

Evaluation of the impact of R&D collaborations between public research laboratories and industrial firms : the building of an operational tool based on empirical cases, the example of the CEA.

Serge PETIT

Communication to

European Meeting on Applied Evolutionary Economics

7 - 9 June 1999, Grenoble, France

Organised by the Institute for Energy Politics and Economics
Organisé par l'Institut d'Economie et de Politique de l'Energie /
IEPE, BP 47, 38040 Grenoble Cedex 9, France

And the INRA-Unit of Sociology and Economics of Research and Development
Et l'unité Sociologie et Economie de la Recherche Développement de l'INRA
INRA/SERD, BP 47, 38040 Grenoble Cedex 9, France

« Evaluation of the impact of R&D collaborations between public research laboratories and industrial firms : the building of an operational tool based on empirical cases, the example of the CEA. »

Serge PETIT

PhD student in Economics at I.M.R.I.
(Institut pour le Management de la Recherche et de l'Innovation)
Paris-IX-Dauphine ;
Supervisor : M. Dominique FORAY (Research director at CNRS)

Located at :
CEA/DTA (French Nuclear Authority / Division of Advanced Technologies)
17 rue des Martyrs - 38054 Grenoble Cedex 9
Tel : 04 76 88 93 29 - Fax: 04 76 88 46 66 - e-mail : serge.petit@cea.fr

Paper presented at the
EUROPEAN MEETING ON APPLIED EVOLUTIONARY ECONOMICS
Grenoble, 7th-9th June 1999

Abstract :

Public laboratories are increasingly expected to evaluate the economic impact of their research. Existing evaluation models are poorly equipped to grasp the very characteristics of innovative activities. Expectations concerning the results that evaluations are able to produce are generally unrealistic. Reminding basic properties of technology and suggesting a renewed conception of evaluation, the paper suggests criteria that could serve as a basis to evaluate the economic impact of R&D collaborations more accurately in the future. These criteria come out of the explorative evaluation of twenty two R&D collaborations between CEA laboratories and industrial firms. The links between the explanatory (structural) variables

and the innovation-induced effects are currently being tested and will be made available later on.

Introduction

Research laboratories are increasingly coming under pressure to evaluate the social utility and quality of their research. After a decade of multiplication of technology dissemination devices bount to foster the innovative efforts in firms (1980-90), public policy makers are seeking to evaluate the impact of these means on the economy. The CEA has been attributed a mission of technology dissemination towards industrial companies in order to contribute to industrial wealth creation and unemployment reduction as overall objectives.

Ever since its origins, the CEA has been involved in collaborations with industrial firms : initially through subcontracting, then, and before all, through R&D collaborations within industrial partnerships and within the technology dissemination programme specifically geared towards SMEs. In this mission, the CEA transfers technologies, competencies and research results towards its industrial partners.

The requirement is for an evaluation system of easy-accessible indicators that allow the CEA to assess on a regular basis the economic impact of those industrial innovations it is participating in. The aim of the paper is to sketch out the framework for such impact evaluation indicators.

The field of investigation is limited to formal technology transfers that associate a research laboratory with a private manufacturing company (called the firm in the paper). The other channels through which the CEA disseminates knowledge, such as technological services, patents or the CDT network¹ will not be discussed. The R&D collaborations under investigation are licensing agreements between the laboratory and the contracting firm.

Different impact assessment models are currently being used. CARR classifies them into four different types :

The « out-of-the-door » model is concerned with counting the number of information support devices that leave the research institution and inform the corporate world about new technological achievements. Publications, signed licensing contracts for instance are measures of such a model. Easy to establish and widely accepted as measures of activity, these indicators tell nothing about the quality of the disseminated content.

¹ CDT = « Conseiller en Développement Technologique » : CEA is present in the regional offices of ANVAR which is the National Agency for the Economic Development of Research through 20 engineers.

The « in-the-door » model is a variant of the « out-of-the door » model. It is mainly concerned with measuring the patents that the institution takes. This indicator can be a productivity measure of the research being carried out. The portfolio of patents held stands for the accumulated knowledge base of the institution. To make the number of patents significant, it is necessary to related it to the amount of royalties generated or to the number of commercially exploited patents. Intersectoral and interdisciplinary differences in patenting (pharmaceuticals versus production or software), the fact that a lot of innovations come out of informal R&D activities ask for cautious interpretation of the patenting data².

The « political model » aims through the presentation of particular success stories to show the type of achievements made possible through an R&D collaboration between a lab and a firm. Its value is essentially promotional and as such of limited use for project and programme management.

The « opportunity-cost model » bases on the cost of a technology dissemination programme and tries to establish which other activities could have been carried out with the funds made available. Its drawback is that generally no basis of comparison is readily available.

The « market impact » model aims at establishing the commercial impact of a technology. As commercial exploitation is the ultimate aim of technology, this model is crucial. Several measures can be used to measure this impact :

- a. the amount of royalties due to the institution on the basis of the technology-based business results ;
- b. the technology-induced turnover that the firm generates under the licence agreement ;
- c. the number of start-ups created and new jobs created through technology transfers ;

On the one hand these indicators are easily applicable to a variety of different situations and technologies and that is probably their main advantage. On the other hand, they are past-oriented and none of them takes into account the indirect effects of technological development. Of course, there are drawbacks to each one of the presented models. All have got a dedicated scope of application, and need to be interpreted in their context. But whereas some drawbacks are mere measurement issues (how measure the quality of an article's content ? or a more

² PATEL & PAVITT (1987)

tricky measurement problem : how measure the contribution of the technological component in corporate turnover ?) which can be overcome by interpretation and experience, others are conceptual issues that are deeply anchored in the way technology is perceived and in established evaluation practices.

Our aim is to address these more fundamental issues as they are the ground on which technology evaluation models generally rest. We will introduce the evaluation issue by discussing the way technology is generally looked at in traditional economics and identify some of the reasons why this conception is inappropriate (Section 1).

We then try to show that the assumptions that most evaluations implicitly rely on are out of touch with programme reality (Section 2). Our assumption is that the conjunction of these two factors gives birth to poor evaluations.

With this knowledge in mind and on the ground of empirical research, we try to *build up an operational evaluation tool that should enable programme administrators* to evaluate the impact of the technology dissemination programme more properly than what the current evaluation models can achieve (Section 3).

1. From a traditional to a more comprehensive conception of technological development

The theoretical argument that underlies this paper is the acknowledgement that technology cannot be reduced to a piece of formalised information, but that technology is made of three constituents that can exist independently of each other (LAYTON 1974). The « artefact » dimension is the aspect that firms are generally most concerned about, as improvements in the production techniques are perceived as a condition *sine qua non* in their ongoing search for competitiveness. Evaluations generally take into account the artefact dimension leaving aside the other dimensions of technology, namely skills and knowledge, notwithstanding the fact that technological development draws heavily upon them. By doing so, the traditional evaluation tools underestimate the impact of technology as it enters an organisation. GEROSKI et al. (1993) showed that disembodied knowledge spills over more easily (to other activities) than knowledge embodied in use-specific products. This argument further legitimates the idea that the impact of technology is likely to trespass the scope of traditional evaluation. If you add on

top of that the possibilities of recombining different technological bricks with each other, then technology ends up as a potentially pervasive phenomenon that is improperly captured by prevailing evaluation tools.

A second theoretical argument concerns the way technology is developed. Different features make technological development a very specific topic in economics. Technological development cannot be separated from its diffusion. They are two entangled phenomena that make technological development an essentially decentralised process that associates a whole array of actors linked through different interfaces. As a consequence, it is difficult to separate an individual firm - as unit of analysis - from the subsystem to which it belongs without neglecting the very characteristics of technological development itself.

Technological knowledge is accumulated through a large number of *parallel development processes* that are generated during the execution of firm-specific routines and through the relationships that different actors build up. The results of these development processes disseminate through several interaction mechanisms³. Knowledge is further developed through its use and the associated learning processes and the subsequent innovations diffuse through the interactions between actors.

Evaluation of technological development should take into account these parallel processes in conjunction with the merchant and institutional environment in which they are embedded. As S. METCALFE puts it: «It is not individual innovations which should be the focus of attention but the conditions which draw forth a sequence of innovations from particular design configurations in a process that is strongly shaped by the connected diffusion processes. In short, technology policy should focus on co-evolving technological and market environments, not upon individual innovations⁴». The choice of the unit of analysis is a major issue in evaluation practice. Unfortunately evaluations often focus on a unit of analysis that is narrow (for instance, a firm) or broad (for instance, the impact of public research on the economy).

Innovative products and processes are the result of knowledge inputs from a variety of internal and external sources. The combinatory effort bringing together the different components has to

³ AUTIO, LAAMANEN (1995), p.661

⁴ METCALFE (1995), p.485

conform to multiple performance constraints. These properties make technological development an inherently *complex process*. Theoretical knowledge and scientific laws left on their own are generally unable to supply the knowledge that is necessary to make complex technological artefacts work under real-life conditions. Tacit knowledge accumulated by experience and learning by doing is necessary not only for imitation but also for innovation processes.

PATEL and PAVITT⁵ show that this conception of technology sheds new light on a couple of assumptions that underlie traditional measurement indicators of technological activities.

- a. technology is not adequately described as an easily transferable set of codified information.
- b. barriers to transfer technology between organisational borders are no due to the withholding of codified knowledge, but because the recipient firm has not accumulated the necessary knowledge basis that allows technology absorption and subsequent adaptation to local conditions.
- c. as a consequence of what has been said and because of different corporate constraints, technology is exploited in a variety of different ways, urging the observer to cope with *technologies* instead of *technology*.

Several authors have pointed at the inadequacy of traditional economic tools to deal with technology-related issues (DAVID 1975 ; ZUSCOVITCH 1986 ; LEVIN, KLEVORICK, NELSON, WINTER 1987; PAVITT 1987 ; DOSI 1988). Cumulative and tacit learning, complexity, specific and complementary assets (TEECE 1986), uncertainty and contingency, sectoral patterns of innovation (DOSI 1988) and appropriation regimes (LEVIN et al. 1984) are just a couple of the characteristics of technological development.

Traditional economics generally devotes relatively little attention to these characteristics, treating technology as mere information. More concerned with adjustment processes of existing resources, traditional economics is poorly equipped to deal with creation phenomena as those displayed during innovation processes. Several authors have shown that these creation processes do not correspond to one single way of proceeding, but that there is a variety of innovative behavioural patterns. As ZUSCOVITCH puts it : « In reality, each industry is built on a particular mix of innovative activities ranging from fundamental science to learning by experience, with many different ways of producing and spreading endogenous technical change in between. As different forms of economic appropriation are associated with different

⁵ PATEL and PAVITT (1995), p.17ff

innovative activities (LEVIN et al.1984), it follows that there is no single performance estimator (such as productivity, net worth, assets and so forth)⁶ ».

Traditional economics treats firms as representative agents that lack the capacity of strategic behaviour. Firms are assumed to face the same production constraints according to the relative prices of production inputs. Several authors among which NELSON and WINTER 1982 and DOSI 1988 show that there are « broad differences in opportunities, incentives, R&D investments, and innovative procedures among industries (...) » and that they are important enough to « help explain the internal structure of the complex engine which in modern non centrally planned economies continuously generates new products and production processes⁷ ».

2. Implicit assumptions underlying evaluation practice

This section draws upon the methodological insights of public programme evaluators⁸. It shows that the traditional assumptions - of who benefits from the evaluation (theory of utilisation), how the programme should be valued (theory of value), how evaluation should be done (theory of practice), what knowledge it should produce (theory of knowledge), what utility it should have (theory of utility) - in spite of their shortcomings, survive and shape programme administrators expectations towards programme evaluations.

These assumptions were that programmes are homogeneous and that their objectives are perfectly explicit. The (retrospectively) pretentious perspective was that it be possible to measure programme achievements through feasible and reproducible experiments that allow no biased interpretation. Successful evaluation results were assumed to provoke increases in programmes budgets. Bad evaluation results would provoke fund restrictions. The aim of any evaluation is to track the causal relationships between programme elements and programme effects. Theory should be concerned with identifying the key variables that allow to track the causal implications. If causality cannot be established with certainty, evaluation should try

⁶ ZUSCOVITCH (1986), p.179

⁷ DOSI (1988), p.1150

⁸ For an excellent view of the different approaches to programme evaluation, see for instance SHADISH, COOK, LEVITON (1991).

through correlational techniques to link potentially influencing factors with changes in the programme results. As gained insights should be generalisable, it is important that study samples be random. To the extent that programme outcomes are known beforehand, evaluation is more confirmatory than exploratory.

Each one of these principles has been challenged upon theoretical and empirical grounds.

2.1. Evaluation's utility

The results of the first evaluations have obliged evaluators to review their ambitions.

In the first place, inertia of social reality is such that it is difficult to induce changes through public programmes. Typically unemployment reduction and wealth creation are among the least manageable social problems as the underlying mechanisms are not clearly known. No single programme can be expected to have a decisive impact on an epidemic problem such as unemployment. Hence evaluation should not rate programme performance too severely on the grounds that initial expectations have not met.

The second deception has its roots in the nature of the problem solving process itself. Programme objectives are often vaguely formulated and the problems they are supposed to alleviate are ill-defined. Programme implementation is heterogeneous and hence generates different effects according to local implementation conditions. The effectiveness of programme implementation is an important factor in the impact generation. Thus evaluation of a programme needs to take into account the *process* as a crucial momentum.

Considering these elements, among the range of change strategies that evaluations could potentially provoke, the most likely to be applied by programme administrators is the « incrementalist strategy », i.e. incremental change by improving programme elements that appear deficient. But even this strategy's effectiveness seems limited when confronted with vague objectives and loosely formulated programme assumptions.

2.2. Evaluation's utilisation

One of the great hopes of early evaluators were that their results would induce policy reorientation. The fact is that no programme has ever been stopped because of unfavourable evaluation results. As programme funding is the result of a complex political decision process, it is indeed unlikely that evaluation results alone reorientate the allocation decisions, and if they

did, the influence would not be detectable. Once a programme is launched, by the time it ends offering scope for reorientation, evaluation results are generally too old to be relevant, or influential stakeholders may have moved on. In both cases, evaluation results may be left out of the next decision process. Especially in France, where hardly any innovation-fostering instrument has ever been revoked, the probability that evaluation results have incisive consequences on the very existence of the programme is quite low. The principal benefit of evaluation is probably the implementation improvements that the results suggest to programme administrators.

Although evaluation results probably have a limited impact on the political decision process, the ways in which the results can be recuperated by other stakeholders are quite numerous. As a consequence, it is important that evaluation delivers high-quality results.

2.3. Theory of practice

If all evaluators agree on the fact that evaluation should improve social reality, differences remain concerning the way this objective should be achieved. Three different tendencies exist :

For the « manipulationnists⁹ », it is less important to know why a programme works than the fact that it works. The assumption that implicitly underlies this approach is that existing evaluation techniques can identify operational solutions and that once they are identified, they are largely adopted. Although this approach has been criticised, it continues to be largely used. The ongoing search of central decision makers for « best practices » witnesses that this approach is still widely referred to.

The « explicationnists¹⁰ » differentiate themselves from the first approach. They consider reality as being ontologically complex. They recognise that certain effects may appear under certain conditions and appear in a completely different setting under altered conditions. For them, rather than producing results that can be instrumentally used, evaluation is supposed to produce « lobby-independent enlightenment », i.e. new knowledge about the programme's reality, notwithstanding the fact that the conclusions are unlikely to be applicable as such. Last but not least, whereas « manipulationnists » consider impact identification as a good enough reason for evaluation to be carried out, « explicationnists » aim at answering a wider array of

⁹ SCRIVEN , CAMPBELL

¹⁰ WEISS, CRONBACH

questions trying to establish models that link programme assumptions with effects. This approach rests on the strong assumption that it is possible to construct this kind of theoretical models. What is more, it assumes that the production of a theoretical piece of knowledge guarantees its transfer to other social situations, which is questionable to a certain extent.

The defenders of a third approach¹¹ consider that an evaluation is best used if shaped to the information needs of a particular stakeholder group. Controversy exists around the identity of the group that should be privileged. Defenders of this approach, knowing that there is no such thing as a value-free evaluation, consider that producing evaluation in the « public interest » is ambiguous and inefficient. For them, stakeholders are in the best situation to decide which problems are important and what solutions should be implemented. The question here is whether stakeholders are indeed able to come up with the relevant topics or whether their understanding is not « necessarily » biased towards their day-to-day concerns. If so, the produced knowledge remains local and difficult to transfer to other social situations.

At this point, it becomes quite obvious that there is a gap between the expectations of a programme administrator and the ambitions of the evaluator himself in terms of what evaluation is able to provide and of how evaluation results can be used. Typically the former rather expects evaluation to produce directly operational solutions (« manipulationnist » perspective) even at the expense of the fine details whereas the latter would tend to produce « scientifically-valid » models (« explicationnist » perspective) that are universally true. This gap has some implications for the type of knowledge that is being produced.

2.4. Theory of knowledge

As indicated, programme administrators would rather belong to the « manipulationnist » strand, and evaluators to the « explicationnist » strand. This inevitably creates a tension between on the one hand the need to understand the growing complexity of R&D and innovative activities and on the other hand the demand for easily accessible and general performance indicators. In their efforts to gain insights, evaluators find that :

There are huge variations between different projects within a same programme and this heterogeneity forces evaluators to give up the idea of finding simple causal relationships between programme elements and effects.

¹¹ J. WHOLEY, R. STAKE

It is very hard to capture all programme effects in preestablished evaluation grids. Evaluators increasingly let effects and impacts emerge from the observation of reality. This new conscience fosters the use of qualitative methods and inductive conceptualisation. More generally, there is a growing tendency for evaluators to use a range of different evaluation methods. This creates a certain strain as knowledge production is prolific, but the different contributions do not easily fit into a commonly accepted frame. GEORGHIOU and MEYER-KRAHMER consider that economic evaluation still is in a preparadigmatic phase¹².

2.5. Theory of value

Theory of value is the heart of every evaluation as it raises the question of the evaluand's value. Value setting basically rests on the following four steps :

1. Development of merit criteria that specify what a good evaluand is supposed to do
2. Selection of performance standards that explicit the expected performance
3. Performance measurement to see whether the standards are attained
4. Integration of results into a unique statement that determine the value of the evaluand

The first step is to develop the *criteria* against which a programme is evaluated. At first sight, the objectives of the programme appear as the natural criteria that could be used for a « rapid feedback evaluation¹³» (RFE) that verifies whether agreed-upon performance criteria have been met by the evaluand (i.e. innovation-induced turnover, number of jobs created or maintained, ...). Implicit in the RFE is the possibility to identify ex ante a « menu » of effects that are regularly produced whenever an innovation is produced. Useless to say that there is no such thing as two identical innovation processes generating exactly the same type of effects. The very formulation of criteria implicitly acknowledges that there is a *theoretical impact space* in which the effects of the programme occur. If this assumption may possibly hold in case the evaluand is a well-understood tool from which surprising results are unlikely to emerge, it seems highly speculative in our case.

The formulation of these criteria is quite important. Indeed, if too general, it is difficult to bring them into play and more precise criteria have to be found. A complementary approach is needed.

¹² GEORGHIOU & MEYER-KRAHMER (1992), p.6

¹³ J. WHOLEY

HENDRICKS suggests such an approach, namely the « service delivery assessment » (SDA), according to which the evaluator tries to identify effects as they are produced independently of pre-set assessment criteria (i.e. externalities on suppliers,...).

Different methodologies can be used to carry out a SDA depending on the context, the precision that is requested and the costs/benefit function of the evaluation user. The scope of data collection methods reaches from light forms of evaluation (i.e. check-lists) to heavy and time-consuming methods (i.e. case studies).

The second step concerns the construction of *performance standards*. There are two different ways of constructing these standards. Or, you compare the programme's performance with absolute and pre-set performance standards, or, you compare the programme's performance with the performance of other evaluands. Some authors are in favour of a comparison with other evaluands, others consider that this is a delicate exercise as it assumes different programmes to be functionally equivalent. A question mark remains anyway on which evaluand should be mobilised for this type of exercise.

The third step concerns the *performance measure*. Two different conceptions exist : the first considers that evaluation needs to demonstrate that it is the evaluand and not a correlated force that provokes the changes observed in the criteria. The idea is to measure the performance of the programme itself. According to the second conception, evaluation does not need to establish causality. It is enough that changes in the criteria occur.

Any academic's ambition would clearly be to establish causality. For the programme administrators however, considering the existing impact evaluation indicators and the stakes in the research they fund (the underlying research to this paper), causality seems to be a less important issue than the knowledge about (all) the effects and impacts that are induced through the programme. In any case, independently of the evaluator's ambition and the administrator's requirement, the information that can be accessed through monographs (mainly document analysis and face-to-face interviews) does not allow to build causality. To build causality, very heavy and time-consuming case studies have to be carried out. On-site immersion and real-time observation of the decision processes are probably necessary to establish this kind of knowledge.

The final step of the evaluation is the *integration of results* into a unique value statement. Two different conceptions exist : some authors like M. SCRIVEN consider that evaluation is not ready until all individual measures have be « translated » into a single metric that allows for results to be summed up. Others like CRONBACH consider that it is more effective to leave the results under their original metrics. Indeed, even if individual results are challenged by a reader, and this is likely to happen, this does not scatter doubt on the other results. What is more, the translation into a single metric assumes that aggregated data are homogeneous when they are not. In spite of good arguments against the integration of results, programme administrators generally prefer aggregated measures.

The evolution from early programme evaluation assumptions to a more realistic picture is summarised in the following chart :

	Original conception of evaluation	Renewed conception
its utilisation	<ul style="list-style-type: none"> • policy reorientation • policy makers / programme administrators 	<ul style="list-style-type: none"> • input into political debate • no monopoly on utilisation
its utility	<ul style="list-style-type: none"> • monitoring social problem solving • essentially external utility 	<ul style="list-style-type: none"> • monitoring programme contribution to problem solving • perhaps internal utility
its practice	<ul style="list-style-type: none"> • instrumental use of transferable solutions • input-output focused 	<ul style="list-style-type: none"> • enlightenment in an ontologically complex world • process modelisation
its value	<ul style="list-style-type: none"> • value-free evaluation • programme objectives as performance criteria • absolute or relative performance standards • causality-based performance • integration of results 	<ul style="list-style-type: none"> • choice of evaluation method and perspective introduces value statement • difficulty to identify an impact space ex ante • what comparative basis ? • how establish causality ?
its knowledge	<ul style="list-style-type: none"> • confirmatory • possibility to identify causal links • construction of perfect knowledge 	<ul style="list-style-type: none"> • exploratory • at best identification of general forms • trade off cost-time / quality of information

It is clear that this renewed conception of evaluation can not be thought independently of new methods of data collection and processing. And indeed, evaluators increasingly rely on different methods exploring the scope of methodological variety reaching from statistics to case studies.

After the discussion of different evaluation conceptions, we suggest the kind of evaluation methodology that is necessary to evaluate the industrial impact of CEA's R&D collaborations.

3. Towards an operational tool for the evaluation of R&D collaborations

We aim at producing an operational tool, but there are a couple of theoretical issues that should not get lost in the process. Properties as complexity, cumulateness, tacitness of technology should find some kind of recognition in the evaluation criteria we develop.

3.1 The constraints of programme reality

The construction of such an evaluation tool faces the following constraints :

If one allows technology to reach its full-blown exploitation regime, evaluation cannot intervene right after the collaboration, but it needs a period of « incubation » which we think to last at least for four years. Hence the empirical sample encompasses collaborations that took place before 1995. In the same time, CEA's technological dissemination programme is - in its current form - a recent mission (beginning of the nineties). And although CEA has always been collaborating with industry, its early collaborations were heavily marked by the working procedures and institutional incentive schemes of its mission-oriented nuclear research. Hence, the features of the collaborations in the sample do not necessarily relate to the current programme's general outline.

As a general feature, filed data on R&D collaborations are scarce : beside the archived contracts, there are hardly any data available. To make things worse, the memory of the early collaborations (end 70's - beginning 80's) is progressively vanishing as relevant interlocutors retired or moved on to new jobs. Retrospectively, access to data has proved to be a labour-intensive and time-consuming exercise.

The evaluation suffers from the lack of clearly formulated collaboration objectives. Official programme objectives are too general (contribution to industrial wealth creation and unemployment reduction) to be used for individual project evaluations. The laboratories' working guidelines are exclusively formulated in technical terms. The collaboration contracts between the laboratories and the contracting firms do not mention any specific objectives either. Thus, it is unclear which criteria projects could be evaluated against. Beside the evaluation indicators presented at the outset, no other or former evaluation exists that could help the search for evaluation criteria or serve as a comparative basis. Lacking this experience and considering the properties of technology (see Section 1), it would be peremptory to assume that the technology dissemination programme could yield no surprising results.

Under these conditions, influential evaluators like J. WHOLEY would raise the question whether this programme is « evaluable ». For him, evaluation needs to rely on « Program Design Model » that specifies the allocated resources, the activities that are bound to take place, the expected results and the assumed causal links activities and programme objectives.

As performance indicators do not exist by themselves, but are social constructs on the basis of what is to be measured, they have to be developed. But in the absence of readily available evaluation criteria, the construction of an evaluation tool rests on a phase of previous data collection. Hence, it seems natural for the purpose of our research to carry out an SDA-like evaluation based on an *exploratory investigation of programme reality*.

3.2. Data collection and methodology

Empirical data were gathered through a series of project monographs that included document analysis on the one hand and face-to-face interviews with project leaders on the other hand.

In line with the methodological background of SCRIVEN, we explored the logical requirements of the task rather than relying on a pre-set formal methodology. We investigated general topics (motivations to collaborate with CEA, merits of CEA technology, induced effects, ...) rather than asking responses to precise preconceived questions. The face-to-face interviews allow interviewees to make comments and remarks they would never come up with if they were to be interviewed through other channels such as the phone or a formal questionnaire. What is more, a couple of contextual - often very qualitative elements - that are necessary to understand the relationship between the laboratory and the firm could not be

caught unless people actually meet. These discussions allow to find out about potential impacts and links between R&D and effects.

The empirical data set is made of twenty two « successful » R&D collaborations with industrial partners of different sizes. The collaborations all took place over the last twenty years (1975-1996). The technological domains covered are very diverse (microelectronics, optronics, material sciences, industrial process optimisation, ...). For the details of the sample, see appendices 1 and 2.

The quality of insight into the different companies is uneven, depending on the number of people that we were able to interview (between 2 and 7 per monograph). In order to have the most reliable information possible, the interlocutors were asked to validate the written report of the interview in the aftermath. As a general rule, the targeted interlocutors were the projects leaders themselves, be it in the laboratory or in the firm. Generally speaking, their co-operation was good considering that the collaborations often laid far back in time.

The empirical investigation produces innovation scenarios and a list of induced effects. A rough estimation tells that not all effects are captured by the existing impact assessment criterion (i.e. reputation effects, ...). The latter certainly is a useful indicator as it provides an objective measure to very heterogeneous situations, but it reflects only the direct commercial impact of the collaboration-induced innovations regardless of other effects (indirect effects, long-term benefits, ...) that might have appeared throughout the innovative process.

If this evidence is confirmed, it is consistent with the properties of technology shown earlier on in the paper.

This situation raises two problems :

The first one concerns *evaluation* itself : according to which criteria should these effects be evaluated knowing that beside the innovation-induced turnover and the employment criterion, no other evaluation criterion is readily available ?

The second one is a *methodological* problem : how can this data set be treated knowing that the sample is not representative according to the canons of traditional statistics and that data appears under heterogeneous metrics ?

The following chapter should answer these questions.

3.3. Data collection and construction of evaluation criteria

For the time being, let us leave all effects under their original metric (generally qualitative appraisals on Likert-scales, or free text). The methodological issue can be addressed through the following theoretical argument.

In a paper published in 1989, A. DESROSIERES¹⁴ argues that the traditional opposition between the « quantitative » and the « qualitative » investigation methods, or to put it in a nutshell, between statistics and monographs, is conceptually misleading. It leads people to think that these two records of reality are of the same kind, when they are not. It leads people to think that if more research funds were available, one could produce intensive knowledge on large samples instead of being restricted to small ones. According to DESROSIERES, monographs and statistics are two full-blown investigation techniques that both aim at representing social reality but do so along different lines. The choice which technique to use depends on *cognitive* (what kind of generalisation and of knowledge accumulation do these investigations imply ?) and *political* arguments (what kind of management model of social reality are the investigations associated with ?). As long as one sticks to a purely cognitive point of view, one ends up in an sterile juxtaposition of both techniques that negates their specificities. Defenders of quantitative methods reproach qualitative methods their inability to infer general and transferable knowledge. Defenders of qualitative methods refuse quantitative methods' massive totalisations where heterogeneous elements have to fit into one single metric, independently of local conditions.

DESROSIERES considers that they are two heterogeneous means to *construct totality*. Both techniques mutually reproach each other to miss totality, and both do, but not the same totality. Statistics *reduce* persons and situations *through codification*. Monographs *neglect parts of populations* defined as logically coherent sets.

This is the point where DESROSIERES argument becomes relevant for our purpose.

In fact, the two techniques can (and have already been) be combined in a variety of different combinations. The author shows for instance how French administrative statistics remained stuck in a regional (« départemental ») logic until the extension and the unification of the socio-technical vocabulary and management tools have progressively eradicated the local reference and replaced it by « group- or region-independent » variables. Some of today's

¹⁴ DESROSIERES (1989)

statistical variables still carry these descriptive components that betray their monographic origins (social group, family types, ...). Others are much purer criteria in the sense that any reference to the globality of a situation has disappeared (for instance the number of corporate employees). Factorial analysis or correspondence analysis are examples of statistical techniques that rest on this kind of compromise. They use a set of *intentionally* « *criticalised* » variables and by doing so, they *aim at reconstructing a* « *monographic like* » *totality*. The results are typologies or space descriptions that have a specific coherence logic.

Our way of processing empirical data is exactly the one that has just been described. On the basis of the pile of empirical observations, we built variables that are not the result of an aggregation of individual effects into a single metric, but « conceptual contractions » of innovation-induced effects into a set of evaluation criteria. These « conceptual contractions » rest on our vicarious experience and upon the « naturalistic generalisation » of accumulated tacit knowledge as programme evaluator. R. STAKE¹⁵ defines tacit knowledge as « that which permits us to recognize faces, to comprehend metaphors (...). Tacit knowledge includes a multitude of unexpressible associations that give rise to new meanings, new ideas, and new applications of the old ». « Naturalistic generalization » is arrived at by « recognizing the similarities of objects and issues in and out of context and by sensing the natural covariations of happenings ». Of course, this way of proceeding leaves out a couple of individual effects, but as the elaborated criteria are meant to serve as a methodological basis for future RFEs, a certain level of generality has to be aimed at.

The set of criteria has got two purposes : first, it supplies the lacking criteria for the evaluation of past R&D collaborations, and it suggests the methodological basis for future « rapid feedback evaluations » (RFE) as those programme administrators like to have.

To wrap it up, the SDA-like exploratory investigation of twenty two monographs has produced a pile of heterogeneous data. In order to construct evaluation criteria, the data have been contracted in a process of « naturalistic generalisation » into a checklist of « criticalised variables » that all reflect monographic realities.

The checklist of is made up of two types of criteria :

¹⁵ STAKE (1978), p.6 ; dans SHADISH et al. (1991)

The first set of indicators is supposed to reflect the *project's contribution to the achievement of the programme's general objectives*. Three indicators are retained in this set :

- the innovation-induced turnover related to the turnover of the concerned business unit ;
- the durability of the average innovation-related turnover ;
- the effect of innovation on the firm's employment level.

The second set of indicators takes into account *recurrent indirect effects* of R&D collaborations. Four indicators have been retained in that respect :

- the level of capitalisation on transferred knowledge ;
- the level of intrafirm diffusion ;
- the reputation effects ;
- the effect on the firm's competence building.

This way of proceeding can be heavily contested :

First, one can argue that it is not the evaluator's role to set the evaluation criteria. According to J. WHOLEY, it is programme administrators that should identify and develop the appropriate measures. Evaluators would tend to add 'omitted criteria'.

Secondly, the construction of criterialised monograph-based variables based on « naturalistic generalisation » appears of course as a questionable method as the knowledge it produces depends on the unique experience of the evaluator without being reproducible, but considering the lack of data, there is no better way.

Once these criteria are set, the question of how to treat them has not yet been answered.

3.4. Data processing and general forms

Knowing that the sample is too small to rely on traditional statistics, we turned towards stochastic statistics, and in particular to a data processing technique based on KOHONEN's algorithm¹⁶. It is a method of data classification, representation and analysis that belongs to the family of neuronal techniques. In particular, it is possible to process qualitative data and small samples.

When confronted with a set of empirical observations, this algorithm gathers observations in classes while maintaining the topology of the original observation space. The notion of neighbourhood between classes is defined a priori. The classification proceeds according to an

iterative scheme and converges to a stable configuration. After classification, observations that are close to each other in the space find themselves back in vicinity on the « map » (the physical support of the processed data) and appear together in the same class or in neighbouring classes. The map can then be read as a non-linear projection of the observations. By highlighting the distance between the different classes, the algorithm makes the interpretation of the map relatively easy. This technique offers the possibility to analyse the links between only two (« correspondence analysis ») or more qualitative variables (« multiple correspondence analysis »).

In our case, this technique allows to treat and represent the variables that have been constructed along the lines presented earlier on in the paper. Our investigation is not limited to the description of induced effects. Each R&D collaboration in the sample is described through an array of variables that supposedly intervene in the generation of collaboration-induced effects. These variables describe :

- the profile of the firm (the « absorptive capacity »),
- the characteristics of the collaboration (the « process »),
- the characteristics of the launched innovation (the « outcome »),
- the innovation-induced effects,
- the contribution of the public laboratory in these effects.

The links between the different variables are currently being tested for their significance. The perspective is twofold :

1. try to identify the most influencing variables in the genesis of innovation-induced effects (i.e. is it the profile of the firm or the configuration of the collaboration that matters most ?);
2. identify those cases where the effects lie beyond a purely commercially impact. The result of this research is not an aggregated impact indicator, but a heuristic model that allows to identify those cases in which the innovation-induced turnover is likely to be a « sufficient » impact measure and those where this impact measure is not informative enough. In this perspective, aggregation is not absolutely necessary.

¹⁶ COTTRELL (1998)

Conclusion

The research from which this paper is drawn is among the very first to analyse the CEA's R&D collaborations from the economic point of view. The particularly rich empirical material has remained largely unexplored by economists so far.

It aims at developing evaluation criteria that allow to grasp the impact of technological collaborations between the CEA's laboratories and industrial firms. Within the landscape of existing impact evaluation models, it is meant as a contribution to the « market-impact model ».

Replowing already well-broken ground, the paper reminds that technology is a phenomenon that is improperly treated by traditional economics and it suggests a more appropriate way to look at it. It shows that traditional expectations towards evaluation are overstretched and it suggests a renewed conception with more realistic expectations.

Rather than executing a confirmatory evaluation (lack of easily-accessible indicators), an exploratory evaluation has been carried out on the basis of twenty two R&D collaborations between CEA laboratories and industrial firms. With the empirical observations that emerge from observation, a set of evaluation criteria have been constructed by « naturalistic generalisation ». The relationships and links between these variables and explanatory variables (profile of the firm, configuration of the collaboration, characteristics of the innovation) are currently being tested by a neuronal technique, namely the algorithm of KOHONEN.

It shows that the construction of an evaluation tool is constantly torn between the requirement for *operational* indicators, the *complexity* of the produced empirical knowledge and the *access* to reliable and high-quality information.

This research should open up perspectives for policy makers on how to make a programme such as the technology dissemination programme « manageable for results » (J. WHOLEY).

Literature

COTTRELL, M. (1998), « Nouvelles techniques neuronales en analyse des données : applications à la classification, à la recherche de typologie et de prévision », WP SAMOS.

DE BANDT, J. ; FORAY, D. (1991) « Concurrence des paradigmes et nouvelles dimensions de l'évaluation », Introduction de Evaluation économique de la recherche et du changement technique, 13-29.

DE BANDT, J. (1995) « Research and innovation : evaluation problems and procedures at different levels », International Journal of Technology Management, Special Issue on the Evaluation of Research and Innovation, Vol. 10, n°4/5/6, pp. 365-377.

L'opposition entre deux formes d'enquête : monographie et statistique », dans : BOLTANSKI, L. et THEVENOT, L. (eds), Justesse et Justice dans le Travail, Cahier du Centre d'Etudes de l'emploi, n°33, CEE-PUF, 1989, p.1-9.

DOSI, G. (1988) « Sources, procedures and microeconomic effects of innovation », Journal of Economic Literature, Vol. 26, p.1120-1171.

GEORGHIU, L. (1998) « Issues in the Evaluation of Innovation and Technology Policy », Evaluation 4(1), 37-51.

GEORGHIU, L. ; MEYER-KRAHMER, F. (1992) « Evaluation of socio-economic effects of European Community R&D programmes in the SPEAR network », Research Evaluation 2(1), avril, 5-15.

METCALFE, S. (1995) « The Economic Foundations of Technology Policy : Equilibrium and Evolutionary Perspectives », dans : Stoneman, P. (ed.) Handbook of the Economics of Innovation and Technological Change, Oxford 1995, 409-512.

PATEL, P. ; PAVITT, K. (1995), « Patterns of technological activity : their measurement and interpretation », dans : Stoneman, P. (ed.) Handbook of the Economics of Innovation and Technological Change, Oxford 1995, 14-51

SHADISH, W.R.Jr. ; COOK, T.D. ; LEVITON, L.C. (1991) Foundations of Program Evaluation, Theories of Practice, Sage Publications, Newbury park, CA.

SHRADER-FRECHETTE, K.S. (1985), « Science Policy, Ethics, and Economic methodology »

ZUSCOVITCH, E. (1986), « The economic dynamics of technologies development », Research Policy, Vol.15, p. 175-186.

Appendices

Appendice 1 : The sample construction

Initially sample suggested by the commercial departments of DTA (LETI and M2RI ¹⁷)	44 executed R&D contracts
Number of realised monographs	27 monographs among which 5 start-ups
Number of not realised monographs	17 monographs
Reasons :	
<ul style="list-style-type: none"> • too recent collaboration (< 5 years) or technology transfer not completed : 5 • too old collaboration or no qualified interlocutor available : 5 • too general collaboration (no licensing agreement, but up-stream R&D frame programme) : 2 • too punctual or short collaboration (no licensing agreement, but technical assistance without knowledge transfer) : 3 • hierarchical veto : 2 	

¹⁷ LETI = Laboratoire d'Electronique, de Technologie et d'Instrumentation ; M2RI = Matériaux, Robotique et Rayonnements Ionisants

Appendice 2 : The sample content (code of company name, transferred technology, data of R&D collaboration)

1	ALC1	Procédés de gravure	84-88
2	ALC2	Réacteurs	84-87
3	BRU	Cytofluoromètre	82-84
4	CEN	Calculateur massivement parallèle	91-94
5	COR	Hygromètre	80-81
6	IBA	Accélérateur d'électrons	91-93
7	IC	Tomographie X	83-86
8	SGM2	Calculateur massivement parallèle	93-97
9	SGM1	Mémoires à bulles	76-80//
10	SCA	Jauges de contrainte	79-80
11	SES	Procédé de polissage	94-95
12	SOP1	Electronique de localisation	86-87
13	SOP2	Détecteur rectangulaire	87-88
14	SOP3	Eléctronique spécifique	89-93
15	TCS	CCD monocouche	86-92
16	DEN	Densification de bois	87-95
17	NIP	Revêtement PVD	82-87
18	SGI	PVD	86-87
19	SICN1	PVD	86-88
20	SICN2	PyC	90-92
21	T+C	Jonctions bimétalliques	81-85
22	VIC	Jonctions plaquées	78-84