Workshop Report

Performance Assessment of Public Research, Technology and Development Programmes

Organised by:
The European Commission
in co-operation with the Washington Research Evaluation Network (WREN)

June 17 and 18, 2004

European Commission Charlemagne Building
Rue de la Loi, Brussels

Europe Direct is a service to help you find answers to your questions about the European Union

Freephone number:

00 800 6 7 8 9 10 11

A great deal of additional information on the European Union is available on the Internet.
It can be accessed through the Europa server (http://europa.eu.int).

Cataloguing data can be found at the end of this publication.

Luxembourg: Office for Official Publications of the European Communities, 2004

ISBN 92-894-8167-6

Printed in Luxembourg

PRINTED ON WHITE CHLORINE-FREE PAPER

Foreword

Governments around the world are more aware than ever of the importance of investment in research for innovation and growth and for social and environmental benefits.

The departments responsible for the design and management of publicly funded R&D programmes must therefore improve and strengthen the evaluation of their programmes' performance in order to best meet the requirements for increased transparency, accountability, and cost-effectiveness.

In order to achieve this, we need new evaluation methodologies and tools based on sound scientific foundations; we need comparable indicators and reliable databases to provide high quality evidence for various evaluation tasks, and a worldwide process of collaboration based on regular exchange of experience, good practice and knowledge in the area of research performance assessment.

This workshop was for us an important first step towards these goals. Having as starting point the outcome of the WREN[1] workshop organised by the US Department of Energy in December 2003 in Washington, we supported this workshop in Europe to further explore leading-edge evaluation practice. It has permitted us to take stock of the situation, to understand the new challenges, and to share experiences. Many of the ideas presented here will be used as background to our future work-plans.

This report presents the summaries of and reflects the key points of the very interesting presentations. We feel that its conclusions will contribute to stimulate further debate and the interest of the constituency involved with research performance assessment and to shape future actions in this area.

Peter Johnston

[1] WREN: Washington Research Evaluation Network.

Table of Contents

EXECUTIVE SUMMARY

1. INTRODUCTION

On June 17-18, 2004, the European Commission (DG Information Society and DG Research) and the Washington Research Evaluation Network (WREN) sponsored a two-day high-level policy workshop at the Commission on the evaluation of publicly-funded R&D. The aim was to open up a broad range of issues relating to the design, management and performance-assessment of publicly-funded RTD programmes. The emphasis of the main speakers was on challenging the status quo and gaining a better understanding of current leading-edge practice in the measurement and evaluation of the innovation, competitiveness and societal aspects of the research. Presenters and panelists were drawn from a wide range of interest groups across the research and evaluation community, from both public and private sectors, in Europe, the United States, Canada and Australasia. There were 115 participants in total.

This executive summary sets out the key conclusions of the group and the body of the report contains a synopsis of the speakers' presentations. The secretariat has assembled a compendium of materials from the conference, including a list of participants (Annex 1), and a set of presentation slides is available at

http://europa.eu.int/information_society/programmes/evaluation/ist_rtd/wren/index_en.htm

2. SCOPE AND OBJECTIVES

The aim of the workshop was to provide a forum for high-level experts from the international evaluation community to exchange views on the performance measurement of research programmes and scientists supported by public funds. Its over-arching theme was about capacity-building through networking in the R&D evaluation community. Specific thematic objectives for each day were set as follows:

Day 1: Collaborative networks and dynamics of the innovation system.

Day 2: Linking research to policy objectives.

3. KEY ISSUES AND CONCLUSIONS

A number of horizontal, cross-cutting themes emerged in the course of the workshop:

(1) We urgently need a new theory of innovation. Numerous speakers underlined the need for a much better understanding of the modern innovation 'engine' and its mechanisms before we can measure causalities and visualise new indicators and metrics. In this connection:

– It is important to recognise that the innovation system operates quite differently in different countries and regions.[2] A key question for evaluation policy is the extent to which there are reproducible patterns or core elements, and much more research is needed into this issue. Once we have a better grip on this, there is a rich base of established tools that can be co-opted from econometrics and survey methodologies, and used in complementary ways for analysis and monitoring.

– Historically, there has been an over-focus on product innovation. We now need to reconstitute the old linear, laboratory-to-factory model and open up the concept of innovation to embrace the entire supply chain. Much more work is needed to pinpoint where and how R&D influences corporate performance thresholds at key points along the internationally-extended value chain.

– Much more attention is needed to (i) process R&D, where the body of empirical evidence shows significantly higher returns to investment, and (ii) R&D aimed at the customer end of the supply chain.

– We need a better understanding of the innovation process in the service industries, which have more profile and where the issues are more difficult to get to grips with. Our existing data systems capture very little reliable information about R&D in services. The data is not there to track whatever changes are taking place, we simply miss them.

(2) We need to address the proliferation of evaluation tools and shortcomings in the availability of reliable, internationally-comparable data. As yet, there is no indicator set, or even common language, for cross-border comparison of national efforts. Adding to this, the current tools are often over-complex and slavishly pursue fine granularity at the expense of the big picture. The aim is to catch the first wave effects rather than chasing the tails of the distribution.

(3) We need to be sensitive to the broader geopolitical and social issues in attempting to formulate international conventions. Key factors to consider include:

– Recognising that key performance indicators need to be structured in line with political as well as technology and social objectives.

– Sensitivity to national philosophical differences of governance ethos, which influence and shape the market control and regulatory systems country-by-country, and are often characterised by the dichotomy between principles-based and prescriptive, rule-based approaches.

– The need to distinguish among the different types and stages of research. Evaluation systems need to be flexible and discriminating in order to take account of the different stages and types of ‘R’ and ‘D’. Basic

---

[2] In many country studies, notably those published by the OECD Growth Project, there is a clear correlation between returns to intangible investment as measured by GDP growth and the receptiveness of the business eco-system and its market-regulatory environment.

research, product realisation, and process R&D each require different judgment criteria, which must also be tailored according to size of project and incremental vs radical goals, as well as factors such as the technology domain.

– A path-breaking conclusion of the group was that there is no one-size-fits-all solution for all aspects of the design, organisational and evaluation segments of the innovation supply chain. Nor is there a realistic prospect of a unified holistic model for measuring research outcomes: consensus on a unified global system is a very long corridor from where we are today.

(4) The search for new network theories. There was strong support for stepping up research to advance the frontiers of network analysis, and numerous case studies were presented of individual networks involving the mapping of interactions between economic agents. A recurring theme here was the search for predictive models. Also, the need to infuse ideas from other disciplines, in particular physics and the social sciences, which underlines the need for an inter-disciplinary forum. The leitmotif must be to stretch and extend the boundaries of our understanding of the network phenomenon, and the instruments for dealing with the issues, rather than simply to reduce the issues to what we can comprehend and measure now.

The main presentation by Bill Valdez (WREN) also offered some valuable insights from a recent U.S. study, which indicate that the emergence of entirely new scientific fields can be predicted through scanning for the formation of networks. This raises the prospect of network formation as a leading indicator of the success of public research programmes.

(5) Property rights. As currently constituted, property rights are public goods that are handed out in a fragmented way, especially where multiple jurisdictions are involved. The field is characterised by heterogeneous actors - often competing - and IP rights frequently assume a separate usage that distorts both product markets and the underlying business models of the players. Increasingly, the IPR system is failing to work in the way originally intended and is arguably being abused, especially for patents. As a starting point, we need to address such issues as patent pooling, cross-licensing and blocking patents. There is also a need for new tools to handle collective groups of patents.

(6) A prominent issue in the closing session was the recurring tension between a priori theory development and empirical research work. Ideally, we need a sound theoretical base for practical initiatives - the days of theory-free evaluations are over - and there was general agreement that we need more theoretical models to underpin practical experimentation and trials. However, opinions differed on the extent to which this is a realistic prospect in the short term. There was a broad consensus that the theory is there, but is fragmented among different disciplines:

– The subject needs to be opened up, structured and made more orderly and disciplined.

– Different groups are working in the same area, but not communicating well and the pace of development is being stalled by a fragmentation of effort.

– It is essential to involve actors of all sizes in research, experimentation and trials. There is both a practical and a policy need to develop a critical mass of new concepts, a lexicon and metrics, and this needs to be taken forward on a pan-Triad basis.

– It is also essential to adopt an interdisciplinary approach – i.e. researchers with different experiences and cultural backgrounds, based around a specific core of ideas. New insights and knowledge emanate from the interaction of different perspectives rather than a corpus of eager PhD students. The leverage is in multiple exchanges and inter-disciplinary interactions between high-calibre minds.

4. THE WAY FORWARD/ NEXT STEPS

In general, participants expressed their willingness to continue with the international dialogue on research evaluation and share experiences gained in the field. Many of the speakers emphasised that such co-operation is essential if we are to achieve any real progress on the issues and concerns expressed during the workshop. The overall consensus was that we urgently need to build a community of practice and this policy group is ideally positioned to act as a focal point - a learning machine that is capable of taking a global perspective and hearings from the different interest groups worldwide. There is a need for continuity of leadership to make this happen.

In conclusion, Peter Johnston thanked the participants for their contributions and extended an invitation to the Japanese R&D community to host a follow-up meeting in 2005.

5. LESSONS & LEARNING ASPECTS

While the presentations generally set high professional standards, a significant minority showed a tendency to wander off into polemics or pet topics without relevance to the session theme. In future, each speaker should be asked to illuminate, and answer, a specific set of questions. Moderators need to be more pro-active and willing to intervene, and exercise a closer grip on the content flow, panel questions etc. It is also important that moderators are fully familiar and comfortable with the scope and content of their sessions.

Overall, there is a need for much more interaction and passionate debate. The moderators also need to facilitate brainstorming and creative conflict, and should be chosen for their ability to facilitate and stimulate debate, not just marking time.

SUMMARY OF WORKSHOP PROCEEDINGS

Set out below is a short synopsis of each of the main presentations.

KEYNOTE PRESENTATIONS

1. Frans de Bruïne, Director, European Commission, DG Information Society, Directorate G: eEurope and Information Society Technology Policies

In opening the proceedings, Frans de Bruïne presented a strategic perspective of the EU R&D programme, and in so doing laid out the global objectives for the 2-day event. Essentially, its main purpose was to open up and explore:

– The importance of a proper linking between the technical, socio-economic, and investment objectives in order to appropriately account for the effects of research on productivity, competitiveness, and growth.

– The value creation process of world-wide research and technology development networking, collaboration platforms and partnerships which are developed to promote the exchange of new ideas, techniques and methodologies.

– The scope for harmonisation and "standardisation" of performance and evaluation systems and data among nations.

Mr de Bruïne expressed the hope that the workshop will make its mark as a first step towards genuine international co-operation in improving the methodologies and their exploitation in research evaluation.

2. ***PETER JOHNSTON,*** **HEAD OF EVALUATION AND MONITORING, EUROPEAN COMMISSION, DG INFORMATION SOCIETY**

Set against this context, a second keynote presentation by Peter Johnston of DG INFSO set out to clarify the scope and identify specific objectives for the workshop: namely to explore leading-edge practice globally in four key areas:

(1) Assessing achievements against objectives and expectations.

(2) Clarifying the "causality" links between RTD outputs, results and impacts.

(3) Networks as essential for creativity and effective exploitation of knowledge.

(4) International co-operation and comparability in RTD evaluation.

3. ***BIRGIT DE BOISSEZON,*** **HEAD OF PLANNING, PROGRAMMING & EVALUATION, EUROPEAN COMMISSION, DG RESEARCH**

A third keynote presentation, by Birgit de Boissezon of DG Research, concentrated on the problems thrown up by the proliferation of networks, methodologies and standards in the field of R&D evaluation. The presentation highlighted first the achievements and particularly the expanding knowledge and skills base but then noted how this had been achieved along multiple paths of development. Common

problems have emerged which now require common or joint approaches to finding solutions. The issues include the need for better linkage between evaluation evidence and the policy process, the problem of 'over evaluation' of the research communities, a shortage in supply of evaluation expertise and the comparability of evaluation experience.

Ms de Boissezon summed up by saying that there is a need for new internationally-acceptable tools and standards in evaluation and stressed the following emerging factors:

– The EU evaluation strategy and process has to be flexible enough to follow rapid policy changes within the overall objectives of the Seventh Framework Programme

– Transferability and comparability of data are essential pillars of any evaluation system, as are scaling of evidence from projects and programmes

– The organisational infrastructure underpinning the evaluation process also needs to be addressed in order to improve management effectiveness and transferability

4. *BILL VALDEZ,* DIRECTOR, OFFICE OF PLANNING AND ANALYSIS, OFFICE OF SCIENCE, U.S. DEPARTMENT OF ENERGY AND WASHINGTON RESEARCH EVALUATION NETWORK (WREN) STEERING COMMITTEE

Bill Valdez rounded off the keynote presentations with an opening view of the U.S. perspective: the essential aim of the workshop is the creation of a transnational network, and engagement among the participants - passionate exchanges – was seen as the critical measure of its success.

Mr Valdez went on to draw comparisons from the leading global evaluation models. The EU's evaluation system was seen as very good, as are those of Japan and Korea, but there are missing elements. In this connection he posed the following questions:

(1) How does R&D contribute to the overall national innovation system and, thus, economic growth (measured both in terms of predictable and unintended effects)?

(2) Given that R&D is no longer a zero sum game, what is the global performance benchmark? A wider context is needed.

(3) Another problematic issue is the comparison of scientific performance between nations. Evaluation metrics can only be based on a transnational comparison between countries, and a common "language" and comparable data are needed in order to compare scientific results worldwide.

In conclusion, what is abundantly clear is that we urgently need to build a community of practice and this policy group is ideally positioned to act as a focal point - a learning machine that is capable of taking a global perspective and hearings from the different interest groups worldwide. A second natural forum is Washington D.C., and in this context Mr Valdez highlighted some upcoming U.S. events:

– The annual meeting of the American Evaluation Association (AEA), Atlanta, Nov. 2004.

– The U.S. Council of Competitiveness symposium (Washington D.C., December 2004).

– The March 2005 meeting of the AAAS, which will focus on EU-US co-operation.

SESSION A: Assessing the Impact of Research on Innovation and Growth

5. *MARKUS KOSKELINNA (MODERATOR),* EXECUTIVE DIR., IMPACT ASSESSMENT, TEKES, FINLAND

The speaker outlined the status of the TEKES framework and its use in Finland as a policy evaluation tool for public R&D programmes. His key messages were: (i) the critical importance of multi-disciplinary RTD initiatives and the need to reflect this in the evaluation approach; (ii) the need to gain a sound understanding of the mechanisms for innovation before we can measure causalities; (iii) that there are material differences in the innovation process country-by-country depending on the macro and micro eco-structures and network complementarities in play; and, (iv) the low prospect of a one-size-fits-all solution.

6. *THEO KARAPIPERIS,* COMMITTEE ON INDUSTRY, RESEARCH & ENERGY, EUROPEAN PARLIAMENT

Mr Karapiperis gave a status report on FP5 STRATA project TAMI, which aims to create a structured dialogue within the TA community, as well as between TA experts and policy-makers, and provide technology assessment tools that will both formalise and enlighten the process of parliamentary approval of technology policies, and their subsequent monitoring. The network involves public agencies from nine European countries (D, NL, UK, CH, B, DK, E, CZ, PL) and the European Parliament. The speaker's main messages were: (i) if technology assessment is to have any real influence on the decision-making process it must be robust and objective, (ii) the parliamentary approval process is often undermined by poor communications, (iii) a common language is needed, (iv) professional communication skills should be part of the "toolbox" of technology assessment institutions, and embedded in their methodologies, (v) national, regional and transnational foresight and strategic analysis may become a major feature of technology assessment in the future.

7. *JERRY HAGE,* UNIVERSITY OF TILBURG, THE NETHERLANDS

According to Prof. Hage, much more research is needed to gain a clearer understanding of the 21st century innovation 'engine'. He asserted that:

– Economic growth today is driven more by product innovation than by productivity improvements or technological change.

– The growth process is now more a function of inter-organisational linkages than stand alone firms.

– It also depends critically on the diversity and mix of human capital inputs feeding into the innovation process.

This poses a major challenge for the existing macro and micro accounting models, which he sees as failing to keep pace with economic reality. As a result, he called for (i) a new economic growth theory, (ii) a growth accounting system that is capable of presenting a truer picture of the economic realities, and (iii) a new multi-dimensional, composite growth indicator (GDI). In this connection Eurostat should be brought centre stage since they have taken the high ground in Europe in developing ideas around an innovation index, based on the EU Community Innovation Survey.

In the Q&A session following, Peter Johnston cautioned against an over-focus on product innovation. Services have more profile and are more difficult to get to grips with. Also, the old dichotomy of products vs services is obsolete, and unhelpful. Cultural diversity is now seen as an asset politically.

8. *PIERRE MOHNEN,* MAASTRICHT ECONOMIC RESEARCH INSTITUTE ON INNOVATION AND TECHNOLOGY

Dr Mohnen's presentation provided an academic overview of the R&D research field, from both a theoretical and empirical viewpoint, with emphasis on (i) the new growth theories, (ii) their impact on the economic 'production function', and (iii) developments in the application of econometric modelling techniques. The main takeaways (many of which are familiar) may be summarised as follows:

– The bulk of empirical studies show a gross rate of return to R&D investment in the region 25-50%, depending on rent and knowledge spillovers, and 'receptiveness' factors such as trade-flow intensity, FDI, product market regulation, labour flexibility etc.

– Returns to process R&D exceed those of product R&D.

– The social rate of return of R&D exceeds the private rate by 50%-100%.

– In terms of productivity growth, R&D contributes 50% to labour productivity growth and 75% of total factor productivity growth (Griliches).

– Spillovers occur at local, national and regional levels, and are often asymmetric. For example, U.S. R&D accounts for 45% of Japanese total factor productivity growth over the period of 1965-1986, whereas there was no spillover from Japan to the U.S.

Stepping back, Dr Mohnen's position is that R&D investment boosts innovation and thus has a positive impact on productivity, but that innovation is not the only output of research that influences productivity. The main difficulties arise in defining and assessing the impact of the diverse channels for fostering R&D spillovers. In conclusion, he also supported the development of a new innovation index based on Eurostat's CIS work.

SESSION B: Methods for Comparing Evaluation Data among Nations

9. ***CHERYL OROS (MODERATOR),*** **DIRECTOR, PLANNING AND ACCOUNTABILITY; OFFICE OF THE ADMINISTRATOR, CO-OPERATIVE STATE RESEARCH, EDUCATION, AND EXTENSION SERVICE (CSREES), U.S. DEPARTMENT OF AGRICULTURE**

Mrs Oros presented the Portfolio Review Expert Panel Process (PREP), which has been introduced as response to the pressure of recent legal and U.S. demands for improved performance assessment, including the Government Performance and Results Act of 1993 (GPRA) and PART.

PREP focuses on programme outcomes rather than individual projects or scientists. The analysis is based on self-assessment and peer review panels. The aim is to go beyond project performance assessment per se and judge the relevance of the research portfolio against a broader framework of science policy objectives. The performance of the research is evaluated against the productivity, completeness, and timeliness of the portfolio.

10. ***PARRY NORLING ,*** **CHEMICAL HERITAGE FOUNDATION, DUPONT (RETIRED)**

Dr. Norling spoke about research evaluation from two perspectives: first that of a former Planning Director in DuPont R&D who had been impacted by evaluations of the company's research, and, second, as a participant in several panels studying federally-funded research. He sought to answer the following questions:

(1) When and how can we benchmark Science and Technology performance and make meaningful comparisons between countries and regions?

(2) What are the dimensions that should be assessed and evaluated?

(3) What are the levels of performance where comparisons between nations are best made?

(4) What are the main challenges in measuring performance and value of scientific advances vs. technology developments?

His key messages may be summarised as follows:

– We cannot measure performance in the abstract, but must measure it at the performance level, i.e: (i) individual research efforts (ii) studies by teams, networks, communities of researchers (iii) projects (iv) groups of projects, programs or research funded by one organisation or agency (v) initiatives or focused research problems (vi) results from an organisation, laboratory, or institution (vii) advances in a scientific discipline or (viii) National or regional innovation systems.

– Metrics and measures will be specific to the level of performance. The choice of indicators will vary depending on whether we are dealing with inputs, in-process variables, outputs, or outcomes - that may be retrospective, in the present, or prospective. Examples were given for each level of performance.

– A rich field of evaluation techniques exists, including: surveys, case studies, econometric studies, network analysis, bibliometrics, learning histories, peer reviews, and innovation indicators/ indices. Challenges include (i) the difficulty of measuring research performance due to long time lags from inputs to outcomes (ii) assigning value to knowledge itself (iii) tracing creation of knowledge to some benefit (iv) the ability to compare different studies using different approaches (v) the inability to have true control studies, and (vi) the difficulty of linking measures to complex innovation process.

According to Dr Norling, such emerging techniques as citation mining (map of science), options thinking and network analysis offer the promise of meeting some of the challenges.

11. *SHUJI YUMITORI,* DIRECTOR, RESEARCH AND DEVELOPMENT PROJECT EVALUATION DEPARTMENT, NEDO, JAPAN

Mr Yumitori presented a profile of NEDO's research activities and laid out the current status of its evaluation methodology in terms of framework, process, metrics and quality assurance.

NEDO is a semi-governmental research organisation set up in 1980 with an annual budget of $2.3 billion as a management agency, initially for energy research. Its evaluation guidelines are set by the Office of the Prime Minister.

Evaluation is carried out by external evaluators using a mix of quantitative and qualitative/ narrative indicators. The combination of a rating method and qualitative evaluation by external experts is considered to work well. A key feature is the use of viva voce interviews and open panel debate. The evaluation process is divided into four stages: (i) an ex-ante assessment of the project's potential, (ii) interim evaluations during the life of the project, (iii) an ex-post evaluation, and (iv) a follow-up and monitoring phase which takes place five years after completion.

A second, higher tier of the NEDO evaluation process is designed to help answer strategic questions relating to the receptiveness of the innovation environment, such as the fit and importance of the technology, the effectiveness of intellectual property rights, and the prospects for exploitation and commercialisation of the results.

12. *ERIC ARCHAMBOULT,* SCIENCE METRIX, CANADA

The presentation by Dr Archambault centred on his firm's proprietary R&D rating system, which assesses programme performance by reference to factors such as scientific papers, citations and patents per capita. An illustrative case study was included, based on the use of scientometric and technometric indicators for benchmarking R&D performance in the field of genomics.

SESSION C: Value Criteria through Research Collaboration and Networks

13. BART VERSPAGEN (MODERATOR), UNIVERSITY OF EINDHOVEN

Prof. Verspagen introduced the third session, which dealt with value-creation in collaborative networks.

14. *BILL VALDEZ,* DIRECTOR, OFFICE OF PLANNING AND ANALYSIS, OFFICE OF SCIENCE, U.S. DEPARTMENT OF ENERGY AND WASHINGTON RESEARCH EVALUATION NETWORK (WREN) STEERING COMMITTEE

This presentation by Bill Valdez outlined the results of a recent study of network dynamics by the Office of Science at the U.S. Department of Energy. The objective was to gain a better understanding of the flows of people, funding and knowledge in networks. The methodology was based on qualitative case analysis (interviews with individuals), network depiction and measurement (adoption of social network analysis techniques to R&D) and multivariate analysis (quantitative analysis of qualitative data). The study claims to have achieved a breakthrough in two important areas:

(1) It is possible to assign value to the flow of knowledge.

(2) It is also possible to detect the emergence of entirely new scientific fields through scanning for the formation of networks. This raises the prospect of network formation as a leading indicator of the success of public research programmes.

Overall, the study demonstrated that network analysis is a systematic way to assess the outcome and value of R&D and provides data and tools to assist in managing complex research portfolios.

15. CAROLINE WAGNER, RAND EUROPE

The speaker presented the interim results of the ERAnets project, a study of network analysis as a tool for performance assessment and research evaluation. The project aims to open up the frontier of network theory and develop tools and datasets for use by the Evaluation and Monitoring Unit in DG Research. Initial results confirm the conclusions of previous researchers: that networks develop and cannot be forced; knowledge flows do not respect political borders; and the importance of weak ties. Cases of network clustering based on live FP5/6 projects were shown, and the study finds that networks tend to form around four nodes: people, places, dynamics (e.g. personal links, acolyte activity) and knowledge domains. An important question for the next phase is why SMEs are dropping out of FP6 proposals.

Overall, the study finds that network analysis is a promising tool for research assessment and evaluation.

16. *NICHOLAS VONORTAS,* DIRECTOR, CENTER FOR INTERNATIONAL SCIENCE AND TECHNOLOGY POLICY, GEORGE WASHINGTON UNIVERSITY, WREN STEERING COMMITTEE

Dr Vonotas' presentation described the Center's work-in-progress on expanding the classical Cobb-Douglas economic production function to provide a mathematical model of the workings of the modern knowledge economy and its productivity processes. A second objective is to flush out new innovation indicators with explanatory power. Points of note were that:

– Recent developments based on innovation surveys are beginning to unravel deeper causalities in the Cobb-Douglas transformation function, but progress is slow and has been bedevilled by data problems.

– STI indicators can be grouped into four categories: input, output, innovation and process indicators. In future, more priority should be given to process indicators.

– A new generation of composite indicators is needed to account for the complex formal and informal relationships among the various economic agents involved in the innovation process.

In conclusion, the speaker presented case studies designed to illustrate and map the inter-nodal communication flows in live, active networks. The cases were drawn from both public and corporate sectors.

17. *OVE GRANSTRAND,* CHALMERS UNIVERSITY OF TECHNOLOGY, DEPT. OF INDUSTRIAL MANAGEMENT AND ECONOMICS

Prof. Granstrand rounded off the first day with a refreshing polemic on the drivers and obstacles in the quest for superior performance in technology innovation and transfer, with emphasis on technology collaboration. Opening with the observation that the last quarter of the 20th century witnessed an unprecedented increase, of questionable value, in the number of technological partnerships in the hi-tech sectors, he argued that:

– Josef Schumpeter's mantra of creative destruction is giving way to 'creative complementarities'.

– There is an over-proliferation of ambiguous collaboration modes – we urgently need to frame and classify them in order to gain a better understanding of their workings and develop metrics to distinguish good, bad and indifferent outcomes.

– We are at risk of obsession with fashionable methodologies, such as network analysis.

– Property rights need a fundamental overhaul. At present, IPR is handed out by governments in a fragmented way. The field is characterised by heterogeneous, often competing actors who are using the grants in ways other than intended. We need to return to basics and re-think the business economics and management case. As a starting point we urgently need to deal with recent developments in patent practice - patent pooling, cross licensing and blocking patents would be good starting points for reform.

In closing, he called for government subsidy of collective R&D networks, in the form of tax breaks, licenses/ pooling exemptions, and IPR assistance.

In the Q&A session following, the Prof. Grandstrand's opinions injected a welcome and spirited controversy into the proceedings. The ensuing debate elicited strong support for pushing the frontiers of network analysis and the need to infuse ideas from other disciplines, in particular physics and the social sciences, which underlines the need for an inter-disciplinary forum.

SESSION D: U.S. President's Management Agenda (PMA) and Federal Performance Assessment Toll (PART)

18. *ROSALIE RUEGG (KEYNOTE PRESENTATION):* TECHNOLOGY IMPACT ASSESSMENT CONSULTANTS, INC., AND WREN STEERING COMMITTEE

The presentation by Dr Ruegg outlined the U.S. Government's Performance Assessment Rating Tool (PART), which is a rolling audit initiative under the auspices of OMB Washington aimed at improving the accountability of research expenditure by federal agencies. Currently, 20% of all federal programmes are reviewed annually. As a result, 7% have received an uplift in federal funding and 1% have had their budgets cut or terminated. The PART scores are posted on the OMB website. The key objectives are (i) to help the agencies themselves realise the importance of evaluation in strategic planning, and (ii) to encourage uniform methods and standards of evaluation. The main features of the PART methodology, current and planned, were presented.

SESSION E: Linking Technical, Socio-Economic and Investment Objectives in Evaluation (I)

19. LUKE GEORGIOU (MODERATOR), PREST, UNIVERSITY OF MANCHESTER

Prof. Georgiou introduced the fifth and sixth sessions with an overview of leading-edge issues in evaluating the relationship between technology research and socio-economic policy objectives.

20. GRETCHEN JORDAN, SANDIA NATIONAL LABORATORIES, US, WREN STEERING COMMITTEE

The presentation by Dr Jordan laid out a 'holistic' family of diagnostic tools based on a theory of diversity in research units. This was developed in work funded by the U.S. Department of Energy (DOE) Office of Basic Energy Sciences and done in collaboration with the University of Maryland, along with a sample of empirical test cases. The main aim of the DOE project is to understand the environment in which excellent and relevant research flourishes in order to define appropriate and innovative measures of operational and scientific performance and improve the effectiveness of research expenditure. The Sandia diagnostic maps the various elements of a typical innovation system through its life cycle and provides a scoreboard of performance indicators – operational, management and technical. A sizeable case database has been constructed at U. Maryland. The speaker's main takeaway message was that evaluation systems need to be flexible and discriminating in order to take account of the different stages and types of 'R' and 'D', which are called "research profiles". Basic research and product realization projects each require different judgment criteria, which must also be tailored according to size of project and incremental vs. radical goals, as well as factors such as the technology domain. There is no one-size-fits-all solution.

21. *TOM CASEY,* CIRCA GROUP EUROPE, IRELAND

Tom Casey presented the interim findings of the IST MIPS project, which he described as an integrated data monitoring system for the European Commission's IST programme. The intention is that the system, which is based on key performance indicators, will be introduced operationally. His key messages may be summarised as follows:

– Performance indicators must be structured in line with political as well as technical and social objectives.

– In the context of a changing political environment, the European Commission has to respond to policy goals that are constantly evolving during the life of the Framework Programme. The Commission has also to cope with administrative changes within its internal structure.

– As a result, Commission planning systems have to be able to respond quickly to changes in political priorities, which range from incremental to discontinuous. It has to be able to respond to both at an exacting level of detail.

The MIPS protocol translates political objectives into a broad framework of policy objectives - economic, employment, innovation and research, economic reform, social cohesion, environmental, etc. Performance is assessed using a questionnaire and scoreboard with project monitoring indicators (numerical), process evaluation indicators (judgmental), and impact assessment indicators.

22. ERIK ARNOLD, TECHNOPOLIS UK

Erik Arnold presented an overview of European R&D evaluation practice based on a recent EU-wide study. His main theme was that the era of theory-free evaluation is over. Evaluation today relies heavily on domain-specific tools, methods and principles for designing indicators. As a result, evaluators need to be expert in the specific research field, close to the relevant scientific community, and familiar with prior art. Other points of note were:

– Research policy in Europe is changing - the aim is to create scientific clusters rather than fund lone scientists.

– In R&D evaluation the context is vitally important and there are wide variations in the characteristics of the national innovation systems around the EU. National innovation systems are in effect eco-systems with different relationships and interdependencies and the evaluation process has to be adaptable, and discriminatory.

– We may have to give up on the notion of evaluating point performance along the innovation chain and look for a higher level of abstraction. In this connection, there is some evidence that policy impact analysis (PIA) is set to take the high ground.

– A change of mindset is needed at the systems level – towards continuous evolution rather than optimisation.

– There is a need for much more conceptual work on network indicators, and data gathering on relationships. Social network analysis could be useful here (underlining again the need for interdisciplinary study).

23. DARRELL BESCHEN, CHIEF ECONOMIST, OFFICE OF ENERGY EFFICIENCY AND RENEWABLE ENERGY, U.S. DEPARTMENT OF ENERGY (AN APPLIED R&D PROGRAM'S RESPONSE TO PART)

The presentation by Mr Beschen centred on a case study of the PART evaluation of the U.S. Department of Energy. The key takeaways were: (i) the aim of the exercise was less one of budget-cutting than bringing the performance of underperforming elements up to threshold, (ii) the importance of involving the research community in the design of the evaluation process to ensure close alignment with what they need to run the programme at a working level, and (iii) the value of PART in providing scientists unfamiliar with evaluation (and management) concepts with a system that enables them to frame and communicate their activities to the policy and political communities.

24. YUSUF KOCOGLU, INSEE, CREST

Dr. Kocoglu gave an overview of the results of a comparative macro-economic study of intangible capital formation in France and the United States in the period 1960-2000. The study uses R&D and ICT investment as proxies, and trend statistics were presented showing the growth trajectories of intangible capital formation in both countries. The study also examined a number of growth accounting issues and, notwithstanding the well-publicised conceptual and practical difficulties with Cobb-Douglas as applied to R&D, presented a range of statistics on the relationship between intangible investment and gross value-added in the two economies.

In the Q&A session following, Peter Johnston commented that the evaluation of research programmes at EU level is, quite intentionally, not as detailed or prescriptive as the PART tool, and is used in different contexts. It is applied only to the European Institutions and the Community budget, which represents only 1% of EU GDP compared to the U.S. where 20% of R&D is publicly-funded.

SESSION G: Panel Discussion: Research Assessment and International Collaboration – Challenges and Common Approaches

25. *NICHOLAS VONORTAS (MODERATOR),* DIRECTOR, CENTER FOR INTERNATIONAL SCIENCE AND TECHNOLOGY POLICY, GEORGE WASHINGTON UNIVERSITY, WREN STEERING COMMITTEE

Dr Vonortas outlined the agenda for the sixth and final presentation session, which set out to illustrate the status of technology evaluation methods in four countries - France, Austria, S. Korea and New Zealand.

26. *DANIELLE BARRET,* NATIONAL COMMITTEE FOR RESEARCH EVALUATION (CNER), FRANCE

Dr Barret presented a broad overview of evaluation practice in French universities, research institutes and public institutions. The evaluation system is rooted in the establishment in 1960 of a Comité National du CNRS and a legal requirement for research institutions to establish evaluation structures, which was eventually adopted by the law in 1985. In the period 1986-2003, evaluation procedures for a systematic assessment of public research were put in place. In so doing, the French government aims to foster a better concorde between the science community and society through improved programme management.

The assessment framework is multi-tiered, and evaluation methods vary according to the level of research under review – project, programme, institution, national policy etc. Generally speaking, evaluation of research units is carried out every four years, by a peer review process. Both quantitative and qualitative indicators are used. Criticism of the system stresses (i) conflicts of interest in the peer system, (ii) an over-emphasis on promotion by seniority, and (iii) a lack of transparency. Dr Barret supported the call for international co-operation towards a common approach to evaluation, and proposed the following topics to be studied as a collaborative work:

– The need to infuse more international expertise into the national evaluation process.

– The major difficulties of assessing research quality in the human and social sciences.

– Differences among the U.S. and European countries over the legitimate role of political intervention in the assessment process and systems.

27. MICHAEL STAMPFER, WWTF, AUSTRIA

According to Dr Stampfer, the European Research Area adds a key international element to the national policies of the Member Sates. The problem is how to achieve convergence on a common rationale and improve the comparability of national evaluation systems across Europe. Historically, the fragmentation of effort has led to a proliferation of procedures and practices at different levels. From the perspective of a small country such as Austria, participation in international cooperation is a choice between openness and insignificance. For 20 years, Austrian RTD projects and programmes have been reviewed by external peers. However, today's complex funding programmes require new forms of international peer review, including managerial input. Common standards and migration of good practice are essential to open up cross-border funding and co-operation in the European Union.

Mr. Stampfer also outlined the workings of the Austrian Evaluation Platform, which is a loose network funded by public institutions. It aims at disseminating and implementing evaluation standards and providing expertise for the Austrian RTD Advisory Council.

28. *HEUNG DEUG HONG,* DEPARTMENT OF PUBLIC ADMINISTRATION, KANGWON NATIONAL UNIVERSITY, SOUTH KOREA

Following the establishment of a National Science and Technology Council in 1999, the Government of South Korea has embarked on the implementation of a new R&D evaluation programme. The initiative was driven principally by the recognition of changing research needs and demands for more accountability for public expenditure, and will be fully operational in 2005. It will then conduct a systematic analysis of government R&D expenditure using independent experts. The initial target is the strategic programmes, with budget in excess of $8million, which will be reviewed twice in the life of the programme.

Prof. Hong supported the view that each country needs its own distinctive evaluation methods, in reflection of their different political, administrative and innovation systems.

29. *MARK DINGLE,* DIRECTOR, PERFORMANCE AND EVALUATION, MINISTRY OF RESEARCH &TECHNOLOGY, WELLINGTON, NEW ZEALAND

The presentation opened with a short review of the policy context in New Zealand, where the population count is 4 million, the economy is relatively small, and there is a low proportion of international scale businesses. Research policy is based on a 3-5 year planning cycle, with open, contestable bidding for public funding by all research institutes, regardless of ownership. According to Dr Dingle, "the universities and research institutes compete down to the wire."

The public research programme is targeted to develop scientific expertise in niche areas such as plant and animal genomics. The indicators used to evaluate R&D performance include monitoring the number of publications per 1000 inhabitants and scientific productivity per research dollar spent. In both cases New Zealand achieves a high ranking internationally. International co-operation is an essential element of the research strategy. In 2004, a scientific counsellor was appointed to Brussels, and next year will see similar secondments to Washington D.C. and Melbourne.

The presentation concluded with some watch points on collaborative R&D ventures:

– Involve evaluators early in the process.

– Start with an extensive and challenging peer review process - prior to funding.

– Identify, encourage and support your very best researchers to develop networks.

– Monitor progress throughout – the end is too late.

– Maintain a long-term perspective – the journey has no end!

30. KEN GUY (MODERATOR), WISE GUYS, ENGLAND

In his concluding remarks, Ken Guy underlined the value of networks and international co-operation in improving research evaluation. Networks are an effective way of achieving convergence whilst retaining diversity. It is crucial to share experience and knowledge gained in different parts of the world. This workshop gave a unique opportunity to open up and contrast a rich variety of approaches, methods and practices in evaluation.

The past 10 years has seen a marked rise in the demand for accountability of research expenditure. The result is better tools and more systemisation of evaluation procedure. PART is a clear example of systemisation of assessment procedures. Concurrently, decision-makers, programme managers and scientists have come to recognise the importance of linking policy objectives with programme targets and funding, and this has led to a shift of emphasis from straight accountability to learning.

But, the garden is not all roses. The positive developments are overshadowed by the continuing proliferation of evaluation systems and shortcomings in the availability of reliable, internationally-comparable data. There is no indicator set, or even common language, for cross-border comparison of national efforts. Adding to this, the current tools are often over-complex, rely on subjective judgments, and slavishly pursue fine granularity at the expense of the big picture. The aim is to catch the first wave effects rather than chasing the tails of the distribution.

Looking to the future, the priorities are: (i) evaluation tools, and (ii) improved monitoring and collection of data. The game, certainly in Europe is moving towards policy impact analysis (PIA) and greater standardisation of indicators. The national statistical institutes and Eurostat have an important role to play here as we go forward.

31. *FINAL RESPONSES:* BIRGIT DE BOISSEZON, BILL VALDEZ, PETER JOHNSTON

Birgit de Boissezon opened the closing panel by summarising the key concerns and proposed that the group adopts the following priorities:

– Foster the development of a new generation of evaluation tools.

– Address the proliferation of feedback systems and data comparability problems.

– Stimulate the creation of an evaluation profession globally.

Continuity is essential in tackling these issues, and to achieve this the international evaluation community will require a permanent network organisation. In her opinion, the European Union is lagging the U.S. in the development of effective evaluation tools, although steps have been taken recently to increase co-operation between the Council of Ministers and the European Parliament in the area of impact assessment.

Bill Valdez. In his concluding remarks, Bill Valdez stressed that research evaluation needs to be theory-based. A solid foundation in methodology and science is essential. He was impressed by the array of practical methods and tools presented during the workshop but the first priority, as a group, is to work on theory-based evaluation.

Peter Johnston. In his closing comments, Peter Johnston stressed the need to keep up the momentum on the theoretical issues since the empirical work requires a strong intellectual consensus around a core measurement theory. This group is well placed to provide a focal point and forum for this dialogue, but needs continuity, leadership and infrastructure.

Forthcoming events organised by the American Evaluation Association and others in the United States offer a good opportunity to expand the network and open up new perspectives. Key events of interest include:

– *AEA,* November 3-6, 2004 - Atlanta, USA (***www.eval.org***).

– *WREN Workshop,* December 2004 - Washington, USA.

– *AAAS,* March 2005 - USA.

– *A further EU-WREN workshop in June 2005* **-** The European Commission will be very happy to host the 2005 workshop, but offers from Japan or other contributors such as South Korea or New Zealand would be especially welcome.

– *AEA,* November 2005 – Toronto, Canada.

In conclusion, he reinforced the pivotal role of evaluation - based on reliable data and comparable indicators - in helping the policy community to set realistic policy goals. International collaboration and sharing experience is the right way to build capacity and develop improved methodological approaches.

Clark G. Eustace
Mantos Associates

Annex 1 – List of Participants

LAST NAME	*FIRST NAME*	*ORGANISATION*
Antoine-Poirel	Frank	European Commission
Antoniou	Leonidas	Research Promotion Foundation - Cyprus
Archambault	Eric	Science Metrix
Arnold	Erik	Technopolis
Atorrasagasti	Ignacio	Permanent Representation of Spain
Atsumi	Toshihiro	Mitsubishi Research Institute Inc
Barret	Danielle	CNER
Barret	Danielle	CNER
Barry	Hélène	European Commission
Belle	Alberto	Knowledge Evaluation
Beschen	Darrell	US Department of Energy
Bogason	Sigurdur	European Commission
Bogers	Marcel	Eindhoven University of Technology
Bussillet	Sophie	Technopolis France
Cabirol	Laurent	European Commission
Casallas	Liliana	European Commission
Casey	Tom	CIRCA Group Europe
Charlet	Vincent	Technopolis France
Chatziioakeimidou	Jenny	European Commission
Chrysagis	Kostas	Zenon
Collins	Isabelle	Technopolis Ltd
Constantopoulos	Nick	General Secretariat for Research and Technology
de Boissezon	Birgit	European Commission
De Ghellinck	Elisabeth	UCL (université catholique de Louvain)
De Laat	Bastiaan	Technopolis France
Deug Hong	Heung	Kingwon National University
Devillez	Caroline	Gordon-Gill
Di Paola	Fabiana	banca Intesta
di Pietrantonio	Loris	European Commission
Dingle	Mark	Ministry of R&T, Wellington
Doherty	Richard	Deloitte
Dratwa	Jim	faculté Universiatire Saint-Louis
Duprel	Carlo	Fonds National de la Recherche, Luxembourg
Eustace	Clark	Mantos Associates
Feindt	Sylvie	SF Consulting
Fitzpatrick	Teresa	European Commission
Garrido	Cristina	CDTI
Georgiou	Luke	PREST, University of Manchester
Giannopapa	Christina	European Commission
Giesecke	Susane	VDI/VDE – IT
Giovannini	Massimo	Europa Region Trentino (Tirol Buro)
Girboux	Joel	Faculté Polytechnique de Mons
Gomez	Pilar	ISC
Grandstrand	Ove	Chalmers University of Technology
Guy	Ken	Wise Guys Ltd
Hage	Jerry	University of Tilburg
Hakkarainen	Naija	Tekes (The National Technology Agency of Finland)
Harada	Koh	Ministry of Economy, Trade and Industry
Hirvilammi	Tero	SITRA (Finnish National Fund for R&D)
Hong	Heung Deug	Kangwon National University
Hong	Heung Deug	University of Kangwon
Honma	Noboru	New Energy and Industrial Technology Development Org
Hosaka	Naoko	New Energy and Industrial Technology Development Org
Huidobro	Almudena	Ministry of Education and Science
Johnston	Peter	European Commission
Jordan	Gretchen	Sandia National Laboratories

Kaniewska	Ewa	ISC
Karapiperis	Theo	EP
Kawemba	Eric	ISC
Ken	Guy	Wise Guys
Kergel	Helmut	VDI/VDE – IT
Koerfer	Renata	Lithuania Liaison Office for RTD-LORD
Kordel	Marianne	DLR
Koskenlinna	Markus	TEKES
Kreysa	Eva	European Commission
Kye Sue	Kim	STEPI
Lejeune	Baudouin	Deloitte
Licciardello	Alberto	European Commission
Lyes	Martin	Entreprise Ireland
Marchetti	Marina	Ernst & young
Martin Oar	Luis	EP
Martinez-Vassallo	Javier	Ministry of Science
Martin-Rovet	Dominique	ESF (European Science Foundation)
Masoliver	Beatrice	European Commission
Mendes dos Santos	Antonio	European Commission
Min Hyung	Lee	STEPI
Mingels	Ronny	ARTTIC
Mohnen	Pierre	Maastricht Economic Research institute on Innovation and Technology
MoonJung	Choi	KISTEP (Korea Institute of S&T Evaluation and Planning)
Mueller	Walter	Federal Office for Agriculture
Nakamura	Kazuo	New Energy and Industrial Technology Development Org
Nieto	Virginia	European Commission
Norling	Parry	RAND
O'Gwin	Chris	US Office of Science
Oros	Cheryl	US DOA
Pagani	Walter	EUC Group
Pahl	Dale	US Environmental Protection Agency
Paleologos	Costas	European Commission
Pavia	Agatha	ISC
Peredy	Zoltan	National Office for Research and Technology
Piras	Michele	Europa Region Trentino (Tirol Buro)
Pohoryles	Ronald	ICCR
Raim	Toivo	Estonian Representation to the EU
Reeve	Neville	European Commission
Roffi	Giuseppe	CNR
Ruegg	Rosalie	Technology Impact Assessment Consultants, Inc
Rusin	Vojtech	Slovak Academy of Sciences
Sanström	Ulf	Swedish Research Council
Schubert	Thomas	Robert Bosch GmbH
Silvana	Alberto	European Commission
Siune	Karen	University of Aarhus (Danish Center for Studies of Research and Research Policy)
Stampfer	Michael	WWTF (Vienna science & technology Fund)
Stanbury	William	not indicated
Takkula	Timo	European Commission
Thys-Clement	Françoise	ULB (Université libre de Bruxelles)
Valdez	Bill	US Department of Energy
Van Heeren	Henne	EnablingM3
Varadi	Imre	Brussels Office of the City of Budapest
Verdoodt	Pierre	Ministry of Flanders Science & Innovation Administration
Verspagen	Bart	University of Eindhoven
Vonortas	Nicholas	George Washington University
Wagner	Caroline	RAND Europe
Wilén	Nina	South Sweden European Office
Windy	Mark	KEA

Woo	Chang Hwa	ITEP
Yumitori	Shuji	NEDO
Zachgo	Jochen	BMBF
Zinoeker	Klaus	Joanneum Research

European Commission

Performance Assessment of Public Research, Technology and Development Programmes — Workshop Report

Luxembourg: Office for Official Publications of the European Communities

2004 — 28 pp. — 21 x 29.7 cm

ISBN 92-894-8167-6