

The frontiers of evaluation: some considerations on the European case*

Laurent Bach

BETA (Bureau d'Economie Théorique et Appliquée)
Université de Strasbourg / CNRS UMR 7522, France

Recebido: 22/03/2011 Aprovado: 07/12/2011

ABSTRACT

This paper aims at presenting the “frontiers” of policy evaluation and the related challenges faced by evaluation in the European context, more precisely as regards the evaluation of the impact of European Commission’s Science, Technology and Innovation policies. The analysis stems from recent reports and conferences organised in the field in the last three years, the bulk of the work on evaluation performed in the last 20 years in Europe, and the author’s 20 years of personal experience in the field. Different notions of frontiers are introduced to highlight the relations among the “evaluation world” and policy makers and policy making processes, STI actors and STI processes, and other social science disciplines connected to evaluation. It appears that exploitation of the diversity of approaches, stakeholders, fields of research and empirical studies represents outstanding potential for dealing with the diversity of frontiers.

KEYWORDS | Evaluation; Science; Technology; Innovation; Policy.

JEL Codes | 038, H11, H59, L53.

* The author wish to thank the anonymous referees for their helpful comments and suggestions.

As fronteiras da avaliação: algumas considerações sobre o caso europeu

RESUMO

Este trabalho discute as “fronteiras” da avaliação de políticas públicas e os desafios que a avaliação enfrenta no contexto europeu, mais precisamente na avaliação do impacto das políticas da Comissão Europeia em Ciência, Tecnologia e Inovação (CTI). A análise baseia-se em relatórios e congressos recentes nesse campo nos últimos três anos, em grande parte das pesquisas sobre avaliação efetuadas nos últimos 20 anos na Europa e nos 20 anos de experiência do autor nessa área. São apresentadas noções diferentes de fronteira para realçar as relações entre o “mundo da avaliação”, os formuladores de políticas públicas, os processos de construção dessas políticas, os atores e processos de CTI e as demais disciplinas de ciências sociais ligadas à avaliação. O trabalho conclui que a exploração da diversidade de abordagens, das partes interessadas, dos campos de pesquisa e de estudos empíricos oferece grande potencial para se lidar com a diversidade de fronteiras.

PALAVRAS-CHAVE | Avaliação; Ciência; Tecnologia; Inovação; Políticas públicas.

CÓDIGOS JEL | 038, H11, H59, L53.

1. Introdução

This paper aims at presenting the “frontiers” of policy evaluation and the related challenges faced by evaluation in the European context. More precisely, we will focus here on the Science, Technology and Innovation (STI) policy field, and especially the evaluation of the socio-economic impacts of such policies. However, it is not possible to cover all European countries in this respect, even if we restrict ourselves to the large countries, which would anyway be limitative and somewhat biased if a European perspective is to be taken into account. Therefore we will focus mainly on the case of the European Commission’s STI policies, which essentially (while not uniquely) go through the so-called Research and Technological Development Framework Programmes (EU RTD FP)¹.

In the recent past there have been three series of reports and/or events that have provided some very interesting and diverse insights on related issues. The first was the release of the report on the audit carried out by the European Court of Auditors on the system set in place by the European Commission in order to evaluate the EU RTD FP ((OFFICIAL JOURNAL OF THE EUROPEAN UNION, 2008). The aim of this audit was not only to make a diagnosis of the situation, but also to propose some improvements to the Commission’s approach. The report identified in a direct and even harsh way some limits of the present system. From the perspective of this article, the report reflects “the point of view, for evaluators and evaluation, of a legally-based board of evaluators”, i.e. an external view based on the standard procedures for assessing public policy (corresponding to the mandate of the European Court of Auditors). Different presentations and workshops have also provided stakeholders with the opportunity to discuss these results, and notably the European Commission has officially reacted.

The second was a conference on evaluation organised by the European Evaluation Society and held at the European Parliament in Strasbourg in 2008 (FOUQUE; T MÉASSON, 2009a). The Parliament is clearly willing to take on more responsibility for the design and evaluation of public policies, and gave the conference strong institutional support. The conference was widely open to all fields of policy evaluation and to all sectors, not only to Science, Technology and Evaluation policies. Regional, national, European and other types of international policies were repre-

1 See for instance Rossi (2007) and Møller (2010) for an historical perspective about the RTD FP, and the website <http://ec.europa.eu/research/fp7/index_en.cfm>. For yearly reports on activities such as the one on 2009, see European Commission (2010a).

sented. Two dimensions of this conference are particularly rewarding. On the one hand, the involvement of various fields opened the window on various approaches, issues, problems, methodological advances etc., some echoing the ones at stake in the STI field, others potentially feeding evaluation approaches in that field. On the other hand, insights from the European Parliament on related matters reflected the view of the legislative power (even if this power is not completely developed in the present status of European Union).

The third was the release of the last “large and comprehensive” evaluation reports on the EU RTD: the ex-post evaluation of the 6th Framework Programme (EUROPEAN COMMISSION, 2009). Several events have been organised subsequently to present and discuss the results of this study, notably a conference organised by the EC Research Directorate in 2009.² Again, other points of views were provided thanks to these events. The point of view of the “professional evaluators” of STI policies was provided extensively, these evaluators having been involved in many such studies (or in some studies backing them) in the past (see for instance GEORGHIOU et al., 2002; FAHRENKROG et al., 2002, for good overviews of past studies on the socio-economic impact of EU RTD FP). On the other hand, it was an opportunity to get a “fresh and structured” position on DG Research, one of the actors, if not the main actor, in the European STI policy framework.

This paper is an attempt to identify some of the most striking conclusions of the recent studies on the European case, notably those related to the three events mentioned above; however, we do not claim to have carried out either a systematic survey or a bibliographic or quantified textual analysis of all the papers and contributions to these. Instead, the approach adopted here is to confront and compare them to some personal reflections derived from more than 20 years of experience in the analysis of STI policies at European and national levels and in the conduct of and/or participation in a large number of evaluation studies in the field. The aim is to provide some reflections on the “frontiers” of evaluation as regards the case of Europe.

The presentation is based on a “user-friendly” elaboration of the very notion of frontiers proposed hereafter. Some issues related to the different frontiers are then introduced, especially policy makers and policy making processes, the evolution of STI processes and the involvement of actors, practices and organisations,

2 See http://ec.europa.eu/research/evaluations/index_en.cfm?pg=home for the conference details (Taking stock and moving forward, Conference organized by the European Commission Directorate-General for Research, Brussels, 18 June 2009), as well as a lot of past evaluation studies including the last Interim Report on FP7 (EUROPEAN COMMISSION, 2010b).

and the other disciplines that are connected to evaluation, especially in the broad field of social sciences. In a concluding part, we advocate that exploitation of the diversity of approaches, stakeholders, fields of research and empirical studies represents outstanding potential for dealing with the diversity of frontiers that has been previously identified.

2. Frontiers

In most languages, the term “frontiers” has different meanings, and three of them are particularly relevant in the present context. Taking them into consideration allows us to play on various dimensions of evaluation and the issues related to evaluation.

According to a first sense, a frontier is something to reach or something to move back. This meaning is for instance constantly associated with the progress of science and the progress of knowledge. It is also a familiar notion in US culture, which often refers to the move towards a new frontier (such as the West or space).

But a frontier also obviously conveys the notion of something that delimits a geographic area or a field and thus also defines this area or field. Here the notion of identity, and of what makes identity valuable or useless (specificities, particularisms etc.), is at stake.

Finally, a frontier is also something to be abolished and suppressed. This speaks for itself in the context of Europe, at least when talking about national frontiers.

These three notions are interesting to explore when analyzing the various challenges that evaluation and evaluation studies have to cope with. However, “word games” such as this require a definition of the areas with which evaluation frontiers are related. We therefore propose to distinguish three broad areas that “interact” with evaluation, thus raising various issues related to the different dimensions of the notion of frontiers.

The first is the area of policy making and policy makers. To a large extent, evaluation is a supplier of analysis and results that help to prepare, analyse, monitor, assess, amend etc. policy goals and policy instruments. In practical terms, evaluation studies are most of the time ordered by policy makers to evaluate policy or by “external insiders” to create benchmarks with other policies, for example. The second area is STI, where what is evaluated is the performers of these activities and the processes according to which these actors carry them out, their practices, the modes of organisations adopted etc. Of course, part of this is determined and influenced by policy making. The third area consists of the disciplines connected to

the evaluation, mainly in social sciences. The disciplines may be connected because evaluation studies rely on them for a theoretical basis, general approaches, methodologies and techniques etc., or because other disciplines are “neighbours” of the evaluation and are also mobilised by policy makers in order to gain insights into their actions, the environment in which these actions are decided and implemented, and the effects of these actions.

We will now explore what type of frontiers are at stake when one looks at the interactions between evaluation and these three areas. This will reveal a certain number of issues that have arisen in the recent past, while also shedding light on some issues that have “been around” for a long time and are still crucial for the development of evaluation.

3. Evaluation, policy making and policy makers

3.1. Institutionalisation of evaluation

A first challenge for evaluation is the possibility of institutionalising the evaluation process and combining it in an adequate way with the processes of policy design and policy implementation. Various dimensions are related to this issue; the first two are related to the content of the evaluation (or more precisely the requirements that should be set up on the policy making side as regards the content), while the others deal with the general organisation of the evaluation processes.

The complexity of the policy tools, of the effects they generate and of the processes during which those effects are generated makes it crucial to put forward as “undisputed evidence” the impossibility of using a single figure and a single methodology in order to take this complexity into account. Such single figures and methodologies simply do not exist, and to put it mildly there is very little hope that they could be designed in the future. Thus there is a fundamental need to use different approaches, methodologies and perspectives, and consequently to deploy qualitative as well as quantitative forms of assessment (with different metrics). Interdisciplinary or multidisciplinary approaches are part of the answer (on the “supply side of evaluation studies”), but policy makers should also be open to using the results of evaluations relying on different disciplines (multidisciplinary on the “demand side”).

The scope of evaluation studies results from some form of trade-off between two forces: on one hand, the diversity of topics and levels of evaluation and appro-

aches pushes towards an “academic and research-oriented” proliferation of evaluation studies, with long-term development of sophisticated evaluation tools; on the other, a more practical orientation is supported by policy makers, with a tendency to call for simple, systematic and quickly available information. A balance should then be found between systematic, highly standardized and simple approaches, and exploratory and research-oriented studies. Moreover, the challenge is not only to have both type of studies co-existing, but above all to ensure a feedback loop between them, with today’s exploratory studies feeding the systematic and simpler studies of tomorrow.

The large number of studies running in parallel (the RTD FRP being a good example), especially as regards the systematic gathering of information for the most recurrent studies, raises many practical problems, such as the limit between evaluation and monitoring (see also below), finding time to fit evaluation in between STI activities, the availability of data, and the redundancy of many studies as regards collection of the same data (which pleads for the possibility of more systematic access to data gathered by various studies), leading to the “evaluation fatigue” frequently mentioned by evaluated parties, not to mention the cost of evaluation...

Correspondingly, it is also very important to ensure that evaluation takes place on a programmed and properly resourced basis. The notional figure of 0.5% of programme budget dedicated to evaluation has sometimes being circulated in the evaluation scene, but beyond the accuracy of the estimation of resources required, the very necessity of keeping enough resources for evaluation exercises is essential.

Another facet of the same issue is the wish to have some sort of stability in the evaluation system, notably on the part of programme managers and the beneficiaries of public support. This would clearly help with learning to respond to evaluation enquiries (whatever tools are used) and using evaluation results, while also affording more visibility in the medium to long term for those stakeholders.

More generally, any progress towards institutionalising evaluation should rely on an understanding of evaluation as a “social process” between those who are being evaluated (requiring justification and legitimation, the possibility of learning at the operational level, and a need to gain new supports for public sources), the audience for the evaluation (requiring more accountability, more strictly controlled resource allocation, and learning at the policy level), and those who perform the evaluation (with a mix of academics and consultants or business interests). This also presupposes an emphasis on interactions between academics, practitioners, policy makers and research actors for a better understanding of the scope, relevance and needs of

evaluation. In terms of the types of frontiers, it can be seen that a great deal is at stake as regards institutionalisation.

3.2. Ex-post rebuilding of concepts and goals

A second challenge in evaluation concerns the tendency of policy makers to ask the evaluators to perform ex-post elaboration of the concepts and goals of the policy that has already been designed and implemented. There are many examples, especially in the EC context, in which, well after policy design and implementation, evaluators are urged to propose sound definitions and a deep understanding of notions such as competitiveness, additionality, European Added Value, scale and scope economies, critical mass of research and critical mass of R&D programmes, while these concepts had been used as the basis for the justification and/or design of the corresponding policy. The reason is often that these notions were not precise enough from the beginning, and frequently included some forms of ambiguity and vagueness.

The same applies to policy goals. One recent trend in many calls for evaluation launched at the EU level is to ask as a first topic for the elaboration of what is now often called a “logical model”, i.e. a structured way of presenting the architecture of goals, sub-goals, tools and instruments, and then to connect to that scheme the possible dimensions of effects or impacts and evaluation targets.

To put it bluntly, the issue here is that to some extent the evaluators are asked to do the job of the policy makers. Clearly there is a need to delineate frontiers between the responsibilities and competencies of all concerned. However, a clear definition of the concepts on which the policy relies, and clarification of the logical structure of the policy, and of its aims, expected outcomes and impacts (i.e. the conceptual and analytical alignment between goals, instruments, impacts and evaluation “targets”), are obviously fundamental, but ideally should be carried out long before.

3.3. Towards real and concrete policy learning

Developing policy learning (and more generally promoting the establishment of a system of policy processes guided by “strategic intelligence”) has become a watchword for most evaluation stakeholders. One side of this is related to the dilemma faced by almost all evaluators between the need to provide “easy to handle” results and the aim of avoiding a meaningless list of indicators or scoreboards, i.e. results that are not put in context on the theoretical side (the question being: what is the

underlying theoretical background, what is the “vision” of R&D and innovation processes these indicators are reflecting?) as well as on the empirical side (the question being: what is the relevance of such and such indicators in a given specific area?). Another dimension is the need to set up practical mechanisms for feedback into policy making, which in turn relates closely to the institutionalisation of evaluation evoked in the previous part of this paper.

In the EC context, besides some quick integration of conclusions drawn from evaluation studies (such as the usefulness of involving technology users and producers in the same projects), there are many examples that illustrate the difficulty and the time often required to really include insights provided by evaluation studies (or more broadly innovation studies) in policy design. For instance, for a long time evaluators (echoing the evaluated parties) have desperately called for the simplification of procedures for firms to access programmes and the creation of “one-stop shops” for SMEs (to avoid forcing them to knock on several doors to get information and relevant support), with very small progress despite much caution. Some notions also sometimes turn into commonplaces. For instance, pre-competitiveness has been put forward by the EC as a way for public R&D support to escape from the constraints of competition policy, to put it bluntly, while at the same time evaluations have shown (to the EC’s paradoxical satisfaction) that almost any support for business R&D potentially benefits competitiveness. Another example is the so-called “European paradox”, i.e. the supposedly high quality of basic research alongside low capacity to commercialise the results of this research, which characterises a large proportion of the policy measures designed to foster collaboration and linkages between universities and firms. This notion has been extensively used in some of the partial reviews which were based on an awareness of Europe’s declining position in basic research (the European Research Area and European Research Council are illustrations of these changes) and were undertaken long before the policy was revised.

Two recent examples are provided by the difficulty EC officials appear to face when required to reconsider their views on the importance of patents and the complexity of “new instruments” elaborated for FP6 (and largely taken up again in FP7). Like many other studies, the Innovation Impact Study (FISHER et al., 2009) quite clearly showed that in general (albeit with some sectoral exceptions) patents were not the most frequently used means of protecting innovations, or considered the best. These findings are relevant both for activities conducted in the context of EC projects as well as current business R&D projects. Nevertheless, EC officials

continued to focus exclusively on patenting as a means of protection. The same applies to the adequacy of instruments: according to their answers, companies do not make any essential distinctions between many of the sophisticated instruments set up by the EC. More precisely, they frequently go to EC funds for the same reasons and with the same types of project, whatever instruments are offered. In many cases, moreover, these correspond roughly to the characteristics of the oldest instrument proposed by the EC, i.e. grants for the co-funding of collaborative R&D projects with between three and ten partners, lasting three to four years and amounting to between 0.5 million and 3 million euros. Project objectives are defined by the consortium itself. What determines the choice of one instrument or another is a matter of opportunity, availability at one point in time, the preferences of their partners etc., rather than a clear strategic choice based on the specificity of the instruments. EC officials seem to neglect this point of view, however, as they continue to act as if instruments should be more and more sophisticated and diversified in order to suit the so-called variety of situations. Each instrument is very strictly structured and defined from an administrative point of view, and hence rigid in its own area, while variety in both sectoral and national characteristics (including partner size, technology maturity etc.) is to be addressed by flexibility and day-to-day adaptation of a simple set of rules and instruments. In brief, evaluation should make every possible effort to remove the mental frontiers of policy makers in all these respects.

4. Evaluation and the evolution of STI processes

STI processes have changed considerably in the recent past, and evaluation should obviously adapt to these evolving situations. It is not possible in this paper to present all dimensions of these changes in detail, but some can be highlighted, affecting especially the ways in which STI activities are performed and how the public part is organised. The challenges for evaluation in each of these dimensions are outlined without exploring all their consequences, despite the risk of presenting a “list”. In any case, for the most part what is involved is the need to push back frontiers in the capacity of evaluation to take into account all these changes.

4.1. Performing STI activities

Globalisation and internationalisation of STI activities is one of the most massive and pervasive observable trends. For evaluation, this entails more comprehensively

taking into account the international references, rather than only the impact of actions by national actors in the domestic arena, while making benchmarking between national or local experiences more complex but also more necessary.

STI activities increasingly involve cooperation between actors, linking science and industry, sectors that were not in contact before. Thus evaluating networks is more and more central in all dimensions (rather than just measuring co-publications, for instance).

The increasing problem-solving orientation of a large part of the research is also underlined by many analysts. This often drives evaluation towards a short-term and practical or market-oriented perspective, leaving aside the long-term and more “unpredictable” dimensions of the impact of STI activities.

Interdisciplinary approaches are also more and more frequently put forward as a condition for “modern” research (perhaps more reflected in policy discourse and in programme calls, guidelines and information packages than in actual STI activities). This probably has huge consequences for evaluation, affecting the scope of possible outcomes and impacts, the complexity of interactions among actors, forecasts of research trajectories and corresponding future sources of impacts etc. But in a more concrete and operational way, it calls into question the continuing use of approaches based on “peer review” as extensively as has so far been the case, both in the ex-ante stage (project selection, for instance), and in ex-post stages. The difficulty faced by the EC in setting up a group of experts (who in addition were required to respect basic principles of neutrality, no self-evaluation, turnover etc.) provided an illustration of this problem.

A possibly minor issue relates to the development of ICT directly for the diffusion of STI results. Many consequences are involved here, with ambiguous effects on evaluation. It is increasingly easy to develop and implement bibliometrics, webometrics, patent statistics etc., but with a growing risk of “one-click” illusion, i.e. the absence of a deep understanding of the way the “units” (publications, citations, patents etc.) are actually identified and counted, how they represent actual STI activities etc. Many research papers have recently shown the huge difference in the results obtained using different techniques, while the methods have not been stabilised and will be hard to stabilise given the speed of evolution, but even so almost all evaluation reports now include a chapter or section on this subject. In parallel, publication types proliferate, including e-publications, sometimes with loose scientific control and sometimes with very high quality and demand. In brief, more and more information is available, and there are more and more tools to manipulate

and process this information, which is attractive for evaluators who are keen on easy-to-handle metrics, but it is hard to be at all sure of the content or quality of these treatments, or of their relevance and control etc.

Lastly, the emergence of open innovation behaviour turns out to have a tremendous impact on the way STI activities are performed, on the way their results are diffused, distributed, owned and shared, and on the way these activities are evaluated by the stakeholders. This relates to several of the points raised above, such as globalisation, collaboration and the impact of ICT. Further research is required to analyse the consequences for evaluation, which may not be insignificant.

4.2. Organising (publicly supported) R&D activities

Various trends are clearly observable in many countries, be it due to policy imitation (“policy mimicry”) or to unavoidable adaptation of the organisation’s STI activities.

There is a move towards more privatisation of public STI activities, more activities run under public-private schemes, more contractualized ways of organizing the relations between providers and beneficiaries of public support, and growth in the importance of competition-based programmes in the range of policy instruments. Each of these trends has a specific impact on evaluation approaches, methods and metrics. More globally, this also reflects an increasing role of the private side in the design, implementation and funding of publicly supported STI activities. The point here is not to argue in favour of or against this tendency, but to raise the issue of the legitimacy of public control, auditing and evaluation run from a public perspective (i.e. with a welfare or broad social point of view). It could be argued that the point of view of the evaluators should accompany this tendency so as to reflect the views of private bodies.

As underlined earlier in this paper, the instruments designed by policy makers are multiplying and becoming more sophisticated in their design. As a result, evaluation should also increasingly reflect the shift and adaptation of instruments, as well as adaptation by the beneficiaries of public support to these shifts. This can be seen as an additional level of complexity in the performance of evaluation, which is unfortunately sometimes resolved by a pronounced “reductionism” whereby evaluation relies on a single instrument and thus loses a more systemic perspective.

In parallel, more and more stakeholders are now in the game, and not only at the policy making level but also at the local or regional level. This is sometimes referred to as a multi-level environment, with multiple levels of decision making,

multiple actors etc. This evidently makes evaluation even more complex, at least as regards the identification and assessment of causalities and the point of view to be adopted (what type of impact for what type of actors).

5. Evaluation and social sciences

Evaluation is a field that borrows theoretically, methodologically and empirically from many disciplines, especially in the social sciences. The relations with these fields can be looked at from many angles. Three different perspectives are adopted below.

5.1. The need for a theoretical background

From an academic point of view, the basic requirement of any scientifically sound work in evaluation should be to guarantee the scientific value of the evaluation methods used. While full substantive legitimacy (which can be clumsily defined as evaluating the “true” impacts of STI policy) is often difficult to reach, procedural legitimacy should at least be aimed at, i.e. a way of defining and performing the evaluation that is compatible with scientific requirements. This means that evaluation has to respect many conditions, such as robustness, repeatability, appropriability, transparency, independence of the evaluators, confidentiality, sampling etc., not all of which are always fulfilled in the field of STI policy evaluation.

Another crucial aspect is the “anchoring” of evaluation methods and approaches in theoretical assumptions, visions or frameworks related to STI processes. This is a crucial condition to avoid the trap of providing meaningless and sometimes “perverse” lists of indicators, as suggested above. Moreover, ideally there is a need to make explicit reference to this background when an evaluation approach is designed, if only to ensure that that this background is in line with that explicitly or implicitly adopted by the policy designers, makers and managers.

In all these respects, there is at least a need to abolish the frontiers between theory and practice, and between academic research and expertise.

5.2. Neighbouring activities

A growing number of activities performed “around” STI policies, corresponding to different levels of the policy process, are very close to evaluation because of their nature or because they are experienced as such by some STI policy stakeholders.

Thus it appears to be relevant to delineate the frontiers in such cases. We refer here not specifically to scientific disciplines, but rather to branches of activities which, like evaluation, combine academic research, expertise and services, are partly organized by some kinds of professional group, and sometimes develop widely acknowledged procedures, standard or rules (albeit with different degrees of finality). Examples include the monitoring of STI activities, which is sometimes considered an evaluation exercise. Auditing is also an exercise that shares some rules, procedures and approaches with evaluation, as is quality management and accreditation, which are based on the identification and clear understanding of goals, organisations, instruments etc. of STI policies and this may be used as a starting point by evaluation. Foresight and technology assessment is another area that has to do with ex-ante evaluation. To some extent also, evaluation is part of a larger move towards the development and implementation of new (?) public management procedures and modes of governance. Again, the point here is not to detail the overlaps between these fields and evaluation, and their respective scopes or relevance, but to point out that evaluation, as a field of activities and expertise, should position itself as compared to them, be able to identify and borrow from them whatever is relevant, and be ready to diffuse to them what can be shared for the benefit of all.

5.3. New concepts, approaches, methods

Another point of view adopted here, and one that reinforces the need to push back the frontiers and explore new territories, is that there are still a great many areas in which “basic” and field research is required in order to develop new approaches, new concepts, new methodologies, new metrics etc. Again, this is not the place to list them all or expound them in detail. The issues set out below are regularly raised as key challenges to the development of evaluation in future (for more detailed explanations of methodological difficulties, see RUEGG; GRETCHEN, 2007). Some were identified long ago, and slow progress has been made in operationalising the concepts, ideas and avenues already explored. However, at least some theoretical background has developed (e.g. in evolutionary economics, systems and network analysis, knowledge-based approaches, and further developments of more “orthodox” views or techniques such as control groups or the “option value” approach). Furthermore, the importance of these dimensions and the difficulty of approaching them with “ready-made” quantified approaches is more and more widely acknowledged by policy makers.

- Evaluation of capacities, capabilities, skills and competencies is an enormous challenge. Evaluation has to be based on a standard input-output system in order to be able to take into account changes in the system itself, as well as changes that are not fully or not yet reflected in outputs, or simply cannot be reflected in outputs.
- Evaluation of networking effects is another “classic”. It should be noted that there has been an interesting and pervasive development of quantitative analysis in this field, based on social network analysis. Nevertheless, qualifying what relations are made of, how they evolve and what type of changes in the economy or society can be generated by these networks is still in its infancy, or is only envisaged with classical metrics such as citations, co-publications, co-patenting, licensing etc., i.e. cooperation and transaction around objectified pieces of knowledge.
- The concept of public goods and their relations to public infrastructure (such as large research infrastructure) has regained some emphasis recently. The interesting point is that some new developments around the nature of public goods are being gradually brought into the evaluation field, such as their mixed properties (local goods or club goods), the conditions of access to and ownership of public goods, their intangible/virtual nature, their distributed/networked nature, the issues of irreversibility etc.. This suggests promising avenues for the future.
- The issue of additionality is a last “usual suspect” regularly quoted in such lists of challenges for the development of evaluation approaches. Recent developments in the study of behavioural additionality have been proposed, while some ambiguity still remains to be solved. It should also be stressed that evaluation of STI policy has almost entirely neglected the increasing use of counterfactuals and techniques relating to control groups in many fields of policy evaluation (social policy, health policy, risk/environmental policy etc). This is probably due to the difficulty of using these techniques in an area where levels of uncertainty and the complexity of causalities are so high. It may also be due to the relative lack of involvement of mainstream economics (in which these techniques are commonplace) in STI research in the past.

6. Conclusions: Diversity of frontiers and diversity for dealing with frontiers

This paper underlines the diversity of the frontiers that evaluation has to deal with, in particular when looking at its relations with policy making, STI and the development of social sciences.

It also insists on the need for different approaches, methods, techniques etc. in evaluation, while focusing on the field of STI policy. More broadly, it can be claimed as a conclusion that in many respects the degree of diversity in evaluation is very high, that the opportunities offered by cooperation and interdisciplinarity across this diversified landscape are countless, and that this represents a outstanding opportunity to reach, push back, delineate or abolish the frontiers mentioned above.

Let us very briefly mention some aspects of the differences in evaluation cultures. Especially in the European context, international collaboration, integration through the EU and decision making at various levels requires particular attention to the variety of evaluation cultures, not only in the “geographic” sense (Europe, nations, regions, and possibly metropolises or other local areas). From this standpoint, part of this diversity is related to the level of development of evaluation in policy and in legal and regulatory frameworks. Interestingly, when talking about the EC it is important to bear in mind that there are also different cultures “inside” the EC and even inside different divisions, departments and programmes of a given Directorate. Within the scope of STI, programmes such as ERA-Net (roughly speaking, collaboration between national/regional STI agencies for higher coordination and integration between programmes at all levels, which also induces connections with other directorates than the EC Research Directorate) are helping to increase awareness of the interest and difficulty of mixing evaluation cultures.

Another field of experimentation for the mixing of evaluation cultures is provided by the increasing development of various ways to combine policies (the so-called “policy mix”), involving different policy areas or different lines of public action.

There are also many differences in evaluation cultures across sectors of the economy. The increasing number of stakeholders in STI policy (and in other policies, if we broaden the scope here) also contributes to diversity. For example, there are major differences between the cultures of trade unions, consumer organisations, patient organisations, clubs of academics etc. Evaluators have to gain legitimacy, either because they represent these different stakeholders while guaranteeing a sufficient degree of independence, or because they set up methods and procedures

commonly accepted by the different stakeholders, but in any case this requires mixing and combining the corresponding evaluation cultures.

As suggested by Fouquet and Méasson (2009b), evaluation cultures can be compared and mixed at a “policy level” or a “technical level”, involving different audiences, possibly different forms of learning, and different types of innovation in the evaluation approaches, methods or tools (see Board 1).

BOARD 1

	Within the evaluation community	Outside the evaluation community	Impact of cultural differences	Forms of mutual learning	Types of innovation in evaluation
<i>Policy debate</i>	Defining standards	Citizen participation	Impediment	Recognising differences	Differentiated solutions
<i>Technical debate</i>	Methods of evaluation	Articulating with neighbour activities	Strenght	Exchanges of best practices	Commom methodologies

Source: Fouquet and Méasson (2009b).

What is key here is that the required mutual learning and enrichment between different evaluation cultures goes well beyond a simple “cutting-and-pasting” or “mimicry” of evaluation practices that exist in one sub-field. It also sometimes goes beyond benchmarking, i.e. trying to “select” the best practices and simply adapt them to a new context: indeed, it may require innovating in evaluation itself. Thus the ultimate frontier may perhaps be to make the different cultures work together and benefit from their diversity.

Bibliographical references

EUROPEAN COMMISSION, *Evaluation of the sixth Framework Programmes for Research and Technological Development 2002-2006*, February 2009.

_____. *Annual report on research and technological development activities of the European Union in 2009*. Communication from the Commission, November 2010a COM(2010) 632 final.

_____. *Interim Evaluation of the Seventh Framework Programme*. Report of the expert group, studies and reports, EUR 24569 EN, November 2010b.

FAHRENKROG, G. et al. *RTD Evaluation Toolbox – assessing the socio-economic impact of RTD-policies*. EUR 20382 EN, August, 2002.

FISHER, R. et al. *The impact of publicly funded research on innovation an analysis of European Framework Programmes for Research and Development*. European Commission, Enterprise and Industry, PRO INNO, 2009 (Europe paper, 7).

FOUQUET, A.; MEASSON, L. (Eds.). *Policy and programmes evaluation in Europe – cultures and prospects*. Paris: Ed. L'Harmattan, 2009a.

_____. The futures of evaluation cultures in Europe. In: FOUQUE, A.; MEASSON, L. (Eds.). *Policy and programmes evaluation in Europe – cultures and prospects*. Paris: Ed. L'Harmattan, 2009b.

GEORGHIOU, L.; RIGBY, J.; CAMERON, H. (Eds.). *Assessing the Socio-Economic Impacts of the Framework Programme (ASIF)*. Report to European Commission DG Research, 2002.

MØLLER, K. European innovation policy: a broad-based strategy? *European Review of Labour and Research*, v. 16, n. 2, p. 155-169, May 2010.

OFFICIAL JOURNAL OF THE EUROPEAN UNION. Notice from the European Union Institutions and Bodies – Court of Auditors, Special Report n. 9/2007 concerning Evaluating the EU Research and Technological Development (RTD) framework programmes – could the Commission's approach be improved? together with the Commission's replies, 2008/C 26/01.

ROSSI F. *Innovation policy in the European Union: instruments and objectives*. November 2007 (MPRA paper n. 2009). Available in: <<http://mpa.ub.uni-muenchen.de/2009/>>.

RUEGG, R.; GRETCHEN J. *Overview of Evaluation Methods for R&D Programs – A Directory of Evaluation Methods Relevant to Technology Development Programs*, Prepared for U.S. Department of Energy Office of Energy Efficiency and Renewable Energy, March 2007.

CORRESPONDENCE ADDRESS:

Laurent Bach – laurent.bach@unistra.fr
PEGE - BETA
61 av Foret Noire
67085 – Strasbourg Cedex, France