Science and Technology Policy Evaluation in the Context of Advanced S&T Policy Planning

[Luis Sanz-Menéndez](#)

Instituto de Estudios Sociales Avanzados (CSIC)

State of the art

Evaluation on the field of S&T contains a set of different practices, with diverse origins and trajectories, that converge and co-evolve in a *policy domain,* generally named as "science and technology policy".

Research evaluation and evaluation of S&T policy

Science and technology had been professional practices that from their origins had set up mechanism for internal control. Research evaluation could be seen as a system for control of the quality and relevance of the results of research. The most common this control practices had been *peer review*. When patrons, mainly the State, developed the research funding system formal *peer review* procedures started to be used and developed as instruments for allocation of funds to research institutions and groups. This extension of the internal control mechanisms of science to the allocation of funds for research had become a widely used instrument, in particular in the realm of basic and fundamental research, and a central element of the legitimisation cycle.

The development of the so called *strategic R&D programmes* show the extension of steering activities, selecting priorities, and direct allocating funds by governments. At the same time the increase relevance of technology and innovation issues in government agendas helped the development and consolidation of specialised S&T policy-making bodies and bureaucracies; new actors in the RTD system, that have different needs of information and knowledge about the S&T dynamics and process, through whom new ideas of S&T (policy) evaluation were introduced.

Evaluation is "examining" or "making judgements" (all cognitive process for action include assessments), and *policy evaluation could be understood as part of the historical process of development of tools and information systems for public management.* But the development of evaluation of S&T policies had evolved mainly from the transformation of the professional control practices of researchers and from the specific forms of management of RTD programmes.

The term *S&T policy evaluation* include activities and practices that usually looks back at the past performance of programmes or policies (or sometimes as they are implemented, as in continuous or real-time evaluation) and they are part of the S&T policy cycle as is traditionally described (design, implementation, monitoring, evaluation, redesign). *S&T policy evaluation* refers to "retrospective or ongoing examination" of programme performance or impacts.

There are other practices such as *appraisal* that refers to the activities developed at the beginning of a programme or project (what is sometimes called *ex-ante evaluation*, often related to a selection process for funding or other purposes) or *monitoring.* In general, *evaluation* and *monitoring* can be seen as part of *control activities in the policy cycle*, with few differences in addition to the more or less "judgement" and to their position in the temporal sequence of the "policy process".

The *boundaries* between those activities are fuzzy and subject to interpretations; depending on the country, the *organisational arrangements* to carry on those activities diverge in important ways; the use and impacts of evaluations are highly local and contextual; and they are developed in specific national settings and arrangements.

The national systems: forms and uses of S&T policy evaluation in Europe

From the analysis developed at ASTPP network it appears that no one single practice or use of S&T policy evaluation had been established in Europe. Fragmented national or sectoral (and sometimes regional) arrangements for S&T policy evaluation dominate the scene. Furthermore no clear tendency towards convergence or towards single institutional arrangements appear possible in short term.

The evaluation of S&T policy practices take a wide different forms and functions. Its development is embedded in the research system of the different countries. There is a tied association between the properties of the S&T systems and the uses of evaluation; these arrangements could be understood as a part of the "national systems of innovations".

The forms along which those evaluation practices are developed, the extent to which they are established and institutionalised in the decision or policy making process are very different. That diversity could be linked to the forms that control activities were developed on the specific historical pattern of rationalisation of the different States. It is still difficult to identify typologies of these national models, but there are cases that are relevant for understanding of S&T policy evaluation in Europe:

- UK appears to be ahead in the development of the evaluation; strong pressures from the government and the operators of the R&D system had help in the establishment of procedures and performance standards for evaluation. While in Netherlands, despite the will of the authorities in favour of the evaluation, the role of the operators in the implementation characterises a system with a lot of mutual accommodation and informal evaluation procedures.
- Lack of systematic approach to evaluation is the main characteristic of many of the countries. In some cases, as Germany, a failure to establish a systematic understanding of the role and practice of the evaluation in the policy system, despite the large experience in evaluation, is probably explained by the institutional diversity and differentiation of the policy system.
- In other countries where the State has strong napoleonic traditions, like France or Spain, the S&T evaluation -at different degrees of development- had become characterised as the "guarantor model".
- In the majority of the countries the dynamic forces that favoured the development of evaluation had come from domestic forces involved in the S&T policy process. However in some others, like Greece and Portugal, where the S&T policy had growth associated to the use of the European Structural Funds (FEDER), the evaluation had become a requirement impose by external forces (European Community rules) for legitimisation of the good use of funding.
- Most of the countries had evolved in their approaches to evaluation, as the cases of Finland, Austria or Spain, and had move all long a clear trajectory. But at the same

time in many countries, like Netherlands, Spain or others, appears that evaluation activities are still not linked to a decision making context.

Techniques and methods used for evaluation of S&T policies and programmes are not standard all over Europe. These are developed as a customised activities. Ad hoc procedures are developed to assess the management, performance and impacts of the S&T policies, but still there is a big influence, when assessing quality or relevance, of *peer review* based procedures.

Is this diversity going to survive?. Interaction and ideas determine that professions (all long to the EU scene) could develop common approaches; but institutions matters and we will observe the persistence of diversity, between the different practices of S&T policy evaluation, and TF or TA, in their use in the countries.

S&T policy evaluation and the changing S&T system

Different normative proposals had been made on evaluation in relation to public policy: In a limited context of rational process the evaluation is trying to reduce the deficit of information of the decision-maker and its utilisation is an after-decision problem. (managerial vision). However other approaches had considered that the definition of problems varies and the main function of evaluation confronting the interest of stake-holders is one of enlightenment. (debating on formation and aims). A third approach to evaluation see its potential as strategic management tool, as specific form of co-ordination between actors that help to develop learning loops and strategic considerations in S&T policy process. As programmes and policies had become organisers of interactions and relationship between actors, management and evaluation become part of co-ordination of interactions. That activity of evaluation, when become institutionalised. could be part of the actors' co-ordination mechanism.

S&T policy evaluation is part of a set of cognitive practices, that often become institutionalised, developed in a context defined for the specific forms of interactions between actors. Diversity here is related to the fact that most of the interactions are local and contextual.

S&T policy evaluation developed in the early eighties, in the context of a changing research system, and was one of control practices promoted by different actors. In Europe some of the actors involved in design-implementation of S&T policy have pushed in favour of the development of those new tools. The analysis of the S&T policies and programmes had become relevant at the same time that policy-makers and RTD managers had developed as independent actors trying to improve control on their interventions. In that way the new practice (in their version) become relevant because actors used it in their interactions.

New actors in the research system (especially those associated with the management of the R&D and S&T policy) and outside (the Treasury Departments or Parliaments) had become part of the *advocacy coalition* of the new developments of S&T policy evaluations as analysis of the effects, impacts and performance of the programmes. The transformation of the forms of interaction between governments and research systems explain a lot of the features and characteristics of the S&T policy evaluation.

These practices, and the specific information and knowledge production tools used, are linked to the actors that used them in their interactions in the S&T policy system. Different actors have diverse interpretations on S&T policy evaluation. S&T policy evaluation appears as a local cognitive practice (socially constructed) that is used by some of the actors involved in the policy-making and political systems. S&T policy evaluation, as activity

promoting learning and change, need to develop a coalition of social forces supporting it, a group of committed users. And the fact is that many different configuration of coalitions had been formed in the different countries.

In the world of action intellectual design and social interaction coexist. Problems are discovered, solutions are attempted to solved and reformulated both by social exchange among organised actors and by their efforts to guide events along anticipated paths by purposive actors. But the extension of evaluation of S&T policy had suffer from stop and go process and perhaps many expectations had been located in this activity as source of rationality, control or co-ordination.

Outputs and functions of S&T policy evaluation

The outputs of evaluation exercises, the regular products of those processes are: *policy analysis, recommendations and general advice for decision and policy-making, that is information and knowledge*. S&T policy evaluation is mostly based on *retrospective analysis*, in that way the general *aim* of S&T policy evaluation is to inform post-programme decisions and, because it is often related to the appraisal of subsequent programmes or projects, both activities are closely interrelated. (They also share some of the techniques and tools used for their work, as is the case with *peer review*). The evaluation studies, in the sense of impact analysis of S&T policy programmes, could be understood as trying to prove the achievement of politically established scientific, technological, economic or social targets or to determine the overall performance of the S&T programme

S&T policy evaluation consist in a process of production of knowledge and information by actors embedded in research, S&T policy or political systems; knowledge and information refers to different issues such us the associations between problems and solutions, or the properties and effects of policies and programmes.

But S&T policy evaluation is not the only type of information or knowledge input about past experience in the policy process. Advice could come to the decision and policy process through very different channels and providers.

Most of attention on S&T policy evaluation derives from the need for *accountability* (with was the main force in pushing evaluations at governmental level at the end of the 80s). Different types of S&T evaluations were launched because of accountability problems at the different levels of the policy and political systems. This problem link to one more general of principal-agents relations in the S&T system. Researchers and firms should be accountable to bureaucrats in Government and funding agencies; funding agencies had to be accountable to the Ministry of Finance; Government action had to be accountable to Parliament, etc.

The accountability problem results from the interactions between actors, that use knowledge and information to make interpretations. The uneven development of the actors and their resources could explain part of the national variation (e.g. the growth of professional evaluators, etc.) but in the context of increase complexity in the S&T system (new actors, more interactions between them) there is a increasing need of information and knowledge about the S&T system that could help actors to co-ordinate (mediate and negotiate) themselves in this complex environments.

Once this kind of S&T policy evaluation practices are in motion they produce effects and impacts in the forces modelling the RTD system and in the way of understanding public policies. Then S&T policy evaluation becomes part of the tension between legitimisation and criticism; between conservatism and change; between exploitation and exploration.

The result had been diverse evaluation patters growing in the different countries, in very different institutional contexts, with practices being wide-ranging, and with *systemic functions and users wills* showing an great deal of diversity:

- *improving management, targeting, controlling and fine-tuning* of S&T policy programmes and enhancing the information and knowledge bases of the S&T policies,
- *provision of legitimisation* for the distribution of funding and the demonstration of adequate, effective and efficient use of it by assessing the quality of the management of the programme, measuring the scientific/technological quality or determining the socio-economic or structural impacts,
- *improving transparency* in research funds allocation and establishment of the rules for S&T system in the sense of a government-led guarantee or regulation between diverging and competing interests of various players within the S&T system,
- *creating negotiation spaces and shared interpretations* between the different Government levels in the sense of *moderation*.

Evaluation produce information and knowledge; in addition the availability of information and the process of evaluation provides legitimisation; it offers transparency and fair play between the players in the S&T system; and it creates a shared space of interpretation for the negotiation between conflicting interest involved in the increase complex game of S&T policy in Europe.

Legitimacy, control or moderation are functions, but they are very dependent on how it is organised the social process that any evaluation involves. At the same time the different cognitive practices play different roles in relation to the actors; for example, the *appraisal* produces legitimacy for the programmes, in front of the research communities (because they guarantee that the funding and selection is fair), while *programme evaluation* produces legitimacy for the programme management in front of other actors (as politicians, etc.)

What role for S&T policy evaluation in S&T policy?

The results of the evaluation could be information and knowledge; but how these inputs become a tool for the S&T policy planning?. Is the evaluation a tool for improving S&T policy?

Users of that knowledge want legitimacy, control and fine-tuning or analysis of policy rationale. Evaluations could produce effects in organisations, as creation of units of evaluation, introducing strategic planning or definition of an agenda for research, and future funding. But the effects are only produced in the case that ideas and arguments are taken up by some of the actors.

It have been mention that, in practice, in most of the countries there are not clear links between the evaluation exercices and the management and decision-making levels. More and more S&T evaluation are commissioned but still appears that their uses are far for been integrated in the S&T policy process. That 's not mean that their finding have no effects. Probably the way that evaluation exercises influence the future decision and policy-making is slower than expected.

There are also in many cases serious organisational barriers to the absorption of the evaluation findings. The dominance of traditions more concern with the determination of legality than with the assessment of performance and the lack of institutionalisation in the policy-making process as part of systematic feedback loop, are also limits.

An even greater area of failure of evaluation is the fact the evaluation is almost confined to close circles of decision-makers and, in some cases, to the clients of the programme. General public and users of S&T are almost absent of the actual game of evaluation.

Those weakness of the evaluation of S&T policy in Europe refers more to the environment of S&T policy-making. There are other intrinsic limits to "evaluation". Evaluations are making conclusion on the past performance, establishing causality relations between different elements of the system, and could help to create shared understandings, but that is not enough to deal with the basic elements need to make decisions on future actions. Evaluation alone could only play a partial role in reinforcing the strategic aspects in S&T policy planning.

ASTPP aims are related to the improvements of methods, but also with the development of normative models and prescriptive proposals for practical improvement of the way how S&T decision-make is organised, but those proposal are always trying to be empirically funded and theoretical plausible.

The actual practice of "monitoring and evaluation" of the EU RTD programmes

The European Commission had played, in the eighties and beginning of the nineties, a relevant role in the promotion and support of the capabilities for S&T evaluation S&T. Relevant methodological developments, case studies, and so on were commissioned by the EC through activities organised mainly through MONITOR programme. But when the IV FP was approved a specific program, TSER, included within its objectives the advance of knowledge in "science and technology policy options", including methodological developments. But most of the funding had been subject to a regular competitive bids in the frame of "calls".

At the same time that the practice of evaluation of RTD programmes had become institutionalise, and the evaluation unit had been consolidated in the DG XII, and the unit have recently been allocated in DG XII AP on charge of the Framework Programme.

However that momentum had not been isolated from the general initiatives of the European Commission. In 1994/95 the EC asked a group of experts to review the existing Commission practice of evaluation and recommend improvements. Following from this report, and within the general framework of what is called the "SEM 2000" initiative (Sound and Efficient Management) and internal communication on evaluation, entitle "Evaluation: Concrete steps toward best practice across the Commission", outlined a strategy for action to strengthen the evaluation of EU programs. This general trend had strong "financial management" content because "the results of evaluation of Community actions, undertaken periodically, should be taken into account in decisions concerning budgetary allocations"

After the document "Towards implementation of coherent monitoring and evaluation of Community RTD actions" issued by CREST (1208/95, 19 May 1995) and the Commission initiative SEM 2000, the Commission had produced a communication on "External and independent monitoring and evaluation of the Community activities on the domain of RTD" (COM(96)220 final).

The proposal of the EC for monitoring and evaluation of programs should be accomplish through two types of actions referred to FP and the specific programmes: i) continuos monitoring with the help of external independent experts and preparation of an annual rapport; and ii) five years mid-term assessment, done by external independent experts that have to include the conclusions of the final evaluation reports of the previous FP and specific programmes.

The two activities of monitoring and evaluation have different nature because the "five-years assessment" report of the FP, in addition to the comments of the Commission should be send to the EP, CM and ESC before the presentation of the proposition for the new FP. The aspects that should be included in the reports are: the coherence between the selection of projects and the objectives of the programme; the efficiency of the management of the programme; etc.

The binding approach to evaluation is very much linked to the assessment of the management and implementation of the programme. No obligation appears to be referred in relation to the "impacts or effects of the S&T programmes" as criteria. Those activities, as a practice are not forbidden, but due to the fact that the European Commission is responsible of the monitoring and evaluation of the RTD actions; those type of evaluation become dependent of the entrepreneurial initiatives of the Directorates.

At the level of the S&T policy represented by the Framework Programme the practice of the monitor and evaluation have take the dominant form of "panels" of independent external experts. The "independence" of evaluators could be viewed as a way of increasing credibility, linked to the variety of actors involved and the complexity of the decision-making process, but in practice the S&T programme evaluations are organised with low budgets and not many people (usually 3 for monitoring panels or 5 mid term five years assessment), that are increasingly dependent on the provision of information and support by the implementation directorate. In addition independent evaluators are selected directly by the RTD management unit. From the situation appears no limit to the fact that managers of specific programs could request evaluations from professional evaluators (teams instead of panels) but in practice with low budget it not appears very feasible.

In relation to the role of the European Commission confronting S&T policy evaluation, two different situation emerge. In one side, still appears committed to the promotion of knowledge and methodologies in S&T policy options, because it had been established as formal domain for Targeted Socio-economic Research. In the side of practice, it looks a pragmatic approach for "control", but with asymmetric participation of the key actors in policy design and decision-making. The practice of the monitoring and evaluation in the Community RTD actions is very much oriented to program management issues and objectives. A lot of the present situation of S&T policy evaluation in the Community action is linked to the concerns on sound and efficient management" as a way of legitimisation and control, and specially of the fulfilment of scientific and technological objectives.

The apparent impasse on S&T policy impact studies could be either the result of the fact that this kind of evaluations are dependent on the entrepreneurial initiatives of the management units or that the increasing involvement of countries in close monitoring of the Framework Programmes (through Programme Committees) had created a new attitude of the Commission regarding the production of sensible information that could be used by the national Governments against EC interest. At the end the European Governments are the principals of the Commission. Because one way in which the EC use some evaluations (or other forms of policy analysis) is to resume its initiatives (permanent innovation) and to escape from the control of the countries.

But appears that the results of evaluations or policy findings have a lot of room to be introduced as inputs in redefinition of the S&T policies, mainly through the process of consultation (inter-services and with governments and interest). That is a positive aspects of the institutional constraints of the European Union. And, at the same time, as Governments enter into the implementation procedures of evaluation and monitoring S&T policies (no matter that for controlling the European Commission) more is disseminate the culture of evaluation within them.

Confronting complex foresight environments with strategic S&T planning: synergies between TF, TA and S&T policy evaluation

S&T systems had growth in complexity and many new actors are now involved. In this system there are many agents acting with conflicting preferences, and trying to advances their interests in their interactions. A single actor can not calculate the consequences of its actions because, as a result of the complex interactions, S&T system changes. A new feature of S&T system is that the foresight horizon is complex, because the structure of the S&T world (included the policy-making) is also changing due to the interactions. This kind of world cannot be easily explored because is under construction by the actors interactions.

In this kind of system there are two nested processes with strategic relevance: The first is cognitive: the actors provide identities and populate their world by locating which agents are there and constructing explicit interpretations of what these agents do in what kind of situations. The second is structural: the actors foster generative relationship between agents, mostly through interaction, relationships that produce new attributions of identities that cannot be foreseen in advance. Confronting this kind of complex foresight horizons a key issue becomes the organisation of the processes of exploration and adaptation.

Traditionally strategic had represented a commitment to a planned or established long term course of action. And the selection of a strategy meant to optimise between different alternatives on the base of calculation of future consequences and their probability. But the notion of what is strategic in this new type of environment must also include understanding how changes come about. In these situations strategic means "the means to achieve control". Since the outcomes depend on the interactions with and between many other agents (inside and outside the p-m boundaries), strategy represents an attempt to control the process of interactions. But the problem is that In this type of complex situations control is distributed.

No one can predict what will emerge from the constructive dynamics. But if the forces (actors) in motion that produce the generative relationship are associated in the processes of attribution of meanings and identities with a system of monitoring their relationships, to assess the potential for generativeness, and "ongoing" and ex-post interpretation (attribution of identities and meanings) we could improve our ability to deal with complex foresight horizons. Here are located the challenges that S&T policy planning has to face.

Strategy becomes a process consisting of a set of practices (that could include in a special location TF, TA and S&T policy evaluation), in which agents inside the policy-making process structure and interpret the relationship inside and outside the p-m process, through which they both act and gain knowledge about their world.

Then we can recommend the development of activities that associate both the cognitive and the interactive side of the process. Unpredictability requires ongoing re-interpretation, but making sense is not enough. Agents must monitor their relationships for fostering the generativeness, and those processes in policy and politics are carried out through negotiations, through mutual adjustment. In that way the use of tools and cognitive process that bring together the actors to "negotiate" are basic to guarantee the increasing adaptativeness to changes.

Two concepts are relevant for understanding the role of information in organisation for decisions and actions: *uncertainty and ambiguity.* While uncertainty can be solved (reduced) by obtaining (elaborating) additional pieces of information (in that sense the development of engineering decision tools is central), ambiguity cannot be solved simply gathering information. Ambiguity is a state of having many ways of thinking about the same

circumstances or phenomena, and in that way more information is not directly relevant for solving the ambiguity. The extent that we can come to a consensus about it, depends on a process of interpretation (what is relevant, what is value to give, etc.). But no matter that information could not be relevant for decision, could be useful for exploring ambiguity. Ambiguity can only be resolved by *shared agreement*, reached through a process of interpretation; and interpretation is giving meaning before it can be acted.

In addition also mostly policy situations are characterised by bunches of actors involved *with conflicting identities and preferences*. And the system should allow us to make a decision, but without eliminating the divergence of identities and preferences, without solving the inconsistencies.

There are not single approach on how to deal with uncertain foresight environments. There are to basic models for decision-making, both are useful for describing the empirical world in a realistic way, and both largely reflect different aspects and needs of decision-making process. But any of those two basic ways of understanding the process of decision-making (the *logic of consequentiality,* the consequences that could be used for explaining, and the *logic of temporality,* where the explanations come from the sequence of events) have diverse informational need to operate. Trying to gain "intelligence" requires different thoughts, depending on what logic we apply: the logic of consequence and the logic of appropriateness are equally logic of thoughtfulness, and the cognitive demands for each are substantial.

In the logic of consequence, there are requirements *for knowledge about the future* and for consistency and clarity in *preferences*. In rational models information is a *means of choosing options* of a satisfactory solutions and the information needed is a function of the goals specified. In the case of a logic of appropriateness, there are requirements for knowledge about the *situation* and for consistency and clarity in *identities*. In *decision under ambiguity models* information is useful for exploring problems and solutions, discovering preferences or maintaining lines of communications.

Under appropriate circumstances action based on either logic can lead to achieving outcomes that are judged to be attractive or contribute (over some time horizon) to survival advantage. However, neither rational exchange nor rule-following (and the learning and selection of rule that lies behind it) is assured of being intelligent. The intelligence of each depends on the ways in which their imperatives are interpreted and on the extent to which capabilities for meeting them exist. Here comes the proposal of ASTPP for creating this capabilities through the association of EVA, TF and TA in an intelligence system for strategic or adaptive learning.

European S/T policy is confronted with a number of heterogeneous challenges that call for highly intelligent and well informed "holistic" procedures of policy planning and decision-making. A *"reflexive S/T system"* is emerging - but still fragmentarily, and it turns out that basically all three "pillars" (TF, TA, EV) are involved: they remain different activities, but not independent, and all three are points of access to this emerging system.

TF, TA and S&T policy evaluation have been developed as cognitive process (to provide enlightenment -calculations and interpretations- to the decisions) by different professional communities, supported by different actors and understood in different ways. But if these cognitive activities are to have impact on strategic policy-making process they need to be developed in association with all the relevant actors involved in the process to help to foster the generative relationship between agents that could tackle with the complex foresight horizons that S&T policy planning represents. It must be clear that the role of information and knowledge production systems (about future and past) in policy-making -more than

simply to make choices- is to help, trough the attribution process, in the generative interactions for increasing the adaptativeness to the changing environment.

As it has been stated above any strategic S&T policy has to deal with three important elements: uncertainty, ambiguity and conflict of interest. The three different tools could be used to reduce uncertainty of future (TF), to learn from the past through shared interpretations reducing ambiguity (EVA), and to deal with the divergent preferences of the actors (TA). Through this three process we increase the strategic aspects (adaptativeness) of the S&T policy. The concrete prescriptive proposal by ASTPP of the articulation of EVA, TF and TA, as part of a system of information and knowledge production, could play a mayor role in defining the tools for dealing, in an strategic way, with the challenges of complex foresight environment that S&T policy have to confront. All three practices had developed as independent streams but they have complementary (or synergetic) properties in relation to some features necessary for considering the strategic aspects of S&T policy.

EVA is a central tool for learning, if we consider that learning (in organisations) is a process of encoding inferences from the past that guide behaviour. It can be learning form experience or from others, but also from development of conceptual frameworks or paradigms for interpreting that experience. The process of evaluation could create shared interpretations and understandings between the actors in S&T system. But in its present - even the most advanced forms- EVA lack the capability either to define criteria for future action different that those encoded in the past or the capability to deal with the negotiations of the conflicting interest involved in S&T policy process..

Bringing TF to the strategic approach of S&T policy planning we are dealing (engineering procedures for monitoring the generative relations between actors and artefacts) with the reduction of uncertainty. We open the door for having scenarios of the future that are necessary for an adaptive capability of the S&T policy to the changing environment.

Including TA in the tool kit of strategic planning will help to solve some of the major deficiencies of the present state of the EVA (an in part of TF). We will take its capability to deal will the conflict of interest or the diversity of preferences in a "negotiated" form, in that way we will be able to monitor the generative relations that are basic to the functioning of the S&T system.

But no matter that EVA, TF or TA could bring some specific competencies to the strategic intelligence system we still need work on the methodological aspects of the links.