



EUROPEAN COMMISSION
Research Directorate-General

EVALUATION OF SCIENCE AND TECHNOLOGY IN THE NEW EUROPE

Proceedings of an International Conference on 7 and 8 June 1999, Berlin

Organised by the Fraunhofer Institute for Systems and Innovation Research (ISI),
on behalf of the European Commission, Research Directorate-General and
the German Federal Ministry for Education and Research (BMBF)
in collaboration with the European Science Foundation and
the Volkswagen Foundation

edited by Bührer, Susanne / Kuhlmann, Stefan
Fraunhofer Institute
Systems and Innovation Research (ISI)
Karlsruhe, Germany
e-mail: sk@isi.fhg.de
e-mail: sub@isi.fhg.de

Contents

Foreword ..... **i**

Opening Session **1**

Introductory Remarks: Evaluation of Science and Technology in the New Europe (*Richard Escritt)* **3**

Evaluation of Science and Technology in the New Europe *(Erika Rost)* **5**

Key Note Speeches: Maximising Return on Investment: A European RTD Challenge **9**

Facilitating Change in the German Research System: Capabilities, Collaboration, and Competition *(Wilhelm Krull)* **11**

Maximizing Return on Investment: A European RTD Challenge: Some Comments and Suggestions *(Antoni Kuklinski)* **23**

How to create an Evaluation Culture **29**

Chairman's Opening Remarks *(Tony Mayer)* **31**

Creating an Anglo-Saxon Innovation Culture *(John M. Barber)* **30**

French Culture *(Rémi Barré)* **43**

Finnish (Nordic) Culture *(Terttu Luukkonen)* **49**

Evaluation in the Context of the Fifth EU RTD Framework Programme **55**

Chairman's Opening Remarks *(José Vianna Baptista)* **57**

EU RTD Programmes Impact/Results Assessment: A Manifold Task in a Complex System *(Y. Dumont, L. Durieux, I. Karatzas, L. O'Sullivan, H. Teuber, G. Stroud, G. Fayl)* **59**

Evaluation as an ongoing Process - the View of a Practitioner *(Niels Busch)* **70**

Impact Assessment: Limits and Possibilities *(Angelo Airaghi)* **74**

After Dinner Speech **79**

The Berlin Science Landscape *(Ingolf Hertel)* **81**

East-West Co-operation in the Field of Evaluation: Major Recent Experiences **85**

Chairman's Opening Remarks *(Norbert Kroo)* **87**

Czech Republic & the European Science Foundation *(Peter Dukes)* **89**

The Impact of Evaluation on Formation of R&D Policy in the Baltic States *(Helle Martinson)* **93**

The Impact of Evaluation on Formation of R&D Policy in the Baltic States *(Carl-Olof Jacobson)* **95**

The Impact of EU Programmes on the Familiarity with and Use of New Technologies in Firms in CEE Countries INCO-COPERNICUS PROGRAMME: The Case of Greece *(Yannis Katsoulacos)* **101**

Evaluation of the EU Programmes and Cooperation with EU Partner's Impact on Innovation within Romanian Firms *(Steliana Sandu)* **103**

Round Table I: Mutual Learning in Evaluation East/ West **105**

Hungarian Experiences with Applied R&D Project and

Evaluating RTDI in the Cohesion Countries: Relevant Findings Programme Evaluation *(Tomás Balogh)* **107**

The Slovenian Point of View *(Edvard Kobal)* **111**

Learning by Doing Evaluation: The relevant Approach in Transition Context *(Kostadinka Simeonova)* **115**

Evaluating RTDI in the Cohesion Countries: Relevant Findings for Future Support *(Lena Tsipouri)* **119**

Towards Integration of Evaluation in RTD Policy-making **125**

Societal Challenges for Evaluation *(Arie Rip)* **127**

Accountability to the Public and Value-for-Money *(Luke Georghiou)* ... **133**

Distributed Intelligence for Innovation Policy Planning: Integrating Evaluation, Foresight and Technology Assessment *(Stefan Kuhlmann)* .. **137**

Round Table II: Mutual Learning in Evaluation of Science & Education **147**

New Challenges for the Evolution of Academic Research Systems Four Hypotheses on the Evaluation of Academic Research *(David Champbell)* **149**

Main Points of Intervention at Round Table II *(Inge Knudsen)* **155**

Mutal Learning: Evaluation of Science & Education *(Sergio Machado dos Santos)* **157**

Evaluation and Accreditation of Science and Higher Education in Slovakia *(Pavol Návrat)* **161**

General Conclusions **163**

Concluding Remarks I *(Manfred Horvat)* **165**

Concluding Remarks II *(Enric Banda)* **169**

Closing of Conference **171**

Evaluation of Science and Technology in the new Europe: Towards a European Evaluation Culture *(Gilbert Fayl)* **173**

Further Contributions **175**

The New Face of Science and R&D Evaluation in the Academy of Sciences of the Czech Republic *(Ladislav Pivec)* **177**

Evaluation of Science and Technology in Slovenia: Present and Future *(Dasa Bole-Kosmac)* **181**

Comments on Selection Criteria for EU Research *(Angelos Manglis)* **185**

Contribution to Round Table I *(Christian Dambrine)* **187**

Evaluation in the Romanian Academy *(Ionel Haiduc)* **189**

Assessment in Science and Science Assessment *(Ladislav Tondl)* **191**

Foreword

Since the early 1990s, we have been witnessing quite dramatic political, economic and cultural changes, not only in Central and Eastern Europe, but also within most of the Member States of the European Union. Many of these changes are also affecting the Science and Technology (S&T) systems and their ability to produce more and better knowledge, and thus their capability to solve human, social, economic, or technical problems. Although this applies more dramatically to Central and Eastern Europe, where many researchers desparately struggling for their professional survival, it also affects scientists and engineers in Western Europe, where challenges provided by the digitalization of knoweldge creation, and the increasing trend towards globalisation not only of the manufacturing process, but also of research and development activities as well as by cuts in public spending, are leading to a re-examination of S&T policies at national as well as at transnational and supranational levels.

It is in the light of this wider context that a conference organized by the European Commission and the German Federal Ministry of Education and Research was planned which focussed on both, the transition phase in which the respective S&T systems are in, and on the role and function of evaluation at various levels of decision-making. At the forefront were questions such as: How can the S&T systems respond adequately to the challenges provided by globalisation as well as those of the New Europe? What is the best way for the national S&T systems to achieve their objectives? Which are the opportunities for transnational synergies? What has been achieved by the evaluation and restructuring activities which took place in the early 1990s (e.g. in East Germany)? To what extent have we achieved the implementation of coherent approaches to quality assurance? What should be the future role of monitoring and evaluation exercises?

Commission of the European Union
Bruxelles

BMBF
Bonn

Opening Session

Introductory Remarks: Evaluation of Science and Technology in the New Europe

Richard Escritt
Directorate General XII, Science Research and Development, European Commission
De Meeûs Square 8, 1050 Brussels, Belgium

May I first warmly welcome you all, meeting as we are in this great city of Berlin in the heart of an enlarging Europe.

Berlin is in many ways a symbol of the new Europe. As it happens, it also contains a symbol of relevance to our work over the next couple of days, namely, how best to evaluate and make tangible the uncertain fruits of research. The newly re-opened Reichstag with its glass cupola has rightly attracted enthusiastic attention. What is perhaps less well known is that this environmentally friendly solution for maintaining an equable temperature and humidity in the debating chamber below is based on research carried out within an EU-funded research project.

The general message is clear: research and technological development is the key to solving many economic and social problems. Moreover, public sensitivity to research is growing. So despite the inherent complexity of the research process, we have to continue to develop our research programme evaluation methods so as to provide the best possible information on the quality of the work and on its potential economic and social returns.

This is a mission which the European Commission takes very seriously. We have had the great good fortune to be supported by the enthusiasm and knowledge of many independent experts. A good number of them are present here today and I am particularly glad to welcome them and to acknowledge our debt to them.

The 5th research Framework Programme has just started. With its greater concentration of resources and its 'problem-solving' approach, it poses a challenge to those implementing it as well as to those who in the future will evaluate its results. Our evaluation methods must therefore take full account of these changes.

It is a particular pleasure to welcome colleagues from the candidate countries. Here too, the 5th Framework Programme marks an important step forward. With the imminent conclusion of the necessary association agreements, they, along with the other associated countries, will be fully equal partners in the programme. This is a momentous step. To facilitate it, the Commission for its part has taken a number of practical measures, notably to spread widely the message of the opportunities for research collaboration. And at this conference too, we can learn from these colleagues' experience of research evaluation just as they can benefit from ours.

In conclusion, let me express my thanks to the Federal Ministry of Education and Research for their support for this conference and in particular to the Fraunhofer Institute for

Systems and Innovation Research for their hard work in making it possible. So once again, a very warm welcome to you all: I am confident that we shall have a very successful conference.

Evaluation of Science and Technology in the New Europe

Erika Rost
Federal Ministry of Education and Research (BMBF) Z21
Heinemannstr. 2, 53175 Bonn, Germany

On behalf of the Federal Minister of Education and Research, Ms. Edelgard Bulmahn, I would like to welcome you here in Berlin, to the „Berlin Evaluation Conference". The conference „Evaluation of Science and Technology in the New Europe" is part of the activities of the German presidency of the EU.

International linkages have a long and productive history in the area of S&T. During the last ten years, co-operation in Europe has developed into a new priority as a result of fundamental political, economic and social changes. These changes have tremendous effects on the (education &) S&T systems: internationalisation, globalisation, the information society, structural change, and financial constraints are challenges to education and research policy. In this context, evaluation and evaluation methods acquire even more importance and relevance.

Today's workshop links the issue of transformation of east European countries with the role of evaluation in politics and decision-making: Building the New Europe - It is a chance to learn from one another and one which we must take.

The Federal Government places a high priority on education and research. Education and research belong together: in their interaction, they contribute significantly to management of structural change in our economy, to the solution of society's problems and to cultural regeneration. A knowledge-based economy is shaped by investments in human capital - in general education and vocational training, in initial training and in lifelong learning.

We will only be able to meet the challenges that confront us at the threshold to the next millennium, if we strive to achieve an international strong position in research and technology.

To this end, government-funded research creates new incentives. Central elements are competition and empowerment: government-provided basic funding of research institutions has been restructured to promote a programme-oriented and performance-driven allocation of funds. The systematic evaluation of research institutions has been intensified. At the same time, the institutions themselves have been empowered through greater administrative flexibility.

We have to use education and research to contribute to solutions to the problems of our society.

Our efforts in education and research policy, in the coming years, are guided by 5 principles:

- Equal opportunity
- Creativity through greater responsibility
- Research for people
- Global sustainable growth
- Accelerating structural change.

Quality is central to this policy.
In the process of safeguarding the quality of our science and technology system and improving the performance of the institutions, evaluation is a key element. Evaluation supports:

- Reorganisation and development of the education and research system to ensure quality and effectiveness
- Programme development and setting of priorities
- Responsible allocation and use of funds.

In their joint responsibility, Federal and Länder governments initiated the process of reorientation of the German research system to ensure that the areas of competence of publicly funded research institutions meet, extend or set international standards.

Important elements are the system evaluation of the FhG and of the DFG and MPG and international experts have presented the results of their studies recently. At the moment, the Science Council evaluates the institutions of the so-called Blue List. Questions are, for instance:

- Do the 83 institutes meet the high standards of quality and
- How many of such institutions can we afford on a long-term basis?

Federal and Länder governments aiming for structures that are less complicated and more effective. Conclusions are being reached and they have to be acted upon.

Our education and research system consists of many different areas - the appropriate evaluation criteria and instruments for each area must be used if we are to achieve the goals set by our policies.

In the higher education sector, the situation is complicated. Efficient higher education institutions with an international reputation are the crucial places for educating and training young scientists and university graduates. By generating, using, applying and passing on knowledge, universities constitute one of the pillars of the knowledge-based society. Safeguarding of quality and evaluation have to take into account the specific demands of education and research in higher education institutions. The Science Council has taken important steps in this context.

The area of project promotion is of growing importance now. New instruments with relevance for the institutions and the system are being introduced. Within BMBF, the instruments to ensure quality will be further developed and fostered. We are aiming at an integrated approach to safeguard quality - evaluation is a key element of a future strategic information system. Evaluation fosters more transparency, efficiency and effectiveness

and more quality. Evaluation supports the reorganisation and modernisation of our systems and helps in the reduction of bureaucracy. To achieve these goals, more efforts are needed. Two aspects are of special importance:

- the evaluation process, the procedures, criteria and methods themselves have to meet quality standards, and
- we have to develop an evaluation culture.

The two aspects are not independent!

The climate in which an evaluation process is performed, the credibility of the evaluation experts, the reliability of the results and, finally, the chance to use the results, depends on high standards of the whole process and its acceptance.

There are no common standards in evaluation. We need to find the appropriate approach in each country and for each problem area, we need to design the evaluation process properly. We need to learn from one another!

Working together, building the New Europe - today's workshop contributes to this.

On behalf of the Federal Minister of Education and Research, I wish you success in your deliberations and I look forward to your discussion and your results.

Key Note Speeches:
Maximising Return on Investment:
A European RTD Challenge

Facilitating Change in the German Research System: Capabilities, Collaboration, and Competition[1]

Wilhelm Krull
Volkswagen-Stiftung
Kastanienallee 35, 30519 Hannover, Germany

"Our capital is the capability of our people."
Roman Herzog

I. Some Observations on Current Efforts to Reconfigure German Universities

Changes and Challenges for Universities

- from traditional ways of acquiring information towards the digitalisation of knowledge
- from predominantly disciplinary structures towards problem-oriented, transdisciplinary approaches
- from bi- or trilateral internationalisation towards globalisation
- from public support towards new modes of financing, especially private initiatives
- from input-related planning processes towards output-oriented assessments.

Prerequisites of Success in Research

- a broad spectrum of different, interactively operating disciplines
- a minimum of hierarchies and a high degree of horizontal communication
- a medium-sized university with ample opportunities for interaction with practitioners
- a strategically focused leadership and clear-cut governance structures
- a clear emphasis on research in the overall objectives as well as in the distribution of funds.

The Volkswagen Foundation's Programme on University Reform

1. Scope

- eight universities are supported; DM 20.6 million have been granted since 1996; two applications are still pending.

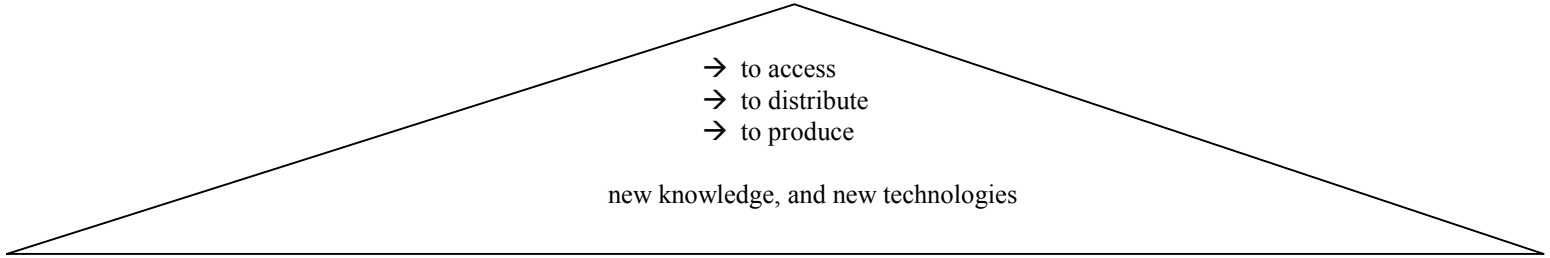
2. Major Objective

- Each university should find its own way in order to increase efficiency and effectiveness, and to improve quality by granting greater autonomy, and by putting collegiality to work in favour of change.

---

1 Paper presented at the Norwegian Institute for Studies in Research and Higher Education in Oslo on 1 June, 1999, and at the Bank of Sweden Tercentenary Foundation in Stockholm on 3 June, 1999, as well as at the Conference on "Evaluation of Science and Technology in the New Europe" on 7 June, 1999, in Berlin.

Pillars of Success of a University



Vision	Culture	Organisation	Resources
→ develop corporate identity → define long-term objectives → create a climate of change	→ keep cultural heritage → improve quality → create a climate of mutual trust	→ improve governance structures → enhance openness and transparency → make management processes more effective	→ recruit high qualitiy personnel → provide sufficient funds → secure high standards in infrastructure

3. Links between Autonomy and Accountability

- to stop diffusing responsibility, and to make specific persons accountable for their decisions
- decision-making competences, commitments, and obligations must be developed at the level at which people can actually take on the responsibility for their consequences
- to develop a common understanding of the objectives of the institution, and to stimulate the readiness to improve its performance at all levels, e.g., to

- make better use of the resources available
- access new resources
- improve the decision-making processes
- intensify communication and cooperation
- invest in personnel development.

"Pro-Uni" at the University of Hamburg

I. Objectives

- to improve quality of teaching, research, and services
- to facilitate interaction and cooperation
- to increase the degree of internationalisation
- to improve local and regional networking
- to enhance individual responsibility.

II. Process

- to develop a corporate identity ("Tor zur Welt der Wissenschaft")
- systemic approach to organisational development
- to move towards contract management and performance-based resource allocation
- to reorganise decision-making processes
- to develop an adequate controlling system.

III. Projects

- to develop, negotiate, and implement contracts
- to strengthen the management capacity at department level
- to reorganise the central administration
- to evaluate teaching and research performance
- to establish a new reporting and controlling process.

The University of Heidelberg: De-centralized responsibility for resources

I. Objectives

- to improve quality of teaching, research, and services
- to improve resource allocation at all levels
- to establish a new cost-benefit-awareness
- to focus university leadership on strategic issues and quality control
- to decentralise all aspects of specific decision-making.

II. Processes and Projects

- to provide the institutes with lump-sum budgets
- to transfer the decision-making capacity to the institute level
- to develop new modes of cost-related budgeting
- to create a market for resources and services within the university
- to establish an internal information system for budget allocations as well as a new reporting and controlling process.

"Organisational Learning" at the University of Bremen

"We want to change our university."

I. Objectives

- to improve quality of teaching, research, and services
- to enhance efficiency and effectiveness of resource allocation through decentralisation
- to transform the institution into a learning organisation
- to establish a new cost-benefit awareness
- to improve local, and regional networking.

II. Current Projects

- to implement new decision-making processes
- to allocate resources on the basis of performance indicators
- to improve internal and external communication
- to implement contract management
- to reconfigure the interface between ministry and university
- to focus on personnel development

Performance Indicators Used in Contract Negotiations at the University of Bremen

- amount of ‘soft money’ (scholarships, grants, and contracts)
- publications in refereed journals
- degree of interdisciplinarity of research activities
- number of conferences, workshops, etc.
- impact of research on other domains ("transfer")
- number of foreign visiting research scholars

- participation in international networks, professional associations, etc.
- degree of involvement in international teaching activities
- participation in advisory, or decision-making bodies inside and outside of the university
- amount of cooperation with other universities, research institutes, companies, etc.

Problems and Prospects

- to link up 'top down', with 'bottom up' processes
- to avoid deadlock situations in decentralising resource allocation
- to combine organisational learning with personnel development
- to redefine the interface between departmental, and central administration within each university
- to reconfigure the interface between ministries and universities.

II. A Brief Report on the Systemic Evaluation of the German Research Association (DFG) and the Max Planck Society (MPG)

The Terms of Reference

The Chancellor of the Federal Government and the Prime Ministers of the 16 state governments asked the International Evaluation Panel to draw conclusions based upon international comparisons concerning the following questions:

- Does the MPG and its institutes have, and apply, appropriate principles, processes, and opportunities for establishing new institutes, restructuring, or closing down existing ones as well as for continuous quality assurance?
- Does the DFG have appropriate principles, processes, and instruments for the assessment of proposals as well as for taking the right funding decisions?
- Does the cooperation with other partners within the research system, especially the universities (in particular the support for young researchers, and the development of new structures) and the economy at large (in particular the innovation process) function well?

The Composition of the International Evaluation Panel

Prof. Dr. Richard Brook, Chief Executive des EPSRC, GB (Chairman)
Prof. Dr. Jan Borgman, former Chairman of NWO and ESTA, NL
Prof. Dr. Gerhard Casper, President of Stanford University, USA
Prof. Dr.-Ing. Hubertus Christ, ZF Friedrichshafen/ President of the VDI, D
Prof. Dr. Dr. h.c. Reimut Jochimsen, President of LZB NRW, D
Prof. Dr. Jean-Marie Pierre Lehn, Collège de France, F
Prof. Dr. Helga Nowotny, Ph.D., ETH Zürich, CH/A
Prof. Dr. Israel Pecht, Weizmann Institute of Science Rehovot, IL
Prof. Dr. Eda Sagarra, Trinity College Dublin, IRL
Prof. Dr. Heinrich Ursprung, formerly President of ETH Zürich, & Junior Minister, CH

Changes and Challenges for a Globalised Knowledge Society

The increasing pace of knowledge production, distribution, and absorption vs. the structural conservatism of institutions, processes, and funding modes

Changes are needed in order to

- increase the degree of flexibility and permeability of the entire system
- avoid petrification of individual segments by creating unified funding modes for all non-university research institutions
- strengthen the ability of universities to manage their own business
- improve the regulatory framework in order to create a more research-friendly environment (especially in labour law)
- reconfigure the concept of competitiveness, and to evaluate the Helmholtz Association of German Research Centres (HGF), as well as the Gottfried Wilhelm Leibniz Association (WGL).

Universities

The largest potential; it provides the basis for all other parts of the system.

Changes are recommended in order to

- facilitate opportunities for young researchers to conduct their research more independently (and to abolish the 'Habilitation' as a prerequisite for becoming a university professor)
- open up the strong disciplinary orientation, and to create more flexible temporary organisational structures which favour inter-, and transdisciplinary research activities
- enable the universities to autonomously manage their own affairs
- develop new, more effective ways of monitoring and evaluating performance, including external participation
- improve the collaboration with non-university research institutions, and the chances for establishing 'centres without institutional walls'.

The German Research Association (DFG)

The most important funder of basic research activities in German universities; annual budget exceeds 2 billion DM; almost entirely based upon responsive mode funding schemes.

Changes are recommended in order to

- complement the responsive mode approach by a new portfolio of strategically oriented programmes
- open up the peer review system for younger people, a greater participation of women, a stronger orientation towards international standards, and for the support of transdisciplinary projects (e.g., by appointing more review panels)
- reconfigure the DFG offices, and to develop a more flexible, strategically minded organisational approach
- introduce a lump-sum budgeting system and to develop an adequate funding portfolio

- actively monitor quality assurance and programme development.

The Max Planck Society (MPG)

74 institutes; 233 scientific members and directors; 39 junior research groups; 15 Nobel Prize laureates since 1954, of which 10 since 1984.

Table 1: Development of the Most Important Funding Programmes of the DFG 1970 - 1998

Year of Introduction	Programme	Proportion of grants in %						
		1970	1975	1980	1985	1990	1995	1998
	Normal processes	36.6	31.6	43.5	43.9	42.9	40.6	40.1
1953	Main focus processes	21.3	18.5	12.7	18.2	16.2	14.3	12.6
1957	Auxiliary research facilities	1.3	1.9	2.3	0.9	1.3	0.8	0.8
1962	Research groups	0.6	1.7	2.6	1.4	1.9	3.5	3.9
1968	Special research areas (SFB)	20.6	34.5	30.8	28.9	28.3	25.1	27.4
1971	Computers/main frames (HBGF)	(14.5)	(2.6)	(1.3)	(0.6)	(1.3)	(1.6)	(1.5)
1978	Heisenberg Programme	-	-	1.6	2.0	1.4	1.6	1.1
1986	Leibniz Programme	-	-	-	-	2.3	1.4	1.1
1988	Gerhard Hess Programme	-	-	-	-	0.2	0.4	0.4
1990	Graduate colleges	-	-	-	-	0.3	4.3	5.8
1996	Transfer areas in special research areas	-	-	-	-	-	-	-

Source: DFG Annual Reports (translated by Fhg-ISI)

Changes are recommended in order to

- open up new opportunities for a closer cooperation between Max Planck institutes and universities
- create a more flexible approach to establishing new lines of research (beyond the institutes' level) through Max Planck research units in universities
- increase the number of junior research groups, and independent working groups
- improve the strategic capacity by implementing an external presidential advisory group
- link up strategic planning processes with the appointment of new directors
- implement a new scheme of comparative evaluations and synchronize it with the renewal of leadership functions.

Table 2: Number of Applications, Volume applied for, Amounts granted and Proportion of Grants in Percent in the Normal Procedure of DFG, 1974 - 1998

Year	**Applications assessed**		**Grants made**		**Proportions of grants in %**	
	Number	m DM	Number	m DM	Number	Volumen
1974	4,413	249.8	3,730	191.5	84.5	**76.7**
1975	3,827	287.9	2,938	189.5	76.8	**65.8**
1980	6,227	615.8	5,099	381.1	81.9	**61.9**
1985	6,406	806.4	4,972	462.9	77.6	**57.4**
1989	6,541	1019.6	4,594	498.4	70.2	**48.9**
1992	8,034	1431.7	5,113	575.9	63.6	**40.2**
1995	8,751	1759.5	6,026	808.8	68.9	**46.0**
1996	9,293	1860.1	5,973	775.8	64.3	**41.7**
1997	10,218	2202.1	5,996	779.9	58.7	**35.4**
1998	11,525	2418.1	6,513	877	56.5	**36.3**

*) Without research groups
Source: DFG

Before 1974 a different method of payment was used, so that the data from earlier years would not correspond to the post–1974 data. The time series therefore begins with 1974.

Table 3: Scientific Members and Directors of MPG

Number by Sections 1968 - 1998

Section / **Year**	Chemistry-Physics-Technics	Biology-Medicine	Humanities	**Total**
1968	65	86	20	171
1972	80	67	25	172
1976	97	73	26	196
1979	97	70	26	193
1984	100	75	23	198
1988	107	69	25	201
1992	111	72	27	210
1996	111	74	30	215
1997	108	78	29	215
1998	106[5]	94[5]	33	233

Source: MPG[2] (translated by Fhg-ISI)

Table 4: Junior Research Groups of MPG

Number by Sections 1969 - 1998

Section / **Year**	Chemistry-Physics-Technics	Biology-Medicine	Humanities	**Total**
1969	-	4	-	4
1972	-	10	-	10
1976	-	8	-	8
1980	-	15	-	15
1984	-	19	-	19
1988	-	27	-	27
1992	-	26	-	26
1996	4	23	-	27
1997	5	28	1	34
1998*	3	35	1	39

2 From 1.1.98 the MPI for Biophysical Chemistry (8 scientific members) belongs to the Biology-Medicine Section, no longer the Chemistry-Physics-Technics Section.

Table 5: Core-funding of MPG 1968 –1997 (in m DM)

Type **Year**	Share of funding	Special funding	Project funding	**Total**
1968	245.9	-	17.9	**263.8**
1972	384.3	15.2	34.6	**434.1**
1976	520.6	27.8	26.6	**575.0**
1980	648.1	22.4	40.4	**710.9**
1984	779.0	11.3	67.6	**857.9**
1988	913.0	22.2	148.7	**1083.9**
1992	1127.9	18.5	186.6	**1333.0**
1994	1318.5	26.9	149.2	**1494.6**
1996	1571.5	30.2	164.0	**1765.7**
1997	1533.9	35.2	173.3	**1742.6**

Source: MPG (translated by FhG-ISI)

III. Concluding Remarks: The Way Ahead

Looking back 10 years; dramatic changes, not only of the European political landscape, but also of the way we perform S&T.

5 major challenges:

- not only to cope with, but to manage the changes ahead proactively
- to turn our institutions into learning organizations, and to develop the personnel accordingly
- to define and attribute clear-cut responsibilities at all levels, and to make people more accountable for their performance
- to strike a balance between ‘top down’ strategic interventions and ‘bottom up’ participation (especially in research foresight, and priority-setting)
- to open up new opportunities for young researchers to pursue their own ideas at an early stage in their career.

Maximizing Return on Investment: A European RTD Challenge: Some Comments and Suggestions

Antoni Kuklinski
University of Warsaw, European Institute of Regional and Local Development
Krakowskie Przedmiescie 30, 00-927 Warszawa, Poland

Introduction

This is a difficult, comprehensive and controversial topic, which could be envisaged in different theoretical, empirical and pragmatic perspectives expressing different approaches and value judgements.

In this paper I would like to organize my thinking along the following thematic lines:

1. Two streams in the transformation of Europe,
2. The challenge of a single European RTD,
3. The glory and misery of the Framework Programmes,
4. The investment view of the European RTD,
5. Maximizing return on investment in European RTD – Feasible reality or an Utopian dream,
6. The evaluation of the European RTD – Some suggestions and proposals.

There are two streams in the transformation of the old divided Europe into the new united Europe.

1. Two Streams in the Transformation of Europe

The first stream – the prevailing classical stream of the second half of the 20th century - was driven by the integration process creating new efficient links of multilateral cooperation among individual countries, individual national societies, individual national economies and individual national R&D communities.

The second stream is transgressing the limits of the integration process and opening the stage of construction of totally new European identities in the form of a single European market or a single European currency leading to a *de facto* Federal Europe.

2. The Challenge of a Single European RTD

To my mind – the real challenge of the 21st century is to create a new identity of a single European RTD as an important actor of the global scene of science, innovation and enterpreneurship. The main question concerns the creation of an equivalent of EMU in the field of European RTD. Naturally, *omnis comparatio claudicat*, but it is worthwhile to discuss this analogy.

The first step in the emergence of the new European RTD Space is the concept of "the European value added" firmly incorporated in the 5th Framework Programme.

The ETAN Report is presenting the following interpretation of this concept:
"Community RTD initiatives are supposed to create "European value added". In this context they must follow the subsidiarity principle, so as to select only those objectives which are most efficiently pursued at Community level. More precisely, projects should – beyond their quality in terms of scientific criteria, partnership and project management either:

1. contribute to create a "critical mass" of human and financial resources across all the Member States: or
2. guarantee a significant contribution to the implementation of one or more Community policies: or
3. address problems arising at Community level including social needs, or questions relating to standardization or the development of the European area.

Moreover, Community RTD programmes and projects should contribute to the economy, science and technology in ways that will encourage the harmonious and sustainable development of the Community as a whole."[3]

The second interpretation is outlined in the Second European Report S&T Indicators:

"In addition EC policies contribute European added-value to the research and innovation system by reducing the transaction costs of international partnerships and by increasing the scale of enterprise activities in R&D. Such subsidies therefore serve to eliminate a form of market failure arising from underinvestment. The report echoes the earlier R&D Impact Study by concluding that research and innovation partnerships are a means of risks and costsharing."[2]

The third interpretation is indicating the scale of the operation in terms of personnel engaged:

"About 8% of total personnel engaged in R&D in the fifteen Member States, i.e. 125 000 people, are involved in one way or another in research projects funded by the Framework Programme. If account is also taken of other Community activities, notably the Structural Funds, we easily arrive at a figure of 150 00 people."[3]

The 150 000 participants create a large network including three groups of participants[4]:

- the technology and research leaders
- network followers
- peripheral partners.

In the Second Report we find rich materials concerning the regional structure of European RTD. These materials could help to answer the question – to what extent the RTD Programmes can be seen as contribution to the reduction of the centre – pheripery gap in

---

3 Options and Limits for Assessing the Socio-Economic Impact of European RTD Programmes. Report to the European Commission, DG XII, Evaluation Unit by the Independent Reflection Group (set up in the context of the European Technology Assessment Network, ETAN), January 1999.

4 Second European Report on S & T Indicators, December 1997, p. 573.

5 P. Caracostas, U. Muldur, Society, the Endless Frontier. A European vision of research and innovation policies for the 21st century, European Commission, DG XII, Brussels 1998, p. 31.

6 See Second European Report, op. cit.,p. 673.

Europe. All these achievements create only the first step in the emergence of the new European RTD Space. Unfortunately, there are major weaknesses of the institutional machinery in this field.

In the Report – "Society – The Endless Frontier"[5] we find the following observation:

"Thus, the Davignon Report, whilst praising the quality of the research undertaken under the Framework Programmes and network of capabilities set up as a result of them, puts its finger squarely on the major weaknesses of the institutional machinery of Community research: the detailed substance of Community measures is the subject of an unwieldy procedure of co-decision by the European Parliament and the Council of Ministers, the latter giving its final assent only after the Member States have voted unanimously to approve them."

This method of decision-making slows down the implementation of a common truly European research strategy, often resulting in unsatisfactory compromises on objectives and priorities. It makes it very difficult to adapt European initiatives regularly to changes in the scientific, technological and socio-economic scene and prevents the European executive from modernizing and streamlining the operational management of these initiatives.

The power in the creative management of European RTD is immensely weaker in comparison with the power of the management of EURO. We have a long way to go to reach the horizon of a Single European RTD Space.

3. The Glory and Misery of the Five Framework Programmes

Future historians will probably describe the rich and comprehensive experience of the Five Framework Programmes as the first stage in the processes which have created a single European RTD of the 21st century.
The glory of the Five Framework Programmes is well-documented in the path-breaking Davignon Report[6].
However, this report is concentrating attention on the institutional pattern of the European RTD, while two factors are not analysed in the report:

1. the extremely modest scale of the Framework Programmes in comparison with the scale of the Common Agricultural Policy. The Framework Programmes are just too small to create a real new big push on the European and the global scene,
2. the substantive composition of the consecutive Framework Programmes.

---

7 P. Caracostas, U. Muldur, op. cit., p. 35.

8 European Commission – Five Year Assessment of the European Community RTD Framework Programmes. Report of the Independent Expert Panel Chaired by Viscount E. Davignon and the Commission's Comments on the Panel's Recommendations, Directorate-General Science, Research and Development, Brussels 1997.

In the final analysis, the Framework Programmes are a relatively weak actor on the European scene which is dominated by the phenomenon of the growing deficit of the European Union in the balance of trade in the field of high technology products.

4. The Investment View of the European RTD

I do not share the point of view of the ETAN Report – that "Despite the attractions of an investment view of RTD, it is widely accepted that normal rate criteria do not apply to RTD".

According to the ETAN Report[7]: "There are three main problems":

- "RTD has many effects",
- "Evaluation results are often needed early" and
- "Effects are not always uniquely attributable to a single project or participant".

This situation is emerging not only in the field of RTD and other intangible investments, but also in the field of multipurpose tangible investment, i.e. in the technical infrastructure.

The OECD contributions[8] on the growth and management of intangible investment are creating a new background for the investment view of RTD.

This point of view is, to my mind, indirectly supported by a recent paper on "Evaluation of Research and Technological Programmes: a Tool for Policy Design"[9]. In this context we should also review the Seibersdorf report[10].

5. Maximizing Return on Investment in European RTD. Feasible Reality or a Utopian Dream

Following the concepts of "efficacy", "efficiency" and "effectiveness",[11] we can outline the typology of investment returns:

a) the teleological returns,
b) the cost-benefit returns,
c) the social returns.

The teleological returns are related to the field of interaction of stability and change in the domain of long-term goals and medium-term targets.

The cost-benefit analyses applied in micro-mezzo and macro scale. The mezzo scales are creating a bridge leading to the identification of social returns related to the improved performances of local, regional, national and European communities and societies.

---

9 ETAN Report, op. cit.

10 OECD – Technology and Economy, Paris 1992.

11 G. Fayl, Y. Dumont, L. Durieux, I. Karatzas and L. O'Sullivan, Evaluation of research and technological development programmes: a tool for policy design, "Research Evaluation", vol.7, nr.2, August 1998.

12 Seibersdorf Report – Evaluation of Austrian Participation in Community RTD Programmes, January 1997.

13 G. Fayl et alia, op. cit.

The whole methodological, empirical and pragmatic *universe* of the discussion on the maximisation of return on investment in European RTD can function only in the context of long term diagnostic and prospective studies analysing the transformation of Europe at the turn of the 20th and 21st century. The stagnation in this field should not be tolerated any longer. For example, a comprehensive and multidimensional answer should be given to the question concerning the performance of Europe in the three decades (1990-2020), described as the Fifth Schumpeter Wave[12].
The recent DG XII publication[13] can be seen as a move in this way. The paper of Roman Galar can be also useful in this context[14].

6. The Evaluation of the European RTD: Some Suggestions and Proposals

It is well known how important the role of great charismatic leaders was in the creation of the theory and practice of a United Europe. These personalistic and heroistic approaches and the best practice approaches should be applied also in the evaluation and promotion of the European RTD programmes. For example, the evaluation of the Fourth Framework Programme should include the selection and presentation of fifty grand personalities – the outstanding heroes maximizing the returns on investment in the Programme, who were responsible for the most successful projects in scientific, entrepreneurial, social and cultural dimensions.
Such a rich and colourful gallery of personalities will be a more convincing argument for the academic and entrepreneurial community as well as for the European public opinion at large than many volumes written in dry institutional language.

Conclusion

This is a tentative paper on incomplete or even insufficient knowledge[15]. *Audaces fortuna iuvat.* I hope, however, that some bold judgements presented in this paper will stimulate the discussion developed during our conference[16].

The full version of this paper will be published in the volume – The Knowledge-based economy – The Challenge of the XXI Century, A. Kuklinski (Editor), Warsaw 2000.

---

14 "Innovation has become the industrial religion of the late 20th century" see "Innovation and Industry", "Industry gets Religion" – The Economist, February 20th 1999.

15 P. Caracostas, U. Muldur, op. cit.

16 R. Galar, Europe as a Continent of Regional Innovation Systems (in) A. Kukliński, K. Pawłowska (Editors), Innovation – Education – Regional Development, Seria Sadecka, vol. 1, Nowy Sącz 1998.

17 This paper represents only the individual opinions of the author.

18 Evaluation of Science and Technology in the New Europe – Evaluation Conference, Berlin 7/8 June, 1999.

How to create an Evaluation Culture

Chairman's Opening Remarks

Tony Mayer
European Science Foundation
1 quai Lezay-Marnesia, 67080 Strasbourg Cedex, France

In approaching this session, the questions arise as to what is an evaluation culture, is there a particular European culture or is it a mixture of national evaluation cultures.

Before attempting to answer these questions, we need to be clear as to what are 'evaluation' and 'culture'. We all know what a theatre or film critic is, but do we know what a 'science and technology' critic may be?

Evaluation is the understanding of the value of an activity and, in modern parlance, the need to evaluate activities in terms of the original specification, its outcome and products and its impact. It is not only an exercise in monitoring and accountability but it is also a guide to future decision-making and must also be used to create a feedback mechanism for those concerned in order to improve and build on past experience.

But what is 'culture' in this context? 'Culture' may be defined as either 'a trained or refined state of the understanding and manners and taste' or 'the state of understanding prevalent at a time and place'. This latter definition is perhaps best used in seeking to understand whether there are distinctive regional approaches in Europe to the whole issue of evaluation in science and technology.

Chairman's Concluding Remarks

The three speakers have explained three very different approaches around the same theme. These have covered a wide range of issues. The UK has developed a more market-driven or problem-orientated approach, which also tries to distinguish quality and relevance. In France, there is a strong emphasis on career and group development within evaluation, while the Finnish (Scandinavian) approach has stressed the need to involve international comparisons and has also laid considerable importance on openness and transparency in the process. These different approaches reflect not only different attitudes and policies, but are also a function of differing national S&T structures. What they have shown is the common basis of professional evaluation based on peer and merit review and the need for international comparisons, especially within Europe. We have heard of this approach in the recent major review of S&T structures in Germany and this indicates the future for European-wide evaluation process.

Creating an Anglo-Saxon Innovation Culture

John M. Barber
Department of Trade and Industry
151, Buckingham Palace Road, Room 359, London SW1W 955, Great Britain

Introduction

1. This paper describes the origins and the current nature of the present system of evaluation of expenditure programmes for supporting Science and Technology (S&T) activities in industry used by the UK Department of Trade and Industry (DTI). Since DTI was an early leader amongst UK government departments in carrying out systematic ex ante appraisal and ex post evaluation of S&T programmes, and is still seen as an example of best practice, its approach has spread to other departments. Thus to a considerable extent DTI's methods of evaluating S&T programmes can be seen as typical of the UK approach to evaluating government intervention to support industry and commerce. However, the UK approach to evaluation in, for example, the areas of social and employment policies, is often quite different as is the evaluation of academic research programmes. I would also hesitate before claiming that DTI evaluation of S&T support programmes was typical of the approaches used in other 'Anglo-Saxon' countries. All practitioners in the field of S&T evaluation should learn as much as possible from each other, adapting what they learn to make it appropriate to the circumstances of their own particular countries.

2. After a brief discussion of terminology and the 'Customer-Contractor' principle, this paper provides a short history of appraisal, monitoring and evaluation of DTI programmes for supporting S&T and innovation-related activities of UK business. It then describes the current organisation of evaluation DTI which is designed to ensure that evaluators are simultaneously
   - Independent of those responsible for the control and management of the programmes being evaluated but seen as competent and fair by the latter;
   - Knowledgeable about the nature and operation of those programmes;
   - Thoroughly professional in their knowledge and application of evaluation methodologies;
   - Free to express their considered judgement of programmes and know that these will be taken into account in future policy decisions.

DTI procedures for evaluation are then described in more detail.

Terminology

3. The term "evaluation" is used in more than one sense amongst S&T policy makers. In the UK and throughout this paper evaluation refers to the investigation of the outcome of S&T support after the event. Consideration of possible new programmes is referred to as "appraisal" while collection of data on the progress of programmes which are currently running is referred to as "monitoring". The appraisal, monitoring

and evaluation of DTI S&T support schemes are carried out as a unified system for ensuring 'total quality' and 'value for money' in S&T policy-making.

A Brief History of S&T Evaluation in DTI

4. The publication of The **Rothschild Report** in the early 1970s has had a significant, if indirect, influence on the way in which DTI S&T programmes are managed and therefore evaluated. The Report recommended that funds for more applied research should be held by government departments themselves who should act as intelligent customers in procuring such research from universities, research institutes etc. Although Rothschild's recommendations were not primarily about support for the S&T activities of industry, they did help to establish a culture in which UK government departments on S&T should have a clearly stated rationale and objectives.

5. The first significant S&T evaluation which DTI carried out was a pilot evaluation of the Product and Process Development Scheme (PPDS) which was completed in May 1984; PPDS was an early form of Support for Innovation (SFI) which provided support for industrial R&D projects (both single company and collaborative), related investments in high technology equipment, and technology transfer from the early 1980s onwards.

6. In 1983 the UK government launched the Alvey Programme which supported around £300 million of collaborative long-term research into Information and Communication Technologies funded jointly by industry and government (DTI, Ministry of Defence(MOD) and the then Science and Engineering Research Council (SERC)). The Director of the Alvey Programme, Dr. Brian Oakley, invited three groups of academics; the Programme of Research into Engineering Science & Technology (PREST) at Manchester University; the Science Policy Research Unit (SPRU) at Sussex University and the London Business School to carry out an evaluation of the Programme in real time. The evaluation of the Alvey Programme continued until 1991 three years after the programme ended, in order to allow time for the results of the research to be embodied in new products and processes.

7. The evaluation of Alvey made many contributions to the development of methodologies for evaluating S&T support programmes. Several members of the PREST and SPRU teams became acknowledged international experts and made a major contribution to the evaluation of the EC Framework Programmes and of similar collaborative research programmes in a number of other countries. DTI's present methods for evaluating collaborative research draw heavily on the lessons learnt from the evaluation of Alvey as do our systems of programme monitoring.

8. In 1985 the DTI carried out a major review of industrial support with the UK Treasury. Two of the results of that review were the development of a detailed rationale for government support for industry based on market failure and the agreement by DTI to develop a proforma which would set out the rationale, objectives, methods for project approval/appraisal, monitoring arrangements, and outline evaluation plan for all new DTI industrial support programmes. This proforma was

developed into a document called a ROAME (Rationale, Objectives, Appraisal, Monitoring & Evaluation) statement which is still the basis of our systems of approving new S&T support programmes, monitoring them during their life and evaluating them afterwards. The current use of the ROAME system is described in detail below.

9. The market failure rationale for government support for industry was based on the view that governments should not intervene unless the intervention could bring about some economically or socially desirable outcome which the free play of market forces would not bring about unaided, or would not bring about within an acceptable time-scale. To show that this was the case, proposals for new programmes had to show the existence of one or more of an identified list of market failures. Also, government support should only be offered where there was a reasonable expectation that the benefits from the support should exceed its costs. Thus a proposal for a new scheme should show how it would realise net economic (or social) benefits which would not otherwise be obtained. This requirement for 'additionality' has been a key condition of all DTI support to industry for S&T and innovation activities and identifying whether a scheme has yielded, or is likely to yield, net additional benefits is a vital part of any evaluation.

10. The market failure rationale for S&T support has been considerably refined since it was originally drawn up in 1985. The list of market failures has been added to and refined. Appraisal and evaluation of a wide range of support programmes together with analysis of a wide variety of policy problems and the results of much academic research has indicated that markets function in a number of ways not fully described in economic textbooks. For example, much resource allocation takes place within large commercial organisations which may perform indifferently for quite a long time before market forces bring about improvement or collapse. Also industrial S&T and innovation performance is affected by a wide range of institutions e.g. commercial law, business culture, standards (measurement, product specification, health and safety, environmental etc.); education system etc. and institutional failure which must be taken into account as well. More recently we have been moving towards a systems approach which defines our strategy for S&T support in relation to the performance of the UK National Innovation System.

11. Between 1984 and 1986 evaluation of DTI S&T programmes was mainly undertaken by the department's economists. In 1986 a multi-disciplinary Assessment Unit (AU) was set up within the Research and Technology Policy Division to undertake evaluations of S&T programmes. As a result DTI economists gave up much though not all of their role in S&T evaluation.

12. In 1994 as part of a reorganisation of DTI the economists and statisticians working on science, technology and innovation and certain high technology sectors were brought together with the Assessment Unit to form the Technology Economics, Statistics & Evaluation Directorate – TESE (see paragraph 16 below). This brought

together the two groups of professionals who had been responsible for evaluation of S&T programmes in the past.

13. A review is now underway of DTI strategy towards programme and policy evaluation. A more strategic approach to innovation policy is likely to require evaluations which are less programme-based and more concerned with the success (or failure) of broad areas of innovation policy. This might involve, inter alia, evaluation of groups of related schemes and programmes using statistical surveys aimed at the types of firms which receive support and comparative evaluations which compare the outcomes of similar schemes. More use will probably be made of social science research as a means of informing policy design and setting the counterfactuals against which the outcomes of industrial support programmes can be judged. Changes to the system of managing public expenditure as a whole are placing much greater emphasis on quantitative indicators of success and the government departments are now being required to achieve a set of measurable objectives.

How Evaluation of S&T Programmes is organised in DTI:
Expertise and Independence

14. All evaluations of S&T programmes encounter the problem of finding evaluators who are both
- sufficiently expert to carry out the evaluation effectively;
- independent of, and without connections to, the people or organisation responsible for the inception and running of the programme concerned.

This problem is particularly acute in the case of programmes designed to support specialised areas of advanced technology where it can be difficult to find a national expert who is not connected with the programme in one way or another. Using experts from abroad does not always solve this problem, as researchers in a particular area of science and technology often form quite close international communities or networks and an expert in one country may be the acquaintance or friend of researchers in a number of other countries.

15. This problem is most acute where an evaluation is designed to determine the relative quality of the research and expertise in a particular area of science and technology. It is less acute where the main purpose of the evaluation is to determine the economic benefits flowing from a particular area of technology and how successful national firms are in exploiting that area of technology. Most DTI evaluations of S&T programmes focus on the latter.

16. Evaluation of DTI S&T programmes are carried out by the Technology and Standards Assessment Unit (AU) which is now part of my Directorate, Technology, Economics, Statistics and Evaluation (TESE). The AU contains 5 professional evaluators, most of whom have a scientific or technological background. The present background of the AU evaluators includes chemistry, chemical engineering, environmental science, and information technology. Several of the evaluators also

have experience in budget and programme management. In carrying out evaluations the AU draws on the expertise of the five professional economists in TESE and in some cases one of the senior economists may take the lead in undertaking the evaluation. Advice on surveys and statistical matters is available from one of two professional statisticians (for more detail see http://www.dti.gov.uk/tese).

17. The economists and statisticians all work on providing advice on science, technology or innovation or on information communication and electronic technologies, aerospace or space. Economic and statistical advice on other sectors are provided by a separate, Directorate, but we try to involve these other colleagues wherever possible.

18. TESE reports to the Director of Technology and Standards who controls (the 'Budget Holder') the Innovation Budget. Apart from standards (see the table in paragraph 2) and a few other special areas of technology support, the Innovation Budget Holder and his staff are not themselves users of the Budget. Instead the Budget is spent by 18 other Directorates elsewhere in DTI who run S&T, Business Best Practice and Sectoral Support Programmes. The paramount interest of the Budget Holder is therefore to secure the best value for money from expenditures under the budget and to ensure that programme evaluation is carried out in a thorough and professional manner.

19. The quality and independence of evaluations of S&T programmes is also guaranteed by other administrative arrangements in DTI. Plans for forthcoming evaluation are submitted to peer review by an Evaluation Methodology Group (EMG), whose membership includes evaluators from the whole of DTI, a representative of the DTI Finance Directorate, and an independent chairman who is a senior economist from the central Economics and Statistics Directorate. EMG also considers all reports of completed evaluations. When an evaluation of an S&T programme is complete, the chair of EMG will write formally to the Innovation Budget Holder setting out the conclusions and recommendations of the evaluation regarding:
- the programme itself;
- any possible follow-on programmes;
- S&T programmes more generally.

The Innovation Budget Holder is required to reply, stating how he proposes to implement the evaluation's recommendations. In the case of Aerospace Evaluations, the Chairman of EMG currently writes to the Head of the Engineering Industries Directorate and in the case of evaluations of Space Programmes to the Director of the British National Space Centre.

20. The Evaluation Methodology Group (EMG) reports to the Evaluation and Policy Improvement Committee (EPIC) which has overall responsibility for evaluation of policies and programmes within DTI. It is EPIC that decides which programmes are to be evaluated and when. EPIC is the final arbitrator of any dispute between evaluators and the managers of the programme which is being evaluated. The Chairman of EPIC

is the DTI's Director of Finance who has the oversight of all the Department's budgets. It is he who will assure the Treasury and the National Audit Office that DTI has a proper system of programme appraisal, monitoring and evaluation.

21. Because TESE reports to the Holder of the Innovation Budget we not only have responsibility for evaluation of completed programmes, but we also play a major role in the appraisal of new programmes and help supervise monitoring of programmes currently underway. All new programmes must submit a ROAME statement to the Innovation Programmes Committee and will only receive funding if (a) the criteria of the ROAME system are satisfied and (b) the necessary uncommitted funds are available from the Innovation Budget. TESE plays a dual role. We will be one of the examiners of the new programme ROAME statement in the Programmes Committee, but we will have previously provided advice to the directorate proposing the new programme on how it should be drawn up.

22. The Assessment Unit (AU) also oversees the system of programme monitoring. Programme managers are required to send regular monitoring reports to the Programmes Committee Secretariat and the AU which has responsibility for chasing up any programmes which fail to satisfy this requirement. From time to time the AU presents the Programme Committee with reports on the state of monitoring of the Innovation Budget.

23. The relative amounts of manpower which TESE (including the AU) devotes to programmes appraisal, monitoring and evaluation can be seen from the following table which shows projections for the financial year 1999/2000

	Number of person years
Appraisal of new programmes	0.7
Continuous monitoring of programmes	1,2
Programme evaluation	3.1
Research associated with evaluation	0.8
TOTAL:	5.8
Of which	
Assessment unit (AU)	5.0
Economists and Statisticians	0.8

The details of how we undertake appraisal, monitoring and evaluation are described in the next section of the paper.

24. The AU also has a budget of up to £200,000 to engage outside consultants (either academics or commercial consultants) to help us in our evaluation work. We use consultants (a) where we need a particular type of expertise which we lack ourselves

(b) if we have more evaluation work than we have the manpower to undertake in-house.

25. The arrangements described above give us both the knowledge and expertise to evaluate S&T programmes and the independence and bureaucratic power to make fair and authoritative judgements on the success or failure of S&T Programmes. In particular:

(a) Our dual reporting role to the Innovation Budget Holder and the Director of Finance mean that programme managers (and their superiors) know they must co-operate with us and that we can judge the effectiveness of S&T programmes without fear of bureaucratic pressure.

(b) The ROAME system allows programme managers to know by what criteria they will be judged and to contribute to the determination of these criteria. By this means the evaluation system is seen to be fair.

(c) The undertaking of evaluations within a specialist unit, the AU, enables the building up of expertise in evaluation methodologies. This expertise is subject to peer review and benefits from the expertise of the economists and statisticians which are co-located with the evaluators.

Procedure for evaluating S&T Programmes in DTI

STEP 1 - Drafting of the Programme ROAME Statement

26. The procedure for evaluating S&T programmes starts with the drawing up of the ROAME statement when the programme is first proposed. Before the programme goes ahead, the Individual Programme Committee (IPC) chaired by the holder (controller) of the Innovation Budget agrees a ROAME statement for the programme. The ROAME system is an integrated approach to the

- approval of proposals for new programmes;
- continuous monitoring of live programmes; and
- evaluation of programmes, usually on or near completion.

27. The ROAME statement which embodies this system incorporates the following elements:

Rationale The rationale contains a clear statement of the overall purpose of the programme and a description of the market, organisational or institutional failure which means that the provision of DTI support will bring about economic benefits which would not otherwise be obtained. It will also contain a broad outline of the way the programme will address the market, organisational or institutional failure and how this will bring about the desired result.

Objectives The objectives section should refer to the overall DTI objectives and indicate which of these the proposed programme aims to achieve/contribute to. Programme objectives should provide strategic objectives for the activity. The ROAME statement should also contain more specific detailed objectives which the managers of the programme have agreed to address. These lower level operational objectives should be quantified, measurable, have deadlines for completion and will be the main basis against which the progress of the programme will be monitored.

Appraisal Depending on the type of activity proposed, the appraisal section will explain how individual projects will be chosen for support or how the department will decide upon a particular contractor to run the whole programme. Criteria will be listed against which both contractors and individual projects will be judged. This criteria will be specific for each programme, although there will obviously be some common themes, such as value for money.

Monitoring The arrangements for collecting information about progress towards achievement of objectives will be set out in this section.

Evaluation Information regarding possible timing of an evaluation will be included here. If the programme sponsor has any specific issues to be investigated, they will be included in the evaluation section. A brief outline of the evaluation plan may be included.

Evaluators have a role at each stage of the ROAME process which enables them to build up expertise in the programme areas which they evaluate.

28. **Approval of New Programmes** Any new programme proposal has to cover each of these ROAME elements and is subject to peer review by interested parties within the department. One object of the ROAME procedure is to agree with programme sponsors and managers the basis on which the programme concerned is to be monitored and evaluated. This is important. Evaluation is a social process and programme managers will be much more receptive to the evaluation process if they know in advance the criteria against which they will be judged and that the whole process is seen to be fair.

29. The approval of new programmes against stringent criteria means there is much less chance that those prospective programmes which are unlikely to achieve their objectives and lack a valid rationale will go ahead. This means in turn that there are fewer evaluations which are highly critical and that programme managers are more likely to see evaluation as a process which helps them perform their job better, rather than an inquisition designed to expose their faults. It is very important that the process be seen as transparent and fair. Under the innovation budget, all programmes over £1m are subject to the same level of scrutiny and are often amended and refined as a result of the discussion by the Innovation Programme Committee (IPC).

30. **Monitoring** Once the proposed programme is approved by The Innovation Programme Committee and DTI Ministers, it will be implemented. Monitoring of the progress of the programme is carried out during its lifetime against the objective (mainly the operational objectives) set. The aim is to require programme managers to monitor progress in a way which facilitates effective management of the programme and 'in flight' correction of any deviation from plan. This has the advantage of confining monitoring to that which a good programme manager would undertake anyway which should ensure that monitoring is done properly and is not seen as a unwelcome and unnecessary burden imposed from outside.

31. During the lifetime of live programmes (usually three years), monitoring ensures the quality of the programme is maintained. This monitoring information also feeds into any future evaluation. Monitoring progress towards agreed testable objectives allows action to be taken if the programme falls behind its targets.

32. Programme managers are required to send regular monitoring returns to the Assessment Unit who use these returns to provide the IPC with annual monitoring reports for each active programm (there are about 60 to 70 active programmes at any one time). Provided a programme is broadly on target and not experiencing significant difficulties, these reports are for information, only but they serve as a useful record for the purposes of subsequent evaluation.

STEP 2 - The Evaluation Planning Paper

33. At the beginning of each year the Evaluation and Programme Improvement Committee (EPIC) agree a programme of evaluations for the year ahead. Programmes are evaluated depending on a number of factors including:
- the timing of any potential follow-on activity;
- the size and importance of the programme and the amount of funding provided;
- its degree of novelty and whether a similar programme has been evaluated before;
- the extent of any potential lessons for DTI support for S&T as a whole;
- if it is a long standing programme the length of time since it was last evaluated.

The timing of an evaluation can pose some difficult trade-offs. Most DTI S&T support programmes run for three years. If an evaluation is to contribute to the decision as to whether there should be a follow-on programme, then the evaluation will need to commence at the beginning of the third year of the programme. This is fine for many technology transfer and spread of business practice programmes, but for programmes which provide support for R&D projects, it may be too soon to determine what the economic and commercial effects of such support is likely to be. In the case of programmes which support collaborative research it may be many years before the final outcome of such support can be known.

34. The formal evaluation process begins with the preparation of a planning paper. This is prepared by the evaluators and describes the methodology to be used, the main issues to be covered and an indication of the main sources of information to be deployed. Once the planning paper has been drawn up it is presented by the evaluators to the Evaluation Methodology Group (EMG). This is a group of evaluators from across the Department who have expertise from a wide variety of the DTI's areas of support and who meet regularly to review the evaluation activities of the Department.

35. EMG will discuss the planning paper to ensure the quality of the proposed evaluation approach, i.e. that the evaluation will be able to address all the relevant policy issues associated with the programme, as well as the validity of the rationale, achievement of objectives etc.

STEP 3 - The Evaluation Itself

36. Evaluation of S&T programmes is undertaken to
- demonstrate that programme has met its objectives and satisfied its rationale so that tax payers' money has been well spent;
- improve the effectiveness of current programmes;
- inform the design of future programmes;
- to increase the knowledge base on which S&T policies are founded.

The Assessment Unit carries out evaluations in teams of at least two evaluators. As stated in paragraph 16 above they are helped by advice and guidance from the economists and statisticians within TESE.

37. The methodology used to evaluate a particular programme will depend on the nature of the programme. For example, the evaluation of a collaborative research programme such as a LINK programme might include the following elements:

a) An examination of how the invitation to firms and universities to apply for funding and a draft business plan for the programme were derived from the ROAME statement:
b) An examination of how applications were processed and appraised to determine whether this was effective, timely and fair.
c) Comparison of the final set of chosen projects to see how far this corresponded to the original intentions expressed in the programme ROAME statement.
d) Analysis of the monitoring data for individual programmes to see how far they met their technical and other objectives etc.
e) Face to face interviews with programme participants designed to cover as many projects and as much programme expenditure as possible, while attempting to be as representative of the different types of projects and participant (large firms, small firms, universities) as possible. In large programmes these interviews might be supplemented by sending a written questionnaire or conducting a telephone survey of the managers or monitoring officers of the remaining projects. The evaluators may also talk to outside experts who are knowledgeable about the programme, to those who had applications rejected (some firms and universities may have had both successful and unsuccessful applications) and, possibly, to firms or universities who might have been expected to apply but did not do so.
f) Examination of the programme files and interviews (probably several) with the DTI programme manager, the programme co-ordinator if the DTI manager has appointed one, and members of the programme steering committee.

Evaluation of other types of S&T programme such as technology transfer programmes or support for R&D projects by small and medium-sized Enterprises (SMEs) will necessarily involve a different mix of evaluation activities.

38. The evaluators will endeavour to discover how far the research undertaken under the programme would not otherwise have taken place, at least in the same form and timescale (additionality). Even if firms did not spend more on research in that technological area as a result of the programme there may be substantial gains in terms of economies of scale and scope in combining their efforts and those of the participating universities. The evaluators will also investigate the degree and nature of collaboration between project participants and any barriers to collaboration such as difficulties over intellectual property rights. They will also look at the extent of information and other technology flows between projects and at any particular difficulties faced by SMEs who participated in the programme. They will also reach a judgement about how well the programme has been managed.

39. The evaluators will attempt to discover what plans firms may have to exploit the research supported under the programme and whether firms and universities intend to carry out follow-on research. In the case of long-term research, the extent of

exploitation may be far from clear at the time when the research has just been completed. On the other hand, if the evaluation were carried out several years later, other matters of interest such as additionality and programme management will be far harder to investigate. Once programme participants move on to other things, evaluation becomes that much more difficult.

40. Towards the end of the process a draft of the evaluation report is passed to the Programme Manager so that he can point out any factual errors.

STEP 4 - Internal Review of the draft Evaluation Report within TESE

41. The report is then subject to an internal discussion within the Technology Economics, Statistics and Evaluation Directorate (TESE), so that other members of TESE, including myself as Director, can comment. This provides a further in-depth review of the economic issues covered in the report and ensures the report adheres to agreed Assessment Unit standards of analytical quality and presentation.

STEP 5 - Presentation of the Draft Evaluation Report to the Evaluation Methodology Group

42. All evaluation reports are considered by EMG which provides further quality assurance and can arbitrate on differences of opinion between the evaluators and the programme manager (see paragraphs 24 and 39 above).

STEP 6 - Presentation of the Evaluation Report to the Innovation Programmes Committee

43. Evaluation reports are also presented to the IPC which originally approved the programme. The chair of EMG writes to the chair of the IPC and the head of the programme directorate, to explain the recommendations from the evaluation report. The IPC discusses the report, may suggest amendments to it and decides what action should be taken, and reports back to EMG. The IPC also decides whether the evaluation report should be published; publication is taken for granted unless there are very good reasons for not doing so. A list of published evaluation reports can be found on the TESE web site http://www.dti.gov.uk/tese.

STEP 7 - Policy Feedback

44. The conclusions of evaluation reports are fed into discussions of possible new programmes by the Innovation Programmes Committee - members of the AU and economists from TESE attend the IPC's meetings. From time to time, summaries of the results of evaluations are fed into strategic discussions of DTI budgets. Copies of evaluation reports are sent to all DTI programme managers who are likely to find them relevant and to participants in the programme concerned.

This closed loop feedback process ensures the quality of, not only evaluation results, but also current and future S&T support programmes.

French Culture

Rémi Barré
Observatoire des Sciences et des Techniques
93, rue de Vaugirard, 75006 Paris, France

Introduction: the functions of evaluation

The function of evaluation refers to the decision-making processes which regulate and structure the 'scientific enterprise'. It has relationships with other functions such as technology assessment and foresight, which are as well decision-making processes in the realm of science and technology.

Evaluation deals with the regulation and decision-making in two different realms:

- matters which are *internal* to the scientific community: the key concept is here the one of 'peer review', upon which the definition of scientific quality and scientific relevance are based, and
- matters which are *external* to the scientific community, which amounts to the regulation of the relationships between science and society. The concept of 'social demand' is here central, which defines the societal relevance of research.

At this stage, we can define evaluation as the function in charge of the quality and relevance of research, both on the scientific and in the societal grounds.
What is the linkage with culture ? What is the linkage with the 'new' Europe?

Evaluation cultures and their diversity

Culture can be defined as the result of the interactions between human 'nature' - and a social context, historically and geographically defined. It is thus a blend of universal and specific characteristics.

We suggest here that an 'evaluation culture' is the particular institutional, legal and organisational form the function of evaluation will take in a given country. The forms of evaluation will result from the interactions between the basic concepts underlying the scientific activity and its internal-external regulation principles, on one side, and the particular set-up of the national research and innovation system, the rules, regulations and institutions which structure higher education and public research organisations, including personnel management. For evaluation, such interaction will determine what is evaluated, by whom, with what consequences.

Obviously, the 'evaluation culture' produced by the dominant research systems - which happen to be also the dominant countries in geopolitical terms - will also be 'dominant'. It means that the Anglo-Saxon evaluation culture has become, quite naturally, the canonic model for evaluation. In this model, the central element is the financing of competitive grants by an agency to university teams in competition; thus the basic elements of evaluation are project and programme evaluation.

The French evaluation culture: some characteristics

I was to bring testimony here of the existence of other evaluation cultures, one of them being the one corresponding to the French institutional and legal set-up regarding higher education and public research.

One single rule has major consequences, which is that, in France, all personnel working in a university or a public research organisation is either a student or a civil servant. To put it simply (a bit too simply, because there are a few exceptions), there are no fixed term contract personnel, no 'post docs' in France. It follows that there are no funding agencies (like NSF or the research councils in the UK), because someone having a grant would have to hire personnel on fixed term, which is illegal (at least on a standard and general basis). Instead of an agency which allocates funds, we have an agency which allocates its (permanent) personnel to constitute academic laboratories along with the best faculty members (this agency is CNRS). We also have a fairly large set of finalised research organisations with permanent personnel, more or less linked to CNRS and university teams.

In such a setting, the evaluation culture is not oriented towards project and programme evaluation, but on individual, laboratory and institution levels :

- individual level: the question is to apply the rules of public service to the specific context of researchers; researchers have to be hired, promoted, oriented, counselled, and this is the basic task of individual evaluations, which is performed by panels which are partly elected by the researchers, partly nominated by the government.
- laboratory (or team, or unit) level: the task here is to organise a relevant programme of work through interaction with senior researchers and allocate the corresponding human and financial resources for that programme; the key issues are the collective scientific strategy of a group of researchers to be assessed, followed and accompanied by the evaluation structure, which is also made of partly elected, partly nominated members.
- institution (public research organisation, or university) level: the issue here is to assess the overall functioning and strategy of an organisation; these kinds of evaluation are performed by ad hoc structures which have only the capacity to give advice.

The problematique of evaluation in the French culture is therefore to recruit good people and manage their life-long activity in the structure, placing them in well-organised units with challenging research programmes; it is not to select good projects within the framework of a programme.

The challenges to the evaluation cultures and the coming of a European evaluation culture

Having said this, it certainly does not follow that nothing should or could change! To understand the reason for something and why its set-up is specific, does not imply it should not be discussed and changed.

Indeed, the French evaluation culture, like all evaluation cultures I presume, is fundamentally challenged by the major changes taking place both within the scientific arena, and also in the relationships between science and society. Accountability, relevance,

multidisciplinarity, flexibility, interactive capabilities, creativity are criteria to be taken into account, and clearly the evaluation culture has to evolve to produce decision-making processes coherent with such objectives.

What are the possibilities of Europe to face such challenges ?

Clearly, there are in Europe a variety of evaluation cultures, and each one must face drastic changes to perform its mission in a satisfactory way; this is the first challenge. But there is only one Europe, and research will increasingly be performed within a 'European scientific space' in the making. It is crucial that some sort of European evaluation culture emerges, in linkage with a European research and innovation system. This is the second challenge.

In fact, both challenges are linked, because the evolution of each evaluation culture can most meaningfully be done by interacting with the others. The starting point is to recognise all evaluation cultures must be reassessed and reshaped. From there comes the need to go beyond one's culture to enlarge the spectrum of experiences and possibilities; in turn, this can bring better and deeper understanding of one's own culture, of what is fundamental and what is socially shaped (and therefore changing).

Conclusion

In doing that, both challenges are addressed, and innovation in the evaluation cultures can emerge, with a definite European flavour, the flavour of a European evaluation culture. This culture will be in tune with the new research styles, the new relations between science and society and - last but not least - in tune with the new Europe. Its function will be to make quality, relevance and efficiency in science possible, in a European context. What we have heard of the evaluation such as it is organised for the FP5 undoubtedly points in that direction.

This is not wishful thinking, but a research and action agenda which, I suggest, makes sense, both for the politicians and the researchers.

Finnish (Nordic) Culture

Terttu Luukkonen
VTT Group for Technology Studies
Tekniikantie 12, 02044 VTT, Finland

Introduction

The Nordic[1] countries have joint features in their evaluation cultures exemplifying the fact that local (national) cultural features affect the way in which evaluations evolve and showing that the evaluation cultures of countries adopt features of the local culture. These joint features include in particular 1) openness and publicity and 2) the use of foreign experts.

In their evaluations, the Nordic countries have adopted a general principle that evaluation reports are published and available for everyone interested. To my knowledge, this principle is also applied in practice, with a very few exceptions, such as internal evaluations carried out in research organisations or very rarely, evaluation reports that turn out to be controversial and are contested by those commissioning them. The principle of publicity is related to a legitimation of evaluation activities through their transparency. It is also an application of the general principle of publicity for official documents and the right of individuals to be acquainted with the criteria used in decision-making concerning their own case. The principle of publicity has, aside from great advantages, also some disadvantages: e.g. critical evaluation of individual researchers' work is subject to public exposure, thus creating additional psychological pressures[2]. However, so far the advantages have been judged to be stronger than the disadvantages.

As to the second common feature, the use of outside experts, the Nordic countries are small and afraid of provincialism. They are therefore willing to call upon outside advice and to use independent experts in evaluation. We may argue that even much bigger countries have not truly independent experts for evaluation. In these fairly small countries, the problem is greater. In most research areas, there would not be independent experts within the country, which means that evaluations use international experts from other countries. By independent, I mean that if the evaluators are experts in the field to be evaluated, it is important that their own research activities will not be affected by the consequences of the evaluation, e.g. changes in resource allocation.

As referred to above, the fact that the Nordic countries have joint features in evaluation cultures is related to their common cultural heritage. It has also a more specific reason. In the 80s, there was a diffusion of evaluation models from one Nordic country to another. This is not surprising, since there is a long tradition of mutual collaboration in the conduct, funding and promotion of research activities among the Nordic countries.

---

19 By Nordic, I am here referring to four countries with strong evaluation cultures: Denmark, Finland, Norway, and Sweden.

20 Luukkonen, Terttu, The impacts of research field evaluations on research practice, Research Policy, 24 (1995) 349-365.

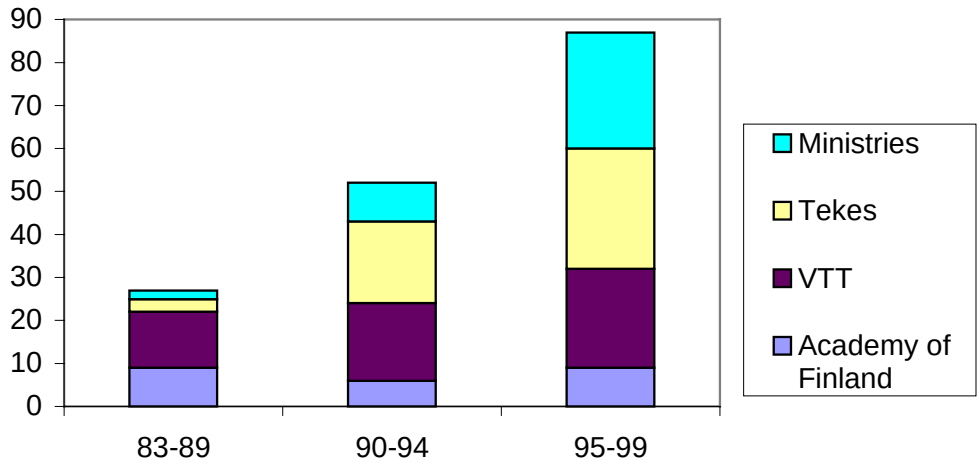
Collaboration and exchange of ideas include policy initiatives to support R&D activities, and these countries have diffused models of research funding and organisation to each other.

In the Nordic countries, evaluation is nowadays a routine activity and applied to most public initiatives in the area of research funding, research programmes, research centres, and new initiatives in research and technology support. As mentioned above, evaluation criteria, methods and uses have multiplied. The Nordic countries have developed an evaluation culture where evaluation is one of the mechanisms by which public accountability and transparency of public R&D funding are being provided. They have other functions, too, as will be seen below in the Finnish example.

Sharp increase in evaluation in Finland

Finland has witnessed a great increase in evaluation activities since the 80s (Figure 1). By evaluation here and above, I mean science and technology policy evaluation, which is characterised by evaluation of larger entities than individual researchers or their projects.

Figure 1. Research evaluations in Finland in the 80s and 90s



The most important organisations that have commissioned evaluations are TEKES, the Technology Development Centre of Finland, a major funding agency for technology development in companies, VTT (Technical Research Centre of Finland, a contract research institute in technology and engineering with nearly 3,000 employees), and Ministries. The Ministry of Education has been particularly active in commissioning evaluations of universities. The organisation that introduced science and technology policy evaluation in Finland, the Academy of Finland, i.e. the research council system, has been relatively less active in recent years. Its evaluation efforts have particularly been devoted to an overall review of Finnish science. The first round of such an exercise was completed in 1997, and another round is just starting.

How to create an evaluation culture: the Finnish example -
Factors that have contributed to the rapid growth in evaluation activities in Finland

(a) Dissemination of evaluations in Finland

The growth of evaluation activities can be seen as a process of dissemination of an innovation. The innovation here is evaluation as a management tool in science and technology policy.

In Finland, there was a visible start for the dissemination process related to the principle of publicity. As already said, the Academy of Finland first introduced science and technology policy evaluation to Finland in the early 80s. The first evaluation, that in inorganic chemistry, completed in 1983, was highly critical and outspoken and attracted a lot of attention, not only in the evaluated field, but also in other fields. It even got wide media coverage. Researchers in other fields, research organisations and research funding agencies could not fail to take note of this new tool.

There was a snowball effect: evaluations were experimented with and adopted in various ministries and funding agencies. A really rapid growth, however, took place in the 90s: TEKES has been commissioning evaluations for all its research and technology programmes on a routine basis since the late 80s. Similarly, VTT adopted the practice of evaluating its research programmes in the late 80s, and in the 90s, started evaluating its research units systematically. At the moment in Finland, all major initiatives to promote R&D are evaluated, whether or not they are research programmes, a major increase in R&D funds or institutional change.

(b) Evaluations in support of science and technology policy strategy

An important background reason for the fact that evaluations have gained wide use is that they support the country's science and technology policy strategy. In Finland, science and technology policy has long been among the priority areas in government policy and its value has not been questioned. The aim is to raise the technological level and to transform the technological basis of Finnish industry and the economy, as well as to benefit other policy sectors and the whole societal development. According to the title of a report by the Science and Technology Policy Council of Finland in mid 90s[3], the goal is to make the country *a knowledge-based society*. The effects of this strategy are also seen in the rapid growth of R&D expenditures[4].

Evaluations have played a part in the strategy by contributing first to the *internationalisation* of Finnish research and to *quality improvement* of research activities through the use of international experts as evaluators and the international forefront of research as the main evaluation criterion. Fostering high quality research has been seen as an important

---

21 Science and Technology Policy Council of Finland, Finland: A Knowledge-Based Society. Edita, Helsinki 1996.

22 In 1999 R&D funds are estimated to be about 3 % of GDP.

element in the improvement of the R&D infrastructures and capabilities. Evaluations have also played a role in *reorganising research institutions* in an effort to make them better adapted to changing needs. If outside experts suggest institutional change, such suggestions have carried more authority. In some cases, evaluations have been commissioned to get more support for already planned institutional reorganisation.

Evaluations can also be seen as a process through which the perceptions and needs of a wider group of participants/stakeholders in the policy process can be articulated and introduced into public discussion. Evaluation panels have hearings in which they collect systematically the views of the researchers to be evaluated or those of the stakeholders in the policy process. The number of people thus heard exceeds that in the more traditional governmental committees. Evaluation is a new kind of *negotiation process* within research policy.

In Finland, evaluations have been carried out during times both of rapid and of slower growth in research funding. In the past twenty years or so, there has been a continuous growth of R&D funds with periods of quicker growth and periods of slower growth. There has been a demand for evaluations in both types of periods. Interestingly, however, there has been an increasing pressure for *accountability* and for proving *the usefulness of public investment in R&D* during the period of rapid growth of public R&D funds, like the last few years in Finland. Increasing demands for accountability and evaluation have led to new evaluation efforts taking place simultaneously with the new investments, the effects of which are being evaluated.

(c) Advocacy of evaluation

Aside from the fact that evaluations have contributed to the science and technology policy strategy, there has been strong institutional advocacy of evaluation in Finland. The Cabinet level advisory body, Science and Technology Council of Finland, has been an active promoter of evaluation in Finland. The Council is chaired by the Prime Minister and has the ministers of the most important research funding ministries amongst its members, as well as the Finance Minister, and also includes prestigious scientists, university rectors, heads of important research institutes, and representatives of business enterprises. This high-level and prestigious body has repeatedly paid attention to the need for evaluating different parts of the research system and has given mandates and guidelines for such activities.

It should, however, be remembered that it is not only a question of a top-down approach, but rather - or in many cases - of mutual negotiation. For instance, the Academy of Finland got its mandate to carry out evaluations in research fields beyond the research areas it funded itself from the (then) Science Policy Council in the early 80s. However, the Academy had asked for this mandate and had (informally) suggested to the Council that it be given.

(d) Evaluation champions in promoting evaluation

There have been two active evaluation champions in Finland. By evaluation champion, I mean a person or persons engaged in active promotion of evaluation. Institutions do not function without people who take initiatives. Thus people, or individuals, matter. In our case, the two long-standing planning officers of the Science and Technology Policy Council have been our evaluation champions. They have kept evaluations on the agenda and provided the institutional memory for the Council, which is re-appointed every three years.

(e) Institutional and personal self-interest

Last but not least is the fact that many groups (researchers included) have realised the opportunities offered by evaluation to promote the status of - and potentially to argue for more money for - the field or institution. This was particularly evident in the early phase of evaluations. In some instances, evaluations have become a tool in struggles over resource allocations (esp. in the universities). Thus institutional/personal self-interest has promoted evaluation.

Another side of the coin is that people (institutions) to be evaluated have had to accept being evaluated. There has not been strong resistance to evaluation in Finland after the first shock in the early 80s in the field that had been critically evaluated. By contrast, peer review types of evaluation procedures are widely accepted. The use of quantitative indicators in evaluation and resource allocation has met with some criticism and resistance, particularly at universities. In spite of this, indicator data are used as criteria in the allocation of university funds among the universities, as well as in internal allocations within the universities. Actual research evaluation relies either on modified peer review or the use of evaluation consultants who use a variety of data collection methods, maybe combined with modified peer review.

Evaluation fatigue can, however, be noticed in some parts of the research system, parts, which have been subject to frequent evaluations. Particularly scientists who work in centres of excellence have been subject to frequent evaluations. Evaluation fatigue in some parts of the research system has, among other things, contributed to the fact that the research councils of the Academy of Finland (Figure 1) have not been particularly eager to launch new evaluations of research fields.

Conclusions

Evaluation is not a goal in itself. It is costly and time-consuming. Drawing on the points presented above, the spread of an evaluation culture is facilitated if:

- Evaluation contributes to policy goals and is seen as useful by the stakeholders in the policy process. As in Finland, a specific science and technology policy strategy to which evaluation can contribute is important in this respect.
- There is an institutional advocacy of evaluation giving it a longer-term basis.

- Knowledge of the methods and uses of evaluation are disseminated, creating a need to educate potential users and providers of evaluation data.
- Those carrying out research activities with public funding accept an obligation to be accountable for their use of the funding and to co-operate in evaluation efforts. If they perceive evaluation as useful for their own agendas, they are more motivated to participate in evaluation.

Evaluation in the Context of the Fifth EU RTD Framework Programme

Chairman's Opening Remarks

José Vianna-Baptista
ICAT – Instituto de Ciencia, Aplicada e Tecnologia
Campo Grande, 1700 Lisbon, Portugal

EU RTD initiatives are supposed to create European added value observing the subsidiarity principle. Such initiatives are oriented to selected objectives which are most effectively pursued at Community level. They should produce significant contributions to the development of Community policies.

Since the First Framework Programme dating back from 1984 up to the present Fifth Framework Programme, the objectives have been broadening. While the initial orientations were supplier-oriented and focused on the effect of technology push, the following ones became market-oriented and aiming towards European competitiveness. The Fifth Framework Programme is society-oriented, focusing on sustainable growth. The obvious consequence was an increasing complexity and diversity of the programme contents, becoming more heterogeneous.

While the expected results were supposed to produce impacts of growing importance both in qualitative terms as well as in the time horizon, the evaluation of the Programmes have become more complex.

Redefinition of policy aims entails modifying instruments and methodologies adequate to effective and timely evaluations. Since the RTD objectives are formulated in terms of socio-economic concerns, the evaluation of impacts has to cover a wider scope and a longer period of time. In addition, it becomes more difficult to identify the results produced by the Framework Programme since they can hardly be isolated from the influence of other factors and changes occurring over the same period of time.

The Fifth Framework Programme can be interpreted as a kind of social contract addressing the challenges European Society is facing in terms of development, quality of life and environment. On account of this policy orientation, the Programme has been organised in a different form, essentially based on a concept of key actions with thematic and horizontal programmes.

Broadening the objectives and focusing on socio-economic relevance imply a much more complex evaluation, recognising a wider range of impacts and identifying a greater diversity of actors and stakeholders.

Clustering of projects will probably become more frequent, promoting enhanced network effects which will also increase the complexity of the analysis necessary for conducting the evaluation exercises.

It is also important to bear in mind that during the period covered by the Fifth Framework Programme a number of projects of the Fourth Programme will still be under development and generating effects which can not be measured immediately.

In any case there is a time lag between the conclusion of a project and the emergence of the first results. Such a time lag is much longer whenever we look to impacts of socio-economic nature, such as those related with the implementation of policies for sustainable growth, environment protection or employment.

In such a context the evaluation of the Fifth Framework Programme will be a very demanding exercise. It will require innovative approaches, eventually the use of multiple models and methodologies and even diverse contractual rules.

EU RTD Programmes Impact/Results Assessment: A Manifold Task in a Complex System

Y. Dumont, L. Durieux, I. Karatzas, L. O'Sullivan, H. Teuber, G. Stroud, G. Fayl
European Commission, Directorate General XII, Science, Research and Development
De Meeûs Square 8, 1050 Brussels, Belgium

1. Background: a complex structure

The European RTD policy is implemented via a relatively large and complex operational tool known as the Framework Programme. The latter covers most of the major scientific domains with annually more than 10.000 running RTD projects involving over 30,000 participants (research groups, individual researchers, industrialists, universities, …) from the 15 Member States and a number of other countries (associated or not to the Framework Programme). From an administrative viewpoint, 10 out of 24 Directorates-General of the Commission are involved in the implementation of the Framework Programme, requiring efficient consultation and co-ordination mechanisms. In order to guarantee the transparency and to ensure that the actual needs of the different stakeholders are taken into account, several advisory bodies, scientific and political, have been set up. This complex system, managed by the Commission services, is dynamic, responding to changing conditions while being under careful scrutiny by the other main European institutions, namely the Council and the European Parliament.

The Framework Programme's complex structure and the pervasive nature of research, including the often lengthy time required for research results to be translated into socio-economic benefits, make the evaluation of the Framework Programme an intricate yet challenging process.

2. Experiences acquired so far: cementing a European Evaluation Culture

During the last two decades, the Commission services have acquired a solid experience in the field of research evaluation. Evaluation has been a legislative requirement for European RTD programmes since the early 1980s. For the first three European RTD Framework Programmes (FP), more than 70 programme evaluations and more than 40 supporting studies have been carried out, all in all involving more than 500 European experts.

The changing S&T environment and the increasing pressure for timely, independent evaluation has led to the current evaluation system, implemented since the beginning of FP4 and which is based on two activities: for each programme, a *continuous monitoring* with the assistance of experts external to the Commission Services, and, at multi-annual intervals, a *five-year assessment* conducted by external experts. These two exercises are integrated in a coherent scheme implemented in a harmonised manner across the Framework Programme (see Figure 1). While this scheme is fulfilling its mission, the need to better demonstrate the usefulness of a European RTD policy increases the demand for

verifiable impact. The forthcoming five-year assessment under FP5[1], which should lead to final reports by mid-2000 as an input to the preparation of FP6, offers the opportunity to take an in-depth look at the impact of the EU RTD programmes.

3. The new legal requirements: setting the objectives

The FP5 legal text sets out the evaluation objectives:

> *"The Commission shall continually and systematically monitor each year, with the help of independent qualified experts, the implementation of the Fifth Framework Programme and its Specific Programmes in the light of the criteria set out in particular in Annex I [2] and the scientific and technological objectives set out in Annex II. It shall assess, in particular, whether the objectives, priorities and financial resources are still appropriate to the changing situation. Where appropriate, it shall submit proposals to adapt or supplement the Framework Programme and/or the Specific Programmes, taking into account of the results of this assessment.*
>
> *Before submitting its proposal for the Sixth Framework Programme, the Commission shall have an external assessment conducted by independent highly qualify experts into the implementation and achievements of Community activities carried out during the five years preceding that assessment in light of the criteria set out in Annex I and the scientific objectives set out in Annex II.."*

The criteria referred to above - which relate to: i) the Community 'added value' and the subsidiarity principle; ii) social objectives; and iii) economic development and scientific and technological prospects – bring the socio-economic impact assessment issue to the forefront of the evaluation process. The forthcoming five-year assessment exercise will cover FP5, FP4 and FP3 activities (as outlined in Figure 2). The challenge for this assessment is to evaluate the past activities in the light of the new criteria and scientific objectives.

4. The evaluation challenge posed by FP5: the design of an effective impact assessment process

The FP5 introduced a problem-solving approach and an emphasis on socio-economic objectives which provide the starting point for the selection of impact assessment techniques. The more the RTD objectives are expressed in terms of expected socio-economic impacts, the more focused and comprehensive the evaluation ought to be. Furthermore, it should distinguish and assess the objectives in their proper context, reflecting the three primary levels of operation: at project, programme and FP levels.

In addition, the collection of relevant information must be organized in a manner suitable for the evaluation exercise. Without ignoring input data (number of contracts, type of participants, etc), more attention is paid to output data which, for FP5 projects, will be col-

---

23 Decision No. 182/1999/EC of the European Parliament and of the Council, of 22 December 1998, concerning the Fifth Framework Programme of the European Community for Research, Technological Development and Demonstration Activities (1998 to 2002).

24 Also attached as ANNEX I to this paper.

lected systematically during and for some time after the completion of a project (3-5 years) in order to capture elements of impact and the developing trends. A schematic representation of the information management scheme is given in Figures 3 and 4.

In line with the recommendations provided by the CREST[25] Evaluation Sub-Committee and the Framework Programme Monitoring Panels, the systematic collection of input/output/impact data is set up in order to facilitate comparisons over time and programmes as well as aggregation at Framework Programme level.

Supporting studies and improvement in availability of quantitative and qualitative data will provide a solid basis for Framework Programme socio-economic impact assessment. Particularly noteworthy in this respect is the integration between the published criteria against which project proposals are assessed before they are accepted for funding and the quality indicators used during the lifetime of each project. Under FP5, proposal assessment, project monitoring and ex-post evaluation are part of a much more seamless whole than in the past.

While the existing evaluation scheme, introduced in the context of FP4, will remain unchanged in its basic principles and components, its efficacy will be improved by supporting activities that will rely on data collection.

5. The key issues to be addressed: assessing socio-economic impact and European added value

In the context detailed in the above sections, the key issues to be addressed when enhancing the evaluation scheme can be briefly outlined with the following list of questions assessing socio-economic impact and European added value. The latter should be read keeping in mind that the enhancement should, in addition to the satisfaction of the legal requirements, increase the usefulness of the evaluation scheme both for the programme managers and the policy-makers. While the 'products' to be delivered will still be the same (yearly monitoring and five-year assessment reports), the work that will support it should provide an increasingly significant added value. This should, besides facilitating optimal implementation of FP5 and future programme design, provide effective support for an EU RTD policy by providing concrete evidence of its usefulness for and impact on society. In more general terms, further improvements in the efficiency of the Commission's evaluation scheme, more than a necessary management tool, should lead to a better appreciation of research outcome at the political level, and greater public understanding of research; namely, the evaluation reports are one of the 'shop windows' of the programmes through which it is demonstrated that the programme objectives are being achieved.

- *What are the different kinds of impact of public RTD programmes (taking into account scientific, socio-economic, European policy-related objectives)?*

---

25 CREST: Committee for Scientific and Technological Research, a body advising the Commission and the Council of Ministers on RTD policy-related issues.

- *What are the limits in assessing these impacts?*
- *What are the implications of these limits on the evaluation of the impact of European RTD programmes particularly from the viewpoint of the annual monitoring and five-year assessment exercises (under the particular requirements of FP5)? Taking into account FP5 features (objectives and criteria) and their potential implications for evaluation, what will be the most useful focus?*
- *Is it possible to isolate the FP impact from other factors (e.g. as far as employment is concerned), including other research efforts? How is the FP's European added value best evaluated? What is the most appropriate time scale for such an evaluation?*

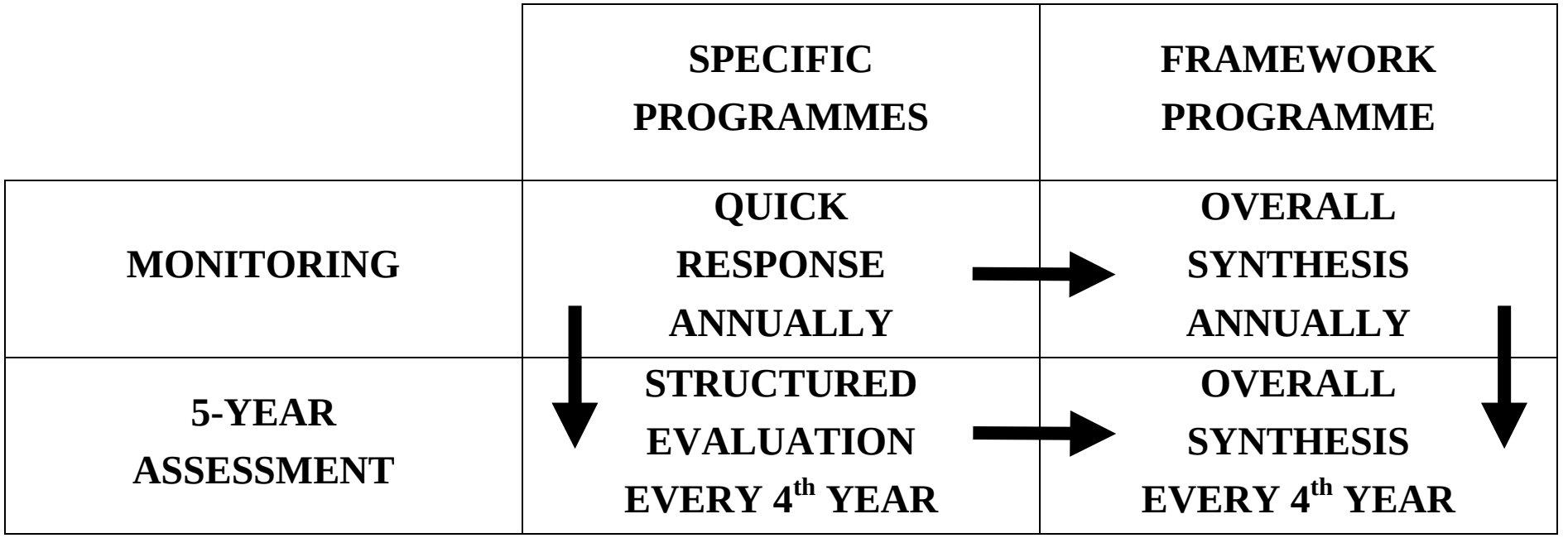
6. Enhanced evaluation scheme: a framework for decision-making

In order to address the above questions, the Commission, in the course of 1998, set up under ETAN (European Technological Assessment Network), a group of experts that issued in January 1999 a report on "Options and Limits for Assessing the Socio-economic Impact of EU RTD Programmes". As the title suggests, this report discussed the limitations of impact evaluations and underlined the fact that policy-makers can act for the best only if they understand what can and cannot be achieved. To this end, the authors proposed a basic model (Figure 5) of support to the evaluation efforts in the context of the Framework Programme. The model combines studies on sectoral impact, questionnaires, scientometrics with reviews in the changes in the S&T environment, and RTD econometrics as major inputs to the evaluation process.

Lastly, as FP5 has been designed in order to provide for maximum flexibility within the legal decisions, decision-makers and programme managers should have at their disposal timely information on actual performance (see point 4 above). This implies that an efficient self-monitoring system will complement in real-time the annual monitoring carried out with the assistance of independent experts.

Figure 1

INTEGRATED EXTERNAL MONITORING AND 5-YEAR ASSESSMENT



ANNEX 1

CRITERIA FOR SELECTING THE THEMES AND OBJECTIVES OF COMMUNITY RTD ACTIVITIES[1]

1. The European Community's RTD policy is directed towards strengthening the scientific and technological bases of Community industry and encouraging it to become more competitive at international level, while promoting all the research activities deemed necessary by virtue of other Chapters of the Treaty. It shall also contribute to promoting the quality of life of the Community's citizens and to the sustainable development of the Community as a whole, including the ecological aspects. Its implementation is based on the twin principles of scientific and technological excellence and relevance to the abovementioned objectives.
   Moreover, in pursuit of a cost-benefit approach dictated by concern for optimum allocation of European public funding and in accordance with the subsidiarity principle, themes for the Fifth Framework Programme and the related objectives are selected on the basis that the Community shall take action only if and insofar as the objectives cannot be sufficiently achieved by the Member States.

2. In application of the foregoing principles, the Framework Programme shall be defined on the basis of a set of common criteria, divided into three categories:
   - Criteria related to the Community "value-added" and the subsidiarity principle:
     - need to establish a "critical mass" in human and financial terms, in particular through the combination of the complementary expertise and resources available in the various Member States,
     - significant contribution to the implementation of one or more Community policies,
     - addressing of problems arising at Community level, or questions relating to aspects of standardisation, or questions connected with the development of the European area,

     so as to select only objectives which are more efficiently pursued at the Community level by means of research activities conducted at that level.
   - Criteria related to social objectives:
     - improving the employment situation,
     - promoting the quality of life and health,
     - preserving the environment,

     in order to further major social objectives of the Community reflecting the expectations and concerns of its citizens.
   - Criteria related to economic development and scientific and technological prospects:
     - areas which are expanding and create good growth prospects,
     - areas in which Community businesses can and must become more competitive,

---

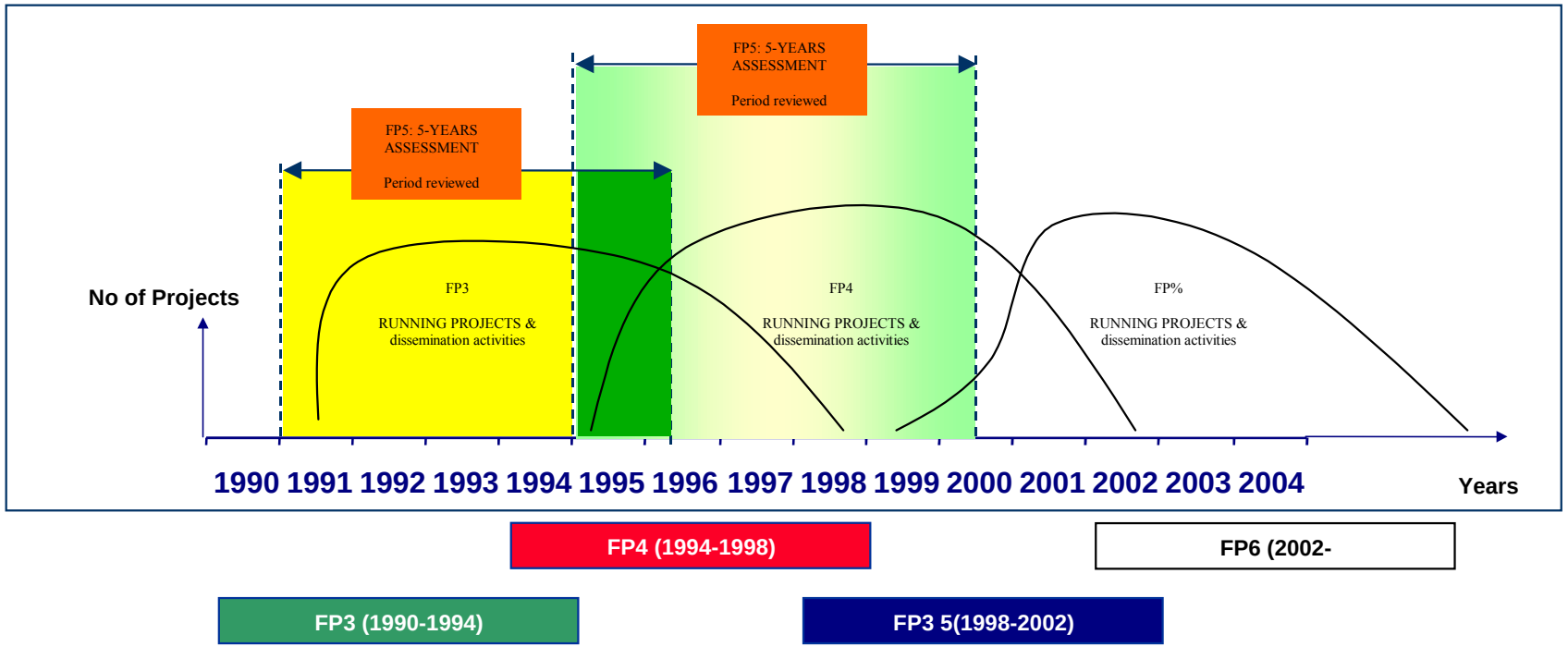
26 Criteria as listed in Annex I of the "Decision No. 182/1999/EC of the European Parliament and of the Council, of 22 December 1998, concerning the fifth Framework Programme of the European Community for Research, Technological Development and Demonstration Activities (1998 TO 2002)."

- areas in which prospects of significant scientific and technological progress are opening up, offering possibilities for dissemination and exploitation of results in the medium or long term,

in order to contribute to the harmonious and sustainable development of the Community as a whole.

The criteria referred to in paragraph 2 will be used, and where necessary supplemented, for the purposes of the implementation of the Fifth Framework Programme, in order to define the specific programmes and select research and technological development activities, including demonstration activities. The three categories of criteria will apply simultaneously and must all be met, although to a different extent from case to case.

Figure 2: The assessment scheme and Framework Programme implementation: overlap and continuity



1. At any given moment there is considerable overlap between the Framework Programmes which are active in funding RTD projects. For results assessment purposes the overlaps are even greater.
2. The FP4 5-year assessment was carried out during the period mid 1996 - early 1997. As far as the FP5 5-year assessment is concerned, it is expected to be carried out during the period mid 1999 - mid 2000.
3. In interpreting the graph, it should be borne in mind that it is not to scale: its main emphasis is on the timing of project activity and not on the actual numbers of projects or projects per FP.

Figure 3

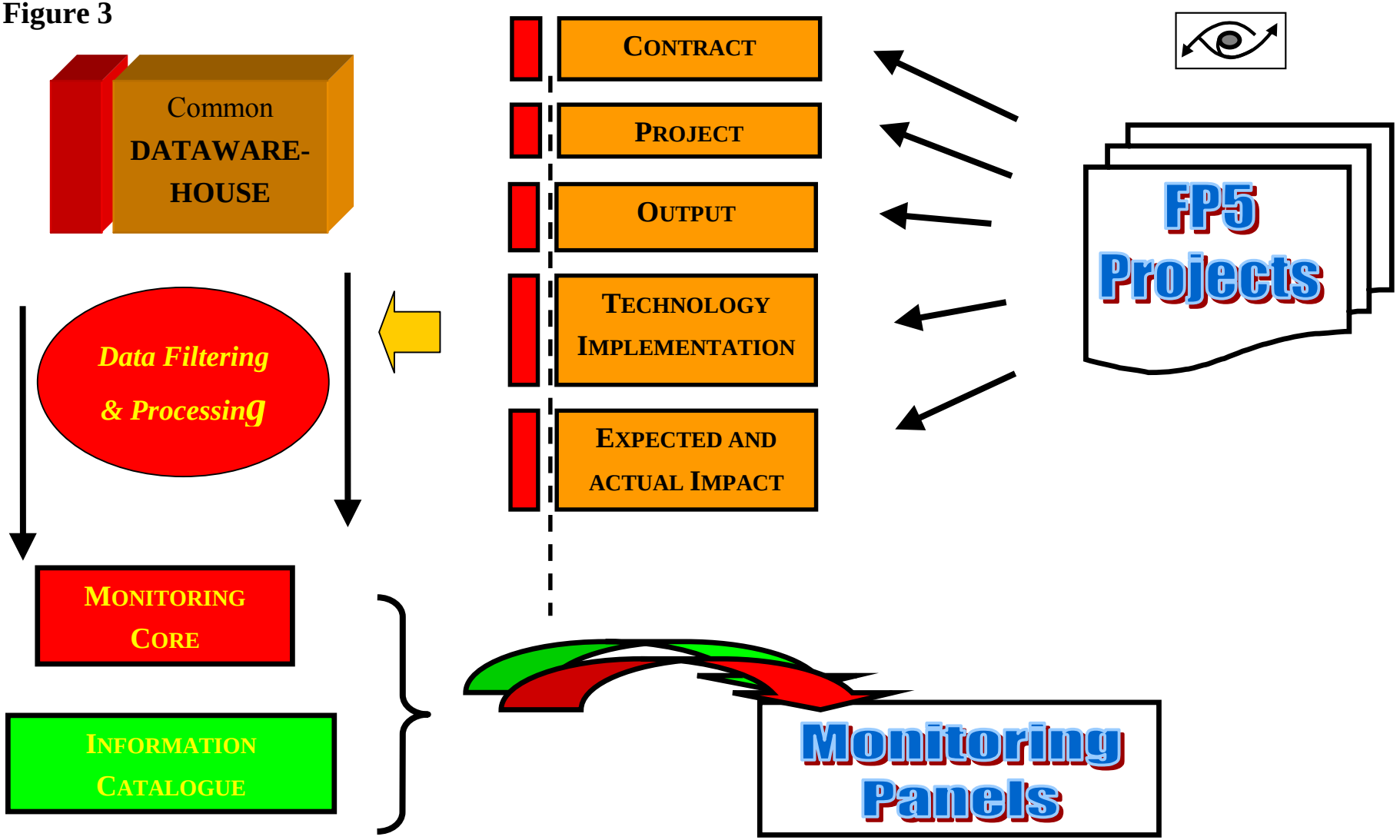

Common
DATAWARE-
HOUSE
Data Filtering
& Processing
MONITORING
CORE
INFORMATION
CATALOGUE
CONTRACT
PROJECT
OUTPUT
TECHNOLOGY
IMPLEMENTATION
EXPECTED AND
ACTUAL IMPACT
FP5
Projects
Monitoring
Panels

Figure 4

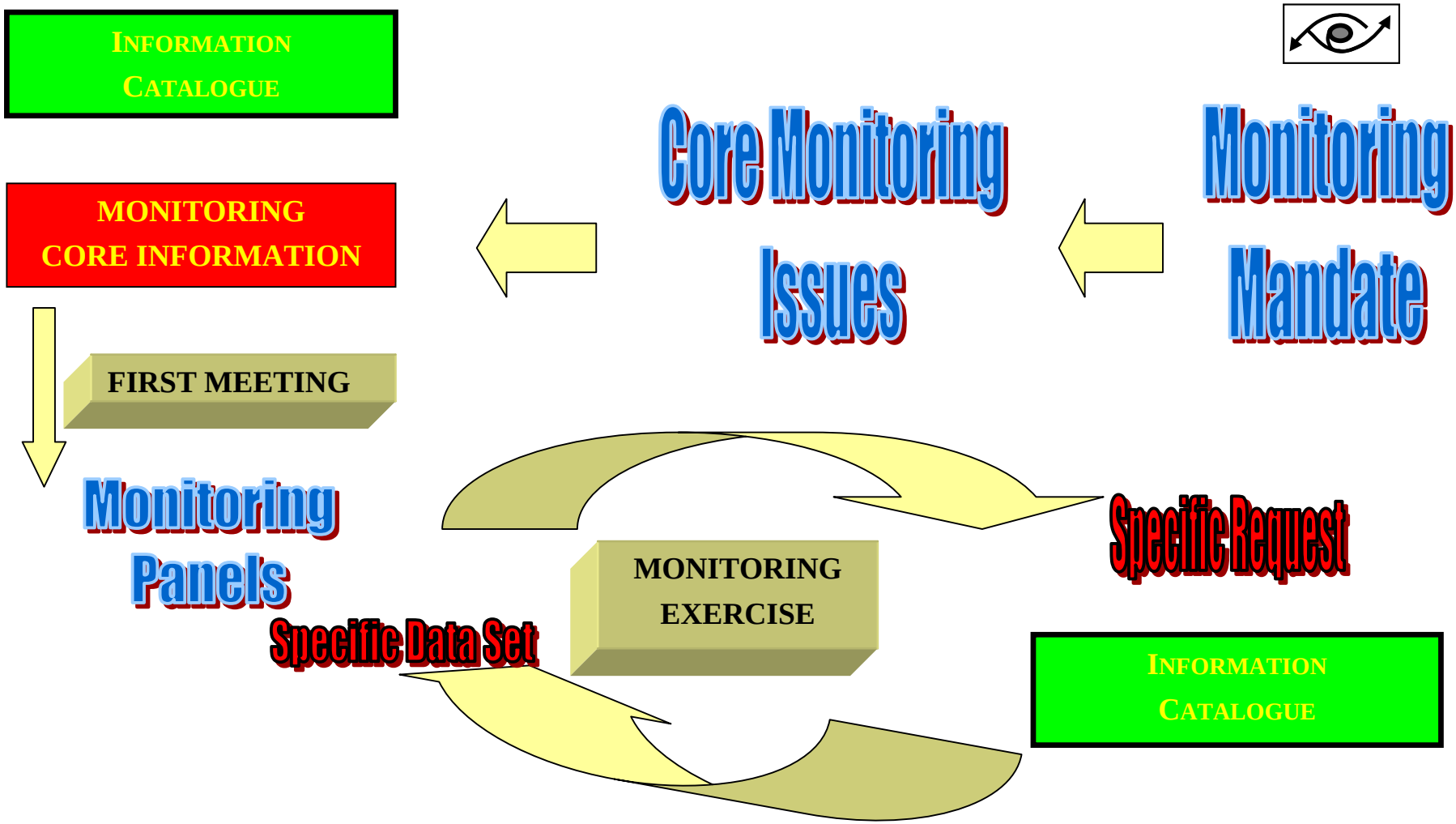
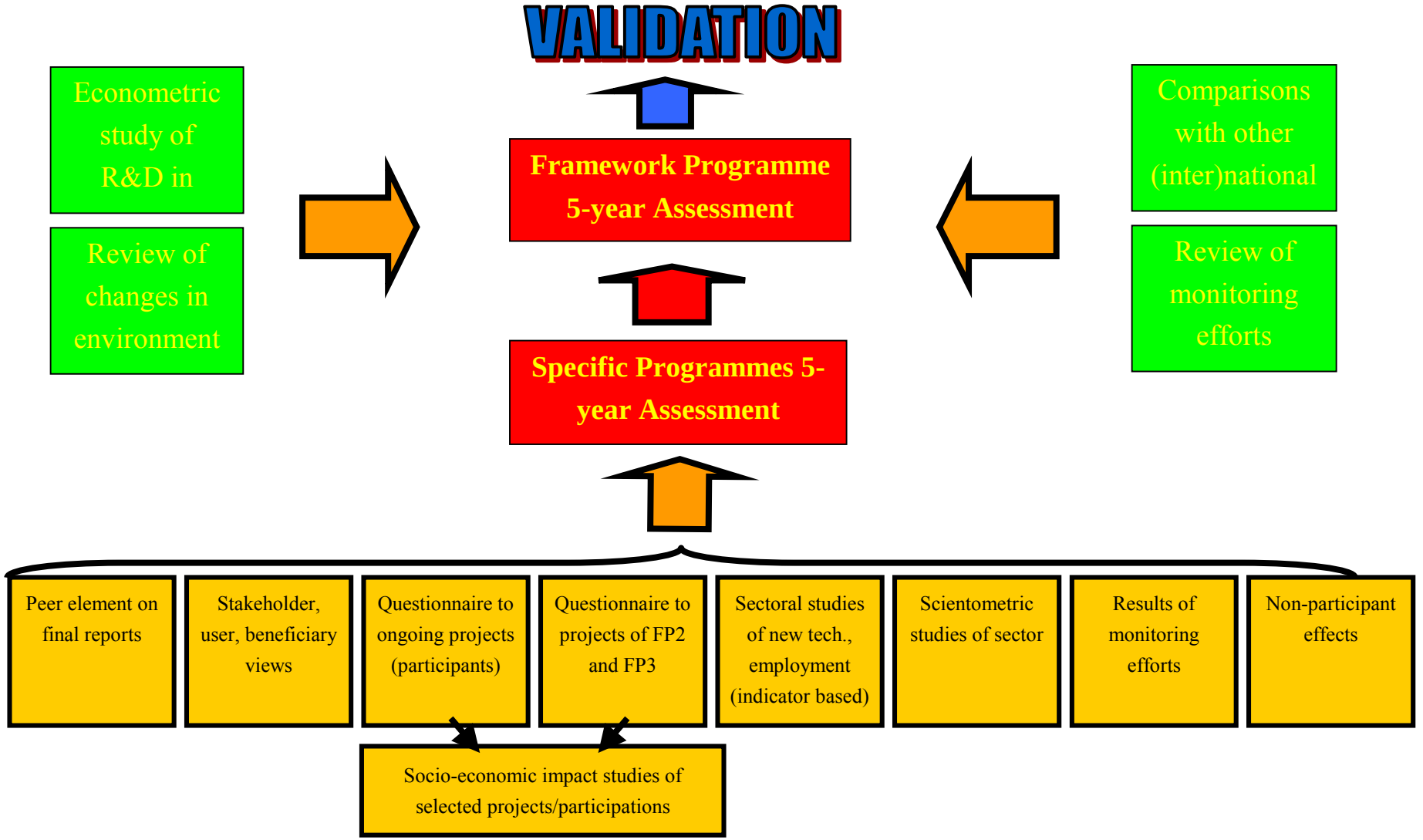

INFORMATION CATALOGUE
MONITORING CORE INFORMATION
FIRST MEETING
Monitoring Panels
Specific Data Set
MONITORING EXERCISE
Specific Request
INFORMATION CATALOGUE
Core Monitoring Issues
Monitoring Mandate

Figure 5: Potential assessment steps for the Fifth Framework Programme



Evaluation as an ongoing Process - the View of a Practitioner

Niels Busch
Busch&Partners
Mandalsgarde 4 floor, 4th, 2100 Copenhagen, Denmark

Introduction

This contribution is by a practitioner. We shall therefore take off from empirical evidence of the immediate past and start with a brief review of the latest monitoring of the Fourth Framework Programme (FP4). In principle, we are then discussing monitoring and not evaluatiom and assessment, but because 1998 was the last year of FP4 before the start of FP5, the Framework Programme Monitoring Panel (the FPM Panel) took a broader view of its task than justified by the definition suggested by the ETAN Reflection Group[1]: Monitoring is a process by which information on the progress and direction of ongoing RTD actions is generated mainly for management purposes.

The ETAN Group spent considerable time discussing suitable definitions - also of evaluation and assessment - and agreed that they depend much on choice and semantics. The Group found that: evaluation is a process by which the quality, implementation, target relevance and impacts of RTD programmes are investigated, interpreted and examined, while: assessment is a synthesis of facts, which arise from the evaluation process, and judgements.

The ETAN Group also agreed that strict adherence to definitions would be meaningless, and that no single methodological approach would work on all scales. We shall return to this later in order to emphasize that this does not mean that any subprogramme or activity is on its own. Rather it means that subprogrammes and activities may have to develop special tools (models), but must apply them within the constraints of a hierachical system and with respect for continuity in time and space.

Monitoring of FP4

The final report of the 1998 FPM Panel has not yet been published. It will, we hope, be discussed in other fora and is not subject for discussion at this conference. We do believe, however, that some of the observations and conclusions are of general interest in the context of this conference as they point to issues that must be clarified if the EU RTD programme evaluation process is to continue its promising development.

The summary is formulated in two sections of which the first concerns the European dimension and the second moves into the more mundane but important management domain.

---

27 Options and Limits for Assessing the Socio-Economic Impact of European RTD Programmes. Report to the European Commission DG XII, Evaluation Unit by the Independent Reflection Group (set up in the context of the European Technology Assessment Network, ETAN, January 1999.

The European dimension

The FPM Panel observed that European added value is a strong concept but a difficult one to formulate in generic terms, particularly at the overall Framework Programme level and in a top-down fashion. More realistic and useful is a bottom-up approach. A shift from science push to social pull is obvious during the FP4 and clearly witnessed in FP5's social contract between politics and science. The FPM Panel noticed a shift in European added value from bridging of geographic and scientific frontiers and creation of scientific "critical mass" towards technical and commercial risk reduction.

The successful creation of a European research community is a major achievement on which the Framework Programmes can pride themselves, but it is obvious that insularity would be counterproductive. We have achieved better complementarity in research, more efficient joint use of expensive facilities, promoted European standards, and shown ability to establish rapid response research (e.g. on BSE - mad cow disease). However, synergy and coordination between programmes, between programmes and policy DGs are found wanting, and synergy and coordination with national programmes is still weak. So is public awareness. Wider knowledge about programme objectives and wide dissemination, application and exploitation of programme results - not the least to policy-makers and the public in a broad sense - should be given higher priority.

Negotiations with potential new members of the EU have been pursued quite successfully. The prospects are interesting and the challenges considerable. Finally, the FPM Panel worried about the amnesia that has a tendency to strike whenever a new Framework Programme is launched. We should not forget to follow up on previous Programmes. The customers have very long memories.

The management domain

It has been stated before, and the FPM Panel does it again: a greater degree of budgetary flexibility and delegation of responsibility is needed in the Framework Programmes, especially as they - and the EU - grow bigger and more complex. The Management Information Systems[2] (MIS) are not adequate - to put it mildly. The MIS problems make efforts at good management and effective monitoring very difficult. The variety of good management practice that has developed should be systematically spread across the Framework Programmes and transfered from one to the next - meaning that presently it is not. A more systematic, uniform and continuous approach to high quality management may be on its way, including one unified Management Information System, better data collection, rationalised core indicators, annual management targets and management reports.

Other points, however important, such as provision of ex-post information for impact studies, staffing issues, and contract administration shall only be mentioned in passing. So shall the issues - dicussed, but not reported on - of open call for proposals in order to even

---

28 Management Information System (MIS): provides ready access to basic data and its analysis which allows the user to assess the status of a project, specific programme or Framework Programme on a real time basis. Analysis of historic data should also be available to yield trends.

out the workload on the Commission staff and of full monitoring, i.e., extension of the annual monitoring to include central support services.

General comments on the 1998 Monitoring

Generally speaking, the EU programme evaluation and monitoring system is in place and developing quite satisfactorily. It is facing serious challenges arising from the emphasis on socio-economic impact evaluation, the need for an empirical base for evaluation of European added value, the demand for demonstration of synergies, and the dynamics of an expanding EU with growing RTD programmes.

The answers to these challenges are clearer policies backed by well-founded assessments supported by to-the-point evaluations, well-documented monitoring, and managements that are well informed about implementation and impact through a Management Information System with running, real-time data collection, analyses and interpretations.

The "climatological" approach

We all have our weaknesses. Mine is that I am a fluid mechanic that sees everything through meteorological glasses. Two colleagues of mine published a book the other day: "The Harmonious Enthusiasm. The Nature of Physics - The Physics of Nature"[3]. They claim that it is more important to ask the right question than to concentrate on the answer. The right answers will come along anyway. As a meteorologist I believe that in order to ask the right questions you must have observations - often many over a long time period and from a great number of places. Quite recently the Swedish scientist Nils Gissler (1715-71) was given credit for the discovery of the "inverted barometer effect"[4]. It happened very late in his career, one must say; he had been dead for 228 years. He meticulously measured the sea level and found that when the barometer was high, then the sea level was low. By asking why and asking why it is not always the case, he discovered the atmospheric pressure forcing of the ocean's currents and the effect of wind-driven surface currents.

The climatological approach is based on observations, i.e., systematic data collection, the more data the better, provided that the data are of the kind asked for, of agreed standard, taken methodically over a long period of time and under carefully monitored and documented circumstances (change of instrumentation, calibration, surroundings, etc. - even change of observer may make a marked difference). The data are analysed to reveal climatic characteristics, variability, and trends.

In the next step, global models are constructed, calibrated, validated, and run to provide understanding of interconnections in time and space and to reveal information about sensitivities to outside forcing (CO_2, insolation, volcanic eruptions, change of surface albedo, etc.). Regional and local interpretation of the results of global modelling is

---

29 Ove Nathan & Henrik Smith, Den harmoniske begejstring. Fysikkens natur - Naturens fysik, 1999, Centrums Forlag, København, 431 p.

30 Bull. Amer. Meteo. Soc., 1999, 80, 675-682

provided through a telescoping set of models, if the resolution of the global model is too coarse to permit sensible interpretation at a smaller scale.

At the third level predictions are made and probabilities are attached to their credibility. Mitigating actions are evaluated in terms of effects and economy and policy scenarios are formulated.

Socio-economics is not physics

The atmosphere-hydrosphere-cryosphere-lithosphere-biosphere system may - in a climatological sense - be much less complicated than the politico-socio-economic system we have built on top of it. Self-fulfilling prophecies are rare in the natural system, and answers cannot so easily be dictated. Climatologists are also assisted by the fact that although cultural differences exist between climatologists, the assumptions behind their models are basically the same. This, we believe, is not the case among politico-socio-economists. Nevertheless, being a layman in the field this speaker holds the opinion that the differences are often exaggerated.

Accepting the analogy between assessment of EU RTD Framework Programmes and the assessment of the probability and consequences of climate change, we arrive at a three stage procedure:

1. A continuously evolving data base should be established that operates in real time and conforms to agreed and accepted standards. A Mangement Information System should be put in place that provides programme management with real time information and permits effective control and monitoring.
2. Models for evaluation of programme effectiveness, impact, quality and relevance on various scales should be developed, calibrated, and validated.
3. Guidelines should be drawn up for the synthesis on Framework Programme level of the evaluations such that assessments can be formulated and judgements passed on the effects of past programmes and expectation of future programmes can be substantiated.

Concluding remarks

Development of such a monitoring, evaluation and assessment system will require dedication, time, discipline, qualified staff, and money, all of which - to varying degrees - are in short supply. Some may consider the suggestion fictional and quote the saying that tactics win battles - strategies lose wars. It sounds clever. However, we have seen the successful beginning of such a system. There may be a limit to predictability. There certainly seems to be in climatology, but pushing that limit apparently produces intelligent resource management, so why not try to the limit in RTD programme assessment.

Impact Assessment: Limits and Possibilities

Angelo Airaghi
SVP Finnmeccania
Piazza Monte Grappa 4, 00195 Rome, Italy

THE EU EVALUATION PROCESS (1)

Evaluation has been a legislative requirement for European research since the launch of FP1 in the early 80s.

As a consequence the EU has:
- a large cumulated knowledge
- run more than 70 exercises
- involved more than 500 experts.

THE EU EVALUATION PROCESS (2)

The "evaluation" covers the entire process:
- project selection
- programme management
- ex post evaluations.

More recently, the concept of "assessment" has been introduced.

THE EU EVALUATION PROCESS (3)

- *Monitoring* is a process by which information on the progress and direction of on-going RTD actions is generated mainly for management purposes.
- *Evaluation* is a process by which the quality, implementation, target relevance and impacts of RTD programmes are investigated, interpreted and examined.
- *Assessment* is a synthesis of facts, which arise from the evaluation process, and judgements. Under this definition, assessment is a policy-making tool for the planning of new RTD programmes, not a political instrument.

EVALUATION IN FP5

The current scheme will be retained in FP5, but with increased emphasis on socio-economic impacts.

This is reflected in the new set of objectives of the FP:
- scientific and technological excellence
- socio-economic impact
- European added value.

IMPACT ASSESSMENT (1)

The focus on socio-economic impact requires innovative methodological efforts.

In order to understand:
- what impacts
- what limitations
- what implications
- and what steps

should be taken for the next five years assessment, the Commission has appointed an Independent Reflection Group (IRG).

IMPACT ASSESSMENT (2)

The IRG has delivered a report addressing some general and specific issues, and providing some suggestions.

The economic, social, environmental assessments of RTD projects and programmes have limits both in general and in the specific FP evaluation process.

ECONOMIC IMPACTS - Generic Limits

At project level	→	Time gap
	→	Mix of competences (background)
	→	Additional investments
At programme level	→	Zero sum game
	→	Indirect impacts on non-participants

ECONOMIC IMPACTS - Specific Limits

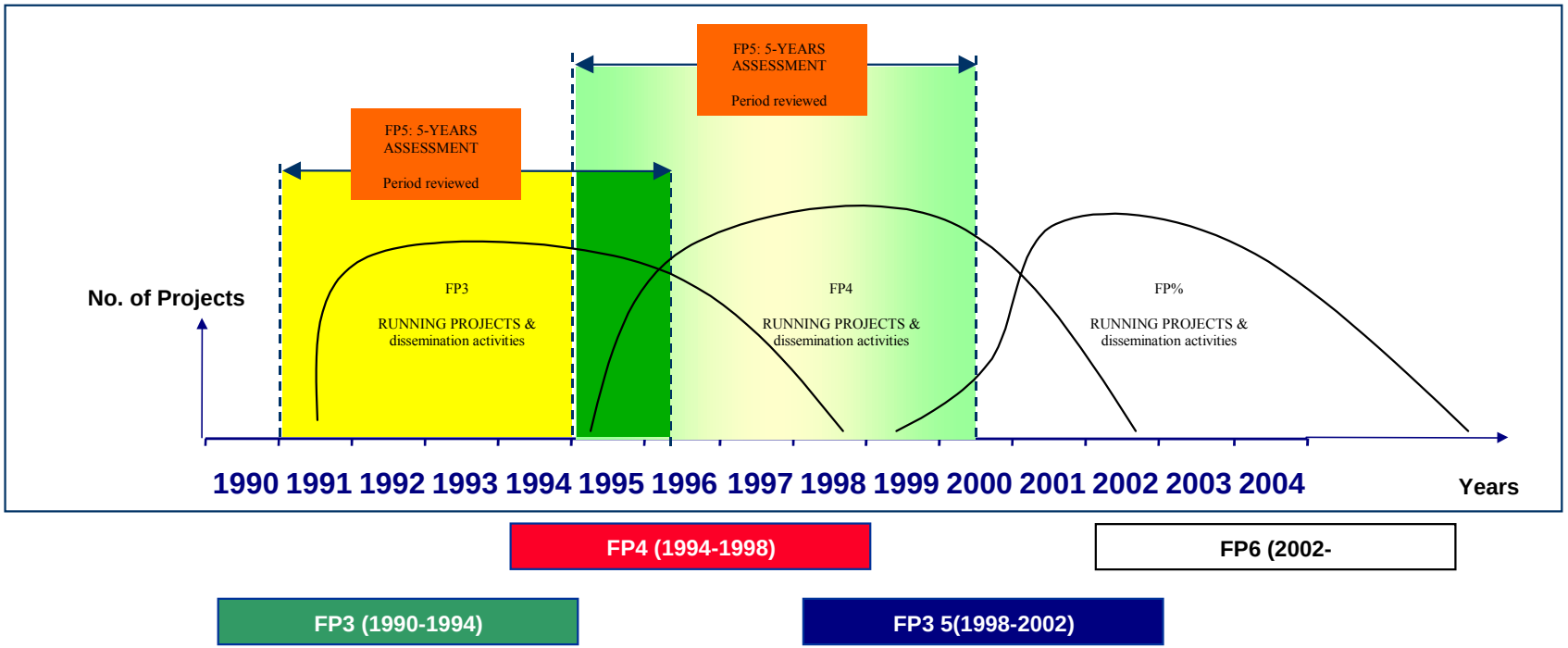
At project level	→	Additionality
	→	Partnership
At programme level	→	Additionality
	→	Marginality

SUGGESTIONS (1)

To overcome these limits IRG suggests to:
- introduce the objectives into the project selection phase
- ask for quantifications at the project level both at the beginning and the end of them
- extend the monitoring well beyond the end of the projects
- support the process with ad hoc studies.

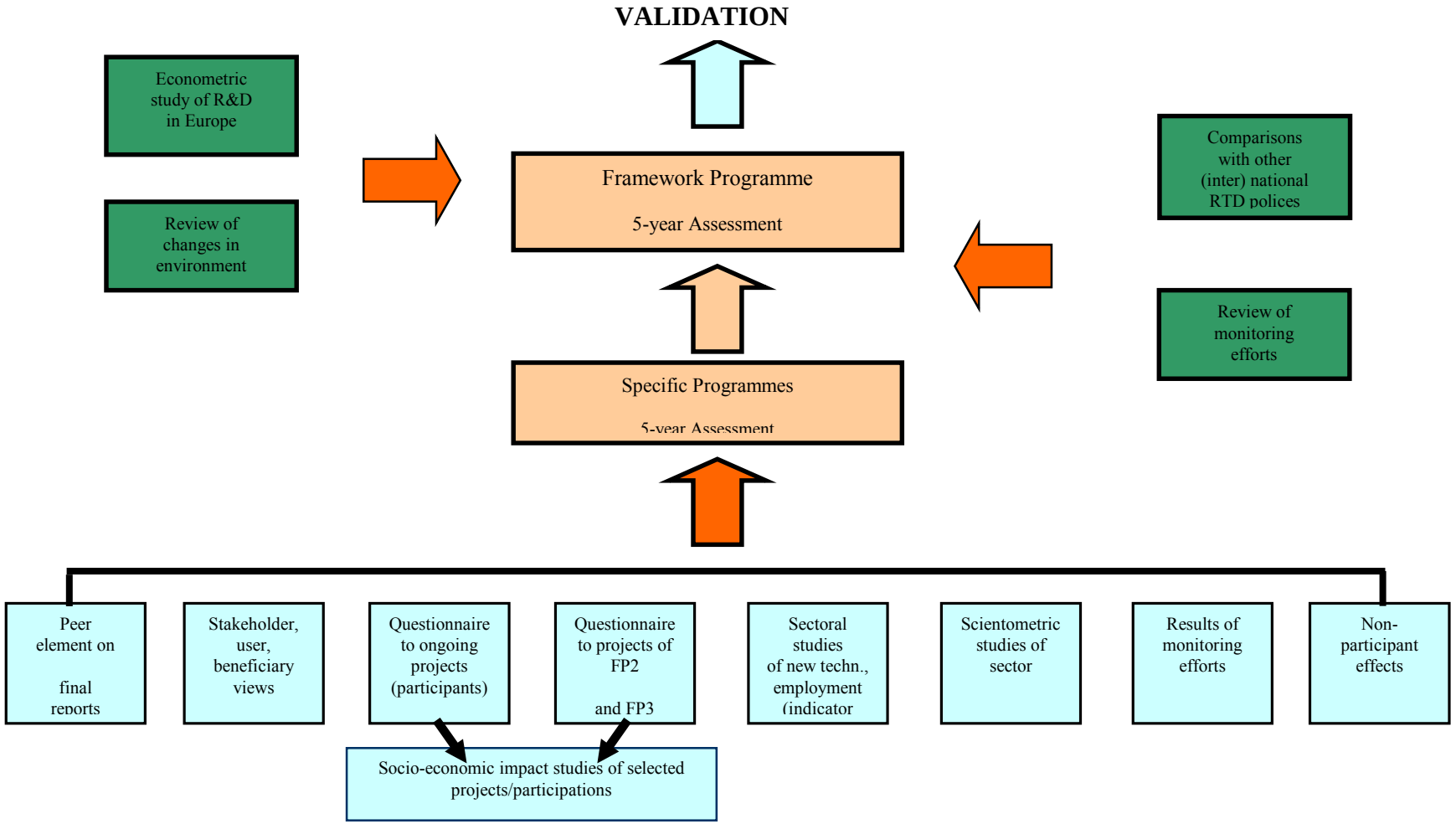
SUGGESTIONS (2)

One of the great limits comes from the fact that the next assessment will try to combine new expectations with old procedures



SUGGESTIONS (3)

A possible approach to the present situation



After-dinner Speech

The Berlin Science Landscape

Ingolf Hertel
Office for Education of Science, Research and Culture
Brunnenstr. 188-190, 10119 Berlin, Germany

Ladies and Gentlemen,

let me start with the attempt to translate the name of this restaurant „Zur letzten Instanz“: it is certainly not the „last instance“; perhaps „to the last resort“ or „to the uppermost court“ - in the sense of the court which in a specific lawsuit makes the final ruling. I have been told that people went here after their divorce to celebrate. But I am sure that this is not an omen for the present meeting, rather I expect this to be the beginning of a wonderful new decade of developing an evaluation culture.

After dinner speeches are not a particular strong tradition in the German culture of public or semi-public speaking. The rhetoric education at our institutions of learned studies is - if anything - rudimentary and we somehow lack the light-hearted and yet very precise Anglo-Saxon humour which is necessary for such a kind of entertainment.

Thus, when I was asked to present an after-dinner speech about the **„Berlin Science Landscape“** on this occasion I was somewhat reluctant - also since the subject is not necessarily one to amuse you after such an opulent dinner - science in Berlin is a quite serious, even though hopefully inspiring subject and - if you talk to any of its representatives, and I am one of them - anything else but too well fed.

Thus, before giving you an exposé about science in a nutshell, let me say a few words about Berlin as such, the new Berlin - now nearly 10 years after the collapse of the Berlin wall, which changed everything in both parts of this formerly strange assembly in the middle of the so-called „red sea“.

In the coming months the German Parliament as well as the major part of our Federal Government moves to Berlin. Berlin is in the process of becoming really the German capital. For my generation and especially for those who have lived in this city for many years during the cold war, this is still some kind of miracle – „the occurrence unheard of“ as Wolf Lepenies phrased it. This German capital which is now emerging will be the capital of a new united Germany as an integral part of Europe. The specific role which Berlin could play in this process toward completing Europe is defined by its geography and also by its history, specifically its post-war history. Let me phrase it this way: in whichever direction you turn from Berlin, you will always be faced with Europe. This is a great challenge and responsibility which the city willingly accepts. But is it already prepared for this challenge? Is Berlin already in good shape for its future?

You must remember: For 40 years both parts of Berlin were breast-fed by their respective national governments. During the cold war Berlin was the shop-window of a divided

world. This created economic and political structures as well as a culture of public spending which turn out to be completely unfit for the stormy waves of a global economy which we are now facing. In addition, the federal support, cash as well as many tax benefits, which the city was enjoying and which kept it alive, was cut dramatically after the unification since Berlin was now considered to be a „normal city“ with the additional benefit of becoming the capital. The result can be studied in 1999.

Berlin, this is about 3.6 million inhabitants, still slowly decreasing in numbers. Berlin, this is a vast construction boom which you can watch everywhere in the city which, depending on individual taste, makes it highly exciting or - unlivable. Berlin is also an industrial desert, having once been (about 100 years ago) one of the roots of Germanys industrial strength, a high-tech boom town one would call it today - which we want it to become again. After the German unification Berlin has lost another several 100 industrial jobs, the present unemployment rate being about 17% (to be compared to 12% in Germany overall and about 19% in East Germany). But we also have special peaks: 30% unemployment in the district Kreuzberg and 20% in Prenzlauer Berg. You can imagine the problems involved. With a gross product of the city of about 156 billion DM the state budget amounts to about 40 billion DM - of which less than 50% are covered by local taxes - and the accumulated public debt presently has risen to 61 billion DM.

Berlin, this is new awakening on the one side, but also nostalgia on the other - nostalgia East and nostalgia West with two different cultures of socialisation which are only slowly begining to merge. We still have our clearly recognisable languages and conceptions we say „Zielsetzung“ in the West and „Zielstellung“ in the East, „Test“ and „Testung“, „Plastik“ and „Plaste“ and you can still recognise the origin of a secretary by her technique of putting together several sheets of paper: by stapling or clamping it.

Berlin, this is also minute reconstructions of large districts with listed buildings and the arising of whole new city quarters out of nothing, with impressive architecture. Including a particularly gigantic one - maybe I should better say impressive - for the presently arriving federal government.

However, most importantly: Berlin is the most densely populated science landscape in Germany, perhaps in Europe, with - and please do not make me responsible for the exact numbers - about 80,000 scientists and more than 200 institutes, colleges, learned societies, archives and museums, of which 80 are involved in one way or other with foreign cultures, languages and international affairs. We have 3 universities, 9 so-called universities of applied sciences and similar colleges, 4 arts colleges, 5 or 6 Helmholz Centres, 6 Max Planck Institutions, 16 Leibniz Institutes, 4 Fraunhofer Institutes, several more research institutes funded by the Land Berlin as well several federal laboratories. Not to mention a wealth of libraries and museums which systematically are counted as part of the cultural sector of Berlin. The total science budget of the Land Berlin is about 3.5 billion DM, i.e. about 8% of the city budget, 2.2 billion of which are in support of our universities and colleges. Science in Berlin is of course international, as science always was and is. We feel this is particularly important for Berlin while developing into a modern metropolis in a world without boundaries. We have, in comparison to other

German states (Länder), with 12% a rather large fraction of foreign students and our universities undertake every effort to further enhance this number, since we are aware that educating young people is the best way to create long-lasting, good and peaceful relations among nations. Our scientists and scholars are involved in more that 400 international cooperations and thus are interwoven with the world.

It is this extraordinary density and plurality which we consider the finest raw material from which we have to create the future of Berlin, the new Berlin. This will be an essential asset for this city in its new role as the German capital and metropolis in the middle of Europe. Politics, business and industry expect that Berlin's science offers a rich ground for innovation in future key technologies but also a high degree of competence in the fields of social sciences and economy as well as knowledge about foreign countries and cultures. We have to compete here with New York, London, Paris and Tokyo and more even so with Bejing and Shanghai.

The numbers I have given you should not make you overlook, however, that all these institutions are presently not particularly well fed. Rather, as a consequence of German unification the buzzwords were „household consolidation", „concentration" reduction of „double offers" and so on. All institutions of Berlin science and tertiary education had to undergo a really radical slimming cure and - of particular relevance to the present meeting - nearly all were subjected to a series of severe evaluation processes.

If you talk about a *culture of evaluation* you can study it here in Berlin in detail.

I will not go into the details of the several waves of „Peers Panels", „Structural Commissions" and „Special Topics Groups" and the rigorous establishment of international advisory boards. Let me simply state that the process was very successful and has led to a science landscape inside and outside the universities which has in most parts now a high caliber quality stamp of the highest German evaluation authorities.

It is fair to state at this point, that the institutional restructuring of the science landscape in Berlin is now complete: The universities have a financial guarantee from the state for several years to finance an official number of 85,000 university places (in practice this will lead still to over 10,000 students in the city). Presently the Senate signs contracts with the universities until 2002. And the non-university research institutes are closely linked into a national funding system which also - we hope at least - offers some long-term stability.

It is our policy to generate a focused strategic agreement: stability and reform, quality and concentration. We have now started a strategic dialogue (the so called Strategieforum Science, Research and Innovation) about our long-term goals which starts with an analysis of strengths and weaknesses and tries to establish a concerted action of all key actors in the Berlin science landscape. Some points are already clear: we will continue to concentrate our efforts on

- the stabilisation and internal reform of our three universities; in addition we will focus on
- the research and technology centre Adlershof, where we want to create a new culture of co-operation between university, non-university research and industry
- and on the biomedical research campus Buch.

Prussia has, a long time ago, once already shown how from the proverbial Streusandbüchse (sand box) plus creative human mind (even if partially imported) a modern flourishing society can emerge. We have no other choice as following this direction and at the same time not repeat the mistakes of the past. And with respect to the limited financial possibilities: it is not a priori obvious that scientific productivity and economic efficiency is directly connected in a quantitative manner.

Cum grano salis we may even discover the advantages of economic modesty: for example, the number of Sonderforschungsbereiche of the DFG in Berlin is steadily growing, even though the number of professors is decreasing. If you take the ratio as measure of quality, as is often done, this would lead to fantastic perspectives if we allow the financial side to continue its present course.

This example may illustrate that benchmarking, quality assessment, evaluation, budgeting, cost-efficiency calculation and whatever else today's miraculous instruments are called have their limitations - as important as the discussion may be. With German accuracy and rigidity we may in a few years have established a complete new profession of evaluators: in public services you may expect the hierarchy of - this cannot be translated into English - „Evaluationsrat, Oberevaluationsrat and Evaluationsdirektor". You see, it will be important to keep a balanced view on these issues in order to make sure that the limited resources are used most efficiently - and this gives you a high responsibility and sufficient work to do!

Thank you for your attention.

East-West Co-operation in the Field of Evaluation: Major Recent Experiences

Chairman's Opening Remarks

Norbert Kroo
Research Institute of Solid-State Physics and Optics, Hungarian Academy of Sciences
PO Box 49, 1525 Budapest, Hungary

St. Stephan, the first Hungarian king in his advice to his son at the beginning of our millenium stated that only a multilingual, multinational country may be strong. Europe is building such a "country" which in my view is going to fulfil this expectation.

European science and technology are also more and more based on this principle, by the way in line with the requirements of a globalized world economy.

The health of the R&D system can be evaluated only on a global basis, any closed system leads to mediocrity, independent of the motivation of the closeness.

The existence of objective evaluation can only be of real use if the available funds to be distributed are above a critical level.

European research funding should be based on efficient mechanisms of evaluation in order to favour competition and enable the selection of the best research proposals.

Peer review by external impartial experts is the basis for the decision-making process for the allocation of resources mainly at project level.

Grant-based financing is efficient only if the basic institutional finances are secured.

Solely grant-based financing of research leads to a fractured system in which strategic goals cannot be realised.

Therefore some resources are needed to realise strategic R&D goals. This could also be achieved through the support of centres of excellence.

Science policy should set priorities rooted in the needs of the economy of the country (region, European Union) but traditions, the performance of different branches of the R&D system, the financial strength of this economy should also be considered.

Scientific results should also be evaluated in the course of the research and after its completion. In order to assess and monitor the achievements of projects or programmes, the scientific results should be compared with the original research objectives which have been clearly definedat the start clearly defined. This monitoring process should be systematically conducted.

Both the reviewing and selection on one hand and the evaluation on the other have to be transparent and carried out by highly competent experts.

In principle there is no significant difference between best practices in the East and West.

In practice, however, there are smaller or larger differences which are rooted in the isolation in the past.

- Less attention has been paid to world level comparisons.
- Evaluation practices have been more strongly restricted to national level, hurting strongly especially the smaller countries.
- Closeness acted against reasonable priorities and initiated autarchic research structures.
- The lack of healthy competition led to mediocre research structures.
- The artificial separation of research and education resulted in a lower professional level of education and therefore in the pauperization of human potential.

These shortcomings affected the different East - Central and East European countries with different strengths and on different level.

At present most of these countries have similar evaluating systems as their western counterparts but carried out with lower level of consequence and on the basis of a much lower level of finances.

The R&D priorities are sometimes lacking, sometimes not consequently applied.

Peer review is mainly national with the risk of limited objectivity especially in small countries. The main argument against international evaluation is said to be of a financial character.

The participation in all-European programmes eased the situation and helped to lessen the gap between East and West.

The strong competition of the Framework Programme and the participation of eastern countries in it should further contribute to this tendency. In a transition period, however, some preferences should be found to ease the pains of this process.

Czech Republic & the European Science Foundation

Peter Dukes
Medical Research Council
20 Park Cresent, London W1N 4AL, Great Britain

Introduction

In April 1996 the Internal Grant Agency of the Ministry of Health of the Czech Republic (the IGA) hosted a Workshop in Prague for about 70 Czech scientists on writing research proposals and on peer review. Many were members of the IGA's 12 research boards. The Workshop was led by Professor Jiri Dvorak, Professor Cyril Höschl and Dr. Ivan Pfeiffer of the IGA, together with eight facilitators - one from the European Science Foundation and the remainder from four member organisations of the ESF's European Medical Research Councils (EMRC)[31].

Background

The IGA was established in March 1991. One of the challenges it took on was to replace the command and patronage model of distributing research funding to one based on competitive peer review. As one of the elements of a strategy to strengthen research quality and relevance, the IGA arranged through the EMRC benchmarking of proposals by external peer reviewers nominated through member organisations of the EMRC. Another element was to develop the capacity of researchers to write competitive proposals and to be effective assessors.

The IGA Aims For The Workshop

The aims of the IGA/EMRC Workshop were as follows:

- to improve the quality of research proposals submitted to the IGA;
- to harmonise the peer review process (ex ante assessment) with international good practice;
- to stimulate young scientists to undertake high quality research;
- to promote productive interaction internationally.

Outcomes & Outputs

We expressed the outcomes of the Workshop as follows. At the end of this workshop, you should be able to:

- explain the principles of peer review in assessing grant proposals;
- describe some of the *different approaches* of funders to assessment - although they follow the same principles;
- distinguish between *scientific quality* criteria and *strategic merit*;
- distinguish between *objective* and *subjective* judgements of the science;

---

31 Dr. Ingrid Wünning (ESF), Dr Beate Konze-Thomas (Deutsche Forschungsgemeinschaft), Professor Erik Dabelsteen and Professor Jorgen Rygaard (Danish Medical Research Council), Professor Bob van Es, and Dr Edvard Beem (Netherlands Organisation for Scientific Research), and Dr. Heather Easton and Dr Peter Dukes (Medical Research Council, UK).

- provide *feedback* after assessment that is helpful, respectful and clear, and which closes the discussion between the funder and the applicant.

A concrete output was the ESF booklet *An ABC of Medical Research Funding*, a short guide to the appraisal and management of research grant applications with tips on how to write and present them.

The Programme

We organised the Workshop Programme in plenary and "break-out" sessions along the following lines:

- A plenary discussion of peer review, to introduce concepts of *scientific quality*, and the importance for funders and proposers that there are explicit *assessment criteria*, relating to the *quality of the question* and the *quality of the approach*.
- A plenary talk on *research methodology* and the principle headings under which a proposal is usually written. This included identifying the research need (clinical, health service or public health problem), framing the specific research question (perhaps as specific hypotheses); setting out the research design; describing key methods (including statistical methods); demonstrating relevant knowledge of the field, your own relevant expertise and the capacity of your research team, host environment and collaborators to deliver the specified project outputs.
- Breakout workshops, which discussed example (anonymised) research proposals, including critiquing several proposals that failed to meet assessment criteria.
- A second plenary on peer review, focusing on *assessment criteria* and the difference between *scientific quality* and additional *strategic criteria* used by some research councils, and on different approaches to the *peer review process*.
- Breakout workshops in which several high quality research proposals were ranked and feedback messages identified, as if by a grants committee.
- A plenary to synthesise key messages.

Observations and Conclusions

From a facilitator's perspective, there are a number of observations and conclusions:

- During the Workshop, tensions among the participants became evident. Not all were committed to change from command and patronage to peer review; feelings differed about the use of benchmarking and external peer reviewers; and for many the challenges of health service funding and organisation were more demanding on their attention than was research.
- The interactive approach promoted by the facilitators was difficult for some senior figures used to a more didactic approach. On the other hand, it appeared to resonate well with younger researchers and senior scientists with a wide international experience. Crucial to success was commitment from the very top, in this case the minister.
- While at the time the Workshop appeared to be successful, it was not formally evaluated. Projects of this kind should have a formal evaluation mechanism incorporated from the start.
- Nevertheless, subjectively assessed, the Workshop gained considerably from the teamwork and the varied perspectives of different research councils, large and small.

- We were comfortable with diversity in the details of the processes to achieve fair, high quality peer review. We concluded there is more than one route to achieving good practice in peer review, we shared a commitment to common principles of **quality**, **equity**, **respect** and **transparency**.

The Impact of Evaluation on Formation of R&D Policy in the Baltic States

Helle Martinson
Estonian Science Foundation
6 Kohtu St., 10130 Tallin, Estonia

The paper summarizes efforts made in Estonia, Latvia and Lithuania
a) to reform research funding by introducing the evaluation of individual performance of scientists and
b) to design national R&D strategy and to re-structure the national research system considering the recommendations of various internal and external assessments.

The process of moving away from centralised allocation of state budget money to institutions and introducing competitive grants based on research quality assessment, as well as establishing of non-governmental decision-making bodies in the Baltic States is described. Due to different political approaches the switch to peer review procedures was realised at different speeds and to different degrees in every Baltic country. The introduction of pre-funding evaluation of the performance of research groups and individual scientists in the course of switching on to competitive grants in Estonia and Latvia is analysed. In Lithuania the performance-based system for resource allocation does not function yet. State budget financing is still distributed to research institutions. The share of individual grants is only 0.2% of allocations of the total state budget money for science.

The second part of the presentation is devoted to the assessment of major science structures on national level and retrospective evaluation of research results of all science fields in Estonia, Latvia and Lithuania by panels of experts, appointed by Swedish, Danish and Norwegian Research Councils in 1991, 1992 and 1995.

The research results were evaluated according to the principles and standards used to evaluate Scandinavian science, without any price reduction. A significant part of this undertaking was the assessment of national R&D systems and research environments, resulting in a number of general recommendations on reforming the research establishment. The opinions of outside observers were a valuable tool in political debates for introducing new principles of research funding and building up new structures.

It is shown how the principal recommendations of these assessments have been followed and to which extent they have influenced the course of the reform of R&D systems in the states under consideration. In Estonia and Latvia, the general recommendations of Swedish and Danish evaluation have been followed to a great extent. In Lithuania, the attitude of science administrators towards evaluation in general and towards re-structuring of the research system has been different from the other two countries. The foreign experts pointed out the urgent need to change the principles of research funding in Lithuania and to optimize the network of universities and research institutes.

These operative extensive *ex post* evaluations of the whole science and research establishment in the starting years of transition were unique undertakings which gave an objective picture of the state of affairs in science of the Baltic states. The results of these assessments have served as a basis for making comparisons and taking political decisions of future development of R&D in this region.

The third part of the presentation is devoted to evaluation of R&D systems of the Baltic states in connection with preparation of accession with the EU, initiated by the EC in 1997. This scanning was aimed at the analysis of the state of the whole RTD system, first and foremost of technological development and involvement of business enterprise sector into RTD in these countries. The appraisers analysed the key outcomes of the reform of the research establishment. They approved the principal changes which have taken place in Estonia and Latvia and stressed the need to implement the recommendations of Norwegian experts to foster the reform of the R&D establishment in Lithuania.

The main concern of appraisers was the weakness of technological development and absence of links between research community and business industry sector in all three states.

The observations and recommendations of this assessment were aimed at increasing social impact of RTD, at finding the ways for animation of technological development as a base for economic growth and productivity improvement. The significance of promoting partnerships of scientists with industries, governmental institutions and foreign organisations and enterprises to take part in international co-operation was especially stressed.

As a conclusion we can confirm that evaluation of science and scientists has been a means for making decisions on reforming the system of research funding and re-structuring of R&D enterprise. The character and profundity of institutional reforms implemented in every Baltic state vary from one state to another in dependence of political atmosphere. It is evident that the radicalism and the results of reform are in direct connection with the speed and decisiveness of implementation of the practice of evaluation on both levels – of individual performance of scientists and science or scientific establishment.

The Impact of Evaluation on Formation of R&D Policy in the Baltic States

Discussion of Helle Martinson's paper by Carl-Olof Jacobson
Royal Swedish Academy of Sciences
N. Rudbecksgatan 13, 75236 Uppsala, Sweden

The content of Professor Martinson's paper is most interesting and, as far as I can judge, a very good and most interesting description of the development and the situation when it comes to Research & Development in the Baltic states.

My own experiences in the field emanate from the Swedish evaluation of Estonian research quality in 1991. It may be of common interest to observe the impact of this evaluation and of those made by research organisations in Denmark and Norway concerning Latvia and Lithuania. I will briefly comment on the development from the viewpoint of Nordic science.

The Swedish commission

In the spring of 1991 the Royal Swedish Academy of Sciences was asked by the then new Estonian Science Council to perform an evaluation of all research paid by the government in Estonia, and to proceed as fast as possible. The Academy agreed to act as coordinator of the endeavour. To be able to handle a speedy process we had to invite the whole Swedish scientific community to participate. Five research councils and four academies were engaged, and the whole procedure with site visits, reading and writing was accomplished in half a year.

The background of the call for assistance was complex, and to fulfill the task in a fair way was not easy, as it turned out. It was agreed that the evaluation should not take into account past difficulties. It should be carried out according to the same principles and standards as if it had concerned Swedish science. In other words, the evaluation should indicate whether or not the Estonian projects were close to the international research front and whether the results were relevant to the international science community. This may seem harsh in the light of the years in comparative isolation for Estonian research, but it was felt that (and now I quote the general comments given by the Natural Science Research Council) "our mission was to identify healthy and productive parts of Estonian science and to indicate where isolation and the resulting lack of international feedback had led to substandard research traditions and suboptimal organisation".

The state of Estonian Science in 1991

The former Soviet republics did not have research systems adjusted to local conditions, since the pre-1990 research priorities were relevant for the needs of the entire Soviet Union and not necessarily for the Baltic states. As is well known, Soviet competence in different research areas was often concentrated in certain institutes spread all over the Union. We must remember that Soviet state money for science was sufficient, often generous. The Centres of Excellence as well as the more regional research institutes were

giant bodies, most often with one or two thousand employees. So, cuts had to be made to get institutes of relevant size for a smaller country and a much smaller economy.

Some research areas thus could show excellence, while, on the other hand, vital research areas for the country, like Information Technology, could be missing or weak. To observe and report on such weaknesses was of high relevance. The evaluators had to look at the total system and make recommendations that should bring competence and balance to the research and education system.

A difficult matter, that sometimes made the evaluators uncertain of the fairness of their judgement, was the fact mentioned by Martinson that part of the research results were reported in Russian or Estonian, not English. If most or all of the published material was inaccessible for the evaluators, the site visits and the discussions with the group members were the only tools.

Swedish recommendations

The recommendations of the Swedish evaluators have already been reviewed by Professor Martinson. I will, however, point to some of the reported observations.

- The quality of the research was very uneven, a statement that goes for all areas. Unexpectedly, many areas were covered and amazingly many research groups had reached a high international level. Even in such a resource-demanding field as molecular biology high quality was found.
- Too many of the results was published in national journals without peer review.
- The bureaucratic fragmentation of the research body in institutes, departments and groups prevented flexibility in cooperation.
- The poor economy did not only mean shortage of equipment and laboratory material. A serious consequence was the cancellation of subscriptions of international journals.
- The evaluators of mathematics found that the overall high Russian level of that subject had not influenced the Estonian mathematicians. The recommendations were to maintain, or rather to create good connections with Russian maths. On the other hand, the evaluators were really impressed by the fruitful collaboration between mathematicians and other scientists.
- Ecologists and environmental scientists had suffered a lot from the fact that their results often had been classified as secret and thus were not allowed to be published.
- Martinson mentioned already the vital need for a rejuvenation of the cadre of researchers. Young people willing to give their life to science is a commodity in short supply in many corners of the world, but this is felt especially in eastern Europe, where money is found easier elsewhere than in libraries and laboratories. Among the recommendations aiming to encourage young people to go on to graduate studies I will mention a few:
  - In some areas, for instance in engineering sciences in Lithuania, the output of doctors was low compared with the number of students. This underlines the necessity to work close to the graduate students and to encourage them in all possible ways. Needless to say, this can be a difficult task, since many of them have no financial means and thus probably try to earn their living outside the universities.
  - The outside world should widen its efforts to assist. We know by experience that what we call the "sandwich" grants can help young graduate students and post-docs

through many difficulties during their early years as researchers. Such a grant could run, let us say, for three years, and allow them to spend part of every year in a good research environment in the west with western salaries. The rest of the year they can spend in the home laboratory with some extra money to make it possible to continue the work and keep in contact with the host institution abroad. Such a scheme is at least to some degree preventing brain-drain (although we shall not forget that the brain-drain mostly has to do with the 'escape' from the universities to business in the own country).

- The evaluators made several suggestions about how the science community in the west, and especially in Sweden, could best assist. Among many ideas, the possibility was mentioned to arrange graduate courses or summer schools for young scientists from both east and west in the new eastern countries. This would make it economically possible for the eastern young scientists to participate and it would give opportunities for future cooperation between research groups from east and west.

Generally speaking, money spent on joint projects between east and west would mean a quality raise for both parties, since the high potential of science in Eastern and Central Europe would be better utilized.

The isolation from the West

A major negative consequence of the Soviet system was the wish for isolation from the west. It was, however, possible to observe that the wall had a lot of slits used by scientists. Several of the evaluated research groups had been able to maintain close relationships with science in the west. Concerning social science our evaluators had expected that the Marxist thinking should have been much more visible and that quality according to international standards should be lower, than was actually found. Obviously good contacts with the world outside the wall and opportunities to read international literature existed, at least for a fraction of the Baltic scientists. Another consequence of the overall lack of contact with the west was poor knowledge of the English language, the 'lingua franca' of modern science. But this is a handicap, which I understand is undergoing a rapid recovery, at least in Estonia and Latvia, the countries I know best.

Research institutes contra universities

Without having evaluated the university educational system, it was commonly felt that it had a sound base that could be relevant for a rapid development of the economic and social welfare of the region. However, all evaluators suggested steps to a better functioning university, steps like reducing the teaching load for university teachers and bringing the research and the education systems closer together, in order to create contact between students and quality science. Strong integration of research and education is a basic principle in western countries, and gives a win-win situation: the students will obtain an up-to-date view of thinking before going out in society, and the research units will be able to recruit good students into science. It is remarkable that the speed of development differs so much from country to country in the former east when it comes to merging research and education institutes.

Breaking down the giant research institutes

One main purpose of the international evaluation was, as I mentioned, to get advice concerning how to break down the immensely large, inflexible and expensive research institutes. In the new situation with its lack of resources they had to be reorganised. This made the task to speak the truth about quality especially difficult and serious. Of course all evaluations of research have a less comforting side: a bad review may mean that a project will be closed and maybe that people are fired. In this case it probably meant that a large number of scientists, technicians and administrators would have to leave the relatively comfortable life behind them and walk out into a non-existing or at least problem-filled labour market. The delicate situation was recognized by all evaluators. I do not think that this led to less honesty of the reviews, but together with the feeling that we measured the result of one system against the standards of another, it created an apparent human dimension of the assignment. The more remarkable was the almost complete lack of hostility among the scientists to be evaluated.

To change an old system - difficulties and driving forces

Today, as stated by Martinson, *all* state budget money for research is allotted to certain projects after evaluation in Estonia and Latvia. In Lithuania only a minor share of the state money goes to projects and the main bulk is still distributed to the institutes. Few western nations have gone as far as Estonia and Latvia. However, as Martinson pointed out, the switch to the financing of projects facilitated the breakdown of the large institutes. This elimination would not have been possible to perform without strong decision-making bodies situated outside the existing research structure. The early creation of the Estonian Science Foundation Council (the body that makes all decisions regarding research funding) certainly was an instrument of the kind that was asked for by some evaluators. I quote from the evaluation of Political Science:

"There seems to be a lack of central leadership as to fundamental organizational changes in the Estonian higher education/research system. It is unlikely, from what we generally know about the tendency of academic institutions to guard their own turf, that the Academy, the University and the other higher education institutions are able themselves to achieve a fundamental restructuring of the system. An active involvement is needed from outside. (...) The difficulties for the government to step in does not only seem to be affected by its internal organization or the structure of the political life in general, but also, and this was an impression obtained during a short stay, by an ideology saying: A government should not try to intervene and direct matters of this kind. This is an ideology formed in protest against an earlier extremely centralized system. (...) Still (...) at the same time as universities and research institutions are to be guaranteed a high degree of autonomy in their daily activities there also often exists a need for governmental interference as to the basic organization of the higher education and research system of a country."

Based on Swedish and other experiences I am not sure that *governmental* interference is the best of roads, even if the economical framework of course has to be decided by the politicians responsible for public funds. A continued international influx of ideas taken care of by national bodies mainly consisting of university and university-related people

have proven able to suggest and realise new thinking that has revolutionised university systems.

Let me make one final remark on the financial system: I am not quite sure that a system where *all* governmental money is given to projects after quality evaluation is superior. Some money always has to be available for small-scale activities going on over a broad spectrum of science and for science outside the modern trends. We never know where a break through might appear and competence in "classical" areas outside the trends might be of value, not least to secure a broad university education.

Let me also comment on Dr. Martinson's mentioning of the absence of links between the research community and business industry sector in all three Baltic states. This was also pointed out in the 1991 evaluations, but it was remarked from the agricultural evaluators that in this vital area of the Estonian society they found a too weak basic research, research which they considered highly necessary for future development of the country's principle industry. They suggested new links between the agricultural research and basic biology research of the natural science institutes.

Danish, Swedish and Norwegian recommendations

Generally views and recommendations from the three Nordic countries to their Baltic neighbours contained the same message: Reform the research system and unite the institutes and the universities! Move as much as possible of the decision-making downwards in the system! Lower the teaching load of the university teachers! Widen the international contacts! And do not forget to keep the good contacts with leading scientists in Russia, the Ukraine and other former members of the Soviet Union! Several of these scientists are among the world leaders in their fields. Rejuvenate the cadre of researchers! Let international evaluations be a regularly occurring process! All the Nordic evaluators ended up with the same bottom line: the potential for rapid recovery and raising the quality to international standards is evident, if sources and the will to change are added.

Final words

The final words in the Swedish evaluation of Estonian natural science were: "Although the very present conditions in Estonia seem chaotic, and the economy of the country is temporarily destabilized, we should look to the future. It is our firm belief that, if proper measures are taken - some of which we have suggested here - Estonian science has both the intellectual and technical potential to be reformed and modernized within not too many years. But it is also clear that in order to speed up this process it must be helped by rather large scale foreign aid." I think Professor Martinson's talk well demonstrated that this statement was correct. Estonia has already come far along on the road that has been staked out. Latvian science is well on its way. Lithuania was for different reasons a slower starter - but it is only four years since the Norwegian evaluation took place. It is only to hope that they are aiming in the same direction as their northern neighbours.

The Impact of EU Programmes on the Familiarity with and Use of New Technologies in Firms in CEE Countries
INCO-COPERNICUS PROGRAMME
The Case of Greece

Yannis Katsoulacus (presented by Asimina Christoforou)
Center for Economic Research and Environmental Strategy
25, Metsovou Str., 10682 Athen, Greece

I. Introduction

The main characteristic of R&D activities in Greece is the low level of R&D expenditure and the pervasive public presence. Greece is in many ways similar to countries with a much lower per capita income, which are now forming their national innovation systems.

In the following we shall present the results of a study on the impact of EU R&D Programmes on the innovativeness of Greek firms. This study was conducted by Prof. Yannis Katsoulacos and Ms. Kelly Benetatou within the scope of the *INCO-COPERNICUS Programme*, which was completed at the end of the previous year. This work should be considered as a starting point and show some directions for further analysis and empirical research.

II. Outcomes

A. *R&D Activity and Innovation*

1. 67% of the respondent firms have R&D activities. These activities are concentrated in sectors with *low technological* level, i.e. Food and Drinks, and in industries with *export potential*.
2. The surveyed firms identified as reasons for the change in their employment levels (between 1992 - 1995) the *development of new products*, the *extension to new markets* as well as the *increase to their turnover*. All employees that had left had *average qualification levels*.
3. Just over half of the respondent firms have R&D activities that relate to *product and service innovations* (rather than process innovations), which may indicate that Greek enterprises follow a strategy of *adopting or copying new technologies*.
4. Most of the firms *had not changed their products* during the period 1993 - 1995, which can be considered as an *indicator of a low degree of product and process innovation*.

B. *Knowledge Exchange and Co-operations*

1. The majority of R&D co-operation is with *universities* and other *firms* and *research institutes* which were sited either in *Greece* or in an *EU country*.
2. From the Greek firms that had reported contacts and knowledge exchange with foreign firms (i.e. 50% of the surveyed firms), only 12% reported contacts with CEECs,

namely Bulgaria, Romania, Albania and Russia. Contacts are mainly taking place with EU countries and, in particular, the UK, France and Germany. The reason for this pattern may be that the Greek industry is considered *relatively underdeveloped* in the field of applied technological research, compared to the western EU countries. Thus, Greek firms are interested in forming collaborations with high-technology firms, in order to re-organise and modernise their production systems and increase their competitiveness. On the other hand, the domestic industry's *trading contacts* with CEECs could be considered a first step towards further cooperation.

3. There is a negative effect, though not very large or significant, of firm's size, sales and exports on knowledge exchange. This means that in Greece *small firms are more likely to seek contacts with foreign firms in order to attain knowledge*. This could be justified only with having in mind the Greek reality: the crisis Greek industry faces since accession to the EEC seems to have more severely influenced the large firms, which have marginally managed to find their way out of it. So it is the SMEs and newly founded firms that appear to seek collaborations in order to acquire R&D and knowledge.

C. *Familiarity with the EU Programmes*

1. The instruments for the stimulation of R&D activities mainly consist of *subsidies and financial support from governmental or EU programmes*.
2. From the surveyed Greek firms, 88% are familiar with the existence of EU subsidies and 55% consider them important for the proceeding of the activities. However, we can say that not many firms use these programmes. Only 49% of them have applied for a subsidy. The main obstacle for requesting subsidies is the *complexity of application and procedures*.

Evaluation of the EU Programmes and Cooperation with EU Partner's Impact on Innovation within Romanian Firms

Steliana Sandu
Institute of National Economy
28-30 Bud. Magheru Sector 1, Bucharest, Romania

Innovation activity has an important role to play in developing the technical level and competitiveness of Romanian industry. The structure of industry set up in the past without taking into account international specialisation and selective development criteria, according to the Romanian comparative advantages. Restructuring of Romanian industry and improving its performance require a close co-operation with EU countries based on principles of additionality and complementarity. Without opportunities for contacts western Europe, Romania's research institutes experienced many difficulties before 1990.

In designing the new innovation policy, Romanian policy–makers need to be guided by the best experiences of EU countries and by EU programmes. The stronger orientation towards innovation and competitiveness of EU programmes offer the general guidelines and support for increasing the effectiveness of national innovation policy. As companies pay greater attention to the processes of renewing products and improving technologies, science, technology and innovation become areas in which policy-makers have an urgent need for better information. Consequently, improving the knowledge about innovation in the EU countries can help policy-making to be able to map and understand the real dimension and implication of EU programmes and contacts between Romanian firms and EU partners on innovation activity within the industrial firms.

- The aim of the joint research project was to provide the relevant data on characteristics and objectives of innovation activity at the level of company, to emphasise the main obstacles facing to innovation in order to help monitor firm's innovation in Romania according to European methodological standards.
- Using the data gathered by the questionnaires and interviews in high quality analysis helped us to design the future strategy of innovation and diffusion of new technologies in Romania in comparison to EU countries, as the Netherlands, Greece or countries to be accepted in EU such as Hungary.

The contacts that firms and research institutes have with partners of EU countries represent an important factor in intensifying innovative activity in Romania.

This factor had a positive influence over the 1990-1997 period, as 52.9% of the interviewed firms and 62.3% of the R&D institutes declared they had successful contacts more frequently with partners from Germany, Great Britain, Italy, France, and the Netherlands. The number of contacts per firm or per institute expressed the intensity of those contacts.

In comparison with other central or east European countries, prior to 1990, Romania's economic and scientific isolation led to specific difficulties in starting the contacts with various partners from western Europe. To initiate the EU contacts it was needed to

overcome certain mentality and psychological barriers, moreover that the Romanian firms or R&D institutes had a weak or no presence in the scientific landscape or goods and services markets in western Europe. In that way it can be explained that in 40.3% of cases of firms and in 35.6% of R&D institutes, the original effort to contact a partner has been made by the Romanian firms or institutes. Only in 8.6% and 15.8% of cases, respectively, a mutual intent to get in contact was registered.

We observe that 45% of the firms and 36% of the R&D institutes interviewed gave no answer to those questions; it is not surprising having in view that up to 1990, a positive answer to such questions would have been appreciated as a political offence.

With the view to the knowledge transfer, the Romanian partners are in the following three positions:

- Only to adopt the know-how proceeding from EU partners, having nothing to offer instead (47.1% of firms and 15.8% of institutes);
- To offer know-how (none of the firms and 2 R&D institutes);
- To perform a mutual knowledge exchange between the partners. Making no mention of that offered or received more, we consider the Romanian partner has got an advantage. This is the case of 5-7% of firms and 44.5% of the R&D institutes.
- The significant frequency of those who categorically answered that no knowledge transfer took place with the EU partners (47.1% of firms and 37.6% of institutes) was also a symptomatic sign of the small experience the firms have in co-operation with western partners.

The different kinds of EU contacts were financed from own sources mainly in the case of firms and from EU sources (programmes and funds) in the case of R&D institutes, because the majority of them are public institutes, non-profit organisations.

The motivations most mentioned for having EU contacts were: finding new markets for their final products or for buying raw materials (case of firms) and increasing of income for the research personnel (the case of R&D institutes). This situation is due to the very low level of payment in the R&D sector in Romania.

The EU programmes are better known than the Romanian government stimulating means. For example, 91% of the total number of interviewed firms and 97% of the institutes declared that they had known the PHARE programme, while only 57% of firms and 60% of institutes had been aware of the government stimulation means.

The most important obstacles in requesting EU subsidies are the lack of information and the difficulties in identifying the suitable partner.

Round Table I: Mutual Learning in Evaluation East/ West

Hungarian Experiences with Applied R&D Project and Programme Evaluation

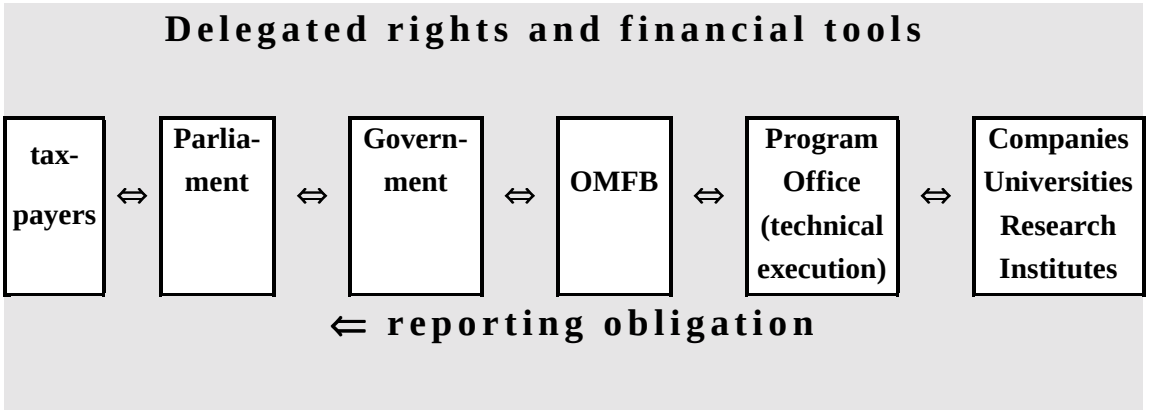
Tomás Balogh
OMFB (National Committee for Technological Development)
Szervita tér 8, Budapest, Hungary

Evaluation of R&D as scientometrics or measurement of the output of individual researchers, R&D teams or institutions has been carried out regularly for some decades in Hungary. Project evaluations at application and finishing of the projects have also long traditions. However, today's culture of systematic ex-ante and ex-post programme evaluations based on performance indicators has been established in 1995 only. OMFB, the National Committee for Technological Development, plays a leading role in this kind of evaluations in Hungay. Our pilot programme evaluation, concerning our largest programme was carried out in the first half of 1996. Mr. Torbjörn Winquist (NUTEK, Sweden) provided a high level professional background during this action.

As the representative of OMFB, I joined the TAFTIE Task Force on Evaluation in December 1995. In my opinion, this was and still is a very useful network where continuous exchange of information among innovation agencies is going on. TAFTIE (The Association For Technology Implementation In Europe) had a lot of experiences already, based on the knowledge of its member organisations: ANVAR (France), CDTI (Spain), ENEA (Italy), FFF (Austria), EI (Ireland), IWT (Flanders), NUTEK (Sweden), OMFB (Hungary), RCN (Norway), SE (Scotland), Senter (The Netherlands), Technopol (Brussels Region), TEKES (Finland), TTGV (Turkey), VDI-VDE-IT (Germany). TAFTIE issued *Guidelines on the Evaluation and Monitoring of Projects and Programmes* in 1997 which was also a background for our institution evaluation strategy.

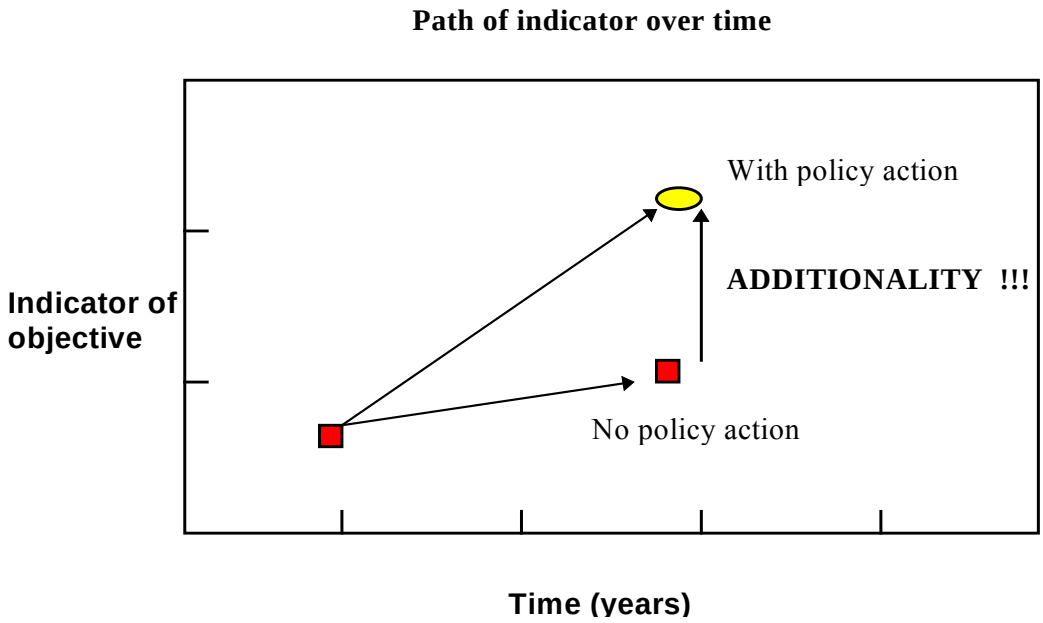
The basic principles of our evaluations are transparency and accountability on the one hand, and feedback and learning on the other. The basic principle of accountability is featured by Fig. 1.

Fig. 1 The basic principle of accountability of public support in the case of OMFB.



All R&D support is additional to the own sources of the partners. As a rule, 50% financing is given, in some programmes as a soft loan, in others as a grant. All support is decided by independent expert bodies where civil servants are only executors of the decisions.

Fig. 2. The meaning of the additionality principle in the evaluations



The OMFB evaluation strategy that was mentioned earlier has an Appendix covering the indicators in four groups: financial rate of return, competitive position, externalities and additionality. After the indicators there are several sample questions from which the interview plans and questionnaires can be compiled, according to the objectives and needs of the specific support scheme. It is emphasised in the document that the indicators and questions can be varied flexibly, according to the nature of the programme under consideration.

Based on the OMFB evaluation strategy issued in 1996, several ex-post programme evaluations have been carried out up to now (the last two ones are still going on):

1. *Applied R&D Programme (672 projects, 1995-96)*
2. *Hungarian participation in the EUREKA Cooperation (52 projects, 1996-97)*
3. *Programme for the Development of R&D Infrastructure (3,138 projects, 1997)*
4. *"BALATON" French-Hungarian bilateral cooperation (74 projects, 1997-98)*
5. *Programme for Developing Competitive Products (202 projects, 1998)*
7. *Hungarian participation in the COST Cooperation (110 projects, 1998-99)*
8. *Hungarian participation in the ESA/PRODEX programme (6 projects, 1999)*
9. *OMFB Policy on Small and Medium-sized Enterprises (policy evaluation, 1999)*

Every evaluation report is published in Hungarian and in English in some hundred copies and distributed by OMFB.

In 1998, there was a first attempt to carry out a portfolio analysis study for the detailed assessment of a programme, taking also the cross-references of data into account. It seems to be an adequate policy tool. There was also some effect of evaluations on the programmes: for instance, in one of our programmes the application deadlines became more frequent and advantageous for the partners, based on the experiences and proposals of the independent evaluation.

Now the first policy evaluation is going on, concerning the behaviour of OMFB towards the SMEs. All policy documents, programmes and measures will be analysed by an independent evaluation panel. We think this kind of large-scale analysis will be useful for the formulation of the national innovation policy as well as for our institutional professional and financial planning.

On 21-22 October 1999 the 3rd TAFTIE Evaluation Seminar, titled "Technology Policy and Evaluation" will be held in OMFB, Budapest. The main goal of the Seminar is to provide a forum for in-depth and interactive discussions on actual and practical aspects in the field of "monitoring and evaluation of technology programmes". The 1999 Seminar will be focussed on the following topics:

- **Evaluation strategies** (workshop session). Why are agency evaluation strategies necessary? How are evaluation strategies linked to technology policies? What are the best practices of formal or informal evaluation strategies for innovation implementation agencies/institutions? How can these strategies be utilised in practice?
- **Impact of evaluations** (workshop session). Papers are expected on the utilisation and feed back of evaluation results (intramural and extramural), including the best practices on communicating evaluation results and/or producing evaluation results that can be better communicated.
- **Socio-economic indicators** (round-table discussion session). Ideas and proposals are welcome to form a common platform for the Evaluation Network on the most important aspects of socio-economic indicators of R&D and innovation programmes representing increasing importance at the member agencies.

The language of the seminar will be English. The seminar is open to all TAFTIE members and to other persons interested in evaluation of technology programmes.

The Slovenian Point of View[32]

Edvard Kobal
The Slovenian Science Foundation
Štefanova 15, 1000 Ljubljana, Slovenia

Introduction

Research evaluation is becoming an increasingly important element in decision-making and the creation of scientific policies. Evaluation can be understood as a methodological analysis or the study of research activities, institutions, policies and their instruments, and research affects. The evaluation process tries to maintain certain scientific standards as well as to preserve the autonomy of science. The basic aim of evaluation is to support and ensure quality in research.

If we accept the hypothesis of the internationally known sociologist Professor Veljko Rus, who claims that evaluation is not a form of political directing of research but merely a method through which research activities can be regulated, one can then state that the methods of evaluation must be subordinated to the object of evaluation. The evaluation method must be compatible with the nature of the research activity. The attributes that the research activity has, on the other hand, are:

- a high level of unpredictability, and therefore a high level of risk;
- a high level of complexity or specialisation;
- highly professionalised activity[33].

What should be the proper forms of regulation of such activities?

Experiences in other countries indicate that the self-regulation evaluation model is more productive than the state-directed model, because it allows not only autonomy, but also the self-initiated regulation of researchers. (Self-regulation of researchers is initiated by the researchers themselves.) Throughout the 1990s, Slovenia has been faced with a discrepancy between the ministers of science and technology and leading researchers. The ministers insist on maintaining the existing state-directed national model of evaluation; the researchers would like to introduce a new model, that of regulated self-regulation. To this date, the ministers of science and technology have not tried to implement or introduce the suggestions given to Slovenia by the group of experts from the EC which have been published in the White Book. The only suggestion which was carried out was the establishment of the Slovenian Science Foundation in May 1994. Unfortunately, the SSF was not given a great deal of authority in science; therefore, most of the activities and programmes are still under the direct control of the Ministry of Science and Technology of the Republic of Slovenia.

---

32 Note: Some of the references in this text are taken from the magazine Raziskovalec (The Researcher) that was completely dedicated to the views of Slovenian and foreign researchers about the evaluation of research. Unfortunately, the magazine is published only in Slovenian language.

33 Veljko Rus, Uvod, Raziskovalec, št. 1, L. 29, May 1999, st. 8-9.

The Processes of Evaluation in Slovenia

According to Dr. Stojan Sorčan[34], processes of evaluation in Slovenia have existed since the beginning of scientific research in our country. The content, forms and strength of the evaluation process have been changing over the years. The analysis of the intensity of changing the evaluation processes in Slovenia would probably show an interesting side of the professional development of science here. Among the examples that Dr Sorčan states are the following three:

- An examination of the post-war Jožef Stefan Institute would probably show a strong correlation between qualitative human resources politics and the introduction of internal evaluation mechanisms that promote qualitative scientific research work.
- In the 1970s we were able to notice qualitative personnel changes in the field of medical science, which was the result of the introduction of international bibliometric standards in Slovenia.
- At the beginning of the 1990s the number of articles written by Slovenian researchers in international publications increased, because appearing in those publications made it easier for a researcher to receive state funding.

Each year the Slovenian Ministry of Science and Technology publishes a number of open calls. Following the closing of open calls, the evaluation of project proposals is carried out. In the first stage of evaluation, the project proposals are evaluated by one Slovenian and one foreign evaluator. Following this first stage, the proposal is also appraised by a group of Slovenian evaluators. The criteria used by the evaluators are in accordance with scientific and economic criteria; those criteria depend on the type of research that the proposal suggests. When evaluating the proposals from the viewpoint of basic research, emphasis is placed on scientific criteria (e.g. relevancy, reference). In applied research, emphasis is placed on information, how the results can be used, and whether co-financing of research is also granted by possible users of the research results. However, when evaluating the project proposals for technological development, the emphasis is placed on the economic value that the research will bring, its technological value and the proposer's references.

The expert system consists of 50 national co-ordinators. Expert bodies are divided into councils, boards and other bodies. The most important bodies for the evaluation process are councils (for instance: Scientific Research Councils, the National Council for Scientific Research, the Council for Science and Technology of the Republic of Slovenia, the Technological Development Council, the Scientific Research Council for the Nature and Culture of Slovenia, and the Programme Council for Targeted Research Programmes). Because of the transition from project financing to mixed financing that is, to project-programme financing of research groups (350) the institutional evaluation was also indirectly performed. This evaluation was conducted with the help of the Slovenian evaluators.

At the beginning of 1999 the Ministry of Science and Technology and the Slovenian Academy of Sciences and Arts signed an agreement on the implementation of the

---

34 Stojan Sorčan, Opomba iz uredništva, Raziskovalec št. 1, l. 29, May 1999.

research evaluation programme for the period 1999 - 2003[35]. Within the programme framework, the methodological basis for changing the development of science should be developed and the study of the research evaluation situation and implementation of the National Research Programme should be conducted.

In order to achieve the compatibility of evaluation systems and approaches through the transfer of knowledge and experience/good practice, the following must be done in Slovenia:

- Experiences that Slovenian experts gained as evaluators working for the EU or other organisations must be transferred into practice in their own country.
- Different funding agencies across Europe should strive to introduce similar rules in their evaluation systems (for instance, compatibility of evaluation systems would help individual researchers be more competitive when writing proposals for the EU or other organisations on the international level).
- Compatibility could be also achieved through the exchange of expertise and experience regarding the planning and implementation of evaluation processes between experts from countries having similar scientific policies and experiences.

In order to improve and further develop evaluation activities in Slovenia, we would welcome the following types of support or co-operation from other countries:

- The inclusion of experts from other countries in the process of evaluation of project proposals.
- Organisation of international roundtable discussions and/or other meetings that would help set up common guidelines of evaluation processes, which could be then adopted on the national level.
- Strengthening of contacts and exchange of experiences between similar kinds of institutions (e.g. with a similar number of researchers).
- The training of experts and science policy-makers.
- The establishment of networks on the European level.

- **The EC could certainly play an important role as a facilitator of co-operation and mutual learning by:**
- Promoting and escalating the professional quality of ex ante evaluations and the effectiveness and importance of external ex post evaluations.
- Promoting the importance of the principle of compatibility of methods of evaluation on the subject of regulation.
- Promoting and encouraging the development of institutions/networks for evaluation of research activities in different countries.
- Promoting and advancing the establishment of national science foundations and (or) agencies for funding of research in Europe.
- Helping to establish mechanisms for the exchange of experiences through providing some form of guidelines to national evaluation systems.

---

35 Dogovor o izvajanju raziskovalnega evalvacijskega programma za obdobje 1999 - 2003, Ljubljana 1999.

In the future, the following actions could be taken at the national or international level in the area of the evaluation process:

- Training of national experts (evaluators). Training could give the experts insight into the significance of the particular criteria during the evaluation process; for instance, what is the role of scientific, economic and social criteria in certain types of research when evaluating a particular project proposal, and particular research when evaluating certain types of project proposals (e.g. basic research, applied research, technological development).
- A study of the effect of evaluations on the trends of national scientific/technological strategies and politics, on the development of particular scientific disciplines, on preserving a "critical mass" of researchers in a particular field of science, etc.

Observations and Conclusions

To increase the co-operation of Slovenian researchers in the EU's 5th Framework Programme, Slovenia needs to speed up the upgrading of the evaluation system to be compatible with the systems and processes in the EU and its implementation in practice.

1. It is important that the experience gained by Slovenian researchers, acting as evaluators under the first few EU framework programmes, be used to advance the national evaluation system.
2. More foreign evaluators should be included in the process of evaluation of the project proposals, as well as in the process of evaluation of the research activities and implementation of the National Research Programme.
3. It is necessary to provide financial and other support for: i) the development of science evaluation institutes on the national level, and ii) an appropriate European network of evaluation.
4. It is important to organise the training of prospective evaluators on the European level.
5. It is necessary to strengthen the work of the international teams of experts to improve the development of evaluation models for research activities on the national and European levels.

Learning by Doing Evaluation: The Relevant Approach in Transition Context

Kostadinka Simeonova
Centre for Science Studies and History of Science, Bulgarian Academy of Sciences
4 Serdika, 1000 Sofia, Bulgaria

1. The contextual problems of learning

Central and East European countries (CEEC) are the special case in the evaluation of R&D not only because their practice is much behind that of the western developed countries. The "differencia specifica" is the context, which attributes the main role of the evaluation for the ongoing reforms and restructuring of S&T systems in the region. This means that special function in the actual S&T policy is attributed to the evaluation. Generally the countries in this region are less prepared to apply the methodology and organisation of the evaluation process. That is why the experience and the transfer of knowledge and practices from western countries are of utmost importance. At the same time we have to keep in mind this difference between the function of evaluation in so to say "normal" situations and the "crises" in which the S&T system is in our region. In particular, the balance between "rational" and "political" aspects of the process of evaluation could be distorted due to this peculiar characteristic of the transition and expected outcomes of the process.

The concept of evaluation presupposes the following internal structure and system of categories important to examine the peculiarities in question:

- **Goal**: the evaluation is not a simple description or scientific analysis of the given situation, but rather the process by which goals are explicitly defined at the iniation. The goals in transition context are specific for CEECs, as they have to bring to the new S&T system: to create new institutions and to abolish some ones, to restructure the research fields and re-orient towards new priorities etc. The explicit formulation and articulation of such goals is a rather difficult task and implies the broadening of the discourse from technology mostly to ideology of evaluation.
- **Comparison**: the evaluation contains always the normative aspect as it is based on the "ideal" model or pattern. In particular cases the other real objects can play the role of the "model" for comparison. There is no single, or ideal practice to be referential for all countries in the region. Therefore the analytical approach to the normative structure (criteria) is a pre-condition for appropriate learning. In the last years not only developed western, but some of the countries in Central Europe, which are much ahead in restructuruing of their R&D systems become a "model" for those in the eastern part of Europe.
- **Action**: as a rule the evaluation is aiming not at knowing, but at action. A process or object is submitted to evaluation if there is a possibility to shift its development or situation in the intended direction. Which are the possibilities to act and to bring to some changes after evaluation in transition context? Who takes the responsibility for such changes? Do scientists themselves expect some positive development, or is their

attitude negative because of lack of such expectations? So, the space of action must be identified before the evaluation is intended. Political will is a key factor in this respect.

- **Effectiveness and efficiency**: the evaluation is analysed from the corner of the ratio between achieved results and explicitly defined goals (effects) and ratio between the results and the means used (efficiency). It is quite natural that without action hardly any effects can be expected from evaluation, i.e. this is closely related to previous characteristics. However in the process of learning one of the effects is the absorption of specific "culture of evaluation" from the respective community and this will create more favourable conditions for the successful evaluation in the future. Substantial intangible effects would be as well the facilitating of the adjustment of the respective S&T system to the patterns of EU countries.

2. Some lessons from R&D evaluation in Bulgaria

In Bulgaria traditionally evaluation is considered to be the "self-regulation" mechanism of the scientific community. It was based on internal professional criteria and the "peer review" system, broadly in use for the promotion of individual scientists, in the operation of scientific councils, editorial boards etc. This mechanism still is keeping its role as a base to enhance the research capacities in the given discipline and field of research.

The evaluation as a mechanism incorporated in the decision-making process and as an instrument of science policy is a new phenomenon in the country. Since 1990 alongside with radical political and economic reforms, evaluation is considered as a main tool of the policy of transformation of S&T system in the country.

By now in Bulgaria we still are far from systematic application of the evaluation. Comparing the practice of some other countries in the region we have to admit that in this country it is still in the initial stage of application. The project evaluation is the most successful example related with the new system of funding, but its impact is limited due to the very small share of project funding in total GERD. The evaluation of an organisation is represented by the Bulgarian Academy of Sciences (BAS), which initiates the evaluation of its units. Unfortunately, it was done on the base of internal capacities of the respective communities without participation of any foreign panels. No research fields or discipline was submitted to evaluation in the country.

Which are the main difficulties to rationalise the evaluation:

- The *monitoring of R&D is a crucial precondition* for good evaluation. This has been one of the biggest difficulties, to assure comparison with the ideal models and has brought to incorrect conclusions. The methodology of measurement must be unified, the statistical indicators compatible. Even after the improvement of indicators according to the OECD Frascati Manual since 1994 there are still the questions about the procedure behind the acquisition of respective status etc**.**
- The explicit *formulation of the goals* of evaluation could be done in a situation of well defined policy and strategy for S&T in the country, with the articulated priorities in different sectors of the economy, which unfortunately is missing because of delaying reforms. The general political instability in the country in last eight years was unfa-

vourable for a certain degree of continuity of the S&T policy, necessary to fulfil the tasks of evaluation. This resulted in difficulties to choose the relevant criteria of evaluation. Aiming only at the reduction of potential, evaluation could not bring to significant impact on the R&D system.

- A good evaluation needs full transparency as regards all four elements mentioned above: goals, means, process and results which have to be defined clearly and to be accessible. One of the important lessons is that the very fact of evaluation needs the change of the attitudes not only of the object of evaluation, but of the political and other bodies initiating evaluation and using its results for the management of R&D. Some "enlightenment" of the managers and key people initiating the evaluation on its the systemic character vs. ad hoc and partial approach is an important precondition.
- The *links between various indicators* is seldom analysed. For instance sometimes qualification structure is assessed positively, but it has to be projected on the age structure to assess whether the potential is "real one"; then age structure must be projected on the disciplinary one to consider the fast changing paradigms in natural sciences, or selected fields of research and the slower ones etc.
- The multicriterial approach is needed to be applied in the evaluation of the multifunctional organisations. It cannot be replaced by one-dimensional criteria for evaluation purposes. For instance, the radical change of the goal orientation of BAS towards exclusively fundamental research *done overnight created unacceptable change in previous policy and goals of the Academy.* As a result, a substantial part of the institutes have lost their advantages and position, being the result of a decade of policy impulses to perform applied research and experimental development. Many of younger and more active researchers, couldn't meet the new requirements and preferred to leave the respective institute. The same was the result when the commercialisation activity of the institutes has been assessed as inappropriate for fundamental orientation of BAS. Recently this policy was re-assessed and reflected in the Code of the BAS in which the amendment was proposed for the orientation of the Academy towards the applied research. However, this year the Academy is seemingly to start the second round of evaluation in which it enters much better prepared - recently the criteria and procedures for evaluation of the institutes and scientists were adopted by the General Assembly, so the process will be better structured and justified.
- *The action as follow up of evaluation was limited*. The most expressed case we have in the country - the evaluation of R&D units initiated by BAS shows that it was not supported by relevant policy on the national level, for instance to disseminate the practice and the space for action of the Academy was limited to one possibility to reduce the number of institutes and staff (in fact by over 30% of employees). Neither the funding nor other research conditions of the remaining part of institutes changed positively. The same holds true for the Ministry of Education and Science (MES) because of its limited possibilities for action. It did not play any significant role in promotion of the evaluation with the exemption of project evaluation, incorporated in the activity of the two Funds in the MES.

It has to be pointed out that the participation of the country in the Fifth Framework Program has a positive impact on evaluation. In the number of seminars and meetings the evaluation procedures of EC has been presented and discussed with experts and scientists.

To increase the competitiveness of Bulgarian scientists the pre-feasibility projects were introduced in the annual call for projects in 1998 from the MES. More concern for evaluation is shown by the authorities in BAS as well in term of broadening of the adopted criteria and indicators, in particular related to application and usefulness of academic research.

3. Some proposals to the round table discussion

The more effective mutual learning in evaluation east- west in order to achieve compatible evaluation systems depends on the following:

- *The selective transfer of knowledge*. It is clear that there are different sources of knowledge as the practices and degrees of development of evaluation differ considerably among the western countries. In my view an important question is from whom will it be more appropriate to learn? I think that *for the smaller CEE countries it is important to learn from the developed smaller countries from the periphery*. In principle, the problems and the coping strategies are very similar in the smaller scientific communities. In particular the dual nature of evaluation as a rational, but also political process in the smaller countries shows the bigger role of the interests of different actors. From the corner of the regional south-east position of my country and important cultural factors on evaluation the experience of the southern EU countries seems to be very relevant to our case.
- The *improvement of the monitoring system and methodology of evaluation* in CEECs needs a co-operation in terms of (a) regular access to the relevant manuals, instructions, working papers and so on on these topics, and (b) participation of the scientists from the pre-accession countries in elaboration and discussion of such unified instruments for evaluation as opportunity to share experience and practice.
- The workshops like the present one in which the broad issue of evaluation is discussed has to be followed by a *series of seminars and working groups*. It seems that the effectiveness of such meetings will increase if their topics would be more focused on specific issues such as evaluation of different objects - programmes, research fields, project, organisations etc. This will bring more detailed and specified knowledge transfer for those countries with no or limited experience. For example, evaluation of research fields in a smaller country with participation of foreign experts will be anextremely important issue for Bulgaria and examine the possibilities how it could be assisted and helped by EC, having in mind the high costs of such exercise.

As a conclusion I would like to point out that recently more attention is paid not only in my country, but in the other countries *to differentiate between the output and outcomes,* the first usually measured by quantitative indicators and later having qualitative characteristics which is broadest and has some intangible dimensions such as impact on the development of knowledge, emergence of the new research fields, initiation of small enterprises, educational and social impact etc. The outcome is much more difficult to evaluate, some time needs more time for evaluation, but its legitimation power is immense and deserves special attention.

Evaluating RTDI in the Cohesion Countries: Relevant Findings for Future Support[1]

Lena Tsipouri
University of Athens, Deptartment of Economic Sciences
5, Stadiou Str., 10562 Athens, Greece

1. Introduction

All recent evidence from the academic literature, the OECD recommendations and the European Commission planning processes attribute technology a key role for competitiveness and welfare. As a consequence an increasing part of the Community intervention, either in the form of Structural Funds (CSFs) flowing to the less favoured regions (LFRs) of the member states or in the form of support to the Central and Eastern European Countries (CEECs), goes into research, technological development and innovation (RTDI). Flows to the former have a track record of a decade and have reached an order of magnitude of 5-6% of the total CSF. The design of RTDI policies for LFRs has been based in the experiences of advanced countries, while the measures used in CEECs are transferred from both advanced regions and less favoured ones.

It is suggested in this paper that the decision to promote technology as a means of development has been correct for all types of current and future regions in the EU. Yet the implementation of this policy is more difficult than one would expect. While the transfer of experiences for physical infrastructure is easy from one place to another, intangibles are less easy to imitate. Because of differences in structures, administrations and path dependencies the transfer of best practices is more difficult than it appears at first sight. In a recent study examining the impact of RTDI in the Objective 1 and Objective 6 regions, several lessons were drawn that can be of relevance for the future of structural RTDI policies both in the LFRs and the CEECs.

The results of the thematic study

The evaluation was carried out by a team of international consultants during 1998 and early 1999. It involved fieldwork and data collection in all Objective 1 and 6 regions. Interviews with senior personnel and key informants in Member States and regions, case studies on examples of good practice, innovation profiles for the regions and a survey of beneficiary small firms, comprised the core of the fieldwork.

The results varied substantially from one region to the other. Some regions, like Ireland, could use their new technology effectively to link with development and improve their

---

36 This paper draws substantial ideas from the "Thematic evaluation of the impacts of Structural Funds (1994/99) on research, technology development and innovation (RTDI) in Objective 1 and 6 regions" prepared and co-ordinated by CIRCA-Dublin, Lena Tsipouri-University of Athens and PriceWaterhouse-Coopers, The Hague.

competitiveness, while others, like Crete, succeeded in building up an excellent research system, though internationally rather than locally rooted. Still others were unable to demonstrate success or even absorb the funds earmarked for RTDI. Ambitions, conditions and results were unique in each one of the about 60 regions studied.

Overall results demonstrate certain tendencies, though strongly differentiated:

1. *The future growth capacity of Objective 1 and 6 regions is being strengthened by the support from the Funds* for research, technology development and innovation (RTDI). While the direct economic impacts are still modest a much needed strengthening of the long term capacity of Objective 1 and 6 regions for economic development is occurring, with the assistance of the RTDI investments being provided by the Funds. The availability of the Funds is speeding up the enhancement of RTDI capabilities, and improving both the pace and the scale of RTDI development.
2. *Complementary actions taken by Member States have not been adequate or sufficient in order to maximise the Community added value from the Funds.* Process inefficiencies were only addressed in exceptional cases, like the Irish overall policy design and specific performance rewarding schemes in Spain. As a general rule the administrations were unable to change their style and live up to the expectations of voluminous and ambitious programmes. In particular it is needed to provide an effective policy delivery system, have effective evaluation and monitoring systems in place and improve the co-ordination between the main Ministries involved at the national level. Better integration of RTDI policies with other non-RTDI policies, so as to enhance the complementary assets needed for full exploitation of RTD potentials, is also essential. The quality of the underlying regional strategies for RTDI is a major weakness (with some exceptions) to be addressed by all participating regions in the next round of the Funds. In this context, the RTP/RIS/RITTS[2] exercise has provided a commendable model for consideration.
3. *The enhancement of RTD capacity by the Funds has been important for its contribution to the development of human capital and skills, but capacity enhancement has been overemphasised during this period of the Funds***.** Greater attention now needs to be given to making full and effective economic utilisation of the existing installed RTD capacity in the Objective 1 and 6 regions. Despite this basic statement there is a case for strengthening RTD capacity only in the faster growing regions. In these, demand has to be anticipated, otherwise bottlenecks in skills and expertise will slow development. But this process must be demand driven. The criteria to be tested in judging capacity enhancement proposals are 'anticipated demand' and 'utilisation potential' by the enterprise sector.
4. Unless there is evidence that the absorption capability of the enterprise sector for RTDI support can be significantly improved, there would be little point in any significant increase in the Funds available for the next round. *Improving the absorption capacity of the enterprise sector in the regions should be a key objective for any future negotiation*. In this aspect regions differ in their formal and informal rules and the propensity of enterprises to adopt change. A typical mistake is to consider the absence

37 Demand oriented regional policy design exercises based on the consensus of all relevant regional actors.

of demand as synonymous to latent demand. Only the latter can be put in the basis of a growth oriented RTDI.

5. *Policy design also needs to become more ambitious and to follow some universal rules, despite the diversity of the regions.* Management inefficiencies and discontinuities are reducing the effectiveness and impacts of the Funds. New and less complex management systems are needed. There are discontinuities between some agencies responsible for management of the Funds at national level which need also to be addressed, and there are some overlaps between programmes, while others which should be mutually supportive, appear to lack co-ordination. Performance monitoring needs to be strengthened - everywhere. Improved real time monitoring arrangements are essential.

2. Lessons learned

In parallel to these generalised conclusions several good practice examples were identified. The term "good" instead of "best practice" was used to denote that there is no optimal or best concept for any given problem. Some examples were positively evaluated in their environment, but before trying to imitate them one should carefully study the conditions of transferability. It is suggested that intervention through a system of *fine tuned and tailor made incentives* can help regional productive systems to be put in motion. Thus on the one hand the literature suggests certain general guidelines and principles to be adopted universally at the Obj. 1 level, and then specific incentives for the national level (change the formal institutional framework) and regional level (change the type and scope of intervention). Incentives need to be very strong if they are expected to change formal and informal rules, as already pointed out by Hirschman in early development theories.

From the variety of very similar schemes implemented in the 1994-99 period, the following were found to be good practices:

On agencies and schemes supporting the management of RTDI resources
- Good management for policy implementation by the Irish Forbairt
- Good linkage between State and Regional Authorities in the French Scheme of Regional Delegates on Research and Technology

On industry support schemes
- Measure 1 in Ireland

On the idea of shifting from grant to equity or repayable loans
- CDTI in Spain
- Equity considerations in Ireland

On the activation of the private sector
- The Federation of Industries of Northern Greece
- The Industrial Research Group in Ireland

On the way to improve capacity utilisation through Cupertino
- The Irish PATs
- The Northern Irish ATCs
- The Portuguese Research Associations in the region of Norte

On clustering as the best means to increase spillovers
- Clustering in Saxony

Schemes very similar to those (clustering in many other areas, public research associations in other countries or the mobilisation of industrial federations in other areas) were less successful or even failed. The key question is then why. Regions or countries designing new RTDI policy measures have to look not only at what has been successful elsewhere, but mainly "from whom" to learn and "what" to learn. It is suggested that it is better to be more modest in terms of ambitions and adopt schemes that have worked in regions with similar administrative and/or productive structures, rather than adopt the most ambitious schemes that one does not have the resources to manage. By this it is not suggested not to be ambitious, but to measure ambitions and endow exceptional efforts with the necessary means.

Good practice has to be put in context. If contexts are similar it is worthwhile trying to replicate it. The separation of policy and implementation, the introduction of an evaluation culture with continuity and feed back and the trust relationships among agents have worked in Ireland because there is a culture of co-operation among agents and the skills of the public administration are ahead of other LFRs. If other regions want to imitate this model it may mean a thorough shift of power, that will need very strong political backing to be implemented. In that sense the Forbairt example is not just a managerial good practice but represents the gradual shift of power and practices.

By the same token, the appointment of research delegates only makes sense, if a real network is created and the national authorities can benefit from network externalities. If only few regions respond to the concept then the regional level remains the only appropriate as regions cannot benefit from the exchange of information and common standards and support. At the same time if the relationship between the State and the region is unequal (in terms of skills or power) then the delegates will find difficulties to co-ordinate their action.

Similarly a key message is that the public authorities should introduce frequent evaluation and not be afraid, based on the evaluation results, to redirect and fine-tune their support. For industry to apply for such support with the aim of becoming more innovative rather than just improve its liquidity, firms need to have confidence to the mechanism: transparency on dates and criteria, abilities of the evaluators, commitment on novelty etc. The establishment of a body similar to IRDG is also very helpful. Yet, overall the Irish success is based also in the strong GDP growth, that increased demand and availability of funds in industry, a condition that cannot be replicated based on RTDI policies alone. Countries or regions without this background cannot imitate the Irish success.

Equity participation, mainly in the form of venture capital, and repayability by bank loans, has been the motor behind the development of new technology in the most successful regions. Their use was generated by the market and the culture and ambitions of companies and financial sources permitted this successful matching. In LFRs the basic approach was that, as development should be based on stronger incentives[3], the idea of grants was introduced. Yet the proliferation of grants in the EU, which probably constitute the most generous transfer scheme for RTDI studied, have attracted proposals for companies that are more interested in the short term benefits of the grant and less in the incorporation of technological change in their strategy. Thus it may be time to consider, at least for part of the grants repayability and equity as a means to both assure the incorporation of strategy in the firm and use it as a means to increase the overall funding available for company support.

There are though important cultural and managerial factors that make it difficult to transfer the CDTI experience and even more the equity participation, as the Irish case demonstrates. CDTI was not efficient in running small R&D projects, thus it can only be transferred in regions and exercises with a broader policy vision that can afford the cost of setting up such a mechanism. The other problem of transferability of the CDTI experience is the acceptance of the repayability clause by industry. In earlier periods this approach was considered as a "punishment for success" and was gradually abandoned by most policy makers. Re-introducing it in a period of financial austerity may be sensible but is likely to encounter resistance.

Finally the idea of linking capacity is mature in all LFRs. It is unlikely to follow the models from core countries where a high number of institutes are linked under the same administrative umbrella[4] because the overall size of the scientific community does not justify such an intervention. But the idea of connecting institutes in the same or complementary fields of research and build capacities that are more likely to serve a wider range of interests, following the model of the U.K. research associations makes sense even for smaller scale activities.

The programmes have been very concise because of the size of the region and include practically two types of programmes offered in other regions: enhancement of university capacity and industry support. Bigger regions and multi-regional programmes are unlikely to be able to follow the same pattern. Their merit lies in good management and a rapid creation of a good reputation. Following the Northern Irish case all actors can learn from the model for the management structure of the university projects, including especially the *Industry Steering Group and the commercial project manager,* as well as by the *project appraisal procedures*. Again the rigorous technical and financial appraisals combined with a well-structured weighting and scoring system for assessing individual proposals made the difference to other similar projects.

---

38 Based on the seminal contribution of A.O. Hirschman on the nature of economic development

39 Like the Fraunhofer Gesellschaft in Germany, the CNRS in Paris and CNR and ENEA in Italy

3. The role of the Commission, the national and the regional governments

Taking these brief remarks under consideration, the Commission can play a crucial role. Going beyond the, as yet invaluable, contribution of requesting the transfer from hard to soft measures of regional support, it can now

- ask regions to design their RTDI policies as function of their capabilities and structures, rather than as an imitation of success
- benchmark the most common support scheme to give regions a possibility to measure themselves against feasible targets
- help the diffusion of good practices, not by choosing one best scheme to promote, but discussing families of similar schemes, their merits, difficulties and local circumstances and finally
- continue the path of clarification of objectives, training and policy minded studies.

Towards Integration of Evaluation in RTD Policy-making

Societal Challenges for Evaluation

Arie Rip
University Twente, Department of Philosophy of Science and Technology
7500 AE Enschede, the Netherlands

Start with societal challenges for science (RTD) rather than challenges for evaluation:

Transparency: Science as Goose with Golden Eggs

That time (of Science, The Endless Frontier, after the title of Vannevar Bush's 1945 Report to the USA President) has gone.
The promises of science/RTD to contribute to economic growth (competitiveness) and sustainability have to be *delivered* now – that's one reason for societal interest in RTD evaluation (let's see what we are actually getting). The link between RTD and outcomes is not immediate, though (cf. below), and the new regime of Strategic Science allows scientists to limit themselves to delivering new options, rather than actual innovations.
There are more societal challenges to science (RTD): society is less fatalistic about impacts, risks (cf. molecular biology and genetic modification) and wants some TA done; and society wants expertise (up to "sound science") even in the face of large uncertainties.

Transparency: ozone layer (early warning & expertise, as well as TA)

Given these challenges to science (RTD), what role for RTD *evaluation*? Start with present roles and trends, and then come back to the challenges.

Evaluation in general has its roots in accountability (cf. role of General Accounting Offices in evaluation) and in so-called punitive evaluation (an evaluation is called for to justify a decision to close down something). In some areas of policy, like welfare and education, evaluation of programmes has become important (and a sector of activities in its own right).

Evaluation of RTD has evolved in a similar way, and was shaped by the particularities of the national and international research and innovation systems (cf. special issue of *Research Evaluation* with country reports). In addition:

> "The major rationale for evaluations has shifted and evolved from a desire to legitimate past actions and demonstrate accountability, to the need to improve understanding and inform future actions. Correspondingly, the issue focus of evaluations has broadened away from a narrow focus on quality, economy, efficiency and effectiveness, and towards a more all encompassing concern with additional issues such as the appropriateness of past actions and a concern with performance improvement and strategy development." (Final Report of the EU-TSER Network on Advanced Science and Technology Policy and Planning, June 1999, p. 37)

(This last part is argued for by evaluators and some policy-makers, but not yet widely practiced.)

Before I discuss these trends in somewhat more detail, I note two particularities of RTD with important implications for the style of RTD evaluation and its integration in RTD policy.

- RTD is open-ended, non-routine
- The institutional landscape of national RTD systems, with its academic institutions, big public laboratories, RTD stimulation programmes, and special research centres
- and the need to maintain such a system and keep it healthy but also to modify and improve, sometimes overhaul the system.

The first particularity is the open-ended nature of RTD: it is not a routine activity with specifiable outputs. As Cozzens (in Centre for Research Policy (Wollongong, Australia) report 1998, p. 85) phrases it (when discussing GPRA, cf. below):

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

For academic research, performance indicators are then often replaced by activity indicators, like the number of articles. For applied research, intermediate performance indicators are sometimes used, for instance the very dubious measures based on patents. As Cozzens notes, there are studies showing the economic and strategic importance of RTD in the aggregate, using some such indicators. But the link at the level of projects and programmes is uncertain.

> "[..] [T]o-date, not a single U.S. funding programme has adopted any of them to measure its achievement of programme goals. The reason is that the links between programme activities and economic outcomes are too complex and unpredictable, especially for fundamental research."

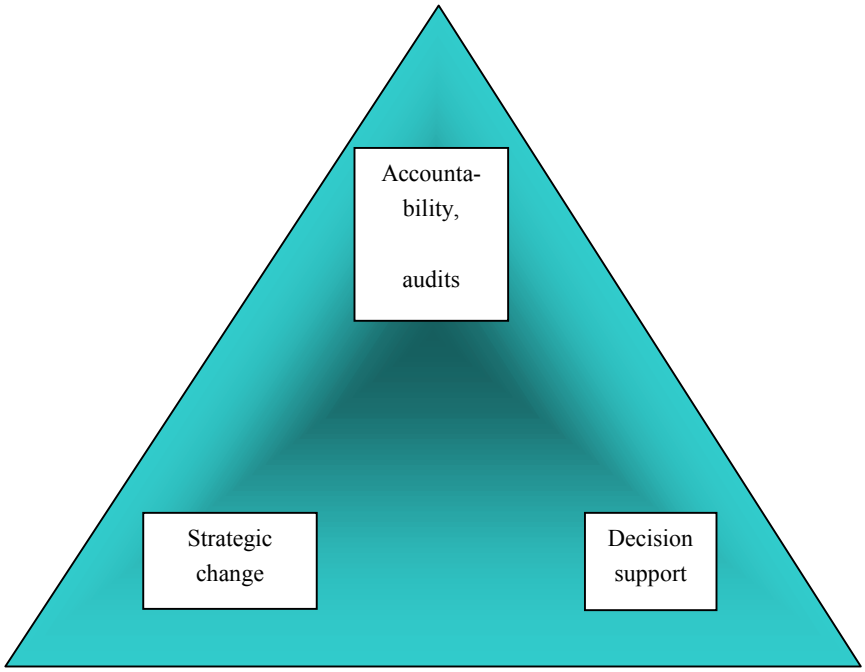
Another issue deriving from the open-ended nature of RTD is the occurrence of goal shifts. In the standard view, evaluations should take the original goals as the standard against which to measure goal achievement. While learning will occur, also about the feasibility and desirability of these original goals, and some adaptation is acceptable, one cannot allow any adaptation (actors could then redefine whatever they did as goal achievement). In the case of RTD, the argument for accepting shifts in goals is stronger, although there must be some limits. (Example of Alvey Programme, where collaborative culture was taken as an important goal when it became clear that the original goal of making British ICT industry competitive could not be achieved.)

The second particularity is the institutional landscape of national RTD systems, with its academic institutions, big public laboratories, RTD stimulation programmes, and special research centres, and the need to maintain such a system and keep it healthy.

***Transparency:* concentric layers in national research system**

Then the tension with the need to modify and improve, sometimes overhaul the system. RTD evaluation is located in this tension, and the same approaches will be viewed and applied differently, depending on whether they are positioned at the side of maintaining what exists or at the side of trying to change it. Evaluations of the system of university research, and evaluations of public research institutes are examples.

Having said this, I can locate the policy functions (functionalities?) of RTD evaluations in the triangle of evaluation.



The three corners are:

(1) *accountability* (what did you do with the money?), which leads to *audit-type methods*. For RTD, e.g. public institutes, regular audit methods can be used, but the bookkeeping is often extended to include research inputs and research outputs. In the 1990s, accountability itself has taken on a new complexion, with the establishment of government Performance and Results Acts in a number of countries (the terminology used here is that of the 1993 USA Act). This has created tensions, especially for the basic research institutions; in the USA, the National Science Foundation was allowed to use adapted performance indicators. The existence of such Acts implies that criteria and sometimes measures will be specified explicitly, and become binding (to a certain extent). In the present political and policy climate, relevance to society will be an important criterion, and there will be pressure to operationalise it. Such pressures occur also without there being an Act, for example in the university research assessment exercise in the Netherlands (not in the UK!): it is a combination of diffuse

credibility pressures and strategic action of scientific entrepreneurs and science officials.

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

(3) *strategic policies* (often trying to effect strategic changes in the research system, in the direction RTD is developing, in the organisation of research performing institutions), which leads to an interest in evaluating appropriateness of policy goals and of promises of actual and possible directions of RTD and the RTD system.

One can locate specific RTD evaluation approaches and their integration in policy in the triangle.

Underlying the analysis is the diagnosis of a changing RTD system, linked to overall changes in the world. When we add this as a sort of additional baseline to the triangular figure, we create a further perspective on RTD evaluation: changes and challenges may be about transformation, which creates a focus of evaluation on strategic changes (left-hand side), or on productive functioning of the evolving RTD system (right-hand side).

Given the changes in the research system and how it is linked with society in general there are a number of specific challenges for RTD evaluation. These will be taken up by my colleagues Georghiou and Kuhlmann. Here, I will mention only a few which relate to "societal changes".

I see several challenges, and block them as two about development of research system as such, two about changes in the way the research system relates to society, and a fifth challenge to evaluation approaches which can no longer abstract from the question of "sound science".

T**ransition to modern research system including peer review**

- audit plus (broader performance measures)
  - gradual internalisation of external goals into the working of the RTD system
  - more involvement and interference of old and new stakeholders: fragmentation or a new mode of knowledge production?
    - quality of expertise in public arenas, “extended” peer review

Accountability to the Public and Value-for-Money

Luke Georghiou
PREST, The Victoria University of Manchester
Oxford Road, M13 9 PL Manchester, Great Britain

Accountability provides probably the longest standing reason for evaluation to be performed and persists as a major motivation along with legitimation and learning. However, in broader political science the concept of accountability has a range of meanings. The most basic is that of political accountability whereby ministers are accountable to electors for government actions. As government has become more complex, and particularly in association with public management reforms introducing private-sector style practices, notably semi-autonomous agencies, accountability has also become more managerial in concept and practice. It is as part of this process that evaluation and the use of performance indicators have become strongly institutionalised. Under this model ministers are deemed to remain responsible for policy, but responsibility for management is delegated to agency heads.

Some have argued that the sheer complexity of these arrangements obscures the trail of accountability[1], while others acknowledge that policy and management issues are not always separable[2]. The role of parliamentary bodies as a source of independent scrutiny in most countries has been extending beyond their traditional domains of financial and political scrutiny. Put another way, such evaluations have moved beyond *economy* and *efficiency* to include *effectiveness* (i.e. what was achieved). The UK National Audit Office includes all three in its extended definition of 'value for money' and sums them up respectively as spending less, spending well and spending wisely.

Chelimsky perceives evaluation conducted under an accountability perspective as being structured on the basis of strong causality, that is to say a policy intervention is made and the evaluator aims to establish what resulted in consequence (and how such effects corresponded to what was intended)[3]. She contrasts this perspective with development (focus on process and performance improvement of an institution) and knowledge perspectives (academically motivated evaluators normally focussed on in-depth understanding of a specific area). One way in which these differences are manifested is in the use made of evaluations. Independence is prized above other characteristics for those working from an accountability perspective which may sometimes prevent an engagement with policy improvement. However, as Chelimsky argues, use of evaluation is itself a complex matter and may be divorced from the findings per se, resulting for example from anticipation or deterrence. It is the experience of many evaluators that major policy shifts result from an accumulation of what might be initially unpopular findings.

---

40 Rhodes RAW, *Understanding Governance*, Buckingham: Open University Press, pp101-103

41 Day P. And Klein R, *Accountabilities in Five Public Services*, Tavistock Publications, London, 1987

42 Chelimsky E, Thoughts for a New Evaluation Society, *Evaluation* Vol 3(1):97-118

Research and innovation policy has evolved in a similar direction to the rest of government, with some differences arising from the need to preserve the autonomy of scientific judgement. However, the growing reliance on socio-economic rationales to justify the allocation or resources to research and innovation policies (with the consequent large if not adequate volume of resources) has opened these areas to more conventional scrutiny. Some major barriers remain for those who wish to approach research and innovation policy solely from an accountability perspective. The most fundamental is that the relationship between research and its socio-economic consequences is not fully understood and such understanding as does exist tends to show that there are multiple paths, long-timescales and unexpected effects which may be difficult to attribute or even to trace in specific circumstances. While most economists are agreed that research in general produces high private and social returns, the degree of precision required to arrive at a rate of return (the ultimate expression of value for money) is not achievable for most projects and programmes. Where there is a strong drive for accountability, proxy indicators such as production of outputs (papers, patents etc) may be used, but those most expert in the use of these normally caution that they are instruments best used at an aggregate level and in combination with other information. Less expert use of these indicators has produced highly misleading results, usually as a result of making inappropriate comparisons across fields of science and technology, industrial sectors or types of research. As in other domains there is an uneasy relationship between evaluation and performance indicators. Perhaps the best way to resolve this is to insist that a performance indicator regime can only be established when evaluation studies have resulted in a deep understanding of context, rationale and the specific model which links research to effects.

What then can evaluators tell the public about the investment it makes in research? A well-performed evaluation of a programme (or, with greater difficulty, an institute) should be able to catalogue the outputs resulting from the research. For nearer-to-market research, where industry or users are involved, evaluators should also manage to trace the nature of their direct impacts upon participants. Participants may be able to estimate the magnitude of these impacts (sometimes with the help of evaluators), but only if they make assumptions about the relative weight of the policy intervention compared with knowledge from other sources and the various complementary assets needed to realise the impacts. Thus the evaluation accepts the participant's implicit model of the relationship between research and socio-economic effects. A most important service the evaluation community can offer is to elucidate the nature of this relationship. Indeed, evaluation studies have played a leading role in developing an understanding of the indirect, networking and human capital benefits arising from research and which underpin interactive models of innovation.

While satisfying some of the demands of accountability, restricting evaluation solely to an accountability perspective deprives society of what is probably its principal benefit, the opportunity for policy learning and improvement. It may be argued that the two perspectives are incompatible because learning involves a level of engagement with the programme being evaluated which compromises independence. This argument does not withstand close examination. So long as a structure exists to ensure that the evaluators

maintain their independence without an expectation of reprisal or reward arising from the outcome of their deliberations, there is no reason why society should not benefit from their findings. The most sterile of evaluations offer only a score or financial consequence even when they have amassed sufficient evidence and analysis to improve the situation which they examined. This line of thought has been behind the strong tendency for peer review mechanisms even for funding applications to offer feedback to unsuccessful applicants. While an evaluation without an accountability perspective might lack legitimacy, an evaluation without a learning element is an example of poor value for money.

Distributed Intelligence for Innovation Policy Planning: Integrating Evaluation, Foresight and Technology Assessment[1]

Stefan Kuhlmann
Fraunhofer Institute Systems and Innovation Research (ISI)
Breslauer Strasse 48, 76139 Karlsruhe, Germany

1 Introduction

It is the year 2005 and the Managing Director of Biomat has a problem. Her spin-off firm builds replacement human organs using metagenic technology, but after two years of success Biomat is at a crossroads. Should it stick to metagenics, which is costly and prone to production problems, or should it use the latest ultragenic approaches - still unproven but likely to yield great cost reductions?

She switched on her videophone and traced the local head of RIB, the Regional Innovation Bureau of ENDBITS, the European Network of Distributed Bureaux of Intelligence for Technology Strategies. RIB helped her to prepare a videonote on technology options. It ran a standard search on the European Foresight Bank, an electronic tool which logged all of the world's foresight outputs and used AI algorithms to cluster the results and build scenarios. Recent expert assessments all looked good for ultragenics, but RIB advised her not to rely solely upon foresight results – however positive. Social and regulatory problems were also possible, and the Bureau had heard of some problems in Austria.

RIB used the Technology Assessment directory to identify the main Austrian experts in the field and confirmed that ultragenics had been subject to ethical challenges from a local religious foundation. RIB then called for more information and scanned the recordings of the Consensus Conference. Relief! The objections were based on a misunderstanding of the procedures for ultragenics (which unlike earlier approaches did not depend upon foetal cells) and the citizens' jury had come out in favour of the technology.

Biomat was ready to launch its ultragenics research programme, but was worried about the cost. RIB pointed out that all three of the European Research Framework Programme agencies offered support, but noted that recent evaluations praised the Prague office for its fast turnaround of proposals and claims.

"Thanks RIB" said the Biomat MD. "Life without ENDBITS just wouldn't be the same".

---

43 This paper is based on a report produced by members of the Advanced Science and Technology Policy Planning Network – a network set up as part of the Targeted Socio-Economic Research Programme of the European Union: 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"*, Karlsruhe (Fraunhofer Institute Systems and Innovation Research, ISI).

The basic premise of this article is that better access to relevant information can facilitate decisionmaking. Science and technological innovation are neither costless nor necessarily benign activities, and society has a right to become involved in determining the directions and the resources of public innovation policies. Specifying appropriate policies, however, is not an easy task: whether located in the public or private sector, decisionmakers concerned with research and technology choices, strategies and policies need a wide range of high quality intelligence inputs in order to make sound decisions.

Traditionally, policymakers have used a number of "intelligence" tools and techniques to provide them with the data they need. In the public sector, for example, innovation policy formulation has been improved in recent years via the use of policy evaluation exercises (EV), technology foresight (TF) or technology assessment (TA). In future, however, more will be needed: there is a growing need for *improved intelligence tools* and access to the results of related exercises carried out across the globe. Therefore, an *infrastructure of European "Distributed Intelligence"* is suggested which could provide policymakers with access to strategic intelligence outputs produced in different locations for different reasons. The aim is not an overarching, monolithic information system but a flexible "bottom-up" architecture linking multiple sources of expertise rooted in a variety of innovation policymaking "arenas" on different levels.

2 Complex Innovation Systems and the Need for Improved Strategic Intelligence

Specifying innovation policies is not easy. Analysts in the field have abandoned simplistic models of how innovation and innovation processes work. It is increasingly recognised that the dynamics of so-called "innovation systems" are complex and difficult to understand, and that scientific and technological communities, not to mention the "users" of their products, face a number of challenges, both now and in the future: (1) *The nature of technological innovation processes is changing*. The production of highly sophisticated products makes increased demands on the science base, necessitating interdisciplinary research and the fusion of heterogeneous technological trajectories. New patterns of communication and interaction are emerging which researchers, innovators and policymakers have to recognise and comprehend. (2) European innovation policymakers have to co-ordinate their interventions with an *increasing number of actors* (e.g. European authorities; numerous national government departments and regional agencies; industrial enterprises and associations; trade unions and organised social movements etc.). (3) The growing cost of science and innovation is also likely to accelerate the *international division of labour in the European research system*, a development which will increase the need for a highly strategic, though not necessarily a centralised, European innovation policy.
Policy-formulation in these circumstances is not straightforward. There is increasing pressure on policymakers to:

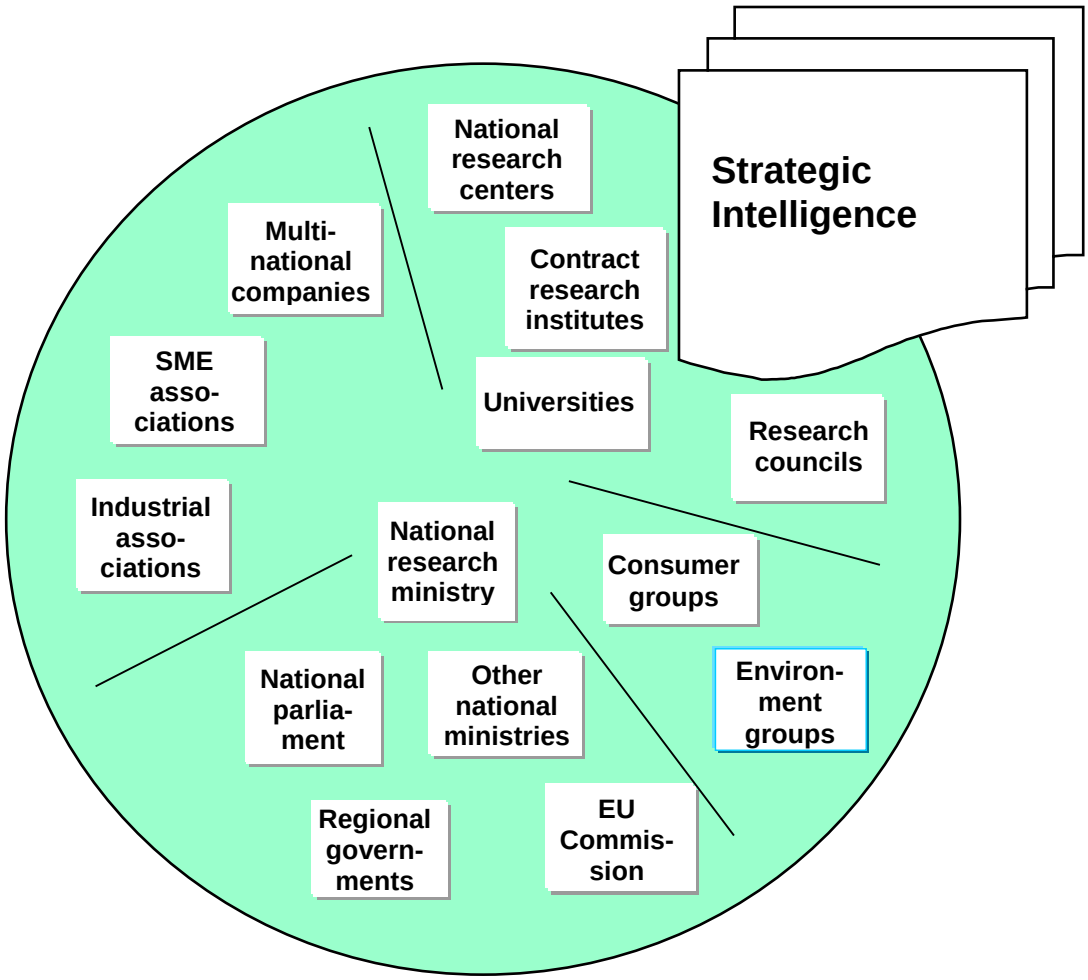
- *increase efficiency and effectiveness* in the governance of science and technology;
- *make difficult choices* in the allocation of scarce resources for the funding of science and technology;

- help preside over the establishment of an *international division of labour in science and technology* acceptable to all actors involved;
- integrate "classical" innovation policy initiatives with broader socio-economic targets, such as reducing unemployment, fostering the social inclusion of less favoured societal groups and regions, as claimed in particular by the 5th Framework Programme of the European Commission;
- acknowledge, comprehend and master the *increasing complexity* of innovation systems (more actors, more aspects, more levels etc.);
- adapt to *changes in the focus of innovation policies* between international (growing), national (declining) and regional (growing) levels.

Over the last two decades, considerable efforts have been made to improve inputs into the design of effective science, technology and innovation policies. In particular, formalised methodologies have been introduced and developed which attempt to analyse past behaviour (EV), review technological options for the future (TF), and assess the implications of adopting particular options (TA). Achievements in these areas have been impressive. As a complement of EV, TF and TA, other intelligence tools such as comparative studies of the national, regional or sectoral "technological competitiveness", benchmarking methodologies etc. were developed and used. Policymakers at regional, national and international levels have all benefited from involvement in these processes and exploited their results in the formulation of new policies. Analytically, one can identify a couple of structural *factors boosting the function of Strategic Intelligence (SI)*:

- A linear model of policymaking as a consequential process (typical steps: formulation, agenda setting, decisions, implementation, evaluation, formulation ...) is no longer appropriate, at least not in the field of innovation policies. Here, all typical *steps are more or less interacting*. The emergence of SI knowledge as a policy resource on the one hand and structural and institutional preconditions of using intelligence activities on the other influence and transform each other. Often it is external pressure on policy actors and the related arenas that gives the impulse for the production and application of advanced SI.
- Innovation policy is rather (and increasingly) a matter of networking between heterogeneous (organised) actors instead of top-down decisionmaking and implementation. Policy decisions frequently are negotiated in *multi-level/multi-actor arenas* and related actor networks. Negotiating actors pursue different - partly contradicting - interests, represent *different stakeholders perspectives*, construct different perceptions of "reality", refer to diverging institutional "frames" (see figure 1). "Successful" policymaking normally means compromising through *alignment* and "re-framing" of stakeholders' perspectives.
- Contesting and negotiating actors use money, power and information as their main media. Various actors have different shares of these resources at their disposal. *Strategic Intelligence* tools (as EV, TF, TA) use in particular *"information" and knowledge as negotiation medium,* facilitating a more "objective" formulation of diverging perceptions of (even contentious) subjects, offering appropriate indicators and information-processing mechanisms.

Figure 1: Actors in Innovation Policy Arenas and Strategic Intelligence



Increasingly, it has become obvious to both policymakers and the analysts involved in the development and use of SI tools that there is scope for continuous improvement and a further need to exploit potential synergies.

3 Innovation Policy Evaluation, Technology Foresight, Technology Assessment – State of the Art

Roughly, one can describe the basic concepts of EV, TF, and TA in the following way:

- Practices of science, technology and innovation policy evaluation are wide-ranging, and their functions vary significantly (1) from the provision of legitimisation for the distribution of public money and the demonstration of adequate and effective use of the funding by measuring the scientific/technological quality or the (potential) socio-economic impacts, via (2) improved management and "fine tuning" of S&T policy programmes, to (3) an attempt to improve transparency in the rules of the game and the profusion of research funding and subsidies, enhancing the information basis for shaping innovation policies, in the sense of a government-led "mediation" between diverging and competing interests of various players within the innovation system.

- "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, B. (1995):: "Foresight in Science and Technology" In: Technology Analysis & Strategic Management vol. 7, no. 2, 140).
- Technology assessment: In very general terms, TA 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 TA is hidden.

A survey of existing practices and experiences with the integrated use of the three intelligence tools for innovation policymaking EV, TF, and TA in various European countries and the EU Commission leads us to the following conclusions:

(1) Although examples of integration between the three bodies of experiences can be identified in several countries, there is no systematic effort, neither by policymakers, nor 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.

(2) Industry has an older tradition of combining approaches when defining strategies to assess uncertain (technological) developments with potentially wide impacts, both commercial and societal.

(3) Present empirical and well-documented examples of cross-border learning show that it is valuable to learn even from different institutional settings, to avoid repeating the mistakes and to pick up good practice experience more quickly.

(4) There is no "blue-print" of how the tools of EV, TF and TA can be best combined. The configuration should be considered from case to case, depending on the objectives and scope of the policy decisionmaking process in question. We do not advocate integration per se, but an integration 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, we 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.

4 General Requirements for Distributed Intelligence

- Summing up and in order to justify the direction we take in this article, we have to stipulate a number of general principles of Distributed Intelligence for complex innovation systems. *General requirements of Strategic Intelligence* are:
- Organize mediation processes and "discourses" between contesting actors in related policy arena
- Inject policy evaluation, foresight and TA results, also analyses of changing innovation processes, the dynamics of changing research systems, changing functions of public policies
- Realize thereby the multiplicity of actor's values and interests
- Facilitate a more "objective" formulation of diverging perceptions by offering appropriate indicators, analyses and information-processing mechanisms
- Create forums for interaction, negotiation and the preparation of decisions
- Respond to the political quest for democracy vis-à-vis technological choices.

Since innovation policymaking occurs in multiple policy arenas on regional, national, European levels there is a need for "interfaces", linking different systems and related policy arenas. General requirements of Improved *Strategic Intelligence Infrastructures* based on Distributed Intelligence are:

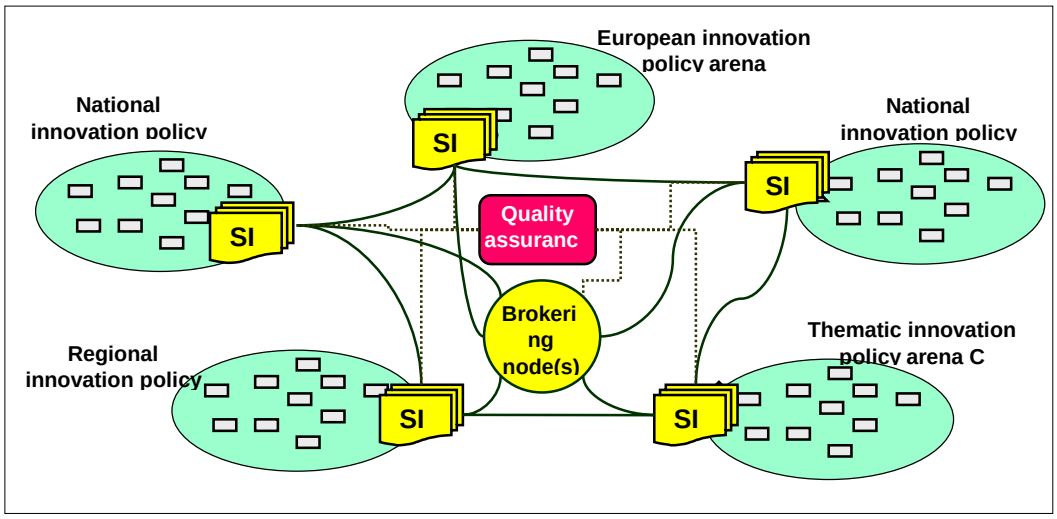
- Create an architecture of "infrastructures" for Distributed Intelligence – but no one unique "system"!
- Link – via SI infrastructure – the existing regional, national, sectoral etc SI facilities, horizontally and vertically
- Build brokering "nodes" managing and maintaining the infrastructure
- Establish a "enabling structure" allowing free access to all SI exercises undertaken under public auspices. Thereby the "pedigree" of information transferred through the infrastructure must be traceable
- Offer a "directory" facilitating direct connections between relevant actors
- Define clear rules to access the infrastructure
- Make the infrastructure robust, able to survive; guarantee adequate resources

General requirements for related *quality assurance* mechanisms are:

- Facilitate repeated and "fresh" exercises (e.g. EV, TF, TA) and new combinations of actors and levels
- Enhance and ensure professional quality of distributed SI production, including registration and accreditation of professional practitioners, and mechanisms to stimulate renewal

To highlight these qualitative transformations we propose to speak of an *architecture of new "infrastructures" for Distributed Intelligence*, to highlight the set of new institutional and organisational arrangements that are required (see figure 2).

Figure 2: Architecture of Distributed Strategic Intelligence (SI)



There are different types of central "brokering nodes": The first type corresponds to *"enabling" facilities*: The objective is to render results arrived at in one place, directly accessible in another without requiring direct contacts between actors in both places. One does not need to contact the promoter of the UK or Japanese TF exercises or the experts they have mobilized to get the results arrived at (e.g. about "ultragenic" approaches). The word "facility" is very important here, since it tells not only about the need for developing and maintaining/updating a bank, it also underlines the importance of "harmonisation" work that is understated in the possibility of entering results and thus facilitating their circulation: "compatibility" issues do not only raise technical "inter-operability" problems, they also raise issues about reliability, i.e. processes through which results are arrived at.

The task of the second type of node is to *facilitate direct contacts*. This is the process function of "technology assessments" or "consensus conferences" as in the "Biomat case" sketched above. One has not only to identify "assessments" or "conferences" which took place on the relevant subject and the compromises arrived at, but to take hold of the reasons why such actions arose, what argumentation was developed by stakeholders, and through what process "consensus" or "dissensus" was arrived at. In general, direct contacts between actors remain important.

The number of "policy arenas" and the variety of problems addressed pushes us to identify a third type of "central node" centred on processes and on the related combinations of instruments. The universe of situations is so vast that the idea of "best practices" does not seem appropriate; it is instead replaced by an idea of a *repertory of practices with indications how and why they functioned.* The repertory includes a growing number of strategic exercises over time, and there must be some structuring and winnowing, supporting *collective learning processes*.

The concept of *"quality assurance"* is used here in a broader sense than usual in quality management, and includes attempts of actors to ensure minimum levels of quality and the assurance, with relevant audiences, that there is some quality control. Thus, it relates directly to issues of trust: what trust can actors in policy arenas have in all the "intermediaries" that are mobilised for preparation or conduct of policymaking. Bottom-up processes are important here, in which professional institutions have a role to play. This includes (hopefully) the notion of "civic duty" to do a good job, as it works, for example, in peer review within science (even if there are exceptions). One could think, in addition, of evaluators, innovation bureaux and the like organising themselves in European professional associations.

5 Enhancing Distributed Intelligence for Innovation Policymaking on the European level

In trying to improve strategic innovation policymaking, one does not have to start from scratch. Current practices in most countries as well as on the EU level, however, have evolved in an uneven, random and fragmented fashion. Individual exercises have rarely been inter-linked either conceptually or politically with one another. This lack of coherent linkages has led to under-utilisation of existing information, knowledge and capabilities in the process of innovation policy formulation. In consequence, this has become a major obstacle to attempts at coherent policy design and practices. For instance, the task set up in the Maastricht Treaty to arrive at a co-ordinated European science, research and technology policy (including regional, national, and European levels) so far has not been fed by the systematic use of intelligence tools.

Given the set of existing institutions carrying out the functions of Distributed Intelligence for innovation policy, there is room for considerable improvement of the functioning along the following principles:

- Better co-ordination of TF, TA, EV along the policy-cycle within the Commission. The role of DG XII as a mediator between other parts of the Commission and national innovation policy actors could be strengthened
- Better co-operation between the Commission and the European Parliament in general and in TA in particular. A stronger role for the Parliament, especially with regard to TA
- Better assignment of tasks of the respective institutions, with a focus of EU institutions on information gathering, synthesising and preparation of policy decisions rather than carrying out the research tasks themselves
- The development and full use of the expertise of national institutions through commissioning, joint projects etc. is a necessary basis for any EU exercise in TF, TA and EV. Information exchange and regular mutual staff exchange between the different communities could be organised on the European level (e.g. in the form of (bi-)annual conferences on TF, TA, and EV)
- The development of interfaces between science and technology actors and the general public (e.g. as the Internet-based "Futur-Prozess", recently launched in Germany extending the Foresight experiences of the 1990s)

In the future, European innovation policies might put an increased emphasis on mission-orientation towards societal problems (while most diffusion-oriented programmes would remain in the domain of the member states), a tendency that has already been emerging with the FP5. In the longer run, new initiatives based on comprehensive considerations of needs and opportunities as well as impacts could be launched to complement current generic programmes and other schemes. This would entail more horizontal activities, and hence different forms of organisation of these activities - e.g. using the model of "task forces". Therefore, the above mentioned principles apply here as well: there would be a need for better co-operation and interaction between Commission Directorates and Member States' authorities dealing with science, technology and innovation policies, and in addition to top-down instruments, bottom-up initiatives must be generated and existing ones supported.

A world of distributed policymaking and Distributed Strategic Intelligence related to it takes into account the changing conditions of innovation processes, the political requirements for a democratic choice of future technologies, and a need to limit public expenditure linked to decisionmaking processes. Distribution means leaning on bottom-up processes, while in order to be effective and trustworthy, standards of quality and quality assurance systems need to be developed. In addition, of crucial importance will be central networking nodes, which facilitate horizontal linkages and the circulation of knowledge between different policy arenas and levels.

Round Table II: Mutual Learning in Evaluation of Science & Education

New Challenges for the Evolution of Academic Research Systems
Four Hypotheses on the Evaluation of Academic Research

David Campbell
Institute for Advanced Studies (IHS)
Stumpergasse 56, 1060 Vienna, Austria

Contemporary advanced societies commonly are described as "knowledge-based", implying that knowledge, know-how, and expertise should be regarded as important factors that determine to a large extent economic performance and economic competitiveness (European Commission, 1997). Crucial in such an understanding is the conviction that research or R&D (research and experimental development) should be regarded as core processes that are responsible for knowledge production. Furthermore, within the context of national (or supranational) research and innovation systems again academic research plays a pivotal role. The term academic research addresses university as well as university-related research. Referring to standardized OECD terminology, university research coincides with R&D that is performed by the higher education sector; and university-related research coincides with R&D being performed by the government and private non-profit sectors (thus "university-related" research can be regarded as a terminological equivalent to the German-speaking term of "ausseruniversitäre Forschung"). Academic research clearly represents a sciences-based and sciences-induced activity, where a major emphasis is placed on basic research and on the combination of basic and applied research. Since academic institutions and universities, in particular, also conduct tertiary teaching and education, academic research is closely associated with the development and build-up of highly qualified human capital (OECD, 1998a).

Focusing on current R&D funding trends of the advanced OECD countries, the following conclusions can be drawn (Felderer & Campbell, 1994; OECD, 1998b): First, when R&D expenditure is calculated as a percentage share of GDP (Gross Domestic Product) — the so-called "national research quotas" —, then R&D expenditure did not expand during the 1990s. This may appear to contradict the postulate of advancing "knowledge-based" societies. An alternative interpretation would be that R&D expenditure developed constantly, implying that the "how" question of R&D funding and research policy might be of a greater importance than the "how much" question of R&D funding. Second, the university performance share of R&D expenditure actually increased in the 1990s (particularly in the European Union and United States). Although university-related research slightly decreased, aggregated academic research successfully sustained its financial importance and weight, when compared with business R&D. This demonstrates that the role of sciences-induced university and academic research is upgraded within the context of national (and supranational) research and innovation systems, which again can be used as a justification for speaking of "knowledge-based" societies. Third, derived from the upgrading effect of university and academic research; more attention should be devoted to the structures and functionality of university and university-related institutions. Evaluations commonly are regarded as means and instruments that address such organizational issues of academia (Campbell & Felderer, 1997; Kuhlmann, 1998).

Evaluations help to optimize academic institutions; evaluations reinforce principles of accountability for academic institutions; and evaluations emphasize the application of explicit and rational criteria for decision-making and policy-making. In summary, evaluations aim at improving national (and supranational) academic research systems, so that they can deal effectively with new challenges that arise during the transformation process of industrialized countries towards advanced information societies.

In the following, four hypotheses are proposed for discussion. These hypotheses point at crucial issues in reference to the proper application of academic research evaluation and the further development and improvement of evaluations.

First Hypothesis on Academic Research Evaluation: **Combining evaluations of research with evaluations of teaching and education**

In the past, a well-defined division of competency existed that functionally separated the universities from the university-related ("ausseruniversitäre") institutions. Universities performed disciplinary-based scientific research, which was mainly basic research; and universities were responsible for conducting tertiary teaching and tertiary education. University-related institutions, in contrast, expressed only a minor competence in tertiary teaching; the research profile of university-related institutions, however, was more application-oriented and devoted to a practical and interdisciplinary problem-solving. During the next years we may witness a gradually increasing functional overlapping between university and university-related institutions. Universities must place a greater emphasis on interdisciplinarity and application-oriented research. University-related institutions, on the other hand, will engage more actively in tertiary teaching and education, implying that university-related organizations simultaneously will perform research and teaching. As a final consequence a growing number of academic institutions will do both, research and teaching, leading also to an intensification of cooperation between university and university-related institutions. Parallel to the functional overlapping between universities and university-related institutions one can also expect the institutional diversification process within the university and university-related sectors to progress.

Once having accepted that evaluations are necessary, *the emerging dual performance mode of research and teaching of academic institutions creates a demand for combining the evaluation of research with evaluations of teaching (and education), since the achievements of academic institutions should be reflected in their functional complexity*. Furthermore, one can postulate that the basic methodological principles and the primary evaluation dimensions of research and teaching are comparable. Two basic methodological principles are: first, peer review, and, second, indicator-oriented strategies. Commonly, the following dimensions of evaluations are classified as primary dimensions: quality; efficiency; and relevance; effectiveness, again, can be modeled in reference to quality, efficiency, and relevance. However, the practical evaluation procedures for research and teaching (education) are separate and distinct. Thus it represents a major challenge to design systematic interfaces and linkages between research and teaching (education) evaluations, for the purpose of comprehensively reflecting the performance of academic institutions. This also refers to the question of

how to use collaboratively results of research and teaching (education) evaluations for a decision-making focus on structural reform and the allocation of resources.

Second Hypothesis on Academic Research Evaluation: **Reconciling the differences of cultural pluralism versus the functional needs of advanced societies**

Evaluation systems of academic research must be open to and receptive for an immanent learning process. One possibility for fostering such a learning process is carefully to compare evaluation systems of different countries, since this enables one to derive or to extract those international examples that successfully can be implemented within the domestic system. One question to be answered, of course, is how to interpret differences in evaluation systems across countries. Clearly, several answers are possible. One approach would be to stress cultural differences: evaluation systems of academic research are different so that they can reflect cultural differences. An alternative approach is to emphasize different functional needs of society, depending on its stages of development: consequently, evaluation systems are different because societies — or specific sectors or subsystems of society — have progressed further or are still lagging behind. Within such a conceptualization differences in evaluation systems reflect maturity differences of academic research systems and also of academic research evaluation.

With regard to the implementation and application of academic research evaluation, *it appears necessary to refer to cultural pluralism as well as to the functional needs of advanced societies, and to reconcile both for the purpose of developing proper evaluation strategies*. There are different national cultures and also different national academic cultures, and they must be recognized by particular evaluation systems. On the other hand, one should be careful not to regress all differences to cultural differences and by this to exhaust culture as a "supervariable" that should explain everything, since this would lead to an undecidable "cultural relativism" (particularly in a multi-cultural and multi-lingual setting, as represented by Europe, this argument is important). Functional demands of advancing and of advanced societes should be taken seriously and perhaps the academic evaluation systems in the United Kingdom and the Netherlands have already progressed further than academic evaluation systems in Germany, Switzerland, and in Finland. However, such developmental approaches — that focus on the concept of advanced societies — may not ignore cultural differences and, furthermore, ought to recognize the implications of path dependency.

Third Hypothesis on Academic Research Evaluation: **Emphasizing the linkages between academic research and the societal environment**

Derived from the observation that the input of R&D funding and also of academic research funding did not increase dramatically during the 1990s, we conclude that the "how" challenge of academic research gains in importance. A team of researchers, under the guidance of Michael Gibbons, attempted to define those principles that determine advanced knowledge production and, consequently, successful academic research. They ar-

rive at a categorization of five key principles, which they summarize under the term "Mode 2" (Gibbons *et al*., 1994): knowledge produced in the context of application; transdisciplinarity; heterogeneity and organizational diversity; social accountability and reflexivity; and quality control. Two core ideas of those principles are: first, that the sustained development of basic research demands the permanent reference of an application-oriented context. Second, research locations diffuse massively across society and many research locations emphasize the concept of research networks. The shortening "time horizons" of (business) innovation cycles and of the applicability of information create the following paradoxical situation: on the one hand, there is a demand for rapidly channeling research results into applications (since time lags outdate the usability of information). On the other hand, research with a bias towards the application end can only improve, but not innovate, and it is primarily basic research that emphasizes the long-term perspectives that are so crucial for knowledge-based societies. This increasing importance of basic research clearly upgrades the significance of university and of academic research and also helps explaining, why university and university-related institutions express such a pivotal functionality for advanced societies.

Therefore, the crucial message seems to be *that linkages should be emphasized that connect academic research with society or the "societal environment",* when society is interpreted as the overall context that embodies the national academic research system. Three issues appear to be important. First, basic and applied research ought to be designed as parallel processes that are connected through interactive linkages (the old concept understood basic and applied research as two processes that were coupled through a sequential "first and then" relationship). Second, networks should be regarded as a structural type of research that will play an increasingly important role for the national research and innovation systems. Since networks connect different research organizations across sectors, that means university and university-related institutions with companies, academic research results can diffuse instantly and focus more precisely on potential user groups. As a by-product of such networking, multiple cross-fertilizations between basic and applied and between disciplinary and interdisciplinary (transdisciplinary) research are also supported. Of course, there is also the challenge of reconciling these network processes with the principles of a fair competitiveness. Third, in the context of academic research evaluations the "relevance of research" should be treated as a crucial dimension. As already mentioned earlier, quite often three dimensions are considered as appropriate for evaluations: quality, efficiency, and relevance (effectiveness, a possible "fourth" dimension, can be modeled upon the other three dimensions). One possibility for operationalizing or indicating relevance is to assess the linkage patterns and frequencies of academic research with its societal environment. When "relevance" is regarded as a crucial dimension for evaluating academic research, this can create the amplifying (and also intended) effect, that academic research communities deal more deliberately with relevance issues.

Fourth Hypothesis on Academic Research Evaluation: **Reinforcing the evolution of academic research systems by applying comprehensive evaluations**

Academic research and academic research systems must change and demonstrate flexibility, so that they can adapt reactively and proactively to new conditions and demands that arise when societies develop and progress. The research intensity of advanced knowledge-based societies demands processes of a permanent remodeling of university and university-related institutions. We already mentioned the challenge of interactively linking basic and applied research in a "parallel" design. *Thus there operates an evolution of research systems and of academic research systems* that can be observed empirically and which also must be demanded normatively, if academic research focuses on sustaining its importance and competitiveness. One pivotal question, of course, is: How should the relationship of evolution and evaluation – within an academic context – be conceptualized? In that respect there are two crucial ideas. First, *the evolution of academic research systems is reinforced and supported by the comprehensive application of evaluations*. Evaluations should be regarded as a proper instrument for supporting the evolutionary adaptation processes of academic research systems, because evaluations generate a systematic feedback on the quality (and efficiency and relevance) of academic research (see also the reference in the "acknowledgment"). Second, *evaluations are exposed to an evolution of evaluation*. Since academic research systems change, obviously also the evaluations themselves, that assess academic research, must change: evaluations, therefore, ought to demonstrate an openness for self-reflexive learning. Such a permanently ongoing "meta-evaluation" of evaluations may be conceptualized as an evolutionary process, emphasizing cross-referential interactions between the evolution of academic research systems and the evolution of evaluation systems of academic research. This still leaves the question unanswered, whether or not evaluations should be regarded as internal or external components or structural elements of academic research systems (both conceptualizations are possible). Therefore, we may conclude: *evaluation processes, that express a receptiveness for an evolution of evaluations, markedly support the evolution — and the evolutionary capabilities — of academic research systems*. This might point at a co-evolution of academic research and academic research evaluation.

Acknowledgment
The idea of fostering the evolution of academic research systems by applying evaluations ("evolution through evaluation") was brilliantly elaborated by Dr. Wilhelm Krull, Secretary General of the Volkswagen Foundation, in the context of a highly stimulating conversation in Berlin on the evening of June 7, 1999, during the conference "Evaluation of Science and Technology in the New Europe".

References

Campbell, David F.J. / Bernhard *Felderer* (1997). Evaluating academic research in Germany. Patterns and policies. Vienna (Institute for Advanced Studies): Political Science Series No. 48 (library@ihs.ac.at).

European Commission (1997). Second European report on science and technology indicators 1997. Brussels (ISBN 92-828-0271-X).

Felderer, Bernhard / David F.J. *Campbell* (1994). Forschungsfinanzierung in Europa. Trends, Modelle, Empfehlungen für Österreich. Vienna: Manz Verlag

(ISBN 3-214-05998-X).

Gibbons, Michael / Camille *Limoges* / Helga *Nowotny* / Simon *Schwartzman* / Peter *Scott* / Martin *Trow* (1994). The new production of knowledge. The dynamics of science and research in contemporary societies. London: SAGE Publications (ISBN 0-8039-7794-8).

Kuhlmann, Stefan (1998). Politikmoderation. Evaluationsverfahren in der Forschungs- und Technologiepolitik. Baden-Baden: Nomos Verlagsgesellschaft

(ISBN 3-7890-5534-4).

OECD (1998a). Education at a glance. OECD indicators 1998. Paris (ISBN 92-64-16127-9).

OECD (1998b). Main science and technology indicators. Paris (ISBN 92-64-05745-5).

Main Points of Intervention at Round Table II

Inge Knudsen
Confederation of EU Rector's Conferences
Aarlenstraat 39-41, 1000 Brussels, Belgium

Research evaluation is the oldest form of evaluation; it is normally based on disciplines; funding is mainly channelled through Research Council; it is often bottom-up in organisation (except in countries where it is linked directly to national research funding).

Teaching evaluation (large was + 4x introduced as 4x a result of mass education; it is often included in the legal framework of higher education (HE) institutions; it is often done per faculty or per department; it is mainly centrally organised (via agencies or ministries directly); it is most often top-down in organisation.

Both types of evaluation are organised nationally, but research evaluation includes international peers, something only slowly being introduced in teaching evaluation. Research evaluation has been accepted from the start, also as part of the prestige linked to research, whereas teaching evaluation has less "attraction" linked to the less prestigious activity of teaching.

Mass education has made the examination process, normally seen as the sole evaluation needed in teaching, insufficient as evaluation and assurance of quality levels in HE. Evaluation of teaching has gradually become a national "must" in most countries.

Institutional management evaluations are emerging as alternative evaluation measures chosen by some HE institutions, i.e. an evaluation of all activities, including research, teaching and administration/information/counselling, always done by international teams/peers. Institutional management evaluations normally take into account evaluations done in research and in teaching in the overall assessment of an entire institution. This type of evaluation gets increasing attention as it assists individual institutions in their aim to increase their standing at global level. Institutional management evaluations are most often voluntary, based on the express wish of the individual institutions. It is mainly organised by international interest groups - examples are the EQUIS accreditation scheme established by EFMD (business and management schools) and the institutional management scheme of the Association of European Universities.

In times where evaluation measures taken at national level, be they evaluations of research or of teaching, have a tendency to substitute evaluation with accounting, many higher education institutions opt for the individual solution of institutional management evaluations to establish themselves in a global setting. Evaluations at national level are no longer seen as evaluations of quality, linked to assessment and assurance of quality, but mere schemes to pinpoint the position of individual institutions on a national budget line.

Mutual Learning: Evaluation of Science & Education

Sergio Machado dos Santos
Vice-President of the Confederation, Universidade de Minho
Largo des Paco, 4719 Braga Codex, Portugal

I will try to answer the four questions put to the panel by giving a brief presentation of the Portuguese evaluation systems. It may be indeed an interesting case study, since the Portuguese universities are developing very fast a culture for quality and for evaluation.

There are two separate systems for the evaluation of education and of research. Although they are very different, both systems are based on the principles of openness and transparency, of using peer review and of introducing international standards by bringing foreign experts into the review teams.

Evaluation of teaching

The evaluation of teaching in the Portuguese public universities was started in 1993, as a pilot experiment launched by the Rectors' Conference, covering four fields of study in all the universities. This bottom-up approach influenced in a decisive way the law passed by the Parliament in late 1994 (Law n° 38/94).

The Portuguese evaluation system closely followed the Dutch model, namely, aiming to improve quality, introducing positive discrimination to help to overcome weak points detected by the evaluation. The evaluation process provides no ranking of universities and has no direct links to funding. It is periodical, on a 5-years cycle.

The system is decentralised and relies heavily on self-evaluation reports prepared by the universities accordingly to common guidelines. The reports are analysed by peer review teams, who visit the institution for 2 to 3 days. Each team usually includes an expert from abroad and a representative of a professional association or of the labour market.

The coordination of the external evaluation is the responsibility of Evaluation Councils, accredited by the Ministry for Education on a protocol basis. There are four Evaluation Councils, one for each sector of higher education (public universities, public polytechnics, private universities, private polytechnics).

The global coordination of the evaluation process, aiming at safeguarding the systems' coherence and credibility, belongs to the National Council for the Evaluation of Higher Education, constituted by representatives from the government (6, including 2 from abroad), from the higher education sectors and Evaluation Councils (7) and from the students (3). The President of the National Council is nominated by the Council of Ministers. The Council is responsible for the approval of the guidelines for the self-evaluation reports and the approval of the composition of the external review teams, for producing overall reports and for the metaevaluation of the evaluation system.

The reports produced by the review teams are public and usually attract great attention from the press.

The self-evaluation process has already had an important impact within the institutions, which became aware of problems that they had not thought about before. External reports are recent and the institutions are now working on them. It is felt that for the public credibility at the evaluation system it is necessary to give visibility to the results and impact of the evaluation.

A weak point of the evaluation system is its emphasis on undergraduate programmes, with little attention being paid to post-graduation programmes and to research, namely the research activities seen as a support for a research-based education. It is expected that some adaptation of the guidelines will be made for the second cycle of evaluation, to start next year, in order to cover the above mentioned points.

Evaluation of research

The main source of funding for the research activities is the funding of projects through open competition.

In Portugal there is a second source of public funding - the plurianual funding programme of research units - designed as a complement to the funding of projects, in order to provide support to the operation of the units, at a low but stable level, independently of the fluctuations in the funding though open competition.

In connection with the plurianual funding programmes there is a system for the evaluation of the research units. It follows a top-down centralised approach and is run by the Minister for Science and Technology.

The evaluation is periodical, on a 3-years cycle started in 1996. The second cycle is already in operation.

The assessment of the research units is done by panels of foreign experts coordinated by a Portuguese member. Each unit provides a report of activities covering the period of 3 years and a plan of activities for the next 3 years, which are assessed individually by each member of the panel. There are on-site visits and a final report from the team, which is public.

The units are ranked in a 5 grade scale - Excellent, Very Good, Good, Fair, Poor and the ranking links to 3 levels of basic funding or no funding at all: Excellent and Very Good units get the maximum level, on a per capita basis; Good units get 516 of the maximum values and Fair units receive 213 of that value. Units ranked as Poor have their financing discontinued.

Each panel is also required to produce an overall report about the state of R&D in their particular scientific field.

The acceptance of this evaluation system is quite good both by the units and the researchers. The fact that the evaluators come from abroad gives credibility and independence to the process and introduces international standards. It may however raise some problems, because sometimes it is difficult for the experts to grasp local or specific contexts, but the principle is positive.

There are some negative points, which will hopefully be solved in the new cycle being initiated:

- There is no linkage to education, namely to post-graduation programmes, which is an important point for universities.
- The units are evaluated as independent entities, without regard for the institutional research policies that some universities are already developing. In particular, the fact that the training of staff is a priority for some new and fast growing institutions must be taken into account.

Anyway, this evaluation exercise is globally positive and represents a very good first step for the consolidation of a culture of quality.

Mutual learning and relation to society

The few minutes available for the presentation only allows me two final points:

- The two evaluation systems are operating in an isolated way and they could learn a lot from each other. It is necessary to establish some links between them and profit from the past mutual experiences.
- The National Council for the Evaluation of Higher Education will play an important role in connection with the relations to relevant partners and to society. The Council is creating standing committees with representatives of the professional associations, which will be helpful in acquiring awareness for the evaluation processes within society.

Evaluation and Accreditation of Science and Higher Education in Slovakia

Pavol Návrat
Vice-Chairman of the Accreditation Commission of Slovakia
Slovak University of Technology
Ilkovicova 3, 812 19 Bratislava, Slovakia

Accreditation Commission

- counselling body of the Government of the Republic of Slovakia (established by law from 1990)
- responsible for
  - accreditation of degree courses of degree-awarding institutions
  - evaluation of higher education institutions
  - expressing opinion on proposals of merging, closing, and establishing of higher education institutions

Monitoring and Evaluation

- period: annual - biannual?
- level: individual, institutional, system
- what is most important?:
  - scientific quality of faculty members
  - educational outcome (how to measure it?)

Monitoring and Evaluation

- method:
  - peer review
- panel of peer reviewers
- on-site visits
  - institution submits information:
    - self-report
    - faculty, its structure
    - international collaboration
    - financing, infrastructure

Evaluation of the Faculty - the Criteria

- extensive
  - monographs
  - scientific papers (international, SCI)
  - patents
- quality
  - basic research:
    - citations, an average impact factor
  - technical research and development

- patents exploited in industry
- unique projects

Accreditation

- aims
  - to define and enforce minimum standards of quality
    - defining:
      - provides a substrate for discussing among relevant institutions
    - enforcing:
      - courses accredited only when meeting the standard degree criteria

Accreditation - Scope

- standards of quality of:
  - a bachelor degree
  - a master degree
  - a doctoral degree
  - a right to nominate
    - an associate professor
    - a full professor

Evaluation of Science and Higher Education Institutions

- Research institutes offering PhD courses are included in the process along with universities - same commission, same criteria, same procedure
- differences:
  - Academy of science institutes
    - stress almost solely on quantitative criteria (e.g., citations)
  - Higher education institutions
    - more comprehensive (educational process)

Evaluation of Science and Higher Education Institutions - Lessons

- What can be learned from each other?
  - From Academy of Science institutes
    - stress on quantitative criteria - transparency, quality is „operational“
  - Higher education institutions
    - more comprehensive - includes various links with society (students, graduates, industry).

General Conclusions

Concluding Remarks I

Manfred Horvat
Bureau for International Research and Technology Cooperation
Wiedner Hauptstr. 76, 1040 Vienna, Austria

Introductory remarks

At the start, it seems appropriate to put EU RTD evaluation into the context of other evaluation approaches.

The discussions on evaluation in the framework of EU RTD activities are very often dominated by experiences stemming from the self-regulating system of curiosity or understanding driven science and research. There, certain approaches to evaluation are applied for (internal) quality assurance and development. Peer review of publications, ex ante evaluation of research proposals by funding organisations, ex ante and ex post evaluations of institutions are well-known measures following accepted standard procedures.

In contrast, EU research is policy-driven and programme-oriented following well-defined strategic objectives like

- strengthening competitiveness of European industry, with a specific emphasis on SMEs,
- supporting other Community policies (environment and sustainable development, energy, employment),
- fostering co-operation between industry and academia.

A legal framework is set by the Treaty for the European Union, by the decisions on the Framework Programme and the Rules for Participation. In the Framework Programme decision the criteria for selecting objectives and themes for RTD are clearly defined. The Specific Programmes are implemented via yearly Work Programmes that are defined by the Commission services with the help of independent External Advisory Groups and the Programme Management Committees. The selection of proposals takes place via Calls for Proposals and evaluation exercises involving independent experts.

A holistic set of criteria is applied for the evaluation of individual proposals:

- scientific quality and innovation,
- European added value and contribution to other EU policies,
- contribution to Community social objectives,
- economic and technological perspectives,
- management and resources.

At the programme level relevance, efficiency and effectiveness are the guiding criteria when programme success and impact are measured and assessed in evaluation exercises at Framework and Specific Programme level.

Often, some concerns are formulated by the scientific community, that in EU RTD evaluation procedures scientific and technological excellence are not sufficiently

recognised, because "additional" criteria are applied, too. However, if science and research are interpreted as human action, the general and specific objectives, the scientific contents and the innovative aspects, but also the social process towards achieving the goals – the management - are important and well justified elements of the overall process and have to be taken into account in evaluation. It may well be, that the evaluation approach designed and applied by the European Commission makes aspects explicit that in other approaches are taken into account implicitly only. However, due to the complexity of transnational collaborative projects and programmes, it is only justified to follow such a holistic approach. In contrary, it may be worthwhile considering what other evaluation systems could learn from the EU approach.

Actors and objects in evaluation exercises

The scenery is characterised by different direct and indirect actors in the evaluation process:

- the European Commission and the European Parliament,
- the Commission services,
- different advisory bodies,
- the Programme Committees,
- the External Advisory Groups,
- other policy-makers and stake-holders,
- disciplinary experts,
- evaluation experts and institutes.

There are different objects of evaluation

- the whole RTD system – at European and national level,
- the RTD portfolio of the European Union,
- individual institutions, departments, etc.,
- the individual researcher,
- RTD proposals and projects.

There are different activities to be distinguished: policy evaluation, monitoring and assessment of programmes and evaluation of project proposals. From the point of view of internal controlling activities, the focus is on quality of management, accountability, legitimacy and justification for using tax payers' money, appropriateness of allocation of resources. With respect to ensuring that the EU RTD system develops into a learning organisation quality assurance – especially in times of structural change, relevance, efficiency and effectiveness are well justified yardsticks for assessing performance and impact of EU RTD activities.

The evaluation system provides feedback to all actors, the basis for continuous reflection on performance, as well as policy advice. The interaction between expert panels and stake holders is an important asset of the process. It supports piecemeal adaptations of programme objectives, contents and procedures as well as fundamental changes when new programmes are being prepared.

Conclusions on methodologies and approaches – some basic requirements and recommendations

From the presentations and discussions one can distil the following conditions for the further development of the RTD evaluation systems at European and national level:

- compatibility without destroying diversity;
- appropriateness to the context – be it national our European – including feasibility and usability;
- acceptability of evaluation approaches both for the European institutions, the governments and the science and research community:
  - you will not improve your system if you don't succeed in achieving collaboration and commitment of those that are evaluated;
  - if there are no – positive or negative - consequences following from evaluation results, no evaluation culture will evolve;
- the general transferability of "good practices" has to be questioned; strategic decisions have to be taken what you want to learn from whom.

Finally, some recommendations can be made, especially with regard to EU enlargement in the area of RTD. It is essential to get acquainted with each other. First steps urgently to be taken will be to integrate the new partners into the CREST evaluation sub-committee, the European RTD Evaluation Network, and the pools of experts for the Five Years Assessment and the Yearly Monitoring Exercises, as well as the project evaluation procedures.

There are a number of necessary activities that should be organised to ensure a positive further development in a new collaborative setting involving the colleagues from the newly associated countries:

(1) Networking

- Supporting the dialogue between policy-makers and evaluation experts (from evaluation panels and expert institutions):
  - ➢ organise a conference every two years!

(2) Awareness, information, training

- Developing the understanding of the specific role of RTD evaluation in RTD management:
  - offer training courses on specific issues of RTD evaluation;
  - develop guidelines on procedures and indicators;
  - integrate research management and evaluation in curricula of engineers and scientists!

(3) Management

- Supporting and coaching for evaluation exercises:
  - provide external experts;
  - provide methodological advice!

(4) Methods

- Supporting methodological development and piecemeal or incremental improvement:
  - develop approaches to meta-evaluation – to evaluation evaluations;
  - utilise experience accumulated by panel members in FP4 together with expertise of specialised institutes;
  - improve understanding of quantitative and qualitative aspects in evaluation;
  - develop approaches to distributed intelligence encompassing technology foresight, monitoring and evaluation, and technology assessment;
  - support the development of evaluation practices as an integrated part a learning system for the benefit of European RTD activities.

(5) Finances

- Making financial provisions for RTD evaluation at European level:
  - ensure a dedicated budget for monitoring and evaluation;
  - follow a pragmatic approach for FP5 following a "rule of the thumb";
  - start preparing the ground for FP6.

(6) The role of the European Commission

- Acting as a facilitator, moderator, supporter – in line with modern politics:
  - nurture a data base of experts;
  - organise expert advice;
  - support conferences, workshops and studies.

(7) Follow-up

- Keeping the momentum:
  - keep the participants of the Conference informed on impact of the meeting;
  - go for the next steps and implement (at least some) of the recommendations – and report on that;
  - nurture the mailing list and inform participants on relevant events and further developments in
  - the area of RTD evaluation, to offer opportunities to meet and stay in contact.

There was a common understanding that evaluation is both necessary and extremely useful for supporting a promising development of EU RTD activities in an enlarged science and research area.

Concluding Remarks II

Enric Banda
European Science Foundation
1 quai Lezay-Marnesia, 67080 Strasbourg Cedex, France

After having attended the meeting from the very start I will try to communicate my feelings and understanding of what was presented and discussed based on notes taken on the spot. Therefore my views can only be taken as personal and in no way a formal list of agreed conclusions.

Scientific research and technological development (R+TD) is now considered not only of intrinsic merit in itself but as indispensable for improving our quality of life and as an engine for economic growth. It relies heavily on public patronage. Therefore, the spending of public funds in R+TD must be justified, if and only if quality is secured and the public can be assured of this. One obvious way to introduce quality in the system is to submit it to assessment and evaluation. Such an evaluation should cover the different stages from policy-making and the setting of priorities to rigorous evaluation of project proposals themselves. Quality must be the underlining principle and will naturally improve the efficiency of the system.

A call for a single European space for science and technology was advocated by some speakers, following the lines already marked by the previous Commissioner Antonio Ruberti and several others. This view should be in the background to the development of European Science and Technology efforts. In turn, this may call for using RTD indicators in a way similar to other macroeconomic indicators that have been successfully used in developing policies to promote convergence of European (EU) economies.

Most of the presentations dealing directly with the procedure of evaluation and restructuring of R+TD systems (including university systems) have shown the extreme complexity of the procedure. Criteria based on trends of globalisation, multidisciplinarity and new funding methods respond to good practise in evaluation, in spite of the attached complexity. The situation is complicated still further by the different evaluation cultures which have been identified. It became obvious at the conference that different countries have different perceptions of the evaluation process and its use for cultural, historical as well as practical reasons. I believe that participants welcomed the aim of keeping this diversity, which corresponds to different cultures, although efforts towards a certain level of homogenisation and understanding will be needed at the European level. A possible European evaluation culture can be built, but only on the basis of diversity and flexibility. However, there is an urgent need for continued international comparison of evaluation procedures and this was apparent from the presentations.

As far as the Fifth Framework Programme is concerned, the perception of the participants was of an increasing openness and desire for improvement in all aspects of the Programme by all concerned. It was felt that the FP5, as the most important transnational

mechanism for funding R+D in Europe, must become instrumental in helping to create an overall evaluation culture.

Stress was put on the convenience of combining bottom-up and top-down procedures. In other words, a balance between scientists and policy managers is necessary in evaluation procedures. Similarly, mutual learning from procedures used by different countries, and notably between eastern and western countries, could only help in the process. Again diversity was felt as a strength if managed properly. New procedures must be built, based on strengths of existing systems in which confidence is proven, with enhanced dialogue is the best recipe towards integration.

It was repeatedly emphasised that evaluation is an activity that requires the recognition that it is an essential overhead and that adequate resources are necessary. However, the budget would be better placed away from decision-makers themselves in order to assure an independence of the procedure.

The evaluation debate turned often to enlargement of the EU and benefits expected from accession states in terms of their participation in the FP. Although it may be politically difficult to digest, the return in strict economic terms can only be expected in the medium and long term. Intangible benefits, business opportunities and protection of the individual R&D systems need to be taken into account in the first stages of accession and in the evaluation of performance.

Finally, there was a strong consensus among participants about the usefulness of the meeting, especially in providing a network between evaluation practitioners and a comparison of procedures. Therefore, there was encouragement for follow-up action to be indicated, perhaps with rather more specific objectives, in which the European Commission should act as a facilitator. In my opinion, this was one of the important conclusions of the conference.

Closing of Conference

Evaluation of Science and Technology in the New Europe: Towards a European Evaluation Culture

Gilbert Fayl
European Commission, DG XII, Science Research and Development
De Meeûs Square 8, 1050 Brussels, Belgium

The exchanges of ideas and experiences enabled by this first meeting on ”S&T evaluation in the new Europe” demonstrated the importance of working together. The 15 countries currently constituting the European Union understood for years the benefit they can get from this kind of collaboration and now realise that it is time to enlarge the circle and in particular to those aspiring to become the future members of the Union.

Differences in the culture, differences in the structures, policy objectives but also in implementation tools, the 26 countries represented at this Conference are in a position to give lessons and to learn from each other. Everybody realises the potential of such a forum where, free of any sort of constraints, the participants are all animated by their eagerness to do better and to do it with the help of others.

At a time where science is in the centre of both major hopes and fears, evaluation of its impact on society and economy, of the efficiency and effectiveness of the publicly funded RTD, becomes more than ever a critical issue. This importance is, in addition, reinforced by the necessity to cope with a relative shortage of public funds that calls for more concrete evidence of RTD policies' usefulness. As the environment of all European S&T policies becomes simultaneously more globalised and competitive, the evaluation tools and schemes should be accordingly adapted.

While being all different and rich of our own history and culture, we all recognise that we share common values. In order to address the challenges ahead, it becomes obvious that we must take advantage of this wealth and progressively build a European evaluation culture that is based on our commonalties, combines our strengths and minimises our weaknesses. This implies to further promote information exchanges, constructive dialogue and to identify best practices that could, subject to adaptation to local conditions, be successfully applied elsewhere.

As we progress stepwise towards a more integrated Europe, combining our human, financial and political resources, we should in parallel work side-by-side on the best ways to better measure our achievements in order to take appropriate corrective actions. The universal nature of science will certainly, in this respect, help us not to be submitted to our differences. We from the side of the Commission are ready to act as a catalysing factor, try to dynamise and nurture the dialogue and links yet established notably by promoting events like this conference. We have identified possible follow-up events to this meeting and we encourage you to continue your participation in the European dialogue in this important field. In this spirit, while thanking you for your participation and valuable contribution, I close the Conference.

Further Contributions

The New Face of Science and R&D Evaluation in the Academy of Sciences of the Czech Republic

Ladislav Pivec
Academy Appraisal Commission, Academy of Sciences of the Czech Republic
Národni 3, 117 20 Prague 1, Czech Republic

Science evaluation at the Academy of Sciences of the Czech Republic has become an integral part of its transformation into an effective, conceptual and democratic science complex. The basic principles and methods of science evaluation of the Academy's 59 research institutes during the 1994-1996 cycle (references, Pivec, 1998, Jenik, 1999) revealed the following:

- independence of evaluators from the evaluated body
- independence of evaluators from the decision-makers and the budget allocators
- international cooperation (47 percent of the evaluators were from abroad)
- peer review ("merit review," A. Teich, 1999) as a basic evaluation method
- multidisciplinary methodical approach to the natural and technical sciences, and humanities.

The Evaluation Committees discussed a number of problems which dealt with harmonizing the differences in evaluation procedures between the natural sciences and humanities, and the limitations of applying quantitative methods (scientometrics) equally in the different scientific disciplines. The special contribution of the humanities to national cultural values was also emphasized as the criterion for evaluation. Much effort was put into the selection of reviewers, and appropriate professional balance was achieved with the domestic staff (mostly invited from non-academic institutions) so as to cover all research topics of the evaluated institute.

General conclusions made from this evaluation course were very positive in documenting the important role the Academy of Sciences of the CR plays in the domestic and international science community. The most frequent recommendations of the Evaluation Committees reflect the present situation where basic research in the Czech Republic (mostly carried out by the Academy) is not yet fully integrated within other cooperating partners, i.e., universities, R&D institutions and industrial enterprises, dubbed as the "Triple Helix Model" of innovation by H. Etzkowitz (1999). The recommendations include:

- to apply for financial support for long-term projects of national and international importance,
- to increase cooperation with research institutions from abroad on joint projects and grants,
- to increase cooperation among Academy institutes,
- to stimulate cooperation between individual research groups within Academy institutes,

- to achieve better conceptual cooperation between the Academy and universities for the solution of common projects and the education of the upcoming scientific generation,
- to contribute to the creation of a new type of interdisciplinary system for the diffusion of the results of basic research into the industrial and social spheres,
- to stimulate dialogue between scientists and society so as to publicize the importance of basic research on the national and international levels,
- to make urgent efforts to obtain additional sponsorship for research from Czech industry and from abroad.

The last two points mentioned and up to date innovation R&D world projects challenge the Academy of Sciences to adopt a new scheme of science and R&D evaluation in the Czech Republic. This scheme is based on scientific fields and problem-oriented evaluation among various institutions (the Academy, universities, research institutes) which can assist them in obtaining institutional funding. To the present time funding for the Academy has been diversified, resulting in a high number of small short-term grants and insufficient support of long-term conceptual projects and centers of excellence.

This new approach to science and R&D evaluation on a broad national scale brings a great responsibility to achieve objective and applicable results. Objectivity needs to be retained across the whole spectrum of evaluation to improve its quality. Inter-institution cooperation between evaluators and those evaluated would be mutually beneficial in promoting contact among all partners involved in world-wide R&D innovation policy. We feel that the new system applied on a broad national scale should build on the experience of the previous evaluation system applied at the Academy. We are concerned that an evaluation which is based only on written materials is insufficient to comprehend the total picture. Personal on-site visits by evaluators are crucial to understand all consequences of a given evaluation result. Since the domestic staff of evaluators is always limited, not to mention occasional conflicts of interest, more effort should be made to attract evaluators from abroad with adequate financial support. We hope that the good scientific and research potential of the Academy of Sciences of the Czech Republic will not only be more fully realized but also make a significant contribution in the “Triple Helix Model“.

There can be no stability in science and R&D when only two of the three partners of the above model are involved. If the government - most important as the third partner in this “game“ - is not actively involved, science and R&D are handicapped in their effort to serve society. We believe the government of the Czech Republic will soon be in a position to be doing more to help science and R&D to serve national and international responsibilities, primarily in the following ways:

- to formulate new policy for science and R&D
- to balance national and international priorities for science and R&D
- to harmonize legislation with the new policy
- to revise legislation which will exert more control mechanisms to subdue loss of taxes, money laundering, tunneling etc.

With the above changes, a greater and more conceptual allocation of money could be made for science and R&D and for the continuing transformation of the Czech Republic.

References

L. Pivec, Basic Research and the Transformation Process in the Czech Republic, NATO Series 4: Science and Technology Policy-Vol. 22, pp 27, (1998)

J. Jeník, Quality Assessment of Research in a Multi-spectral National Institution, NATO Series 4: Science and Technology Policy - Vol.28, pp 138, (1999)

E. Etzkowitz, The Endless Transition: A "Triple Helix" of University-Industry-Government Relations, Workshop on Transformation of Innovation Systems, Budapest 24-25 June 1999

A. Teich, How R&D Budget Decisions Are Made in the United States, Workshop on Transformation of Innovation Systems, Budapest, 24-25 June 1999

Evaluation of Science and Technology in Slovenia: Present and Future

Dasa Bole-Kosmac
Ministry of Science and Technology, Counsllor to the Goverment
Trg OF 13, 1000 Ljubljana, Slovenia

When preparing my contribution for this conference, I was faced with a dilemma as to whether it would be more fruitful to focus on one or two narrow issues of science and technology (S&T) evaluation practices, or to use the opportunity of this international gathering to present a general picture of S&T evaluation in Slovenia.

The Berlin conference is devoted to the issues of East-West co-operation and mutual learning as a means of upgrading, with joint efforts, S&T evaluation in Europe. Since knowledge of each other as potential partners is a basic prerequisite in this process, I chose to do the less attractive yet necessary job and introduce the Slovenian S&T evaluation »landscape« to the broader European public.

Fields of evaluation

So far, Slovenia has not managed to develop a network to comprehensively evaluate national developments and policy in the field of S&T. Nevertheless, several partial measurements and evaluations are underway or being developed.

The Slovenian Ministry of Science and Technology (hereafter the MST) reports annually to the government and to the National Assembly which provides an important opportunity - together with budgetary debates in the Assembly - for the broadest assessment of the national policy in the field of S&T.

Regular ex-ante evaluation of basic, applied and experimental *research projects* has been practised for decades, while foreign evaluators have taken part in the process since 1992. In 1995, ex-post evaluation of applied and experimental research projects was introduced as the beginning of a permanent and more comprehensive evaluation practice for the future. The procedures for both types of evaluation have been fairly well elaborated. The same is true for the assessment of *individual researchers and research groups* (as a part of the criteria for funding research projects).

Annual programmes and work reports of government *research organisations* have also been collected by the MST on a regular basis, but their integrated analysis is only now becoming part of practice. This year, in-depth analysis of research organisations in the government and higher education sector is performed on the basis of data collected through a special questionnaire.

Besides these regular evaluations, other elements of the national S&T system and related policy are evaluated on an occasional basis (see, e.g., Stanovnik 1998, or a collaborative project ISI & IER, 1999). For a better idea of what is currently analysed and evaluated in

Slovenia, let me mention four more projects now in their concluding phase which evaluate R&D activities in the higher education sector (Dr. Sonja Kump et al.), the MST's existing system for evaluating research projects (Dr. Franc Mali), the Slovenian technology policy (Dr. Lojze Socan et al.), and policy related to technologies implying ethical problems and risks (Dr. Andrej Kirn et al.). I will not cite these projects here since they are not available in English.

Evaluation criteria

Three facts deserve special mention under this item.

In the early nineties, the policy of attaining high quality research at international standards prevailed in Slovenia. This led to a system where the number of citations (*science citation indices - SCI and SSCI*) and number of publications in renowned international journals with high *influence factors* attained by the evaluated researcher, research team or organisation were introduced as decisive measures of *scientific quality* and effectiveness. Such evaluation criteria brought about a stronger internationalisation of research in the country which was clearly a good result. On the other hand, this led to an exaggerated orientation of researchers towards subjects currently »marketable« in the international arena, to the detriment of topics relevant from the viewpoint of national scientific, technological and social development.

Apart from this questionable practice which needs to be overcome as soon as possible, a positive development can also be reported in the domain of bibliometric evaluation data. In the early nineties, when new bibliometric criteria were introduced as scientific quality and output measures, they were defined equally for all scientific fields, even though they were based on the outputs typical of the natural and technical sciences. These measures were unjust and highly unsuitable for evaluating the output of other research fields, especially social sciences and the humanities with their own completely different logic of research and way of reporting research findings. In order to correct this, an expert group of the Slovenian Academy of Sciences and Arts recently developed differentiated bibliometric measures adopted to the specific nature of the six broad fields of science (natural sciences, technical, medical, agricultural, social sciences and humanities).

In addition to scientific quality, the *social relevance* of research projects and programmes (meaning their relevance for solving real development problems faced by the country - technological, social, economic and others) is evaluated within regular evaluation procedures. It is to be stressed, however, that criteria for evaluating such relevance are vague and thus arbitrary which is due to the absence of precisely defined national development priorities (only very broad priority technology fields have been defined until now). Preparations for the launch of the first national foresight programme promise to essentially improve this weak point of Slovenian R&D and S&T policies.

Organisation and performers of evaluations

Regular evaluations of publicly financed R&D projects and programmes now take place in the external expert bodies of the MST and within the MST itself. For the field of science

policy including basic and applied research, these bodies consist of renowned scientists [1], while for the field of technology policy including development projects, other experts are incorporated as well [2]. The members of these bodies are nominated by the minister. On the whole, the participation of users in these evaluations is unsatisfactory and so is the self-evaluation practice of research groups, institutes and industry.

Due to the principle of independent evaluators, *decentralisation of the existing system* for evaluation and budgetary funding of R&D is necessary. A new Law on Funding and Organisation of Research, set to be adopted by the year 2000, is to establish among other things the possibility of setting-up independent public and private intermediary agencies following successful solutions used in other countries. Since during the socialist regime, such an independent agency (named the Research Community of Slovenia) already existed and functioned for years in Slovenia, the country has a quite strong tradition in the decentralised type of R&D evaluation.

Social acceptance of S&T evaluation

Slovenia's S&T system and S&T policy are both characterised by a traditionally strong research community, which predominantly influences the administration of the S&T sector. The result is full acceptance of the evaluation of scientific quality of research in the country, but much lower interest or even resistance to the evaluation of the social relevance of this research. The same can be said of the evaluation of S&T policy measures and orientations which has no real tradition in Slovenia. Here, promoting the public understanding of the role which science and innovation have in the development of contemporary societies, and especially sensitising policy-makers in this regard, remain tasks for the future. A real evaluation culture in the broadest sense of the word is still to be developed in Slovenia.

Quantitative data bases and their exploitation

Since the beginning of the nineties, Slovenian R&D and innovation statistics have been fully adapted to Eurostat's and OECD's methodologies and standards and today cover all the so-called main S&T indicators. Further, the MST's annual reports include thorough reviews of budgetary expenses by research projects, programmes and broad fields of sciences and thus provide (together with statistical data and occasional research projects producing complementary technology and innovation data) useful quantitative material for use in the analysis and assessment of the S&T system and S&T policy in Slovenia. Despite certain absences in data coverage and weaknesses in the quality of data, Slovenia does possess relatively good databases for S&T evaluations. The main problem in this regard is

---

44 In 1992, the MST set up a network of external expert bodies - the National Council for Research and Development (NCRD) and six research committees covering the six broad disciplinary areas (Natural Sciences Research Committee, Engineering Sciences Research Committee, Medical Sciences Research Committee, Biotechnology and Agricultural Sciences Research Committee, Social Sciences Research Committee, Humanities Research Committee).

45 The Technology Development Council is the most recent MST expert body. It was established in 1995 to strengthen expert support for the MST in the area of technology development. Its members are appointed by the Minister from among relevant experts, taking into consideration the balanced representation of technology fields and branches of the economy.

not the availability of data *per se*, but the poor capacities for analytically exploiting existing data resulting from insufficient numbers of hired personnel (see detailed analysis in Bole-Kosmac, 1998).

What remains to be done

Some of the tasks for the future:

- improvement of methodologies for evaluating the economic and social impacts of development research projects;
- regular evaluation of technologically relevant research groups with regard to the national strategy of economic development;
- further decentralisation of the evaluation function;
- further development of databases necessary for the assessment of the S&T system and S&T policy measures and strengthening analytical use of disposable data;
- integration of partial evaluations (of various aspects, at various levels, by various subjects) into a comprehensive evaluation system; and
- establishment of mechanisms to integrate evaluation into strategic research management and S&T policy formulation.

In fulfilling these tasks, much is expected from the newly-established *Group for Research and Evaluation of Science* at the Slovenian Academy of Sciences and Arts. This group of researchers is in charge of carrying out the five-year *Science Evaluation Programme 1999-2003* that is based on a contract between the Academy and the MST. For the beginning, the programme's budget extends to no more than 2 FTEs per year, but both the Academy and the Ministry agree that the group should develop into a focal point of S&T policy research and evaluation in Slovenia. It is to co-ordinate existing research on S&T in Slovenia, as well as to promote this specific research field and gradually develop an efficient national system for researching, monitoring and evaluating science and technology in the country. The urgent need to develop such systems in all accession-states in order to support national S&T policies was also stressed in the study commissioned by the European Commission and conducted by a group of technology consultants (Coopers & Lybrand, 1998)

References

Bole-Kosmac, D. (1998), Selected Issues of Monitoring R&D and S&T in Slovenia, Researcher, December 1998, Vol. 28, No. 3, pp. 11-17; see http://www.mzt.si/mzteng/pub/researcher/1998-4.

Coopers & Lybrand (1998), Impact of the Enlargement of the European Union towards the Associated CEECs on RTD-Innovation and Structural Policies, March 1998, p. 34.

FhG-ISI & IER (1999), Bross, U., Koschatzky, K. & Stanovnik, P., Development and Innovation Potential in the Slovene Manufacturing Industry: First analysis of an industrial innovation survey, Fraunhofer Institute for Systems and Innovation Research, Karlsruhe & Institute for Economic Research, Ljubljana, Arbeitspapier Regionalforschung Nr. 16, March 1999, 58 p

Stanovnik, P. & Faleskini, R. (1998), Evaluation of subsidised research projects promoting technological development, Researcher, December 1998, Vol. 28, No. 3, pp. 46-49; see http://www.mzt.si/mzteng/pub/researcher/1998-4.

Comments on Selection Criteria for EU Research

Angelos Manglis
Serectary General HELFTAS, Hellenic Evaluation and Assessment Society
51 Polytechnia St., 54625 Thessaloniki, Greece

It is known that the Fifth Framework Programme starts with the principle "put research at the service of the citizen", and giving primary emphasis to the socio-economic impact of research. However, when studying the selection criteria for Specific Programmes (e.g. IST, GROWTH, etc) the following paradox comes forward:

- In the IST programme, at the RTD one-step or step 2 type of action for example, the "scientific / technological excellence" criterion takes 4 (weight) while the "social objectives" takes only 1 and the "economic development" takes 2. Moreover, if someone looks at the thresholds, then the case is even more diverse: the threshold for "scientific / technological excellence" and "economic development" is 3 while the "social objectives" criterion has NO threshold.
- In the LIFE programme, at all key actions, the "scientific / technological excellence" criterion takes 30% (weight) while the "social objectives" takes only 20% and the "economic development" takes 10%. Moreover, the threshold for "scientific / technological excellence" is 4 while the "social objectives" and the "economic development" criteria have NO threshold.

In the GROWTH programme the situation seems to be more in line with the FP5 priorities and principles. In the KA1 case for example, all selection criteria take 3 (weight) while "economic prospects" take 4.

This is a lot closer to the nature of the new FP and should form the example for all the other programmes. The socio-economic selection criteria should be given equal weight with the other criteria or at least the difference should not be a significant one. Furthermore, there should not be such a divergence between the specific programmes.

Contribution to Round Table I

Christian Dambrine
Association Nationale de la Recherche Technique
15 rue Philibert Delorme, 75017 Paris, France

Some of the speakers, this morning, have described the process of reorganisation of their national research system, and the accompanying evaluation procedures. Among the aims and objectives of such a reorganisation, beyond the search for excellence in the research system, the necessity to promote innovation was stated as a priority.

Innovation is often described as new concepts or ideas integrated into processes and/or products – including immaterial products -, that have reached the market, hence the necessity to involve industry at large - including services - from early on, in the process of reorganisation of a research system and its evaluation.

The transformation of ideas, inventions, etc. into products and services on the market presupposes the existence of interfaces between research and the market in order to link individuals with different mentalities, background or even priorities; these interfaces may be different depending on traditions, cultures, they have in any case to be diversified, at a local as well as at national level; from their efficacy depend to a large extent technology transfers as well as innovation.

Whatever they are, public, semi-private or private organisations, these interfaces are part of the research system; because their role is essential, they have to be evaluated as part of the system if the promotion of innovation is one of the objectives in the reorganisation of a research system.

Evaluation in the Romanian Academy

Ionel Haiduc
Academia Romana
125 Calea Victoriei, Sector 1, Bucharest, Romania

Evaluation in the Romanian Academy was introduced five years ago, after a Prague Conference, where the German and American science authorities (DFG, NSF, etc.) explained how the process works in their respective countries.

In the Romanian Academy the first step involved an overall view of the situation in the research institutes of the Romanian Academy and an evaluation of human resources, financial status, major programmes and performance (publications, etc.). This produced a clear picture of the status quo but was not followed by any measures.

The evaluation was met by (some part of) the science community with mixed feelings and reservations and some researchers were scared by the prospect of closing some institutes or lay-offs. A mobility of personnel occurred anyway, as a result of "brain-drain" (emigration, transfer to universities or to other fields of better paid activities) but the institutes survived.

In order to make the evaluation less scary, a second step was the identification of the centres of excellence among the Romanian Academy institutes. Based upon a set of criteria, a questionnaire was launched and the institutes were invited to apply for recognition as "centres of excellence". Out of 66 institutes 25 applied and 20 were recognized as such, on the basis of the criteria used. These are institutes involved in major research programmes of national importance and/or institutes which are internationally visible (according to publication in main stream journals and citations). The centres of excellence received an increment of their budget allowances. The report was published (in Romanian language) in the ACADEMICA journal of the Romanian Academy.

A third level of evaluation process is just coming to an end now, in which ALL institutes were subjected to a detailed evaluation exercise, largely based upon the 23 criteria (or questions) used in Germany after the unification, for the evaluation of research institutes (Academy, etc.) of the former GDR. This evaluation intended a SWOT analysis of the institutes. The final report is being now written. It is planned to classify the institutes according to their performances and to reflect this classification in further budget allowances.

It is hoped that the experience gained by the Academy in the process will be useful in the evaluation of the research groups in universities and of the institutes of the National Agency for Science, Technology and Innovation system.

Assessment in Science and Science Assessment

Ladislav Tondl
Academy of Sciences of the Czech Republic
Jilska, 1, 11000 Prague 1, Czech Republic

1. The Spectrum of Evaluating Activities in Science

Cognitive activity in science and research is a complex of interconnected and - usually considerably varied - activities whose integral part lies in evaluating and decision-making processes. The very formulation of a research project poses to the scientist the task of selecting procedures, opting for suitable methods adequate to the objectives of the cognitive task in hand, choosing appropriate tools, measuring and experimental instruments, methods, theoretical concepts etc. suitable for the specific project. Consequently, the subject of cognitive activity has to choose from alternatives, having to assess the available and known options in order to select only one alternative for further procedure. He has at his disposal not only his own knowledge (which he has gained in the course of his previous work, previous research activities, knowledge which he has acquired or adopted from his colleagues, from the available funds of knowledge), but also certain values or criteria for optimum selection. Such an assessment is, therefore, associated with his target orientation, with his expectations. As a result, such an assessment also has its forecasting dimensions.

But, figuratively speaking, the subject of cognitive activity in science and research should also keep "looking back", i.e. evaluate the attained results of his own activities or the activities of those whose work he has been developing. Consequently, this particular task involves the process of evaluating attained results, acquired data, proposed methods or procedures, finished projects covering larger data sets, for instance generalizations etc. Moreover, this involves various modes of assessing what are the results of cognitive activity and have the nature of statements, sets of statements or - as is more frequently formulated in terms of information science - the nature of data or data sets of a certain type. The most significant types of assessment of partial results of cognitive activities include the following kinds of procedures:

- verification procedures, i.e. procedures conducive to the determination as well as verification of the results of empirical findings of a singular nature, i.e. results of observations, measurements, experimental results, establishment of the presence or absence of a specific attribute under scrutiny, studied substances or functions,
- confirmation providing for verification or validation of existing generalizations, hypotheses or scientific laws, their domains or competencies, i.e. including the limits to their validity etc.
- evaluation of the functionality, reliability, efficiency and other evaluating characteristics of proposed methods, methodological procedures, algorithms, i.e. general guidelines for the solution of the outlined type of problem situation.

Under the term *verification* we understand a decision-making process that decides about the truthfulness or untruthfulness of an empirical finding with the nature of a singular statement relating to what we call a fact, status, event or situation in a limited universe. This implies that truthfulness, as a result of such a decision-making process, also represents a certain value attributable to the given statement. Moreover, we assume that the given statement is meaningful, conforming to accepted syntactic and semantic rules, and that it is interpretable by other subjects. The decision-making process should be perceived as applying appropriate empirical procedures, i.e. new observations, measurements, experimental findings. While implementing such a process, it is expedient to specify that universe, i.e. the domain relating to the given statement. That universe can be viewed as a component, part or selected portion of the actual and accessible world. But that world also comprises the realm of human constructs, the world of artefacts. Some of them can be seen as models of possible worlds, as images, projects or other types of representations of what is considered, designed, required or projected. Once their universe is specified, it is also possible to decide about those artefacts in analogy to decision-making processes concerning empirical statements.

Decisions are also taken in case of generalizations, i.e. scientific laws, hypotheses or empirical generalizations, even though the term truthfulness is not usually used in such contexts. The term *confirmation of generalizations*, i.e. confirmation vis-à-vis a given set of empirical statements related to an outlined domain, the testing of hypotheses vis-à-vis available empirical knowledge etc., is used more frequently. But the verification of generalizations, their justification as a legitimate component of certain well-defined sets of knowledge, constitutes a very extensive set of different procedures, with the methods of statistical decision-making making up just one partial, although - for many spheres a very substantive - component of verification, and hence also justification of generalizations.

In the course of cognitive processes in science and research we have to deal with many other conceptual instruments that cannot be simply classified among the given types of statements. This applies primarily to guidelines, instructions for the implementation of certain activities or instruments to organize such activities, algorithms, different forms of the “knowledge how“. Those instruments, which usually have the nature of statements with intentional operators, must be evaluated as well. Generally speaking, this *involves evaluation of the functionality* of those tools. This particular group also includes the process of assessing efficiency, reliability, measure of risk, for instance efficiency of therapeutic procedures, certain technological solutions etc. Seen in this light, this is the type of assessment which, although different from verification or confirmation, is no less important in domains utilizing acquired knowledge or application of knowledge.

The values that are attributed in the cases mentioned above have the nature of epistemic or semantic values. By attributing these values, we justify what we understand under the (general) acceptance of the appropriate statements being evaluated as an expression of knowledge. At the same time, such an attribution forms a prerequisite for a sufficient - or at least acceptable - degree of certainty (or - to put it still better - an acceptable level of uncertainty).

When assessing results in science, epistemic or semantic values should be differentiated from information value and information assessment. (Introducing such semantics, it should be emphasized that the term "information" and the related attributes are interpreted here in the sense given by C. Shannon and N. Wiener, i.e. as a characteristic or a quality of data (including generalizations, rules, instructions or guidelines), which is in a position to ensure that such data can reduce the original indeterminateness, upgrade the quality of decision-making or diminish the rate of risk associated with the given decision-making process. This perception differs from the interpretation of the term "information" whereby the concept is understood as synonymous with the terms "statement", "report", "data", "knowledge" etc. Consequently, the interpretation of the term information given here is connected with the concept introduced in scientific thinking by the theory of information, with the concept associated with the generalized term "entropy".) Let us add the following examples for graphic differentiation of both evaluating principles: A physician facing the task of diagnosing a patient and proposing his therapy can have at his disposal a large quantity of empirical findings, the truthfulness of which he has no need to doubt and which he can easily verify. These include, for example, the colour of the patient's eyes, colour of his hair, and, naturally, a large quantity of data on the established symptoms. But only some of them are relevant for the specific situation. If the patient suffered a massive loss of blood, it is essential to know his blood group, even though many other data may also be established. Only some statistical methods are relevant for the task of determining the measure of statistical dependence of the occurrence of two different phenomena in the defined circumstances, while other methods are insignificant for the solution of the task in hand. Therefore, it is generally possible to consider the *rate of information relevance* of certain data, hypotheses, rules or methods, always in relation to other, already known, data or knowledge, and in relation to the goals of the task under solution.

The author of this study, who was also involved in the sphere known as the semantic theory of information, linking up to the works of R. Carnap, Y. Bar-Hillel and J. Hintikka, has coined the term "*pragmatic information*"[1] for the purpose of overall characterization of those approaches facilitating the evaluation of the actual rate of information relevance of data, generalizations or rules, invariably in relation to the already accessible knowledge of other data, generalizations or rules, and in relation to the objectives or requirements concerning the quality of the task or problem situation being solved. (Interpretation of the term "pragmatic information" proceeds from the results of the works of Scandinavian and Finnish logicians, namely works by J. Hintikka, and the author's earlier study on information assessment of scientific results [2]) Pragmatic information is conceived as a measure of transmitted information (transinformation) in view of the particular data or knowledge at the user's disposal and in view of the objectives of the task being tackled, and the requirements of the quality of its solution. Information evaluation of different types of data based on the given approaches is most frequently bound up with those types of scientific

---

46 Tondl, Ladislav (1989), Pragmatic Information. In: M.G. Singh, ed., Systems and Control Encyclopedia. Pergamon Press, pp. 3858-3860. Český překlad: Pragmatická informace. In: L. Tondl, Mezi epistemologií a sémiotikou. Filosofia, Praha, 1996, pp. 68-76.

47 Tondl, Ladislav (1986), Some Methods of Information Evaluation of Scientific Results. Computers and Artificial Intelligence, 5, 1986, pp. 385-394.

procedures in which we operate not only with the results of empirical findings but also with available generalizations or rules, while solving a specific assignment, for instance tasks involving explanation, forecasting, optimizing the solution of a given task, diagnostic tasks etc. Viewed in this light, pragmatic information seeks to specify the rate of information relevance of the data being evaluated, namely either empirical data, different generalizations, hypotheses or rules in a situation where we are solving an outlined assignment (for instance, forecasting, hypothesis testing, choice of variant solution of the given task etc.) connected with a certain target orientation, with the demands concerning the quality of solution or with endeavours aimed at minimizing losses or risks. This is, therefore, a pattern - or rather an abstract model - for the selection of relevant data, namely in relation to target orientation of the specific task under solution, and also in relation to what is invariably a limited information equipment at our disposal. Hence, the following situation in which we tend to find ourselves in the information society is particularly typical: Even though we are flooded with an enormous quantity of data in which we take our bearings only with difficulty, we are aware of an information deficit when solving a given task.

While resolving many scientific problems, we are in a situation where, already before tackling such an assignment, we have at our disposal both sets of generalizations or rules, and, in view of them, appropriate sets of empirical findings. A well-arranged pair of both types, i.e. a set of generalizations and a set of appropriate empirical findings, may then be described as a decision-making basis relating to the given assignment, and the goals and requirements related to its solution. It is not always easy to choose or select such a decision-making basis, i.e. a pair of suitable pieces of knowledge, and empirical findings relevant for these particular pieces of knowledge. The term decision-power may then be coined for the purpose of information assessment of each task's decision-power. Decision-making greatness is the bigger, the more the decision-making basis under evaluation is able to reduce the original indeterminateness connected with the outlined target complex - or rather - the more this decision-making basis is capable of limiting the risk accompanying the attainment of the ultimate goal of the solution of the given problem situation. Seen in this context, decision-power can be defined as a rate of transinformation of the selected and evaluated decision-making basis in relation to an outlined target complex.

Information assessment of the whole decision-making basis, in which the subject of information assessment is a pair formed by a knowledge set, i.e. generalizations, hypotheses or rules, and a set of empirical findings relating to the given assignment or relating to the problem situation under solution, should be distinguished from evaluation proper involving selected generalizations, a type of evaluation which then constitutes a prerequisite for their acceptance and for the acknowledgement of their competence to be applied to a certain domain. A classical example of the first mode of assessing information relevance is the process of selecting hypotheses and empirical data for the solution of assignments, such as - for example - forecasting phenomena in the sky, namely the eclipse of the sun or of the moon. In this particular case, we can hardly make do only with the well-known laws of celestial mechanics. We need to have at our disposal time-specific data on the routes of the appropriate celestial bodies. A physician, who is to

diagnose a patient, who must not only have at his disposal and select from a sufficient body of medical knowledge but who also has to choose adequate empirical findings, i.e. determine the incidence of certain symptoms, finds himself in a similar situation. As a result, he has to evaluate and select an appropriate pair formed by a set of knowledge and a set of empirical findings. Since this is a decision-making process sui generis, we are entitled to speak here of a decision-making basis and decision-making greatness. On the contrary, assessment proper involving generalizations, i.e. scientific laws, hypotheses, rules, empirical generalizations, as an inductive method of processing the existing body of evidence is important for the purpose of specifying their competence and hence, usually, their potential applicability.

A case in point illustrating assessment proper involving generalizations may be found in procedures testing hypotheses, which are relevant for decision-making concerning the plausibility of those hypotheses, for the specification of the extent of their competence and limits to that competence. As for this kind of knowledge, arising from the inductive mode of treatment of the body of available empirical evidence, i.e. such as empirical generalizations, their testing on the basis of new evidence - or rather - broader fields of evidence is of particular importance. As a result, distinct boundaries or clear-cut limitations to such generalizations may emerge, even though these had apparently been confirmed for many generations. This is corroborated by newly arising opinions concerning some of the past behaviour patterns, standards of eating habits or physical and mental hygiene. That is why the positive results of testing hypotheses, based on a body evidence which turns out to be limited, may be proved false by new findings and a new body of evidence.

Another example of information assessment is the evaluation of the relevance of empirical data in relation to tested generalizations. Testing relevance of a set of empirical (or also experimental) data may be considered in this particular case. Testing relevance of a set of empirical data in relation to a set of generalizations is the greater the bigger the rate of a set of empirical data, in which information is transmitted in relation to a set of generalizations; it reaches its maximum if generalizations are logically implied by the given empirical data; and it is minimal if empirical data transmit no information relating to the given set of generalizations at all.

The above mentioned methods as well as other analogous instruments of information assessment of data and knowledge presuppose application of probability or rather statistical parameters, namely in the shape of established statistical distribution patterns, and hence in the form of frequency characteristics, or at least qualified estimates expressed in terms of probability. Acquisition and verification of such characteristics is neither easy nor always sufficiently reliable. Moreover, doubts may later be cast on those characteristics as a result of further shifts in our knowledge or - as K. Popper put it - they may be proved false. But operating within the framework of those and similar situations, we can hardly avoid perceiving not only the hitherto available, but also newly acquired pieces of knowledge through the prism of the still accepted concepts and knowledge. (As a result, this means that we always find ourselves behind a “curtain“ or a screen formed by the still acknowledged conceptual patterns and opinions, i.e. in a situation described by W. C.

Salmon by the term "screening off"[48], while this also applies to situations involving information assessment of data and knowledge).

The given evaluating and decision-making processes involving individual cognitive steps, cognitive activities and their results are, in actual fact, integral components of scientific learning which is, at the present level, unthinkable without permanent control, without decisions about proper selection, acknowledgement and acceptance of individual partial data and partial findings. The subject of assessment is usually the subject of the process of learning itself. Seen in this light, cognitive activities at the current level should be regarded as decision-making processes sui generis, envisioning not only sufficient knowledge equipment, but also an ability to master and apply certain methodological instruments. Since we invariably operate with a certain knowledge and methodological fund the results of each new step should be seen in the context of that particular fund, its consistency and any phenomena or situations upsetting that consistency should be verified. Due to these reasons, when evaluating each cognitive step we take into consideration the hitherto created or accepted system of knowledge, using a systems approach, without being directly aware of it.

2. Internal and External Science Assessment

The process of distinguishing internal and external assessment of activities in science and research and of acquired results has grown to be topical when - ever since the end of the past century and during the 20th century - activities in science and research have ceased to be a matter of individual interests, becoming a highly organized and institutionalized domain. Proceeding from the concept of university chairs for leading scientific personalities, from individual researchers, this branch has gradually gravitated towards what M. Polanyi called "the republic of science"[49], towards a differentiated and well-organized structure of departments, teams, institutes, laboratories. This organized structure of smaller or larger national and international scientific and research communities is also matched by the structured and organized system of outputs, publishing and communication instruments in the shape of periodicals and large editions, a system of congresses, conferences, symposia and working meetings and - in recent years - also a system of direct communication links through state-of-the-art information technologies replacing the traditional "invisible colleges". But the process of organizing, institutionalizing and concentrating scientific and research activities does not mean that creative activity in these branches has ceased to be "a profession", as claimed at the beginning of the century by Max Weber, one of the protagonists of the sociology of intellectually demanding rational activities, that it has stopped being an instrument of qualified self-assertion for creative personalities[50]. As a matter of fact, such an activity has grown to be one of the major sources and stimuli for innovation processes, radical technological changes and, by

---

48 Salmon, W.C. (1971), Statistical Explanation and Statistical Relevance. Pittsburg University Press, Pittsburg.

49 Polanyi, Michael (1962), The Republic of Science. Minerva, vol. 1, pp. 54-73.

50 Tondl, Ladislav (1979), Science as a Vocation. In: Jan Bärmark, ed., Perspectives in Metascience. Göteborg University, Berlings, Lund, pp. 173-184.

extension, standards of human behaviour. These tendencies have been confirmed by the results of the sociology of science, as initiated by Robert Merton, Ben-David and many other authors.

While the focal point of scientific and research activities has been shifting towards social groups, called by today's sociology of science as scientific or research communities, many relevant results in science and research are conditioned not only by the material, technical and economic infrastructure of those communities but also by the intellectual, value as well as moral climate, so typical of the inner life, orientation and direction of those communities in terms of their subjects and special interests. A survey of the fruitful scientific communities which are known to have produced major scientific initiatives and significant works, especially projects awarded by Nobel Prizes, has highlighted the importance of a fruitful, demanding and, at the same time, critical intellectual climate prevailing in such "centres of excellence" or "intellectual clusters". On the other hand, communities dominated by mediocrity have been producing - or rather reproducing - what some authors have called the permutational nature of scientific work, i.e. repeating already finished experiments, discovering what has already been discovered and producing works that are more or less compilations of the already known opinions or thoughts.

Inner assessment and also inner checking of all the results, still before they are submitted to a broader academic community and to the public, are, therefore, a major prerequisite of sufficient standards and quality. However, such internal assessment is known to lay certain demands on the inner relations within the community, on mutual spirit of togetherness and responsibility. This applies particularly to the following factors:

- The highest demands, both in intellectual and moral terms, are placed on top-ranking and managerial personnel. Violations of this particular requirement, introduction of other - for instance - political or ideological criteria in an effort to establish what can be described as "intellectual Darwinism upside down", a situation Czech scientists know from their country's totalitarian past, seriously eroded the overall quality, prestige as well as attractiveness of the entire research centre in the eyes of young talented researchers. That is why the prospects of the whole workplace, research team or otherwise characterized scientific community are heavily dependent on the compliance with the strict rules for the selection of managerial and responsible personnel. This applies - to the same extent - to habilitations and professorial appointments. It is a genuine tragedy when mediocre or academically unsuccessful staff, deriving most of their strength from the "fraternities of the mediocre" or similar pressure groups, try to push their way to top managerial and responsible posts, for instance in boards of management, grant commissions etc., hence posts involving external assessment.
- The quality of a research center affecting internal assessment depends on the overall intellectual climate of that particular institute, especially on its critical climate and - at the same time - an atmosphere of professional collaboration and togetherness. Just as a top-level medical clinic and its leading specialists shall not tolerate doctors and medical staff they cannot rely on and co-operate with, likewise a top-quality researcher can hardly tolerate in his team any "dead souls", i.e. researchers lacking profound

interest and knowledge who simply "drift along", tarnishing the good image of the research center involved. A positive and stimulating environment in a research centre may be seriously damaged by the emergence of inner conflicts, some of which may be brought on by acts of external interference, especially those politically and ideologically motivated. Indeed, the danger of conflict situations has been singled out by international research projects, carried out by individual national scientific communities and managed by the sociologist of science Ben-David, projects in which the author of this study was also involved[51]. A challenging and critical environment is marked by solidarity of the competent, and intolerance towards the incompetent and the irresponsible. Moreover, intellectual and moral solidarity is usually known to transcend the boundaries of a single research center and the frontiers of one state. Indeed, it was the commendable international solidarity in intellectual and moral terms in particular that helped several Czech scientists survive the restrictions, and sometimes personal restraints or even personal persecution, during the rule of the former totalitarian regime, at a time when the ideology of monopoly power and also what the author of this study calls "intellectual Darwinism upside down" dominated in many scientific branches.

- A high level of internal assessment can hardly be secured without good working contacts and primarily without well-functioning information links with the leading domestic and world centers engaged in the studies of similar subjects or similar or related research programmes. Needless to add, such professional contacts may assume different forms, ranging from work on joint research projects and division of labour to mutual exchange of information and attained results. These forms also include exchanges of experts, reciprocal short-term research affiliations, invitations to attend conferences, working meetings and lectures.

Internal assessment does not and cannot result in total uniformity of the research centre involved, in efforts to suppress different approaches or diverse views. There are signs that in the past internal atmosphere of research centers was definitely not improved by the authoritarian approach of its managerial staff, epitomized by the formerly ingrained deplorable practice whereby a team's leading researcher had to be mentioned as co-author of any project or at least had to be extensively quoted in it. On the other hand, a very important factor in this respect is a certain commitment and co-responsibility of all the members of a research community to make sure that each and every result of the work of each individual as well smaller groups associating members of the community should meet the highest possible standards. That is also why the best record of an outstanding researcher or a university lecturer should be seen in the achievements of his or her outstanding pupils, even though their work should modify or transform the teacher´s very concepts from which they have proceeded.

Internal assessment of scientific and research activities that concerns subjects of these activities, their target orientation and thematic focus and, invariably, the attained results,

---

51 Tondl, Ladislav (1970), Conflict Situations in Scientific Communities. International Social Science Journal, vol. XXII., No. 1, pp. 111-126.

too, can neither be dictated nor crammed into any universal formulas. Instead, internal assessment evolves from the prevailing inner atmosphere of the scientific community concerned, from its awareness of solidarity and co-responsibility. Furthermore, internal assessment and its accompanying critical climate, coupled with such an awareness and social responsibility, have grown to be a spontaneously recognized norm and a model of decent professional relations in many scientific and research institutions. In this respect, internal assessment substantially differs from external evaluation which is - to a great extent - institutionalized, being bound up with a number of rules, many of which are formulated as obligatory norms.

In most cases, external assessment does not arise spontaneously. It is launched in order to help in solving specific problems, for instance selecting the best candidates for a managerial post in a competition, worthy grant recipients, assessing the implementation of research programmes. It is also due to these reasons that external experts are hired, and an evaluating or expert team is set up. External assessment is also known to have specific characteristics, the most important of which seem to be the following:

- target orientation of assessment, covering both individuals, candidates, research teams and whole research centres,
- well-defined objects of assessment, for instance results to be acquired and publications of those results to be issued, response to or reception of such publications, formulation of systems of criteria for the individual types of objects of assessment, practical applicability of the attained results,
- a competent evaluating team, for instance a group of experts, a habilitation board, a selection board etc.,
- a required form of the results of evaluation, for instance a preferential arrangement (short-list) of candidates in a competition, results of a points system in a given interval scale etc.

External assessment, particularly if organized by competent bodies and institutions, is not usually purposeless, it sets its sights on a specific problem situation and, therefore, has its own target orientation. This also means that a proposal to initiate such a process to assess an identified or assumed state has been formulated and approved. For the purpose of such an assessment appropriately complete and reliable source materials have to be accumulated, while their scope and quality should be previously given or specified. External assessment is also accompanied by certain expectations, including compilation of a body of background materials for competent decision-making. A major prerequisite for internal or external assessment (external in view of the necessary participation of qualified external specialists) also includes the task of formulating well-defined rules of assessment, for instance, specifying an applicable scale and its interpretation, voting methods or quantification etc.

The process of distinguishing various types of objects of assessment and sets of criteria adequate to those types actually means that considerably different evaluating tasks are involved, if records of individual scientists or researchers, their careers and their scientific production, research plans or projects, and the competence of research teams to solve such projects, and records of entire research institutes are to be evaluated. The differences

and the specific features of such tasks should be reflected in the composition of the evaluating team, appropriate commission or evaluating board.

A key part in any external assessment is to determine the set-up of its evaluating team. Naturally, it is impossible to come up with reliable results if a project is being evaluated solely by the friends of its author or by those whom the individuals or the groups to be evaluated have selected themselves. Still more dangerous are situations where, as a well-known foreign participant in assessments made in the Czech Republic put it, "a fourth-rate team" is to asses a "third- or second-rate research group". Analogous situations are known to have happened in cases involving selected evaluators or boards awarding grants. The rule in such cases is that an irresponsible appointment of jurors or members of evaluating boards actually anticipates the result of such an evaluation. This aspect is also exceptionally significant when specifying evaluating powers to be exercised by various research institutions or universities, especially the powers of doctoral studies, habilitations and professorial appointments. A well-known tenet says that "the average tends to reproduce itself easier and faster", while the repercussions of such a practice are certain adversely to affect the entire academic community, often for many years to come. (This is notoriously known in the Czech Republic as part of the sad heritage or relic of the country's totalitarian past.)

- While it is expedient to outline in great detail the requirements for the necessary background materials to be used for external evaluation, which must always be an expert assessment, it is equally desirable to specify the actual form of evaluating tasks and the actual shape of their outputs. In such contexts, one should make a point of not ignoring the fact that this is a highly sensitive and challenging issue, which should not be left either unattended or at the mercy of totally subjective opinions and ideas. This particular principle should be applied to any individuals whatsoever, no matter how high an office they hold or what membership of top-ranking institutions they have been entrusted with. It is due to these reasons and due to the immense complexity of evaluating tasks that the most advanced countries, noted for a high level of their scientific and research potential, are known to be drawing up recommendations or manuals to govern those forms of evaluation of research institutes which have the character of expert assessment or, as this is often called, the nature of "peer review". Expert assessment seems to have penetrated into scientific life, becoming its inseparable part, just as the democratic forms and principles have been asserting themselves in social life. Typical recommendations for expert assessment of scientific and research institutes reflect the following components (an overview of such recommendations used in Canada is given in[52] in particular).
- Sets of criteria for quality assessment, criteria aimed at permanently safeguarding high standards (quality assurance), criteria for evaluating impacts, application possibilities and especially stimulating and innovation initiatives.

---

52 Stoicheff, Boris P. (1999), Evaluation of Researchers, Research Proposals and Accomplishment. In: V. Pačes, L. Pivec, A.H. Teich, eds., Science Evaluation and Its Management. IOS Press. Amsterdam, pp. 67-74.

- Recommendations for drafting documentation (which most frequently results from internal assessment), recommendations or requirements concerning the content and scope of data in preparatory background studies (usually known as self-studies), scientific biographies, publishing and biographical data, etc.
- Recommendations for the evaluating procedures proper, including a dialogue to be held by evaluators and candidates[53], desirable forms of discourse between leading and rank-and-file personnel and, at universities, dialogue with students as well.
- Recommendations for requirements concerning conclusions of each assessment, evaluation of the feasibility of research programmes and projects, a comparison of the research centre or workplace being evaluated with analogously oriented advanced research institutes, and overall conclusions of the evaluating process.

3. Objects of Assessment and Systems of Criteria

Talking about interpretation of the terms of evaluation or science assessment (these are two terms whose meaning considerably overlaps, as corroborated by the English words - "evaluation" and "assessment"), most people will think of evaluation or assessment of the results of scientific and research activities, namely results submitted in a standardized form, i.e. as texts, calculations or models submitted for public review, for publication or materials already published. In fact, this covers only a relatively small, although very important, niche in that sphere which can be regarded as evaluation or assessment of science and research. It is, therefore, useful to distinguish various types of objects that are evaluated or assessed. At the same time, we have in mind external assessment, i.e. assessment by subjects other than authors or creators of what is being evaluated. We also assume that there are specific systems of criteria for various types of objects. Naturally, this does not exclude the possibility that diverse types of assessment may link up to one another, may supplement or influence one another. Seen from the viewpoint of different types of objects, the following may be singled out as the most important kinds of assessment:

- assessment of individual items always describable and communicable as specific units, namely results, texts, studies, articles, major studies and monographs, empirical or experimental results and discoveries,
- assessment of individual scientific personalities, including evaluation of their scientific careers, their professional prospects,
- evaluation of research groups, teams, divisions and research institutions sharing common subjects, of scientific institutes or otherwise defined scientific and research centres,
- assessment of scientific and research goals, programmes and intentions, research projects in particular.
- Naturally, one cannot rule out various comprehensive types of assessment, synthesizing evaluation of more types of objects, for example assessment of research institutes and scientists working there. A substantial feature is that each of the given kinds of assessment calls for a somewhat different approach, and especially for the

---

53 Koutecký, Jaroslav (1999), Dialogue Between Scientists and Evaluators as a Means to Properly Choose and Assess Research Directions. In: V. Pačes, L. Pivec, A.H. Teich, eds., Science Evaluation and Its Management, IOS Press, Amsterdam, pp. 79-82.

application of different systems of criteria. Similarly, the significance of certain criteria will vary with different types of assessment as well as in various evaluating tasks. (This concerns, for example, the importance of scientometric and quantitative criteria in different branches and in different evaluating tasks.)

Assessment of individual texts, studies or articles for publishing purposes, usually characterized as a "review", is affected by the nature or the desirable standards of the magazine, periodical or publication series involved. It is a matter of course that prominent periodicals and members of their editorial boards tend to lay higher demands, particularly on the following aspects: on the innovative nature of a study considered for publication, whether the work under scrutiny lies within the range of "normal development" or whether it marks a more pronounced shift in the accepted or recognized standards, patterns or conceptual models or, expressed in terms of innovation concepts, whether it represents only incremental or also radical innovations. Seen in the light of contemporary trends in science and innovation policy, great importance should be attached to the evaluation of whether the work considered for publication carries what is called an "innovation charge", whether it opens up new possibilities of application or other forms of fructification of the achieved results. Although this particular form of assessment and evaluation also principally involves considerations covering a large number of criteria, the resultant decision-making process has the nature of a two-value decision, i.e. a ruling on acceptance (for publication) or rejection, or a three-value decision, i.e. ruling whether to accept or reject the submitted work or whether to suggest rewriting, supplementing or otherwise modifying it.

As for individual studies, we usually do not apply scientometric and quantitative methods in general. An exception to the rule may be seen in studies with a topographical focus, seeking to capture the diffusion of various new concepts, graphic and cartographic presentation of diffusion processes, especially processes connected with major innovation initiatives, their further development in time and space. (This particular mode of assessment has been developed notably by T. Braun and R. Schubert in a number of studies published in the journal "Scientometrics"[54].

While assessment of individual works, texts or studies is a process of multicriterial evaluation, this is true - to an incomparably higher degree - of the assessment of scientific personalities. The selection of criteria and the extent of this selection is greatly affected by the nature and target orientation of the evaluating task in hand. Such tasks may, for example, range from a competition to appoint the best candidate for a high-ranking post, habilitation or professorial appointment, accreditation of a post-graduate study and - within the framework of accreditation - approval of guarantors or instructors. Within most of those evaluating tasks, the entire scientific profile and intellectual development of the scientist involved, his or her scientific record, primarily the published results and response to them, their significance for the status and development of both, and their anticipated direction are assessed. Of similar importance is the human and moral profile

---

54 Braun, Tibor, Schubert András (1990), A Topographical Approach to World Publication Output and Performance in the Sciences. Scientometrics, 19, N. 3-4, pp. 159-165.

of the person involved, his relationship to colleagues, scientific teams and schools of thought. It is evident that, at the contemporary level, a comprehensive picture of the scientific profile of such a personality is not feasible without applying scientometric methods, covering not only lists of publications but also their position on the international scientific scene (usually characterized as the "impact factor"), response within the scientific or academic community (expressed by citation indexes), contacts with other authors and scientific schools, with communication networks and innovation stimulation etc. It should also be emphasized that scientometric methods, used to evaluate scientific personalities, cannot be perceived as the only and totally universal method of assessment, but solely as one of the evaluating approaches whose significance considerably varies in different thematic fields. That is why the so-called quantitative methods of assessment are relevant primarily when preparing background materials for the process of assessment or evaluation itself, which should be a multicriterial assessment, notably an evaluation by a qualified team of experts (i.e. the so-called peer review). Furthermore, the values acquired by scientometric methods are mutually compatible only within the framework of a single thematic domain. What is more, these are no additive values in any sense at all (since a positive and appreciative quotation cannot be counterbalanced by a negative or critical citation).

Still more aspects and criteria enter into assessment if research groups or teams, research institutes or institutions are evaluated. As a rule, these bigger units have a vertically and horizontally organized structure, its leading, responsible and rank-and-file or subordinated personnel. What is involved here is not only a sensitive issue concerning the distribution of responsibility and its delegation but also a network of interpersonal relations, accompanied with its problems of co-operation, tolerance and intolerance, and sometimes even outright animosity. Figuratively speaking, a larger research unit incorporating scientists may be a good team, but also a quarrelling group of individuals rife with envy and hostility. It is natural that these "additional" aspects may constitute not only an interesting theme for the sociology of science but also a factor likely to affect the team´s scientific productivity, its stimulating and critical climate and the achieved or achievable scientific and research results which are, quite understandably, the main subject of the process of evaluation of larger research units.

Assessment of such research centers plays a major role in evaluating and deciding about scientific intentions, plans or research projects. In order to evaluate the plausibility of such research projects it is necessary to incorporate, in addition to other criteria, also what is described in technical thinking as "feasibility conditions", encompassing the needed capacities, knowledge and technical prerequisites, and a time horizon. Even the most brilliant projects may fail or end in utter disaster unless sufficient feasibility conditions are taken into due consideration. Regrettably, this is sometimes the cause of failure in the case of accepted projects, especially in situations where less qualified members of evaluating boards concentrate solely on the concept of a project, ignoring a thorough assessment of its anticipated results. Basically, one can agree with the view claiming that the greatest threat to what is called peer review is posed by situations whereby scientists belonging to a lower or less qualified research team evaluate the work of those at a higher level, which naturally does not mean to say that the right or indeed the possibility to

criticize certain scientific positions should be denied to anybody. After all "the rule of the mediocre" is a major threat to a creative and stimulating development of each scientific and academic community.

We have already emphasized the importance of multicriterial evaluation for all levels of assessment, while the selection of criteria and their significance may considerably vary in different evaluating tasks. The same holds true of appropriate combinations of qualitative and quantitative methods. However, one type of qualitative approach is the so-called points evaluation system, stemming from an interval scale whose individual intervals are specified verbally, i.e. in qualitative terms. This naturally excludes - also in the contexts mentioned above - additivity, and, consequently, simple arithmetic operations. Scientometric methods, i.e. the use of registers of publications, impact factor and citation registers, are quantitative methods in their own right. A serious warning for the individual subjects of assessment is primarily a dire shortage or absolute absence of publications and responses to them. On the other hand, minor differences in quantitative data need not yet be indicative of any major differences in the quality of work or scientific results. The fact that varying degrees of what is known as a positive and critical response cannot be distinguished is viewed as a shortcoming of response analysis. Another limitation emerges when citation response actually turns out to be a response coming from one's colleagues, response which occurs predominantly within the framework of routine scientific developments or - to use T. Kuhn´s terminology - which is an evaluation under the conditions of "normal science". This means that more pronounced or radical scientific innovations or - to borrow Kuhn's terminology once again - changes in scientific paradigms are not immediately recognized and duly appreciated. (A classical example is the fact that the theory of relativity or the start of the quantum methodology were eventually accepted only after years of bewilderment and hesitation.) The least responsible model of (what seems to be) a quantitative approach is used when an evaluating team commissions external evaluators to do the job, asking them to use a points system which is, in actual fact, only a qualitative evaluation. The team then feeds the results of points evaluation into a computer, without paying due attention to the differences among the value approaches taken by the variously competent or qualified evaluators. It is evident that besides responsible external assessment, some evaluators may indulge in "playing an assessment game".

4. Science and Society's Value-related Attitude

The idea that science and research development actually takes place away from the public eye, inside universities, research institutes and laboratories or - as figuratively put - "in ivory towers", epitomized by Archimedes´s maxim "noli tangere circuloss meos" (Do not touch my circles!) has long been a matter of the past. Since the middle of this century, applications of modern progress in science and technology have affected the work, activities and interests of virtually each and every citizen. As a result, this has been instrumental in creating a problem area often called "public perception of science" or "public appreciation of qualified knowledge". Indeed, during this century, the social or public perception of science and research, of the values brought about by creative work in this sector has seen many far-reaching changes. The first half of this century was marked

by a spectrum of wide-ranging value-related attitudes with prevailing admiration, optimistic anticipation, always accompanied by perceptions of science and research as something very involved, distant or virtually inaccessible or something which does not concern the life, goals and requirements of the so-called ordinary people. Starting in the 1960s, these value attitudes began to be punctuated by critical voices, fears of a possible abuse of the sector, bringing a special atmosphere which was described by the term "disenchantment with science and technology". Society was realizing, ever more distinctly, the limits to possibilities and the limits to growth, the threat posed to "normal" living conditions, to the entire environment affected by irreversible processes seriously jeopardizing the future of the human race. As a result, many critical words, fears and warnings, including words about the stress of our planet[55], were uttered. No less important were critical voices and apprehensions aimed at some of the trends in biotechnologies, genetic mutations, information technologies and artificial intelligence, and some other fields.

Within these contexts, many questions are being raised, spelling out not only apprehensions and warnings but also demands or subjects for public control to be imposed on this particular problem area, sometimes called "finalization of scientific and research activities". What is also at stake is the process of new values and value structures entering the orientation and direction of scientific and research programmes, the focus of mainstream applications, the scope of responsibility on the part of those who launch such applications. Most leading scientists and cultural figures endorse the view that cognitive initiatives should not be limited. On the contrary, their innovation charge or spectrum of possibilities should be more appreciated. But at the same time it is necessary to specify the control mechanisms of the spheres of application, preventing unbridled experimenting with applications, and especially blocking the start of those applications that pose danger in moral, cultural and humane terms.

Even though the final say in evaluating tasks, which have the nature of external assessment of the results of scientific and research activities, should be given to well-qualified and highly competent teams, citizens, civic initiatives or special interest groups should not be denied the right to express their own evaluating standpoints either. A significant role is played in this respect by the prevailing atmosphere and recognized value structures. This also concerns the task of cultivating and humanizing the existing value structure, placing new accents on the role of moral values, on individual, civic as well as social responsibility. As long as society is dominated by the surviving overblown values of consumerism, primitive types of entertainment with the cult of so-called "entertainers", as long as society is ruled by the values of pleasure-seeking and excessive consumerism, as long as it is permeated by the conviction that material wealth is the main value, and the main source of such a wealth lies in speculations on the capital market, it

55 Mungall, Constance, Mc Laren, Digby, eds. (1991), Planet under Stress. Oxford Univ. Press, Toronto, Oxford Univ. Press, Toronto, Oxford, New York.

will be impossible to appreciate what is known as "intellectual wealth"[56], the values and the beauty of what A. Einstein called "the adventure of learning".

Of great importance for the assessment of the results of scientific and research activities is the standard, the ability of reception or - as it is sometimes characterized - the "absorption capacity" of the reception, application and user sphere. In view of the available knowledge and, by extension, results of scientific learning, society's absorption capacity is dependent on various problems areas of which the following seem to be most significant:

- the availability of knowledge, especially the ease and promptness in acquiring it;
  - knowledge reception, including the capacity to understand, interpret or adequately select relevant data and findings.

Publication is viewed as a routine and virtually traditional form of presenting results of scientific and research activities. This means the type of publication accessible to any potential recipient. As a result, such a statement providing access to the appropriate knowledge fund is theoretically made possible to anyone who has access to specialized journals, research reports, well-stocked special libraries etc. The turbulent developments of present-day information technologies have greatly expanded the spectrum of forms and possibilities of suitable methods of information transfer, for instance the use of electronic mail and Internet-type information networks etc. In this respect, one may refer to the advantages of the information society which offers access to the available knowledge to anyone who is interested in and capable of asking for the information needed in an appropriate fashion. But the given additional notes reflect both the gist and the difficulties of that particular issue. They can be summarized in several questions spotlighting the difficulties and the related horizons of the viability of the information society:

- Are all the members of present-day society genuinely interested in acquiring this knowledge which is likely to help them in coping with their current situation, in solving their problems, in achieving their goals?
- Are all the members of present-day society capable of accepting currently attainable knowledge? Are they always in a position to know that such knowledge really exists at all and that it is somewhere available?
- Are all the members of society capable of understanding the attainable knowledge, adequately interpreting such knowledge or efficiently utilizing it?

Negative answers given to those and other analogous questions will convince us that while everything offered by the information society at the current and perspective levels can, indeed, ensure the general accessibility of all the available results of learning in science and research, it cannot guarantee that these information values will be duly understood, adequately interpreted and, what is particularly important, appropriately used. The maximum that can be safeguarded is the right to the accessibility of those information values, limited, as it were, due to the necessary protection of the integrity of the individual, its privacy and everything that can endanger the man and the citizen in his

---

56 Holländer, P., Koutecký, J., Kurzweil, J., Tondl, L. (1997), Veda, vysoké školy a rozvoj "intelektuálního bohatství". Aula, N. 3, pp. 2-8.

individual life, in his human and social relations and in the exercise of his civic rights. (Needless to stress, these are highly sensitive issues concerning not only health care and medicine but also other sectors, while such limitations and socially imposed exceptions to the rule should be precisely defined, notably in normative terms.)

While the term "information society" is today mentioned together with the postulate of "knowledge society", this definitely means much more than general accessibility of the available knowledge to all the citizens. It is not by chance that the term "knowledge society" has been associated with the value of education, with the need of permanently promoting, upgrading and enriching education or - as is sometimes comprehensively described - the "educative society". Many major postulates are formulated within such contexts, the pride of place among them being held by the following:

- due appreciation of the values of education and learning, values pertaining to what is called "intellectual wealth",
- systematic care for the promotion of knowledge, interest in acquiring knowledge as well as the capacity of qualified reception of knowledge or rather what is termed permanent education,
- efficient distribution of special, professional and often highly specialist knowledge as well as knowledge, which should be respected by all the citizens, including well-organized methods and procedures aimed at distributing relevant knowledge in a transparent and comprehensible form, including an organized system of popularizing learning and conducting a dialogue between the specialist and the citizen,
- resistance on the part of citizens to what is described as "knowledge surrogates", i.e. different forms of quackery, easy and attractive forms of populism and fundamentalism, promising all-encompassing therapies and panaceas, starting with miraculous drugs and ending with the promises of a "bright future" or the possibility of "commanding wind and rain".

Let us take a closer look at the problem of knowledge distribution. It is generally quite well-known that certain forms of knowledge distribution have always figured as prominent components of civilization principles, while some specific interests in a specific form of knowledge distribution have served the interests of society or rather those who officially represent such interests. Already in the past enlightened rulers sought to give their subjects basic education and training, making it possible for them to discharge their elementary duty of producers. It was the industrial society which developed a relatively broad spectrum of such interests, notably the requirement to provide adequate professional training, offering different subjects and different levels of attained knowledge. The industrial society has succeeded in incorporating into its system of knowledge distribution not only the training of highly erudite experts but also of creative personalities capable of distinguishing, supplementing and correcting the attained intellectual and cultural values. Today's system of knowledge distribution and, hence, educational systems basically meet the requirements of advanced industrial society which has also developed the tertiary sector, services, cultural self-assertion, the entertainment and leisure-time industries. Nobody should harbour any doubts that the transition from advanced industrial society to knowledge society should be associated not only with a new vision of knowledge distribution but also sets of other intellectual, cultural and

ethical values, with value structures giving preference not only to consumption and control but also to voluntary restraint, protection and preservation of all the positive aspects, brought about by the development of nature, human culture and civilization. Such a vision is focused not only on a better use of the results of science and research, including those relating to identified horizons, limits to what is possible, potential threats or dangers, but also on the goal of stimulating better ways of learning about the possibilities of such dangers.

It is only natural that many new findings and values are a matter for highly qualified specialists, notably in medicine, natural sciences, in technical disciplines, some branches of the humanities and in other spheres of human managerial, decision-making and evaluating activities. It is also imperative to make sure that what we call "knowledge society" should be a society of adequately informed citizens, informed especially about possible dangers and main risks, a society consisting of citizens well-aware of their own responsibility. Therefore, a certain type of the distribution of knowledge reception should be envisaged, with target orientation being a key to its efficient arrangement. After all, an understanding of the far-reaching impact of Gödel´s theorems on incompleteness of the proof of some formal systems is significant only for a small portion of the academic community. Much broader sections of the community should grasp the meaning of some key principles of the theory of information, the significance of the discoveries and expansion of those discoveries in genetic information. On the other hand, most active and well-educated citizens are in a position to realize the far-reaching importance of certain procedures in genetic engineering. We could go on listing many other analogous examples illustrating a very similar distribution pattern of the reception, understanding and appreciation of the impact of scientific knowledge, of knowledge and technological innovations. All the signs are that a seminal role is also played in these contexts by the mass media, by the prevailing atmosphere as well as different expectations connected with innovation steps in science and research. It is only natural that most citizens are capable of mastering, grasping and immediately utilizing only a small fraction of the current knowledge. But through mediation they can make use of the knowledge fund, while recognizing the seminal role played by those procedures, paths and instruments paving the way for knowledge acquisition, distribution and utilization, i.e. scientific learning, research and all forms of education. That is why a society which may be called knowledge society is characterized by its recognition of intellectual values, intellectual wealth and all the procedures instrumental in its efficient and responsible use.

We have shown that at the end of this century and millennium various forms of assessment, certain complex clusters of evaluating and decision-making processes have grown to be a legitimate and inomissible component of creative work in science and research. Also humanistic values, namely the values relating to the sustainability of the positive developments of the human race, interpersonal relations, respect for man and his rights, and also for the "rights of nature", have now been asserting themselves to an ever greater extent in the spectrum of values applied in science and research. Basically, equal efforts should now be made to promote the system of feedback, to see to it that a substantial portion of the citizens in contemporary society, regardless of its considerable differentiation or structuration, should be well aware of those values, limits, risks and

potential threats posed by contemporary knowledge. Seen in this light, such a feedback system is one of the key goals of this vision called the “knowledge society”.