe situations where the effects of an intervention on a particular individual, group or area are only realised at the extent of other individuals, groups or areas.

Dissemination: The set of activities implemented to make available to a wider audience the knowledge generated in an evaluation.

Double-loop learning: A type of feedback, in which the information compiled by an evaluation is used to call into question the very existence of an intervention or to bring about major changes in its basic orientations. Double-loop learning is almost always the result of summative evaluations.

Dummy variable: A binary variable designed to take account of exogenous shifts or changes of slope in an econometric relationship.

Dynamic optimisation: Maximisation problem to which the solution is a function.

Econometric model: An economic model formulated so that its parameters can be estimated making the assumption that the model is correct.

Effectiveness: Determines to what extent the intervention's impacts contributed to achieving its specific and general objectives

Efficiency: Establishes how economically the intervention's inputs have been converted into outputs and results.

Elasticity: A measure of the percentage change in one variable in response to a percentage change in another variable.

Embodied technological progress: Technical progress which cannot take place unless it is embodied in new capital.

Endogenous: A variable is endogenous in a model if it is at least partly function of other parameters and variables in a model.

Equilibrium: Some balance that can occur in a model, which can represent a prediction if the model has a real world analogue.

Estimation: The quantitative estimation of the parameters of economic models through the statistical manipulation of data.

Estimator: A function of data that produces an estimate for an unknown parameter of the distribution that produced the data.

Evaluability assessment: An attempt to determine whether or not the questions raised by a given analytical agenda for an evaluation are at all answerable by an evaluator using appropriate research methods.

Evaluation: Is the periodic assessment of a project's relevance, performance, efficiency, and impact (both expected and unexpected) in relation to stated objectives. It generally involves an in-depth study using research procedures in a systematic and analytically defensible fashion to form a judgement on the value of an intervention. The process is designed to assess the effectiveness of the project in attaining its originally stated objectives, and the extent to which observed changes are attributable to the project.

Evaluation design: A model which is used to describe an intervention and provide evidence on the effects which may be attributable to it. Evaluation designs are either causal or descriptive in nature. A given design should lead to the choice of one or more data analysis and collection techniques.

Evaluation project: A sequence of logical steps starting out from the formulation of problems and interests motivating the evaluation to arrive at a series of questions that can be addressed in an analytically acceptable way. The aim is to establish a work plan setting out a framework in which the evaluation proper is to be conducted and then to choose the evaluator. There are seven steps involved in elaborating an evaluation project: (i) identifying the goals of the evaluation; (ii) delineating the scope of the evaluation; (iii) drawing up the analytical agenda; (iv) setting benchmarks; (v) taking stock of available information; (vi) mapping out the work plan; and, (vii) selecting the evaluator.

Evaluation report: The end product of an evaluation, the evaluation report must follow a logical structure and meet the needs of the evaluation sponsors and the principal stakeholders. Evaluation reports must include an executive summary of no more than five pages in length. The structure of the expected report is usually specified by the sponsors in the terms of reference.

Evaluation sponsors: The entity responsible for launching the evaluation of an intervention.

Ex-ante evaluation or Appraisal: An evaluation conducted before the implementation of an intervention.

Exogenous: A term which describes anything predetermined or fixed in a model.

Expected value: The expected value of a random variable is the mean of its distribution.

Ex-post evaluation: An evaluation conducted either on or after completion of an intervention.

Ex post facto design: An example of a descriptive design, which can be used where the evaluator cannot select who is to be exposed to the programme, and to what degree. These designs have been used to examine interventions with universal coverage.

Excess burden or Deadweight loss: Unless a tax is imposed in the form of a lump sum unrelated to economic activity it will affect economic decisions on the margin. Departures from economic efficiency resulting from the distorting effect of taxes are called excess burdens because they disadvantage society without increasing fiscal income receipts.

Executive summary: Is a document which facilitates the dissemination of findings by summarising the contents of the evaluation report. It can be distributed as a stand-alone document.

Experimental or Randomized control designs: Is an evaluation design in which the selection for the treatment (those exposed to the intervention) and control groups (those for whom the intervention is withheld) is random within some well-defined set of individuals. In this case there should be no difference with regard to expected values between the two groups besides the fact that the treatment group had access to the programme intervention. Since programme participants are selected randomly, any difference with non–program participants is due to chance. For this reason, experimental designs are usually regarded as the most reliable method and the one yielding the easiest-to-interpret results. In practice, however, this type of evaluation design can be difficult to implement, not least because on ethical and political grounds it might be difficult to withhold benefits from equally eligible individuals. Differences may still appear between the two groups due to sampling error; the larger the size of the treatment and control samples the less the error.

Expert or Review groups: This methodology is a judgement tool which brings together a group of independent eminent scientists and/or research managers not necessarily experts in the field being evaluated which provide an assessment based on their broad views and expertise.

Expert opinion: A data collection technique, similar to a survey, which relies on the necessarily subjective views of experts in a particular field.

External economy or externalities: A direct effect, either positive or negative, on an individual welfare or firm profit arising as a by product of some other individual’s or firm's activity.

Externality: An effect of a purchase or use decision by one set of parties on others who did not have a choice and whose interests were not taken into account.

External evaluation: An evaluation which is performed by persons outside the organisation responsible for the intervention itself.

External validity: The confidence about whether or not the conclusions achieved in the intervention can be generalised to apply within a broader framework. A threat to external validity is an objection that the evaluation design does not allow causal inference about the intervention to be generalised to different times, places or subjects to those examined in the evaluation.

Feedback: The process by which the information compiled by an evaluation is used by decision-makers to either change the way in which an intervention is implemented, or to bring about a more fundamental change in the basic orientations of the intervention, including calling into question its very existence.

Focus groups: Small group discussions led by a trained moderator who introduces a topic and facilitates participation by all group members.

Formative evaluation: An evaluation concerned with examining ways of improving and enhancing the implementation and management of interventions. Formative evaluations tend to be conducted for the benefit of those managing the intervention with the intention of improving their work.

Foresight or Delphi: A technique which can be used to systematise expert opinions on future trends of science and technology. Experts are consulted separately in a number of different rounds. In each successive round, each individual is told the views of the other experts in the previous round. This technique can be used to arrive at a consensus while maintaining diversity.

General objectives: The desired effects of an intervention expressed in terms of outcomes, i.e. the longer-term impact of the intervention on society.

General purpose technology: Is a technological breakthrough that has an effect over an entire economic system, through various sectors in the economy

Goals: The broad aims the project or programme wants to achieve. The significant, longer-term changes that strategy policy makers expect to occur as a result of the project or programme.

Ideal experimental design: A theoretical way of deriving the counterfactual situation, and hence the net impact of an intervention. It involves comparing two identical groups which only differ on the exposure to the policy intervention. Differences between the group which has been exposed (the treatment group) and the group which has not (the control group) are then attributable to the intervention. In the real world, this design does not exist since we can never be absolutely certain that the two groups are identical in all other respects. The potential non-equivalence of the two groups weakens the validity of any causal inference about the intervention. A number of real world evaluation designs are available which each have their own strengths and weaknesses.

Impact evaluation: Is the systematic identification of the effects —positive or negative, intended or not—on individuals, households, institutions, and the environment caused by a given policy intervention (program/policy).

Impact indicators: Are specific information or evidence that can be gathered to measure progress toward program goals and objectives or to measure effectiveness of program activities when direct measurement is difficult or impossible. They measure the highest objectives seek by the policy intervention, for example improved welfare and living standards.

Impacts: A general term used to describe the effects of a policy intervention on society. are those longer-term outcomes that are consistent with agreed upon policy intervention goals and that would not have occurred without having conducted the policy intervention. They evidence the changes in the condition of the target population which generally reflect the primary objectives of the intervention.

Incidence: The ultimate distributional effect of a tax, expenditure, or regulatory program.

Indicator: Is an objectively verifiable measurement which reflects the activity, assumption, or effect being measured. In evaluation, indicators are a characteristic or attribute which can be measured to assess an intervention in terms of its outputs or impacts. I used to measure the progress of policy intervention towards its defined goals.

Inflation: The proportionate rate of change in the general price level, as opposed to the proportionate increase in a specific price. Inflation is usually measured by a broad-based price index, such as the implicit deflator for Gross Domestic Product or the Consumer Price Index.

Input assumptions : The expectations regarding the effectiveness and quality of the project inputs.

Inputs: Measures the resources in the form of material, goods and actions, human and financial resources allocated for the implementation of a policy intervention

Input-output models: Analyse the impact of an intervention using input output matrices which capture the linkages between the different parts of an economy, as the inputs of one industry can be thought of as the outputs of other industries.

Instrumental variables or statistical control: Methods in which one uses one or more variables that matter to participation but not to outcomes given participation. This identifies the exogenous variation in outcomes attributable to the program, recognising that its placement is not random but purposive. The "instrumental variables" are first used to predict program participation; then one sees how the outcome indicator varies with the predicted values.

Interim or intermediate evaluation: An evaluation conducted during the implementation of an intervention as a first review of progress, a prognosis of a project's likely effects, and as a way to identify necessary adjustments in project design.

Internal evaluation: An evaluation which is performed by members of the organisation responsible for the intervention itself.

Internal rate of return (IRR): The discount rate that sets the net present value of a project stream of net benefits equal to zero. When the IRR equals or exceeds the appropriate discount rate, then the project’s NPV will be positive and the project will be also acceptable from the NPV perspective.

Internal validity: The confidence one can have in one’s conclusions about what the intervention actually did accomplish. A threat to internal validity is an objection that the evaluation design allows the causal link between the intervention and the observed effects to remain uncertain. It may be thought of as a question of the following nature: could not something else besides the intervention account for the difference between the situation after the intervention and the counterfactual?

Intervention: A generic term used to cover all types of public actions (e.g. policy, programmes, projects)

Intervention logic or logical framework (Log-Frame): Helps to define the expected causal links existing in an intervention, from its inputs to the production of its outputs and subsequently to its outcomes and impacts on society and the economy. The examination of the intervention logic will be of central importance in most evaluations helping to clarify its objectives. The evaluator needs to establish how the intervention achieves its specific objectives, and how the specific objectives contribute to the attainment of the general objectives and the intervention seeks to address the target beneficiary groups.

Key informant interviews: A face-to-face meeting between a trained interviewer and a person selected to represent a certain group whose knowledge, attitudes or practices are being monitored or evaluated, or with a person likely to offer informed views.

Life cycle cost: The overall estimated cost for a particular program alternative over the time period corresponding to the life of the program, including direct and indirect initial costs plus any periodic or continuing costs of operation and maintenance.

Linear model: An econometric model is linear if it is expressed in an equation which the parameters enter linearly, whether or not the data require non-linear transformations to get to that equation.

Literature review: A data collection technique which enables the evaluator to make the best use of previous work in the field under investigation and hence to learn from the experiences and findings of those who have carried out similar or related work in the past.

Logit model: A particular case of the probit or linear probability model where the values of the dependent variable are constrained to lie within the 0 to 1 probability limits.

Macroeconomic models: Examine the impact of an intervention in models describing the behaviour of the economy as a whole and the evolution of important macroeconomic variables such as investment, employment, productivity and the trade balance.

Management information system: A generally computer assisted tool used in programme management to compile and analyse programme monitoring data.

Market failure: A situation, usually discussed in a model not in the real world, in which the behaviour of optimising agents in a market would not produce a Pareto optimal allocation. The sources of market failures include monopoly and externalities where producers and consumer have respectively incentives to under produce and to price above marginal cost and to buy less Pareto optimal allocation.

Markov process: A stochastic process where all the values are drawn from a discrete set. In a first-order Markov process only the most recent draw affects the distribution of the next one; all such processes can be represented by a Markov transition density matrix.

Markov transition matrix: A square matrix describing the probabilities of moving from one state to another in a dynamic system.

Matching: Is a quasi-experimental method for constructing a comparison group consisting of selecting non–program participants comparable in essential characteristics to participants, on the basis of either a few characteristics or a number of them, using statistical techniques. Evaluations using matching methods are often easier and cheaper to implement than experimental designs, but the reliability of results is lower and the interpretation of results is more difficult.

Mean: The most commonly used descriptive statistic for summarising data. It is a measure of central tendency of a variable.

Microeconomic models: Examine the impact of an intervention focusing on the behaviour of individuals, households and firms in specific industries and markets using equations which represent the supply and demand functions for a particular good or services.

Monitoring: The continuous assessment of project implementation in relation to agreed schedules, and of the use of inputs, infrastructure, and services by project beneficiaries through the ongoing collection and review of information on project implementation, coverage, utilisation of inputs and objectives achievements. Provides managers and other stakeholders with continuous feedback on programme or policy implementation. Permits an early identification of programme deviations from operational objectives facilitating a timely adjustments to project operation. Monitoring also compiles the programme data which is used in evaluation.

Modified peer review: This methodology constitutes an expanded version of traditional peer review which incorporates the inputs of the potential users of scientific and technological research

Monte Carlo simulations: These are data obtained by simulating a statistical model in which all parameters are numerically specified.

Multi-criteria analysis: A decision-making tool which can be adapted to form judgements about interventions. Multi-criteria analysis allows to formulate judgements on the basis of multiple criteria, which may not have a common scaling and which may differ in relative importance.

Natural observations: A data collection technique in which the evaluator makes on-site visits to locations where the intervention is in operation and directly observes what is happening. Observational data can be used to describe the setting of the intervention, the activities which take place in the setting, the individuals who participate in these activities and the meaning of these activities to the individuals.

Needs: The socio-economic problems which an intervention aims to address, expressed from the point of view of its target population.

Net Present Value (NPV): Is a method that calculates the present value of the net benefits of a project as the difference between the discounted present value of benefits and the discounted present value of costs. Two conditions must be satisfied if a project is to be acceptable on economic grounds: (a) the NPV of the project must be positive when discounted at an appropriate rate; and (b) the project's expected NPV must be equal or larger than the NPV of other alternatives.

Nominal Interest Rate: An interest rate that is not adjusted to remove the effects of actual or expected inflation. Market interest rates are generally nominal interest rates.

Nominal Values: Economic units measured in terms of purchasing power at a certain time period. Nominal value reflect the effect of price inflation.

Non experimental designs: Are used in situation where it is not possible to select an ideal control group or a comparison group that matches the treatment group. Program participants are compared to non participants using statistical methods to account for differences between the two groups. Under this situation, regression analysis permits to control for the characteristics of the participants. This evaluation design is relatively inexpensive and easy to implement, but the interpretation of results is not direct and results may be less reliable than experimental or quasi-experimental designs.

Non-statistical analysis: A general term used to describe the analysis of mainly qualitative data which is typically used in conjunction with statistical analysis (of either qualitative or quantitative data). Usually, this includes an assessment of the reliability of any findings derived from such methods.

Null hypothesis: The hypothesis being tested in an econometric model . The hypothesis that the restriction or set of restrictions to be tested does in fact hold (denoted by H_0).

Objectives: The desired effects of an intervention commonly defined in terms of operationalised goals which specify the desired effect of an intervention in terms of output and the level of change expected.

Omitted variable bias: Standard expression for the bias that appears in an estimate of a parameter if the regression run does not have the appropriate form and some relevant variables are not included.

Opportunity cost: The maximum worth of a good or input among possible alternative uses.

Ordinary least squares: The standard linear regression procedure. The parameters are estimated from the observed data applying the linear model $y = X\beta + \varepsilon$ where *y* is the dependent variable, *X* is a matrix of independent variables, β is a vector of parameters to be estimated, and ε is a vector of errors with zero mean.

Outcomes: The longer-term impact that occur as a result of the intervention, usually expressed in terms of broad socio-economic consequences. It includes the intermediate effects, often behavioural, resulting directly from project outputs that may be necessary to achieve a desired impact.

Output Assumptions: Expectations regarding the ways goods and services (outputs) will be used by the target population.

Output indicators: Measure the quantity of goods and services produced by an intervention to the target population. It shows the immediate physical and financial outputs of the project: physical quantities, organisational strengthening, initial flows of services

Panel or longitudinal data sets: Data from a small number of observations over time on a usually large number of cross sectional units like individuals, households, firms, or governments. They include information on the same individuals (or other unit of analysis) at least at two different points in time, one before the intervention (the baseline) and another afterwards. Panel data sets are highly valued for program evaluation, but they can be expensive and require substantial institutional capacity.

Parametric: A function is parametric in a given context if its functional form is known to the researcher.

Pareto Optimal allocation: In an endowment economy, that situation where no other allocation of the same goods would be preferred by all the agents.

Path dependency: The view that technological change in a society depends quantitatively and/or qualitatively on its own past.

Peer review: Methodology to evaluate the quality of scientific output using the perception scientists have of the scientific contributions made by other peers.

Performance audit: Conceptually closer to evaluation than traditional audit, performance audit is strongly concerned with questions of efficiency (of an intervention's direct outputs) and good management. Performance audit and evaluation share the same aim of improving the quality of programmes, but evaluation goes much further. It also looks at issues such as sustainability, relevance and the longer-term consequences of a programme.

Policy: A set of activities, which may differ in type and have different direct beneficiaries, directed towards common general objectives. Policies are not delimited in terms of time schedule or budget.

Population: In statistics, the entire aggregate of individuals or subjects, from which samples can be drawn.

Primary data: Refers to data compiled directly from original sources or collected first hand.

Probability sampling: A statistical technique used to obtain samples from a given population, whereby every unit in the population has a known, non-zero probability of being selected for inclusion in the sample. The conclusions from this type of sample can then be projected, within statistical limits of error, to the wider population.

Probit or Linear probability model: An econometric model in which the dependent variable is a dummy binary variable which can only take the values one or zero and is expressed as a linear function of one or more independent variables.

Production function: Describes a mapping from quantities of inputs to quantities of an output as generated by a production process.

Productivity: *A measure relating a quantity or quality of output to the inputs required to produce it*

Programme A set of organised but often varied activities (a programme may encompass several different projects, measures and processes) directed towards the achievement of specific objectives. Programmes have a definite time schedule and budget.

Programme document reviews: A data collection technique based on reviewing general programme files, financial and administrative records and specific project documents.

Programme group: A group of entities which have been exposed to an intervention. The programme group can be compared with the control group in order to determine whether systematic differences between the two groups may be attributed to the effects of the intervention.

Project: A single, non-divisible public intervention directed towards the attainment of operational objectives, with a fixed time schedule and a dedicated budget.

Propensity score: An estimate of the probability given observed characteristics that an entity would undergo the treatment. This probability is itself a predictor of outcomes sometimes.

Propensity score matching: In which the comparison group is matched to the treatment group on the basis of a set of observed characteristics or by using the propensity score. The closer the propensity score, the better the match. A good comparison group comes from the same economic environment and was administered the same questionnaire by similarly trained interviewers as the treatment group.

Proxy indicator: A measurement used as a substitute when true indicators are too difficult to measure directly.

Pure research: Is research without specific practical ends. It results in general knowledge and understanding of nature and its laws.

Qualitative studies: Assess conditions of the population (often identifying vulnerable subgroups) before, during, and after adjustment policies are implemented through focus groups, interviews, and other qualitative techniques.

Quasi-experimental designs: Evaluation designs that eliminate competing explanations of project effects without the benefit of a true control group. When randomisation is not feasible, a comparison group can still be constructed. Contrary to true experimental designs which require randomisation in the selection of treatment and control groups, the comparison group is defined through a non random process.

Real interest rate: An interest rate that has been adjusted to remove the effect of expected or actual inflation. Real interest rates can be approximated by subtracting the expected or actual inflation rate from a nominal interest rate. Real interest rates determine the investment levels.

Real or Constant values: Economic units measured in terms of constant purchasing power. A real value is not affected by general price inflation. Real values can be estimated by deflating nominal values with a general price index, such as the implicit deflator for Gross Domestic Product or the Consumer Price Index.

Reflexive comparisons: Is a quasi-experimental method for constructing a comparison group where the counterfactual is constructed on the basis of the situation of program participants before the program. A baseline survey of participants is done before the intervention and a follow-up survey is done after the policy intervention. The baseline provides the comparison group, and the impact is measured by the change in outcome indicators before and after the intervention. Program participants are compared to themselves before and after the intervention and function as both treatment and comparison group. This type of design is particularly useful in evaluations of full-coverage interventions such as nationwide policies and programs in which the entire population participates and there is no scope for using a control group. There is, however, a major drawback with this method: the situation of program participants before and after the intervention may change owing to myriad reasons independent of the program. Unless they are carefully done, reflexive comparisons may not be able to distinguish between the program and other external effects, thus compromising the reliability of results.

Regression analysis: A statistical inference technique which can be used to establish the significance of any correlation (association) between variables of interest. In regression analysis, we attempt to establish whether the variation in one variable (known as the dependent variable) can be explained in terms of the variation in one or more independent variables. The dependent variable is often quantitative. Special techniques are available, however, to deal with situations in which the dependent variable is qualitative.

Relative Price: A price ratio between two goods.

Reporting: Takes place when the evaluator transmits the evaluation report (usually in the form of a document, or else through some audio-visual presentation) to the sponsors and when they, in turn, transmit a copy (or a summary thereof) to other interested parties.

Research synthesis: Provides an overview of the current state of knowledge about a socio-economic problem and about the rationale for policy intervention. It permits to indicate the principal information gaps prior to the intervention and therefore help to identify the data collection needs and analysis tasks to be undertaken by the evaluation.

Risk analysis: Is a tool for estimating the probability of project outcomes (or costs) based on estimates of the probability of project parameters. A series of random selections of values from the probability distribution of parameters are made to generate a probability distribution of project effects or costs (monte carlo analysis), from which the mean and variance of project cost-effectiveness or cost-benefit can be determined.

Sample: Subset of individuals or items selected from a given population which is used to represent the entire population so that properties and parameter from the parent population may be estimated or hypothesis tested.

Secondary data: Is data produced after extensive manipulation and interpretation.

Scientific studies: Whereas scientists may undertake research in order to expand the sum of human knowledge and frequently confine themselves to one highly specialised discipline, evaluations are undertaken for more practical reasons. Evaluations aim to inform decisions, clarify options, reduce uncertainties and generally provide information about programmes within their own specific contexts.

Scope: The field of investigation of an evaluation. Typically, this has to be defined from an institutional (EU versus national or local level), temporal (period review) and geographical (part of the EU territory) point of view. In addition, one has to identify the key evaluation issues (relevance, efficiency, effectiveness, utility, sustainability) which will be examined.

Selection bias: Occurs when the differences between the control group and the programme group are due to the initial differences in their characteristics rather than the effects of the intervention being evaluated.

Sensitivity analysis: Is a means of exploring how plausible changes in assumptions about uncertain variables affects the conclusions. The cost-effectiveness (or cost-benefit) of the alternative interventions is recalculated for each new assumption (or set of new assumptions) to see if this affects the conclusions. If it does not, the conclusions are considered robust. If it does, then it is important to specify the conditions under which the different conclusions will hold.

Shadow price: An estimate of what the price of a good or input would be in the absence of market distortions, such as externalities or taxes. For example, the shadow price of capital is the present value of the social returns to capital before corporate income taxes measured in units of consumption.

Single-loop learning: A type of feedback, in which the information compiled by an evaluation is used to bring about changes in the way an intervention is implemented.

Specific objectives: The desired effects of an intervention expressed in terms of results, i.e. the initial impact of the intervention on society.

Stakeholders: Individuals or organisations directly or indirectly associated with or affected by the implementation of the policy intervention and who are likely to have an interest in its evaluation (for example programme managers, policymakers, programme target population).

Standard deviation: A commonly used descriptive statistic, it provides a measure of dispersion of a variable around its mean value.

Statistical controls: Specifies regressions that control for the differences in initial conditions between participants and non–participants in policy intervention. The approach identifies the differences between participants and non–participant entities in the pre intervention period and then controls for these differences to identify the impact of the intervention in performance after the intervention.

Statistical inference: Is a data analysis technique used to test for relationships among variables or generalise findings to a wider population.

Steering group: Part of the management structure for an evaluation, a steering group allows other services (and possibly other stakeholders from outside the Commission) to contribute to the development of the evaluation project.

Stochastic process: Is an ordered collection of random variables.

Subjective data: Data compiled incorporating issues related to personal feelings, attitudes and perceptions.

Summative evaluation: An evaluation concerned with determining the essential effectiveness of programmes. Summative evaluations tend to be conducted for the benefit of external actors (groups who are not directly involved in the management of a programme), for reasons of accountability or to assist in the allocation of budgetary resources.

Sunk cost: A cost incurred in the past that will not be affected by any present or future decision. Sunk costs should be ignored in determining whether a new investment is worthwhile.

Surveys: A widely-used technique for collecting data from a sample drawn from a given population. Surveys are often based on probability sampling, and survey information is usually obtained through structured interviews or self-administered questionnaires. Cross-sectional surveys involve measurements made at a single point in time. Panel surveys involve measurements acquired at two or more points in time.

Sustainability: To what extent can the programme's positive impacts (as measured by its utility) be expected to last after the intervention has been terminated?

Target population: The intended beneficiaries (individuals, households, groups, firms) of an intervention. An intervention may have more than one target population. This term should be distinguished from "population" in the statistical sense.

Targets: Are quantifiable levels of the indicator that a policy intervention wants to achieve at a given point in time.

Technical change: A change in the amount of output produced from the same inputs. Such a change is not necessarily technological; it might be organisational, or the result of a change in a constraint such as regulation, prices, or quantities of inputs.

Technological development: Is the approach through which specific products or process are being designed or tested.

Terminal evaluations: Conducted at the end of a project, are required for project completion reports. They include an assessment of a project's effects and their potential sustainability.

Terms of reference: The terms of reference outline the work to be carried out by the evaluator, the questions to be dealt with and the time schedule. They allow the sponsors of the evaluation to define their requirements and allow the evaluator to understand clearly what is expected of the work to be undertaken (including, often, the structure of the expected evaluation report). Clearly defined terms of reference are vitally important where an evaluation is to be conducted by an external expert, and can also be of tremendous use when it is to be performed in-house.

Thematic evaluation: An evaluation which focuses on one or more themes which are common to several different interventions (programmes or other activities).

Tobit model: An econometric model in which the dependent variable is censored because values of the variable below zero are not observed.

Transfer payment: A payment of money or goods. A pure transfer is unrelated to the provision of any goods or services in exchange. Such payments alter the distribution of income, but do not directly affect the allocation of resources on the margin.

True experimental designs The best real world approximations to the ideal experimental design, in which the evaluator tries to ensure the initial equivalence of the programme and control groups by creating them beforehand through random assignment. Although causal inference based on such designs is usually very strong, true experimental designs are difficult to administer and implement. Also referred to as “randomised experimental designs”.

Utility: Widely construed in economics to be synonymous of welfare, satisfaction and occasionally welfare.

Variance: A descriptive statistic which provides a measure of dispersion. It is computed as the average of squares of the distances from the values drawn from the mean of the distribution.

Welfare function: For an individual, the relationship between his level of well being, welfare or utility and the elements that contribute to it.

White noise: A description of variation which is purely random and contains no systematic elements.

Willingness to pay: The maximum amount of money an individual would be willing to give up in order to obtain the provision of a good or service.

With and without comparisons: Which compare the behaviour in key variables in a sample of programme supported geographical areas with their behaviour in non supported geographical areas (a comparison group). This is an approach to the counterfactual question, using the experiences of the comparison group as a proxy for what would otherwise have happened in the supported geographical areas. It is, however, quite difficult to achieve a true comparison group. An important limitation of this approach is that it assumes that only the adoption of a particular policy or program distinguishes supported geographical areas from non–program areas and that the external environment affects both groups equally.

Within estimator or Fixed effects estimator: A method of estimating parameters from a panel data set. The fixed effects estimator is obtained by OLS on the deviations from the means of each unit or time period. This approach is relevant when one expects that the averages of the dependent variable will be different for each cross-section unit, or each time period, but the variance of the errors will not.

ANNEX

List of the participants in the Final Conference
"Socio-Economic Evaluation of Public RTD Policies (EPUB)"
BRUSSELS, AUSTRIA HOUSE, 21 MAY 2002

Name	Surname	Organisation
From national and regional governments		
Pascale	Dengis	Ministry of the Flemish Community, Belgium
Lutz	Gros	Federal Ministry of Education and Research, Germany
From universities and research institutions		
Jonathan	Eaton	Boston University, Boston, USA
Eamon	Kinsella	The Circa Group, Dublin, Ireland
Giorgio	Sirilli	National Research Council, Italy
Henri	Capron	Dulbea, Free University of Brussels, Belgium
Michele	Cincera	Dulbea, Free University of Brussels, Belgium
Stefan	Kuhlmann	Fraunhofer ISI, Karlsruhe, Germany
Nicholas	Vonortas	George Washington University, USA
Wolfgang	Polt	Joanneum Research, Vienna, Austria
Jaime	Rojo	Joanneum Research, Vienna, Austria
Klaus	Zinöcker	Joanneum Research, Vienna, Austria
Spyros	Arvanitis	KOF, Swiss Federal Institute of Technology, Zurich, Switzerland
Robin	Cowan	MERIT, Maastricht, The Netherlands
Luke	Georghiou	PREST, Manchester, UK
John	Rigby	PREST, Manchester, UK
Elliot	Stern	Tavistock Institute, London, UK
Michael	Stampfer	Technology Impulse Society, Austria
Gernot	Hutschenreiter	Austrian Institute of Economic Research
Ken	Guy	Wise Guys Ltd., UK
Georg	Licht	Centre for European Economic Research, Mannheim, Germany
From the European Commission		
Anne	O'Brien	DG Information Society
Elie	Faroult	DG Research, Development, Technology and Innovation
Neville	Reeve	DG Research, Development, Technology and Innovation
Liam	O'Sullivan	DG Information Society
Alberto	Silvani	DG Research, Development, Technology and Innovation
Graham	Stroud	DG Research, Development, Technology and Innovation
Paras	Caracostas	DG Research, Development, Technology and Innovation
Mike	Rogers	DG Research, Development, Technology and Innovation
Jérôme	Poussielgue	DG Research, Development, Technology and Innovation
Jean Paul	Joulia	DG Trans-European Networks
Gustavo	Fahrenkrog	DG Joint Reseach Centre, IPTS
Alexander	Tübke	DG Joint Reseach Centre, IPTS

CONTACT INFORMATION

EDITORS

European Commission
Joint Research Centre - Institute for
Prospective Technological Studies
Edificio Expo W.T.C., Calle Inca Garcilaso s/n
E-41092 Sevilla, Spain

Gustavo Fahrenkrog
Tel.: + 34.95.448.84.50
Fax: + 34.95.448.83.59
e-mail: Gustavo.Fahrenkrog@jrc.es

Alexander Tübke
Tel.: + 34.95.448.83.06
Fax: + 34.95.448.83.26
e-mail: Alexander.Tubke@jrc.es

Joanneum Research
Institute of Technology and Regional Policy
Institut für Technologie- und Regionalpolitik
Wiedner Hauptstraße 76
A-1040 Vienna, Austria

Wolfgang Polt
Tel.: + 43.1.581.75.20.2811
Fax: + 43.1.581.75.20.2820
e-mail: wolfgang.polt@joanneum.at

Jaime Rojo
Tel.: + 43.1.581.75.20.2815
Fax: + 43.1.581.75.20.2820
e-mail: jaime.rojo@joanneum.at

Birgit Woitech
Tel.: + 43.1.581.75.20.2812
Fax: + 43.1.581.75.20.2820
e-mail: birgit.woitech@joanneum.at

Klaus Zinöcker
Tel.: + 43.1.581.75.20.2827
Fax: + 43.1.581.75.20.2820
e-mail: klaus.zinoecker@joanneum.at

NETWORK MEMBERS

THE CIRCA GROUP EUROPE	The CIRCA Group Europe Limited Eamonn Kinsella Upper Pembroke Street 26, 2 Dublin, Ireland Phone/fax: +353 1 662 5895 / +353 1 637 3986 e-mail: eamonn.kinsella@circa.ie
SPRU Science and Technology Policy Research	University of Sussex SPRU (Social Policy Research Unit) Pari Patel Sussex House, Falmer, BN1 9 RH, Brighton, United Kingdom Phone/fax: +44 1273 68 67 58 / +44 1273 68 58 65 e-mail: parip@sussex.ac.uk
	The Tavistock Institute of Human Relations Elliot Stern 30 Tabernacle Street, EC2A 4UE, London, United Kingdom Phone/fax: +44 171 417 0407 / +44 171 417 0567 e-mail: e.stern@tavinstitute.org
THE UNIVERSITY of MANCHESTER PREST POLICY RESEARCH IN ENGINEERING, SCIENCE & TECHNOLOGY	The Victoria University of Manchester PREST (Policy Research in Engineering Science and Technology) Luke Georghiou, John Rigby and Marc Boden Oxford Road, M13 PL Manchester, United Kingdom Phone/fax: +44 161 2755921 / +44 161 2731123 e-mail: luke.georghiou@man.ac.uk, john.rigby@man.ac.uk, mark.boden@jrc.es
ZEW Zentrum für Europäische Wirtschaftsforschung GmbH	ZEW (Zentrum europäische Wirtschaftsforschung) (Centre for European Economic Research) Georg Licht and Max Keilbach 103443 L7,1 68034 Mannheim, Germany Phone/fax:+ 49 621 1235177/ +49 621 1235170 e-mail: licht@zew.de, keilbach@zew.de
ULB	Université Libre de Bruxelles Unité d'économie spatiale et de la technologie Dulbea – CP 140 Henri Capron and Michele Cincera Avenue F. Roosevelt 50, 1050 Brussels, Belgium Phone/fax: +32-2 6504141 /+32-2 650 38 25 e-mail: hcapron@ulb.ac.be, mcincera@ulb.ac.be
	Boston University and NBER (National Bureau of Economic Research) Jonathan Eaton 63 Westcliff road, 02493 Weston, MA, United States Phone/fax: +1 617 353 4142 /+ 1 617 353 4999 e-mail: jeaton@bu.edu

ISI Fraunhofer Institute Systems and Innovation Research	Fraunhofer Gesellschaft zur Förderung der angewandten Forschung Fraunhofer Institute Systems and Innovation Research (FhG-ISI) Stefan Kuhlmann and Susanne Bührer Breslauer Strasse 48, 76139 Karlsruhe, Germany Phone/fax: +49 721 6809 170/4 / +49 721 6809 260 e-mail: sk@isi.fhg.de, sb@isi.fhg.de
ETH Eidgenössische Technische Hochschule Zürich Swiss Federal Institute of Technology Zurich	ETH (Eidgenössische Technische Hochschule Zürich) Swiss Institute for business cycle research industrial economics research group Heinz Hollenstein and Spyros Arvanitis Weinbergstrasse 35, ETH Zentrum, CH-8092, Zurich, Switzerland Phone/fax: +41 1 632 59 29 / +41 1 632 12 18 e-mail: hollenstein@kof.reok.ethz.ch, arvanitis@kof.reok.ethz.ch
Consiglio Nazionale delle Ricerche	Consiglio Nazionale delle Ricerche Istituto di Studi sulla Ricerca e Documentazione Scientifica Giorgio Sirilli Via Cesare de Lollis 12, 00185 Roma, Italy Phone/fax: +39 0644879214 / +39 064 463836 e-mail: sirilli@www.isrds.rm.cnr.it
MERIT	University of Maastricht MERIT (Maastricht Economic Research Institute on Innovation and Technology) Robin Cowan PO Box 616, Tongerstraat 49, 6200 MD, Maastricht, Netherlands Phone/fax: +31 43 3883891 / +31 43 3216518 e-mail: robin.cowan@algec.unimaas.nl