First Draft
of a paper presented at the
COMMORG Kick-Off Meeting
Rome, May 3-5, 2001

EVALUATING RESEARCH PROJECTS: ISSUES IN THE EVALUATION OF THE *COMMORG* PROJECT

Moses A. Boudourides

Computer Technology Institute and
University of Patras, Department of Mathematics
265 00 Patras, Greece

mboudour@upatras.gr
http://www.math.upatras.gr/~mboudour/

Introduction

That RTD (Research and Technological Development) expenditure is related to social and economic development is widely acknowledged. At first sight, this relation could be traced as follows: "Expenditure in basic research produces discoveries used by applied research to conceive new products or new processes which generate positive social and economic impacts in the mid-long term" (STOA, 1999). But this scenario is deceiving; for instance, it cannot explain the European paradox that Europe invests less than the USA and Japan in basic research, achieves excellent results in this field but is not able to turn them into patents and economic competitiveness.

Thus, innovation paths should not follow such a simple linear model but instead their trajectories should be rather complex, allowing connections among all components, and, thus, constituting what is better called an 'innovation system.'

Clearly, understanding the performance of an innovative system can lead to its improvement and optimisation of its output in relation to its input. In principle, the goals driving its output should be very clear. For example, as it was declared in the Treaty establishing the European Communities, EU expenditure in RTD activities aims to (i) strengthen the competitiveness of the European industry and (ii) to encourage research and technological development.

Now, the need to evaluate innovative systems arises especially when they are consuming scarce public funding, which has to be justified to society and the public (Georghiou, 1997). This is certainly the case with RTD programmes and projects and, thereby, the importance of their evaluation.

In general, RTD projects are evaluated to the extent to which they satisfy certain demands in terms of attained economic return and produced social impact. Such general demands can be formulated in terms of the evaluation criteria for the effectiveness of the funded projects, which, thus, can be seen as investments from both the economic and the social point of view.

For instance, Michel Callon and his co-workers (1997, p. 11) describe their evaluation criteria as the five vertices of the so-called ‘research compass card’:

1. Production and circulation of certified knowledge (through publications in journals and participations in conferences).
2. Creation of innovations (for the advancement of competitive advantage).
3. Contribution to achievement of the objectives of public authorities (such as collective goods, power, prestige and well-being).
4. Provision of embodied knowledge (through training).
5. Extending the public understanding of science and technology and popularising the relevant expertise acquired through the project.

A similar configuration of evaluation criteria is used by European policy-makers for Community funded RTD investments:

[1] Scientific/technological quality and innovation.
[2] Community added value and contribution to the EC policies.
[3] Contribution to community social objectives.
[4] Economic development and S&T prospects.

To the above four criteria, a fifth one is sometimes added:

[5] Resources, partnership and management.

So, from now on, when we are referring to the *EU evaluation criteria* or *perspectives*, we will mean the above ones.

However, Luke Georghiou warns that the “return-on-investment approach usually presumes a linear/sequential model of innovation whereby the benefits of a single research project lead to or are captured by specific innovations. This is rarely the case particularly for collaborative research” (Georghiou, 1997, p. 20). In other words, if the project results could be directly expressed in terms of economic effects, their evaluation could be easily implemented by classical techniques of financial mathematics.

But RTD projects refer to multiple areas of impact (for instance, industrial competitiveness, quality of life, environmental protection, improvement of employment, as far as FP 5 is concerned). “Socio-economic effects may be manifested not only through competitiveness and exploitation/market-related effects (sales of products, reduction of process costs, etc.) but also through individual and organisational learning effects (including partnerships and networking), influencing norms and standards, generation of externalities, and contributions to skills and research manpower” (Georghiou, 1997, p. 20; see also Georghiou & Meyer-Krahmer, 1992). All these impacts can hardly be expressed in economic terms and require the use of a broad set of indicators.

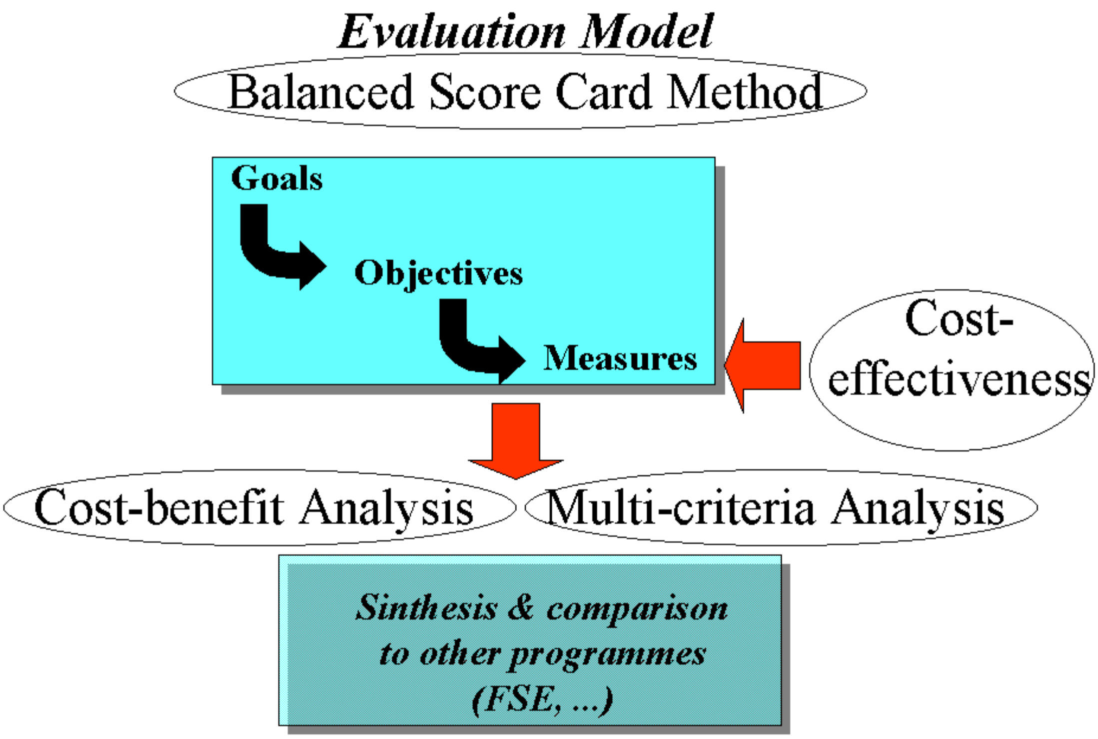
Identification of Measurable Objectives and Performance Indicators

So, to evaluate the project, i.e., to check whether it has met its objectives and it has sufficiently satisfied the evaluation criteria, one needs to possess a formulation of the project objectives in a measurable and verifiable form and to be able to generate consistent indicators, which would benchmark the project performance.

To proceed with the evaluation of RTD programmes EU has developed a project-level toolkit approach known as COMEVAL (Common Methodology for the Evaluation of RTD Results). COMEVAL provides a standard terminology for the collection of data on project aims and outputs and a set of criteria on which to judge project achievements (PREST *et al*., 1996).

In particular, in the evaluation of RTD effectiveness, the EU favours the following methodologies (STOA, 1999):

1. Cost-effectiveness Analysis;
2. Balanced Score Card Method (BSC);
3. Cost-benefit Analysis;
4. Multi-criteria Analysis.



The Cost-effectiveness Analysis evaluates research projects from an economic point of view (both ex-ante and ex-post) by considering the relationship between the achievement of the objectives and resources mobilised. The problem of this analysis is its emphasis on a single criterion, which makes it impossible to deal with programmes containing various research projects with more than one concurrent objective. Rather than an isolated method, the cost effectiveness analysis is usually regarded as one of the performance measures taken into consideration in the Balanced Score Card method and in the Multi-criteria analysis.

According to the Balanced Score Card Method (BSC) method (Kaplan & Norton, 1992), each one of the EU evaluation perspectives has to be equipped with a three-layered structure: goals (or mission), objectives and measures. Goals are defined in the project proposal; objectives too as expected results; and measures should be expressed in quantitative expressions, which reflect the extent to which objectives have been fulfilled. Apparently this structure is estimated from both the output and the outcome of the project. In the case of RTD projects, the outputs are concrete products (such as artefacts, publications, patents, processes, etc.), while the outcomes are the real results of the project, which carry the medium/long term impact of the project. There are four different kinds of measures (or indicators):

?? Outcome measures, which identify the extent to which the project has achieved its goals objectives, met various requirements or professional standards.

?? Output measures, which indicate the number of produced units (such as new products, new processes, new patents, new models, new methods, new tools, etc.) or provided services (such as workshops, conferences, seminars).

?? Efficiency measures, which provide cost (or the amount of other resources) indicators per unit of output (such as cost per participant to seminars, cost per patent, cost per new process, cost per new product).

?? Cost-effectiveness measures, which measure the cost per unit of outcome.

The cost-benefit analysis (Dasgupta & Pearce, 1978) allows the comparison of the sustained project expenses with the quantified project benefits, i.e., the project results translated into monetary terms. In comparison to the cost-effectiveness analysis it is more complex because it quantifies the benefits of all the achieved objectives without limiting the evaluation to a high priority objective. The evaluation of the research is, therefore, carried out comparing costs and benefits and calculating the traditional indicators of economic return of the investment (capital budgeting) such as NPV, IRR, Pay-Back, etc. The cost-benefit analysis approach is suitable when used to evaluate share-cost industrial RTD projects and as such it has been extensively used at NIST (USA). In this context, one usually separates between private and social returns, where the difference between the former and the latter defines what is usually called 'spillover gap.'

The multi-criteria analysis (MCA) (Triantaphyllou, 2000) is a technique for the evaluation of many criteria without translating them all in monetary terms. It has been widely applied in order to assess environmental impacts. Through this technique decision makers are expected to build up a representative model in order to single out the impacts of different investment options on their strategic objectives (goals) and on external variables. Usually in EU evaluations, the same indicators with the BSC method are used in multi-criteria analyses too (STOA, 1999).

As a remark, in the project evaluation process, one should be aware of the need for those objectives and indicators, which could respond flexibly to changed internal and external (to the project) circumstances. For instance, it might happen that unexpected spin-offs are so important that they can modify the context of the original targets (Massimo, 1997, p. 45-46).

Furthermore, one should recall that project evaluation is based on accumulated project data. Two kinds of such baseline data are needed for evaluation: Technical data on the planning, progress and termination of the various project phases and more detailed data on the performance of the project. These data need to be collected systematically and updated regularly in order to be able to map properly the degree of efficiency and coverage of the planned goals, tasks and deliverables of the project. As a result, the above project evaluation methodologies favoured by EU might proceed in different steps and phases of the project according to the produced data at each step or phase.

Finally, let us highlight that project evaluation should play an important role in project management by providing a reflexive, sensitive and accountable mechanism of adaptation to the evolving conditions of the project implementation (although the anticipation of possible significant risks and contingency plans should be also incorporated in the project workplan). In this sense, processes of project evaluation should contribute to an integrated project management. This is exactly the content of the move towards 'new public management' with its general emphasis upon accountability. Such a deployment of evaluation processes was already manifested in the United States through the Government Performance and Results Act of 1993 (Cozzens, 1995) but, in particular, it was applied in the RTD management of information systems (King & Lake, 1989; ? sell, 1987).

Monitoring and Self-Observation of the Project Performance

The idea that any project constitutes a network is not new, although the network approach for the understanding of science-technology-innovation (STI) systems has been applied rather recently (Callon *et al*., 1999). Furthermore, social network analysis has been along ago applied to resolve fundamental sociological questions about social evaluation (Gartrell, 1987).

In particular, as far as the internal evaluation of the project is concerned, we could apply the methodology developed by Saadi Lahlou (1997) in order to measure the network effects in the evaluation of the EU Science and Stimulation Programmes (SSP). This is a quantitative method, which can compare the behavioural relationships between the institutions-partners participating in a project before, during and after the implementation of the project. Networks are described by calculating a 'mean value' of relationships between pairs within the network. Monovariate and multivariate statistical analyses can produce quantitative indicators, which allow comparisons in different phases of the project and, thus, provide useful indicators for the evaluation of the project.

Lahlou's method is based on data collected by questionnaires, which are completed by the participant institutions in the project during different time periods. In these questionnaires the following information about contacts between pairs is solicited (Lahlou, 1997, pp. 365-8):

?? Number of contacts
?? Occasional contacts
?? Personal contacts
?? Exchange of research material
?? Co-authoring
?? Regular seminars
?? Permanent collaboration
?? Sharing funds

Besides methods of multivariate statistics, nowadays it is very popular to use methods of *Social Network Analysis* in order to understand the dynamics of relationships developed in networks of collaboration such as in projects. A social network is a set of actors and relations occurring among them. Actors can be individual people, objects or events as far as certain relations hold them all together; actors can be also aggregate units such as organizations, institutions, communities, groups, families etc.

The very idea of the social network approach is that relations or interactions between actors are the building blocks or the key factors that sustain and define the network (Wasserman & Faust, 1994).

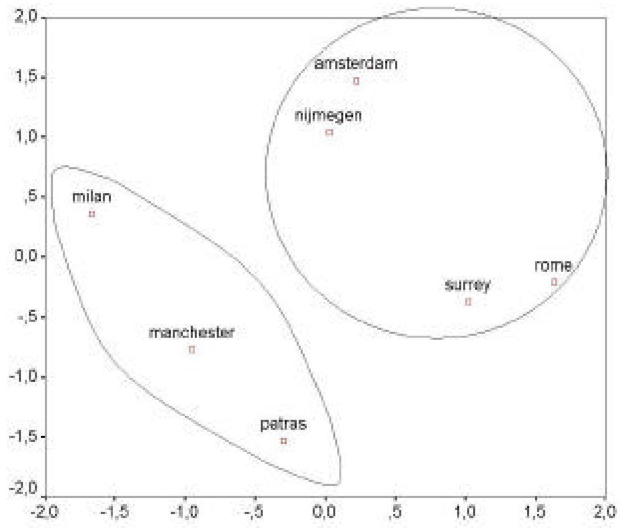
For instance, Haythornthwaite (2000) collected data from four computer-supported distance-learning classes in order to build a picture of the size and composition of personal online networks. The relevant questions were referring to the maintenance of four types of relationships through work and non-work interactions:

?? Collaboration on class work
?? Exchange of information or advice about class work
?? Socialising
?? Emotional support

Furthermore, students reported their frequency of communication with each member of class via each of the available means of communication: Webboard, IRC, Email, telephone, face-to-face scheduled or unscheduled meetings or other media they indicated they used. Apparently, Haythornthwaite's methodology can be applied into a project setting by interviewing project participants on similar questions about their contacts and the uses of various communication technologies in them.

Another approach, inspired from social network analysis that we could follow in the self-evaluation of our project concerns the communicative activities of the partners-individuals inside the mailing lists for the project operation and coordination. As the partners are exchanging e-mails into these mailing lists, they participate in multiple threads of discussions. By processing the archives of the project mailing lists, we can extract data on the partners' participations in such threads of messages and from these data we can derive the corresponding networks (or graphs or sociograms) of interactions among the partners. In these networks, the proximity between nodes (representing partners) is manifesting the extent of their interaction through their participation in a large number of common threads of discussions. On the other side, a possible isolation of certain nodes (partners) will indicate their reluctance to interact within the group of partners through the project mailing lists. Needless to say that this measure of interactivity and, thus, of the evaluation of the project communicative performance is closely pertinent to the object of the studies of this project, i.e., e-mail and its uses and adoption throughout organizational processes.

For instance, we have applied the above ideas to the discussions in the mailing list COMMORG, which took place from September 2000 to March 2001. The communicants in the mailing list were grouped according to the location of the participant institution at the project: Rome (coordinator), Milan, Patras, Amsterdam, Nijmegen, Manchester and Surrey. The following sociogram maps the relational positioning of communicants according to their interactions in the mailing list through their participation in threads of messages (obtained by MDS and the clustering in two groups having resulted by SPSS factor analysis):



Furthermore, the above analysis can be performed in different time periods in order to understand the time evolution of the communicative patterns emerging in the mailing list.

Finally, we should mention that scientometric methods (which in particular can be considered as network methods) have been already used to measure the accumulation of knowledge and to evaluate the dynamic structure underlying various research activities (Callon & Courtial, 1997).

Peer-Reviewing

A usual practice of the European Commission is to use a panel of independent experts as external evaluators. The Commission calls this method 'peer evaluation' by analogy with the scientific 'peer review' method, by which the scientific value of proposed projects is examined by specialists in the same field. In fact, it was in 1978 when the use of the 'peer review' method was first put forward as an appropriate way to conduct the first programme evaluations (Bobe & Viala, 1997, p. 29).

According to Luigi Massimo (1997, p. 42), the choice of panel members constituting the body of experts as peer reviewers is the most delicate part of the evaluation of a project, because it influences both the efficiency of the evaluation but also its credibility. The need to have independent evaluators follows from the respect to democratic principles of decision-making. Certainly, these panel members should not directly benefit from the project and, in order to ascertain their critical disposition, sometimes it is recommended that they represent alternative poles in controversial cases. Through their detachment from the project, they can suggest and introduce new ideas and also challenge the schemes of thought of the project group when the latter seems be stagnant in the implementation of the project.

Besides the evaluation of the scientific output of the project, the panel of evaluators can also measure the impact of the project on innovation, on European policies, on cooperation and on dissemination of the acquired knowledge. In other words, the peer-reviewers may utilize the measurable objectives and the performance indicators, which have been defined and constructed above. This means that they need to evaluate the collected data for the project performance and efficiency to produce the planned goals, tasks and deliverables of the project. Therefore, all preparatory work either on the conceptual construction of the project implementation indicators or on the reflexive self-evaluation of the network structure and effects of the project has to be available to them.

References

? sell, P. (1987). Evaluation of cooperative research. In E. Ormala (ed.), *Evaluation of Technical Research and Development*, pp. 74-87. Espoo, Nordforsk.

Bobe, B., & Viala, H. (1997). Panels of experts of the European Commission. In M. Callon, P. Larédo & P. Mustar (eds.), *The Strategic Management of Research and Technology: Evaluation of Programmes*, pp. 29-40. Paris: Economica International.

Callon, M., Larédo, P., & Mustar, P. (eds.) (1997). *The Strategic Management of Research and Technology: Evaluation of Programmes*. Paris: Economica International.

Callon, M., & Courtial, J.-P. (1997). Using scientometrics for evaluation. In M. Callon, P. Larédo & P. Mustar (eds.), *The Strategic Management of Research and Technology: Evaluation of Programmes*, pp. 165-219. Paris: Economica International.

Callon, M., Cohendet, P., Curien, N., Dalle, J.-M., Eymard-Duvernay, F., Foray, D., & Schenk, E. (eds.) (1999). *Réseau et Coordination*. Paris: Economica International.

Cozzens, S. (1995). U.S. Research Assessment: Recent developments. *Scientometrics*, vol. 34, no. 3, pp. 351-362.

Dasgupta, A.K., & Pearce, D.W. (1978). *Cost-Benefit Analysis: Theory and Practice*. London: Macmillan.

Gartrell, C.D. (1987). Network approaches to social evaluation. *Annual Review of Sociology*, vol. 13, pp. 49-66.

Georghiou, L. (1997). Issues in the evaluation of innovation and technology policy. In the OECD Report, *Policy Evaluation in Innovation and Technology: Towards Best Practices*, pp. 19-33.

Georghiou, L., & Meyer-Krahmer, F. (1992). Evaluation of socio-economic effects of European Community R&D Programmes in the SPEAR Network. *Research Evaluation*, vol. 2, no. 1, pp. 5-15.

Haythornthwaite, C. (2000). Online personal networks: Size, composition and media use among distance learners. *New Media and Society*, vol. 2, no. 2, pp. 195-226.

Kaplan, R., & Norton, D. (1992). The balanced score card: Measures that drive performance. *Harvard Business Review*, vol. 70, no. 1, pp. 71-79.

King, J., & Lake, J. (1989). The role of management information system in the assessment of biological and agricultural research. Communication at the ECE Seminar on Evaluation in the Management of R&D. Madrid.

Lahlou, S. (1997). Measuring the network effects of European stimulation/science programmes. In M. Callon, P. Larédo & P. Mustar (eds.), *The Strategic Management of Research and Technology: Evaluation of Programmes*, pp. 361-383. Paris: Economica International.

Massimo, L. (1997). Lessons from the experience of the European Commission. In M. Callon, P. Larédo & P. Mustar (eds.), *The Strategic Management of Research* *and Technology: Evaluation of Programmes*, pp. 41-53. Paris: Economica International.

PREST/Smith System Engineering (1996). Common Methodology for the Evaluation of RTD Results. Report to DGXIII, September 1996.

STOA Study (1999). Cost-benefit analysis of EU Research and Technological Development (RTD) Programmes. PE 168 439. European Parliament. http://www.europarl.eu.int/stoa/publi/168439/default_en.htm

Triantaphyllou, E. (2000). *Multi-Criteria Decision Making Methods: A Comparative Study*. Dordrecht: Kluwer Academic Publishers.

Wasserman, S., & Faust, K. (1994). *Social Network Analysis: Methods and Applications*. Cambridge, UK: Cambridge University Press.