Introduction

In France, the last decade has been marked by the institutionalisation of evaluation as part of the
landscape of public policies. This movement started with science and technology, which, compared
to other public policies, went through a specific institutionalisation process. In this chapter I shall
present the main transformations and draw some preliminary conclusions on the specific choice
made, i.e. what is often called, following Robert Chabbal’s work on the EC system (1987), the
“guarantor model”.

This approach departs from the view of “evaluation” as an ex-post monitoring exercise: it
focuses on the “proactive” role of evaluation activities and their embedment in the policy-making
process as one of the recognised instruments of “strategic management”. It is therefore difficult to
separate the two aspects and analyse the changing situation of evaluation without at the same time
considering the fairly significant organisational changes that have taken place in the French S&T
arena over the same period.1 This leads to the following methodological stand: these transformations
should not be analysed from an external point of view, i.e. with a given definition of evaluation, and
with pre-set norms about what it should be and how it should work. Instead, the responsibility should
be left to the actors themselves to decide what should be considered as evaluation, and to what
process it corresponds. I propose to address this issue by analysing the position of evaluation in the
French public research system and how it has de facto developed over the last decade.

The OECD has a long tradition of evaluation of national S&T policies, pointing to a need for a
global (if not holistic) view of the situation of the “national system” as a whole (see the chapter by
J.-E. Aubert). For French policy makers there has long been a continuum between “evaluation” and
“advice”, as reflected in the rich and entangled set of advisory bodies whose capabilities and
intervention are of crucial importance in the shaping and understanding of the evaluation scene, the
importance given to indicators (with the creation of the Observatoire des Sciences et des Techniques
– OST), and the central focus on “operators”.

The research institutions, universities, programmes, agencies and procedures which mediate
between the political sphere and scientific activities have been at the core of the French experience in
evaluation over the last decade. The growing importance of this intermediary layer, to borrow the
term proposed by B. Van der Meulen and A. Rip (1994), calls for a radical reappraisal of our “ballistic” (or linear) view of the “formulation-implementation-control” triad upon which evaluation research has long been based. It is interesting to see how evaluation of S&T operators has developed so that they are considered not solely as the agents of implementation of a policy defined at a higher level, but rather as mediators between research activities and policy making with their own strategic, political and managerial capabilities. In this chapter, I will focus on three major settings highlighting the guarantor approach: two committees established at the central political level for the evaluation of higher education institutions (CNE); all the other S&T public operators (CNER); and one internal setting, the Délégation aux audits, which, because of the central role played by the Comité national de la recherche scientifique (CNRS) in the French landscape, can be considered to represent a third “central” initiative. This focus will enable us to highlight the results obtained in two dimensions: the robustness and credibility of the evaluations; and the embedding of “strategic” evaluation in the evaluated bodies, especially research institutions. It will also underline the current limitations, especially in relation to the link between evaluation and policy making at the national level – a situation shared and strongly indicated by the evaluation settings themselves.

Most of the evaluations carried out during this period dealing with research institutions shed new light on “human resources” and on the role of evaluation in the dynamics of scientific activities per se. This reminds us that in France – and this may be due to its long-standing organisational double choice of strong research institutions associated with university enseignants-chercheurs – evaluation rhymes first with the careers of researchers and the life-cycle of the research collectives in which they work. The growing importance of these aspects has led to analysis of the evaluation settings developed to handle them, focusing on two cases in so-called mission-oriented institutions. These two aspects could even be considered to constitute a major element in the changing evaluation landscape in France.

Advice on and evaluation of national policy

At De Gaulle’s return to power, in 1958, he established the Conseil consultatif de la recherche scientifique et technique to assist the newly created Direction générale de la recherche scientifique et technologique (DGRST) in the design of a research policy. Known as the Comité des Sages, this body constitutes a permanent reference for “global” guidance of national policy. Why consider such advisory bodies here since they bear no resemblance whatsoever to the long-standing OECD practice of S&T policy evaluations? The answer lies in the approach taken to evaluation and the balance established between “monitoring” (of past decisions) and “advice” (based upon the lessons learned from past actions). In France, the evaluation of public policies has two main features: it is recent and has developed outside the existing monitoring bodies, the Cour des Comptes and the inspections générales. This departure from usual monitoring practice is not the result of power struggles within the bureaucracy. On the contrary, if we follow P. Viveret (1989) who was in charge of the report which was instrumental in establishing evaluation as part of the policy process, this separation is useful and reposes on the role of evaluation in the democratic process: to act as an “instrument for political intelligence” and thus participate in the collective building of a shared judgement. Evaluation does not exist to deliver good or bad marks, rather it exists to provide political actors with access to past experience to help solve present-day problems and decisions. Evaluation is thus strongly linked with the global views expressed on the national S&T system or policy.

This was clearly the rationale underlying the law of 1982 which replaced the Comité des Sages by the Conseil supérieur de la recherche et de la technologie (CSRT), whose role is to support “major
choices of government scientific and technological policy” and to produce an “annual report on the evaluation of national S&T policy”. At the same time, its composition changed from a small number of “wise men” directly nominated by the government to a quasi-parliament of research with representation of all the major stakeholders and a complex nomination process. A decade later, the need for a “strategy supporting body” was felt, and the Comité d’orientation stratégique (COS), composed of 15 members nominated by the Prime Minister, was created in 1995. The priority task of this committee was to “prepare the annual report on the national research strategy” to be presented by the Minister to Parliament.

If we now turn to the CSRT’s counterpart for universities, the CNESER, it can be easily seen that these high-ranking advisory bodies and their outputs constitute a major element of the French “evaluation scene”. All the more so since, in 1984, Parliament, in the face of this unequal balance, decided to develop its own capabilities and established the Office parlementaire des choix scientifiques et techniques, a joint body with members from both chambers. This description would be incomplete without the periodic mobilisations which took place before each of the above-mentioned changes (the Colloque de Caen in 1956, the Assises nationales in 1981 and the Consultation nationale in 1994). These national consultations constitute yet another dimension of the national scene, although their importance lies less in their direct outputs than in the issues they brought to the forefront and the “cultural” changes they promoted. We shall later refer to some of the ideas promoted by the 1994 Consultation nationale which, for the first time, included dedicated debates on evaluation.

I do not intend to analyse here the effects of these consultative bodies and the periodic national consultations. However, the reader will easily understand that, given this rich context and the frequent use of the word “evaluation” to refer to their outputs, no room was left for yet another screening mechanism of national policy as a whole. This did not mean that evaluation was no longer an issue, instead the debates focused on the best ways to provide these bodies and arenas with robust and credible material. In-house information provided by Ministry services or powerful institutions were regarded as biased and unreliable by the other stakeholders. Evaluation and the building of national indicators appeared to provide solutions to this problem. Both these issues were handled at the same time and along the same lines. The Observatoire des Sciences et des Techniques (OST), in charge of the production of French S&T indicators, was created at the same time as the CNER (see below) as an integral part of the evaluation system. Its task was to provide a link which was missing in France, although efforts had been made since the beginning of the 1970s to develop a French equivalent of the US science indicators. The fact that OST is an independent body, funded by several Ministries and the largest French research institutions, demonstrates a central feature of the institutionalisation of the French evaluation process during the 1980s: “credibility” implied independence from the central administration and use of the “guarantor model” of evaluation.

The issue of robustness and credibility (required by both Parliament at the global policy level and the researchers themselves) led to the institutionalisation of evaluation as the question arose about systematic and periodic evaluation of “operators” – the research institutions, agencies, programmes and procedures that mediate between policy making and research activities. We shall see that it has proven more difficult than anticipated to strike a balance between credibility and relevance to the policy process.
The evaluation of operators: from control to learning

In Europe, evaluation again became fashionable after the rapid development of a new policy instrument: the “technological programme”, focused on so-called new, generic technologies such as information technologies, new materials or biotechnologies. The specific nature of these public actions and the fact that they produced no clear, tangible outputs – as was the case in large development programmes (such as the space programmes) – rapidly boosted the need for evaluation, i.e. making the activity and effects of public action visible. In France this question was brought to light with the creation of the programmes mobilisateurs. The 1982 law stated that these programmes should be evaluated; in the following three-year programming law (1985), a whole item was devoted to the evaluation of programmes and research institutions. However, no implementation structures were put into place. Meanwhile the new law on the universities voted in 1984 required the universities to be systematically evaluated, and established the Comité national d’évaluation (CNE) for this purpose. Its success drove the Ministry of Research to establish, in 1989, along similar lines, the Comité national d’évaluation de la recherche (CNER) for the periodic evaluation of other public research “operators”. At the same time, the CNRS established its own auditing system supported by the Comité consultatif des audits (CCA). In less than five years, a new approach to the evaluation of research activities had been institutionalised. This approach no longer focused on research actors nor on the national system as a whole, but rather on the bodies that mediate between both levels and that participate in the shaping of research policies and strategies. These French public research “operators” cover universities, research institutions and agencies, but the terminology is there to remind us that other operators exist, such as temporary organisational settings in charge of given technological programmes or procedures.

These three committees have now established processes and accumulated experience. Although they share the same principles (independence from the “executive” and quality control mechanisms to ensure robustness and credibility), they have developed along different lines, providing us with three complementary operational models. I shall analyse each in turn. This will enable the different achievements with, in many cases, quite significant “local” effects, i.e. on the institutions being evaluated, to be distinguished. At the same time, this approach will highlight the main problem faced by all three committees: their relationship to the “global” policy-making process. This remains an open issue which may call into question the performance, if not the relevance, of the whole approach.

Universities and the Comité national d’évaluation

In 1984 the “Savary” law changed the status and organisational structure of the universities, transforming them into établissements publics à caractère scientifique, culturel et professionnel (EPSCP). To respond to the then, and still, ongoing debate on university performance, a Comité national d’évaluation (CNE) was set up. Following the success of the Commission nationale de l’informatique et des libertés, the CNE was established as an “independent administrative authority”. Five main features characterise it: it is independent from the Ministry, reporting directly to the Président de la République; the Committee is composed of 17 members nominated for four years through a strict nomination process involving both chambers of Parliament and independent state auditing institutions; its missions are clearly established by law and consist of a systematic and periodic evaluation of all EPSCP; it has full responsibility for its evaluation programme, for the methodologies adopted and for the dissemination of its evaluations. Last, but not least, it has its own budget line covering its staff (at present seven chargés de mission and a
At the end of 1994 CNE officially finished its first tour of the French universities (evaluating on average nine universities per year) and entered a new cycle. The CNE decided to conceive evaluation “not as a comparison between two points in time but as a dynamic analysis of the efforts engaged by universities in response to rapidly growing demand while improving the quality of teaching and research” (CNE, 1993). Meanwhile, to account for the wide diversity of situations, the Ministry developed a new policy based on systematic four-year contrats d’établissement with all universities. The CNE was requested to include the evaluation of these contracts in its own evaluations; this required adjusting the process and rhythm of the evaluations.

The CNE now carries out more than 20 evaluations per year. The “institutional” evaluation of universities and higher education establishments is the backbone of its activities, although the evaluations vary according to whether they concern departmental issues (following on from geography, information and communication sciences and odontology, pharmaceutical studies are currently being evaluated), thematic issues (e.g. university libraries) or even geographic ones (e.g. evaluating the added value of geographical synergies in a pôle, such as that created in the Lyon area).

The process is now well established as an “institutional approach which gives priority to peer reviews over quantitative approaches and indicators” (CNE, 1989). Each evaluation is carried out by a sub-group of the Committee (generally two or three members), which is responsible for the conclusions and recommendations; it is managed by one of the seven chargés de mission who, in addition to the logistical issues, is responsible for the initial information collection and the drafting of the synthesis report. The four-step process lasts approximately nine months.

Information collection is a “preliminary” step, handled by the secretariat. It is based on a questionnaire, which the Committee views as a “check list” of the minimal quantitative elements a university should know about itself. However, most universities still “do not know themselves”, and there have been recent cases where even the number of students was unclear! Annual reports from the CNE regularly emphasize the urgent task of “moving into the light” (CNE, 1993). As members of the secretariat have pointed out, the construction of a set of indicators – a tableau de bord – is a major limitation and a major direction in which to strive. We are clearly faced with the continuum which exists between management and evaluation, and a major lesson from this evaluation approach, and one which the CNER also emphasizes, is that wherever monitoring instruments are missing, evaluation tends to play an active role in their initial development (if only for its own requirements).

The starting point of the evaluation is the visit made by the sub-group of the Committee to the university. As the Committee has begun to return to universities that have already been through an evaluation process, it has changed its approach; its analysis is no longer all-encompassing, covering all the departments and activities of the university. Instead, it focuses on the identified strategic problems faced by the university. The conclusions and recommendations of the previous evaluation, analysis of the information set and the priorities included in the contrat d’établissement which the university has signed with the Ministry of Education, all play a role in the identification of strategic problems. Dialogue with the “government of the university” (to use the phrasing adopted by CNE) is thus central to the identification of “crucial issues” and to the construction of the terms of reference of the evaluation. Biotechnologies in the case of Strasbourg, and medical imaging in Rennes, are good
examples of this approach, which is not always accepted by the university stakeholders, who challenge the relevance of the choices.

The third step, the experts’ report, is central to the evaluation process. The use of individual experts – and this will come as no surprise to evaluation analysts – poses three main problems: the selection of the experts; the development of a common framework for the individual reports; and the assembling of the individual reports into a coherent whole.

◊ All evaluations relying on experts pose a “quality” issue associated with the independence and competence of the experts. The CNE, which uses on average nine experts per evaluation, has chosen to draw its expertise almost exclusively from the French universities: therefore most experts evaluate their colleagues in the knowledge that the situation will soon be reversed. To counterbalance this situation, the Committee has decided to keep the experts’ reports confidential. Nevertheless, the Committee is considering increasing its use of foreign experts.

◊ How can the relevance of the experts’ work be ensured? This is another well-known issue which has led the CNE to develop a sophisticated approach. All the experts participating in an evaluation are invited to a one-day meeting at the CNE to become acquainted both with the nature of the work they will be undertaking (the Committee has drawn up a “briefing note” for participants) and with the situation of the university. Visits to the site are carried out over a short period (less than 10 days) to allow for dialogue between experts. This is fostered by the chargé de mission who is on the site during the whole period to ensure the smooth functioning of the visits.

◊ To help in the formulation of a common view, the experts meet a second time before writing their individual reports, which remain confidential. Under the guidance of the Committee’s sub-group, the Secretariat drafts a synthesis report which is discussed by the Committee. At this stage, the synthesis report does not include any conclusions or recommendations.

The confrontation phase now takes place. The analytical report is sent to the university (as one chargé de mission states, it is “often written rather harshly in order to make them react”), and this is followed by an on-site examination with the Committee group in charge of the evaluation. Taking the reactions into account, the conclusions and recommendations are then formulated by the group and adopted, after discussion, by the CNE. They are sent to the President of the university for an official response which will be incorporated in the published report. The wide dissemination of the report ends the process. With their distinctive blue cover, these reports are now part of the university landscape.

Embedding evaluation in the university landscape has always been one of the objectives underlined by the CNE in its reports to the Président de la République. And, clearly, the CNE considered as a major achievement the recognised need for evaluation expressed by the conference of university presidents: “autonomy and partnership cannot exist in the absence of an external and independent evaluation of universities” (cited in the 1993 report). This did not stop the Committee from addressing the questions of the impact and consequences of evaluations in all of its reports.

The CNE (1989) perceives the evaluation process as a process of “revelation”; it acts as a catalyst to allow problems to be better identified and stated. It also considers that there are two very different categories of problems: some are linked to university policy and action, while others relate to “the desirable or undesirable effects of national policy on a given university”. The
recommendations therefore address both the strategy of the university and the framing and implementation of national policy. Their effects have been very contrasted. Following Nemitz (1993), the core of the tangible effects is linked to the direct take-up of the recommendations by the universities. This can take many forms, with in a few instances the university asking an evaluation expert to help in the implementation of recommendations. The most common effect, however, is contained in the second information package and the university’s handling of the previous recommendations of the Committee (“in the last report you said …., this is what we have done”). On the other hand, the link with Ministry policy remains erratic, even in cases where the law requires a specific evaluation by the CNE before any new action can be taken (as for the so-called universités nouvelles, established for five years with a derogatory status). This situation led Nemitz (1993) to call for better co-ordination between evaluation work and the services in charge of “four-year contracts”, and to advocate a “decision-making-geared” follow-up by Ministry services.

Effectiveness in promoting more coherent local strategies, in supporting individual universities in shaping their policy, in reinforcing the capabilities of the “government” of individual universities, is an achievement which requires a counterpart: effectiveness in the shaping of the national university environment. Clearly the linkages with national policy making (implementation as well as formulation) are at risk. We shall see that this is also a problem faced by the CNER.

CNER and the periodic evaluations of research operators

Although it made evaluation a compulsory feature of RTD policy, the 1985 law did not define any implementation structure. The success of the CNE provided the Ministry with a framework in which to operate. It initially contemplated increasing the responsibilities of the CNE, but in 1988, after four years of existence, the CNE had only evaluated one-third of French universities. Finally, the government opted for replication. The Comité national d’évaluation de la recherche (CNER) was created in 1989 on the same basis as the CNE, i.e. independent from the Ministry of Research and reporting annually to the Président de la République. Constituted of 10 members nominated for six years after a strict nomination process involving the French Academy of Science, the CSRT and monitoring bodies, the CNER has its own logistical means and own budget and complete freedom in the choice of its timetable, methodologies and dissemination policies. One of the key issues was the definition of the operators that the CNER was to periodically evaluate: a pragmatic solution was adopted by including in its scope all the institutions and R&D activities receiving state support from the civilian budget.

The CNER started operations in 1990. In seven years of activity it has evaluated 13 operators, although the annual annex to the budget identifies more than 60 “operators” without taking into account the universities and other higher education institutions. The required time span to evaluate all the operators would appear to be very long (almost 30 years), especially compared with the objective underpinning its creation (a full round of operators for each cycle of the committee, i.e. six years). This may explain why the CNER went through a long interim period and waited over a year for the renewal of its members and the nomination, in April 1997, of its third president. At the same time, its evaluations, especially those of research institutions, are considered robust and have had significant impacts. How can this paradoxical situation be explained? The CNER itself posed these questions in its 1996 report which reflects on its six years’ experience, and the title of which, L’évaluation de la recherche, un enjeu capital, throws light on the difficulties facing the current approach.
Three aspects need to be investigated in order to better grasp the elements which explain the current situation. These relate to the rationale of the CNER, the handling of the process and the identification of and interlinkages with the main stakeholders.

The blurring of the CNER’s rationale

The décret which established the CNER sets out in detail the conditions under which “institutions, programmes and procedures” should be evaluated. It begins by defining the role of the CNER: “to assess the implementation and results of the national research and technology policy conducted by the government”. Although based on the same objective – the evaluation of individual operators – this opened the door to two different options. The initial approach used by the CNER corresponded to a systematic review of all operators, with ten evaluations started in the first 18 months. A different direction was then progressively taken, which drove the CNER to emphasize in its 1996 report that “the objective of the CNER is not limited to its recommendations on evaluated operators... It is to progressively identify the strengths and weaknesses of the whole national R&D system” (p. 5). The report underlines the major issues facing French national policy: the definition of sectoral policies; the mentoring (tutelle in French) of national programmes and research institutions; and the conditions under which these research institutions function and adapt to changing situations. The CNER even dedicated a whole annual report (1994) to the “national apparatus for research and technological development”.

Individual evaluations, however informative and whatever specific transformations they advocate, are a means to an end: to produce a(nother) global view on the national system and the transformations required. For instance, in the above-mentioned 1994 report, the CNER identified three governmental models used in France over the last 30 years:

◊ co-ordination, based on an office at Prime Minister’s level (as was the case with DGRST);
◊ autonomy (with a specific Ministry in charge of research and technology and budgetary responsibility for all civilian budgets, as was the case with the Ministry of Research and Technology in 1984-86 and then 1988-93 under Mr. Curien); and
◊ agglomeration (linking research and technology and other responsibilities, this linkage being considered a strategic issue as with industry during the 1970s).

CNER took a clear stand in favour of autonomy, while agglomeration, with higher education and education as a whole, has been the choice of all governments, the present one included, since 1993, following the German model. Compared to the CNE, a clear first lesson emerges which goes back to the reasons why the CNER was created: to produce robust and credible information on the components of the system in order to foster public debate; not to encapsulate the public debate in the closed room of the Committee, however wise and informed its members may be. Such a setting, and the guarantor model as a whole, can be productive and feed into the political debate only if the other option is chosen, that of the systematic and periodic evaluation of operators. But is this option feasible? The lessons derived from the process adopted by the CNER lead me to think that it is.
The conditions for robustness and credibility: a four-step process

As was the case for the CNE, evaluations by the CNER follow a four-step process. However, two main differences can be pointed out, pertaining to the initial phase and the handling of the experts’ reports.

The determination of its terms of reference was the first issue facing the CNER. Contrary to universities which are all shaped in the same mould, each of the other research operators is an *ad hoc* and historic construction (the CNER calls this “the result of a historical stratification”) which requires “the use of investigation modes adapted to the specificity of the operator under evaluation”. Furthermore, the CNER made it clear in its first annual report that it wished to adopt a proactive approach, focusing not on initial objectives and subsequent potential divergences, but rather on current problems and issues through which to re-analyse recent action and its effects. In order to establish adequate terms of reference and select relevant methodologies, information on the evaluée had to be collected, the main stakeholders identified and discussions organised around key issues. This characterisation phase, as the CNER terms it, rapidly became a very time-consuming task requiring specialists, “professional evaluators”. As was the case for the CNE, the global information production process is managed by internal professionals, CNER’s *chargés de mission*, under the supervision of a sub-group of the Committee. However, increasingly, and this is the main difference with the CNE process, external professionals have been called upon for specific aspects (mailed surveys to actors, economic analysis of direct and indirect effects, institutional analyses, financial and budgetary management audits, etc.). The formalisation of this “preliminary” stage, including agreement by the evaluée on the *dossier de base* before the evaluation questions are formulated, is probably the key dimension of this second version of the guarantor model, a dimension that has had lasting effects. As was the case for the CNE, it highlighted the lack of management and monitoring tools. However, in the case of the CNER, it was, in many cases, instrumental in the development of tools which have since been internalised by the evaluées: what remains an open question for the CNE constitutes one of the CNER’s major achievements. This impact on the tools used for strategic management is a clear effect of the guarantor model.

Both the CNE and the CNER have chosen the same path for the central phase of the evaluation process (the work undertaken to answer the terms of reference): full reliance on external individual (thematic or disciplinary) expertise, with the CNER having more variety in the geographical and institutional origins of its experts. As was the case for the CNE, each expert investigates a specific problem identified in the terms of reference and is required to produce an individual report on his “topic”. However, contrary to the CNE and following the usual practice of evaluation panels (Bobe and Viala, 1995), the experts are also asked to produce a collective report as their principle output (the *rapport d’instruction*). This dual approach to the experts’ report has proven effective both in ensuring, through the individual reports, effective treatment of all the questions raised, and in organising, through the collective report and the necessary compromises which the experts have had to make, a coherent approach to the operator being evaluated. Although the process raises problems about the choice of experts, as do all panel-based evaluations, to date the information base has never been challenged and the experts’ analyses and proposals have never been rejected as irrelevant.

This report serves as the basis for discussions with the evaluée. The CNER accords considerable importance to the “confrontation phase”. The report is circulated to the evaluée and corresponding Ministries, and written comments solicited. This provides the basic material for a dialogue between the experts and representatives of the evaluated body and the ten Committee members. In the words of the CNER (CNER, 1993), these confrontations have been “very fruitful” and instrumental in
promoting “acceptable” or “realistic” recommendations. In the final phase, the CNER’s own recommendations are presented in a short, synthetic document written and unanimously adopted by the Committee. This constitutes the only official document issued by the CNER which, contrary to the CNE, has not opted for a wide dissemination of evaluation results.  

This four-step process has been effective in delivering robust evaluations (i.e. the results have never been challenged as being irrelevant). However, the process is time-consuming, although the CNER admits that the “time lags could be shortened” (p. 8). The question is whether it is realistic to call for a systematic and periodic review, or whether the review process as is stands is necessary to guarantee the credibility and robustness of the evaluations. Compared with the CNE process, there are significant variations: the major difference can be seen in the characterisation phase, although the heavier burden weighs mainly on the secretariat and the external professionals it mobilises. The approach adopted for the “confrontation phase” which currently mobilises the full Committee represents another significant difference, while the involvement of the Committee in the drafting and adoption of the conclusions and recommendations is similar for the two processes. This leads me to consider that there might be an issue of “critical mass”: the secretariat of the CNER, with less than ten people, would appear to be understaffed compared to that of the CNE, and it might be necessary to enlarge the Committee in order to better share the burden between members. Therefore, there does not seem to be any structural reason why a guarantor should not be able to cover the whole body of French research operators within a reasonable period of time – as the CNE does for the universities.

*From “local” to “global” effects: the role of the CNER in policy making*

The CNER was required to periodically assess the follow-up of its evaluations. It has done so for the first three of its evaluations which all concerned research institutions (CNER, 1996). The conclusion is clear: evaluations are effective as long as the evaluated institutions can implement the recommendations on their own. As the Committee noted, follow-up largely “took place during the evaluation, the evaluation of ORSTOM being a good example (of) anticipating conclusions later formulated by the Committee” (p. 8). Evaluations have been quite influential at the operator level, especially with the development of new management tools. However, the CNER did not aim only at “local” effects, however important these may be. The CNER itself stresses that its recommendations have had limited influence on policy making or on modifications to the institutional framework.

Thus, the CNER and the CNE are faced with the issue addressed by the 1994 *Consultation nationale* in its sessions on evaluation: the absence of any take-up mechanism. How can they ensure that their recommendations are implemented or that explicit reasons are given for not following them? One proposal was to link the evaluation process to existing consultation mechanisms and to make it compulsory for the evaluee and the government to provide a follow-up report to the *Office parlementaire des choix scientifiques et techniques*, thus raising the debate about potential modifications to the national research system at the legislative level.

By dissociating evaluation from immediate policy and government management concerns, a gap is created which has two opposite effects. One is to focus the evaluation on effective long-term issues and thus promote the “proactive” and “strategic” dimensions of the evaluation exercise. This may prove highly relevant for the evaluated operators as long as they can handle them alone. However, required changes in national policy have been difficult to implement, and difficult even to propose. The guarantor model, which is effective in producing robust and strategic evaluations, requires an institutional framework to link it to the political decision-making process: it is not sufficient to focus
on the conditions for producing evaluations, their “space of circulation” has also to be established, and this cannot rely solely on the dissemination of reports.

**Evaluation and “strategic” management**

In the same way that it is not enough to grasp only the direct effects in order to analyse the impact of a programme, the indirect or structural effects of an evaluation should also be accounted for. Although direct causality may be questioned, the growing importance of evaluation has been accompanied by three major changes which are shaping the French landscape.

The adoption of a “proactive” approach, moving away from *ex-post* monitoring of past actions, emphasized the need for more long-term, “strategic” analysis to complement the over-dominant role of annual budgeting. From 1989 onwards, a progressive but continuous change has occurred in the relations between Ministries and operators. Four-year *contrats d’établissement* between the Ministry and the universities were established in 1989, after five years of action and evaluations by the CNE. These contracts take into account the diverse situations of French universities and their corresponding specific needs. This movement was reinforced by the growing involvement of the regional councils which, since the decentralisation law of the early 1980s, had become accustomed to five-year *contrats de plan État-régions*. The Ministry of Research progressively implemented a similar scheme which was generalised to most research institutions and agencies by the mid-1990s. This increasingly pluri-annual nature of government action calls for a corresponding revision of its management tools, as proposed by the CNER, which emphasized in its 1996 report the dual role of “strategic thinking” and of “evaluation” in the new process.

The second major change was mentioned above, and deals with the operators’ management tools. Both Committees have continued to insist on the inadequacy of current monitoring instruments and have dedicated significant efforts to starting to develop new evaluation tools. This has had a clear impact on the managerial settings in the evaluated institutions themselves and also more widely: some research institutions, such as CEMAGREF, or research agencies, such as ADEME, have developed functional evaluation and strategy directorates; and the Ministry in charge of equipment, transport, construction and housing research has set up its own evaluation unit.

However, the major change was implemented by the CNRS. The *Délégation des audits*, created to report on issues of crucial interest to top management, has been very active and, through a very rigorous process controlled by its own guarantor, the *Comité consultatif des audits*, has produced evaluations dealing as much with common facilities (*e.g.* the CNRS computer frame) as with interdisciplinary research programmes (such as Imabio or Ultimatech) or, more recently, with the functioning of the *Comité national*, in charge of the management of researchers’ careers. The evaluation process is, again, different and is built into the decision-making process of the CNRS. The issues to be examined are decided by the CNRS. An expert group is then set up, with the task of defining the terms of reference and, with the support of the *Délégation des audits*, selecting the external auditor(s). The findings of the auditor(s) are then discussed with the evaluated body before the expert group presents its results. The whole process is overseen by the *Comité consultatif des audits* which guarantees the relevance of the methodology adopted and the quality of the study. The process terminates with the presentation by the directorate of CNRS of the new “orientations” decided. The whole process – with its five components – is made public. The organisational audit of the *Comité national* (CNRS, 1995) is exemplary of this dynamism which shows that, provided the
government of the institution does make use of the instrument, the loop can be looped and the lessons learnt embedded into the decision-making process.

The above-mentioned evaluation is also exemplary of one issue on which all three guarantors have been placing increasingly greater emphasis: the handling – not to say the management – of “human resources”. An audit of the Comité national would have been unthinkable only a few years’ ago: in the early 1990s, it took more than two years to redefine the borders between its 50 sections! The recruitment, careers and activities of the enseignants-chercheurs is an issue raised in all the CNE reports to the Président de la République. The sub-titles used in the reports on three evaluations of research institutions carried out by the CNER (CNER, 1993b) confront the reader with similar issues: “the problem of the scientific evaluation of researchers”, “the need for a new system of human-resource management”, “career opportunities for young researchers”, “the evaluation of laboratories”, “the co-ordination of a network of laboratories”, etc. And what better recognition of the crucial importance of this dimension than CNER’s description of the “production structures” of the French research system as “disseminated in thousands of research units”, “a mosaic of laboratories” (CNER, 1996, p. 37). Operators’ evaluations and their guarantors take us back to the basic unit of production, to the initial focus of the evaluation exercises: the researchers and their research units. It is therefore logical to close this analysis by concentrating on some recent evaluation approaches which have been designed to tackle this important issue.

**A renewed stake: the evaluation of research actors**

There is no need to present at length the dual rationale of research as a public good which justifies state intervention and the delegation to peers of the reviews on which this state intervention is based. Neither is there any need to pinpoint the long-standing choices that have been made in France to handle this delegation process: it is jointly based on research institutions with their own research staff and on the embedment of research in the duties of university staff. The French public research capacity is currently divided into two almost equivalent groups of professionals: full-time professionals in research institutions (the organismes publics de recherche), and university professionals, who by status are enseignants-chercheurs and are supposed to devote half of their time to research (a situation which is supposed to be taken into consideration when teaching duties are assigned). Both are recruited young and the development of their careers is intrinsically based on periodic evaluations: whether they become professeurs or directeurs de recherche depends on choices delegated to evaluating bodies, which are established as such and are independent from the professionals’ direct hierarchy. The Comité national, linked to the CNRS, and the Comité national des universités, with their numerous disciplinary “sections”, are examples of this situation which was extended to all the établissements publics à caractère scientifique et technique (EPST) by the 1982 law. Not to take account of the debates which have taken place over the last decade, both at the individual level and, increasingly, at the collective level of the “research units”, would be to miss out on a significant part of the French evaluation scene. Scant attention has been paid to these dimensions in international fora during the last decade, although they might well become, and not only in France, a major issue in the coming years.

This is why I shall present two “original” configurations linked to so-called “mission-oriented institutions”: the scientific follow-up of researchers at INRA, the national institute for agricultural research; and the periodic monitoring of research units in INSERM, the French national institute for medical research and biomedical sciences. It is, of course, not possible in such a short presentation, to give due credit to the complexity and richness of these configurations. Instead, I have voluntarily
chosen to limit myself to stylised figures highlighting the conditions under which evaluation has become an integral part of the management and policy-making process of these research institutions.

The scientific evaluation of researchers at INRA

INRA, the French national institute for agricultural research, is organised in seven scientific directorates in charge of some 20 research departments, each covering a theme or an area. A typical department comprises 6 to 12 research units and 100 to 200 researchers, spread over a dozen different locations covering the whole of France. Researchers are recruited in their mid-twenties (after their PhD) as chargés de recherche and become directeurs de recherche in their forties. Apart from these two rendez-vous, researchers’ career paths depend on their hierarchy.

This may explain why, when INRA became an Établissement public à caractère scientifique et technique (EPST) in the mid-1980s, a new mechanism for the periodic scientific evaluation of researchers was established. The 1 900 INRA researchers are evaluated every two years by their respective “specialised scientific commission” (CSS). There are 13 such commissions, whose members (between 10 and 20 depending on the size of the population under review) are nominated for four years with a fixed representation: half are external scientists, the other half are from the institute, split equally between those elected by researchers and those nominated by the directorate of INRA.

The evaluation is consultative, it operates as an early-warning mechanism for both the researcher and his supervisors. But how can such an early-warning mechanism operate without becoming a routine task with, at best, limited influence? I shall try to show that its relevance lies in the small organisational “details” which embed the process into the life of the institution.

First, however, it is important to “measure” the interest of this early-warning mechanism for “mission-oriented” institutions. A simple but robust indicator is the relative importance of identified problems: in 1996, they affected nearly 12 per cent of the researchers evaluated. Examination of these situations further indicates that such problems tend to be concentrated in specific periods in the research life cycle. They overwhelmingly concern the two extremes: “younger” researchers (under 35, i.e. with less than ten years’ experience in the institution); and “older” ones (over 50). The reasons for this “weakness in scientific production” differ significantly between these two groups. In the case of younger researchers, two major reasons emerge, both of which are linked to their participation in the research group: “dispersion” – i.e. too many projects lacking sufficient continuity to enable the academic capitalisation of the knowledge acquired. This is often linked to participation in the numerous contracts with which the unit is involved. On the other hand, “wavering” is often related to the difficulties faced in handling a project which is poorly linked to the other work undertaken in the team and thus suffers from limited backing or room for discussion. It is easy to see how such situations can evolve over a two-year period, i.e. between two evaluations, reinforcing the importance of this “early-warning” system (which then becomes a “follow-up” process). The importance of the system does not diminish for the second group of “older” researchers. Although the situations tend to be more varied, two main patterns emerge: first, “thematic exhaustion”, where an individual progressively exhausts his/her subject without finding sufficient “challenge” (in most cases this is linked to the progressive marginalisation of a researcher who has not followed the changing interests and approaches of his/her research unit); second, the progressive transformation of the researcher into a service provider (often of a complex technical
service), which is useful for the lab or the related economic activity, but which does not correspond to any “new” scientific activity.

How can such a characterisation be obtained? A simple evaluation of scientific production would not allow such a depth of understanding. This is where the organisational setting comes in. In addition to the wide dissemination of the findings, three aspects of the process are worthy of note. First, researchers describe not only their personal research activities, with the corresponding list of publications (of whatever type during the last two years); they also describe their other activities, indicating how their time is split between: teaching (with specific attention to PhD tutoring), “collective” activities (management of a research unit, co-ordination of a European project, etc.) and “transfer” activities (expertise provided to public bodies or the profession, patenting, etc.). In addition, they have the opportunity to address a personal message to their commission; this opportunity was taken by 38 per cent of researchers in 1996. Second, the reasons for the warning must be presented in writing and transmitted to both the researcher and his/her hierarchy: this fact alone has often lead commissions to re-organise the process in order to be in a position to discuss all cases which pose problems and to formalise their justifications. Third, for each case, the corresponding scientific director presents to the commission his point of view and the action he expects to take. These exchanges are central to the process. It is interesting to note the three main situations observed: i) in many cases, the situation had already been identified and the warning was given as a complementary diagnosis in support of the action taken; ii) in a significant number of cases, the relevance for the institution of the researcher’s involvement in other activities is such that the “temporary” effect on his/her scientific activity is judged as a necessary counterpart; iii) finally, in a far more limited number of cases, the process operates as a true warning mechanism, especially for young researchers, and thus helps in the strategic management of the research institution.

The term “strategic management” is not accidental: it is truly a strategic issue since, in France, research institutions recruit their researchers “for life”; the institutions’ capabilities thus largely rely on the adequate “shaping” of their human resources. This example shows that in order for evaluations to play the role of an “early-warning” system they must be linked to the decision-making structures and this relies mainly on the small organisational “details” that have been put in place. These two dimensions – relevance to the core dynamics of the institution and embedment in its management – are central to the performance of any evaluation process, although the long-term effects on the shaping of the competences of the institute as a whole have yet to be seen. It should also be noted that this “early-warning” system, which took over a decade to become established, remains the exception rather than the rule.

**The evaluation of research units at INSERM**

If we follow the results of the analyses by INRA, most of the problems faced by individual researchers are linked to the relationship between the individual and the research unit in which he works. I will not go into the reasons behind the increasing centrality of the “research unit” in the dynamics of knowledge, and shall limit myself to an image we often use at CSI: the “laboratory” is to science what the firm is to the economy – the basic unit of production. It comes as no surprise then that over the last decade fairly significant transformations in the approach to research activities have taken place in many countries, even in United Kingdom where basic funding is now allocated through a grading of university departments and centres. This is a question that French institutions have faced for some time, each developing its own approach. The CNRS has a long tradition of four-yearly examinations of “own” and “associated” laboratories. However, nowhere has this practice been so
Research units are established for four years and their mandates can be renewed twice, meaning that after 12 years they are either “closed” or “recreated”. Creation is a bottom-up process through a yearly call for tender. No thematic priorities are expressed by the institution, but clear criteria have been formulated in order to promote hybridisation: a unit must bring together at least two INSERM researchers and two non-INSERM ones (mainly from university and university hospitals), and it must find its own location (again most often within a hospital or university setting). In 15 years of activity, from 1983 to 1997, the number of INSERM units has remained relatively stable (growing from 236 to 260). Over the same period 189 units have been closed, 70 have been “recreated” and 213 have been “created”. Before examining the conditions under which this level of renewal takes place, let us first analyse the situation.

Three situations need to be taken into consideration: units asking for prolongation after four or eight years of existence (35, on average, over the last two years); units arriving at their twelfth year of life and which thus will be either closed or “recreated” (around 30 per year); and proposals for new units (45 on average, with over half corresponding to units asking to be “recreated”).

The renewal of the mandates of existing units is the overwhelming rule since, on average, only one unit in 20 was closed after a negative evaluation, and around one unit in ten has been placed “under surveillance” with a follow-up evaluation planned in two years. Although more than four out of five units go through the evaluation process without change, the failure rate is high enough for the units not to regard it as a simple routine check.

The process for “closing” units (those with 12 years of existence, of which there were 45 in 1992-95) is, of course, very different. One year before the closure date, a specific call for tender is made to identify whether or not there are candidates to take over the unit as it stands: few candidates take up this opportunity (only six cases out of 45). The units are then given a three-year “temporary mandate” in order to prepare their “recreation”. The vast majority of labs (34 out of 45) choose to directly enter the global competition for the creation of new units. One-quarter failed (nine) which, added to the few cases where no project emerged (five), brings the total number of real closures to one-third of all “closing” units during 1992-95.

The majority of the “closing” units thus enter the basic competition for the “creation” or “recreation” of new units. In 1992-95, more than 90 units were created in this way. To understand the nature of this renewal process, it is important to grasp the origins of the units. One unit in three was a “recreation” (either directly or via a “temporary mandate”). This means that two-thirds of new units were completely new for the institution. What are the processes that foster this high creation rate? “Self” generation (i.e. creation from scratch, mixing staff from various existing groupings), exists but is relatively rare (one in five new units). Two major phenomena of relatively equal importance help to explain the process. The first relates to the units’ research dynamics, with a team progressively developing its own agenda and shifting the balance of its collaborations from within the laboratory to external partners and thus calling for full recognition as a unit (often in conjunction with some of its external partners). This “birth” process can also be generated from outside existing teams. To stimulate the process, INSERM has established a competition to allocate a five-year project to “young teams” to give them the time and opportunity to prepare for the main competition. This has
proven highly efficient, since one-third of all creations in 1992-95 (including “recreations”) stemmed from this preparation process.

How did this encouraging situation come about? Three elements of the organisational setting are worthy of mention.

The examination of proposals is a two-step process which involves the scientific council of INSERM (30 members, renewed every four years) and, as in INRA, a set of “specialised scientific commissions” (CSS). There are currently 11 such CSS for some 2,000 INSERM researchers and 260 research units. While, similar to the process at INRA, the composition of commissions is based on strict rules,42 the number and heading of commissions changes every four years. This is the first specificity of the process. The commissions are not discipline-based (how could one expect to cover the whole spectrum of relevant sciences in 11 commissions?); there is a changing mix between discipline-oriented and disease-oriented commissions. Changes can be quite significant, as in 1995 which witnessed the disappearance of “biomedical engineering” as a full commission and the recognition of “transmissible diseases” as a completely new commission. This movement is all the more important since each proposer, when putting forward his/her project, must also designate the commission that will analyse it.

Commissions are central to the process. The project proposal provides a basis for analysis but, in order to better assess the proposal according to the recommendations put forward by the institution,43 the core of the evaluation takes the form of a classic site visit: for each proposal, the commission sets up a group of six persons, five commission members plus a foreign expert; this group makes a full-day visit to the site, listening to exposés, auditing the potential director as well as each team leader and the technical staff. The report of the group44 is confidential – in order to foster more pertinent comments – which means that the evaluators do not have access to it. Instead, they receive a one-page summary of the recommendation of the commission after a debate on the group’s report. Proposers can appeal by providing supporting arguments to the scientific council which devotes each year a full meeting to the examination of all cases: although the commissions’ proposals are endorsed in the vast majority of cases, this counter-examination process has been instrumental in changing the outcomes of such evaluations. The process can go either way (i.e. the case of biomedical engineering, where the scientific council recommended the closure of a number of units which the specialised commission had proposed to maintain).

A third key element lies in the handling of closed units: the closing unit has a further 18 months of budget to enable the researchers and technical staff to finish their ongoing projects and join a new research setting. For the supervisors, this mechanism is crucial to the performance of the whole system since it provides the time and means for the unit to adequately reposition itself. Judgements on the relevance of the research unit rarely have any negative effect at the individual level: the commissions undertake the same screening of individuals as for INRA, but with a far lower rate of warnings: between 1992 and 1995 warnings were given to only 2 per cent of the full researcher population (although it was screened twice). Comparison of the two rates is illuminating and shows how central the “research units” are to the long-term management of the relevance and orientation of the capacities of the institute. “Periodic” evaluation in its fullest sense – covering ex-ante, ongoing and ex-post exercises – is the major instrument used to promote a satisfactory collective setting in which research work can develop.

These two examples remind us that “human resources” are not only a fashionable issue, but that the handling of the human-resource element is crucial to the performance of national research
systems. Evaluation has long been considered the major instrument for handling these issues. The two cases above confirm this but, at the same time, indicate that, although peer review remains the cornerstone of evaluation, the choice of the “units of analysis”, the process which drives it and its linkage to “strategic decision making” are central to its relevance.

**Striking the right balance between evaluation and management: an ongoing issue**

Evaluation was fashionable during the 1980s, particularly in Europe and in France. However, the fashion has since waned and few policy makers emphasize evaluation in their speeches (or, if so, refer only to EU programmes!). Still, although it is going through a difficult period, the activity has not disappeared. On the contrary, the number of evaluations has steadily increased and we have witnessed a significant institutional spread. This illustrates a lasting problem and the need to clearly understand what the policy has produced in terms of action. The role of evaluation is to foster the dimension underlined by economists as crucial to the success of innovations, the learning processes and the numerous loops, adjustments and redefinitions between policy makers and the numerous and diverse stakeholders concerned by these policies. It is only through this iterative and never-ending process that objectives progressively take shape, implementation structures evolve and policy goals or finalities are revealed and transformed. The problem is then to identify under what conditions “evaluation” can “loop the loop”.

The French experience highlights three conditions which mirror the role of co-ordinating mechanisms in innovation processes (Larédo, 1996). They deal with the intermediaries produced (i.e. the content of evaluations and their interlinking capabilities), their production process (i.e. the conditions which make them credible for stakeholders), and the structural arrangements necessary for confrontations, debates and alliances on the periodic (re)definition of policies. The guarantor model, as used in France by both the CNE and CNER, responds to the first two conditions, delivering robust, relevant and credible evaluations. However, its institutional framing is paradoxically based upon a “linear” assumption which assumes that the mere existence of written evaluations is sufficient to stimulate “political” debate and change. As Callon and Rip (1992) or Finne (1995) have shown, there is a need for the creation of “hybrid fora” or “arenas” for debate. It remains to be seen whether this can be fostered by more adequate institutional framing (as was suggested during the 1994 French Consultation nationale).

Finally, this approach applies not only at the national level; it also concerns the “government of operators” to use the revealing term used by the CNE. Here, evaluation first and foremost rhymes with the evaluation of researchers and their research settings. We have used two recent examples to illustrate the conditions under which a guarantor approach, largely based on peer judgement, can be productive when embedded into the strategic management of research operators.
NOTES

1. One should be careful not to consider such transformations as linked only to broader political issues. The movement towards a full Ministry had been prepared before 1981, and the 1997 political change has maintained the choice made at that time, i.e. associating research and technology with education in general.

2. This is not to say that other evaluations have not played an important role; this is especially true of the work by R. Chabbal on SMEs and ANVAR (1995-96), but they fall into the well-known tradition of mixing *ad hoc* commissioning with the recognised “wise man” status of the individual in charge. This exceptional auditing approach is a long-standing practice of French S&T policy makers. During the period under consideration, it was still being used by policy makers, especially in one case in which two of the three evaluators were nominated as heads of the evaluated institution (CEA in 1986).

3. This focus will also leave out the numerous but still *ad hoc* initiatives of some French *conseils régionaux* to evaluate their own involvement in STI activities.

4. F. Jacq (1996) in his PhD thesis on the history of French research policy shows that the reference is somewhat idealised and that the *comité consultatif de la recherche scientifique et technique* had a major activity and influence only during the first years of its creation.

5. In his introduction to the one-year seminar which took place in 1990-91, the *Commissaire au Plan* stated: “What characterises the evaluation of public policies in France is a delay” (CGP, 1991).

6. I shall not deal here with the specificities of the French situation regarding the evaluation of public policies. Let us simply say that the institutionalisation of evaluation for S&T policy preceded the institutionalisation of general policies. The four principles put forward by Viveret: independence (“to avoid confiscation by any given power”), pluralism, methodological rigour and transparency in its materials, are very much linked to those put forward for the evaluation of S&T policy: timing, relevance, robustness and credibility. In both cases, they are based on the guarantor model, the *conseil scientifique de l’évaluation* (CSE) acting as such in the general case. However, two major differences contrast these two approaches. While in the case of research, evaluation is systematic and periodical and covers the whole spectrum, for other policies it depends on the decision of a government interministerial commission (CIME); the CSE has only an advisory role to “oversee the quality and objectivity” of evaluations (by intervening at the beginning on the relevance of the proposed implementation and methods, and towards the end in order to judge the quality of the work done) while they also have implementation responsibilities in S&T. The CSE also has a more global methodological role (see the recent report on the methodologies for evaluation).

7. To follow the headings of the 1982 and 1985 laws on research and the title of the last report (dated September 1995).

8. The historian will find a firm connection with the short-lived *conseil supérieur de la recherche scientifique et du progrès technique* created in 1953, by P. Mendes-France who also created the first *Secretariat d’État à la recherche*.

9. Although the COS has now been in existence for two years, no such report has yet been produced. The only available report deals with researchers’ activities and careers. Still, it must be underlined that COS was presented by the Minister when it was created as his own scientific council providing only “private” output and advice.
10. For instance, many analysts of the French research system emphasize the role of the *Assises nationales* in the changing relationship between industry and public research.

11. This did not refrain Ministers from requesting such global external views, as was the case with the OECD evaluation of French innovation policy (OECD, 1986).

12. In evaluation circles, we often consider that “quick and dirty” evaluations are of better use than robust evaluations which arrive too late. Apart from the fact that the former have seldom been used (at least in the French context), they have had – in the French context which witnessed a full wave of “audits” at the end of the 1970s – a lasting counter-productive effect, that of complete and utter distrust by researchers, their representatives and research managers of those exceptional one-man exercises which served no other purpose than to promote choices which had already been made.


14. This was carried out in two steps, with the whole structure being fully stabilised at the end of the 1980s.

15. To ensure continuity, half of the members are renewed every two years.

16. It can also evaluate other higher education institutions such as the *grandes écoles* (if they, or the department in charge, request it). New missions can be added by law, as has been the case for the contractual process which the Ministry has entered into with the universities or for the so-called *universités nouvelles*.

17. This check list was developed over time and is now organised around five main headings: students (composition, success rate, further employment), teaching staff (number, residence, effective teaching activities, involvement in research), technical staff (qualifications, duties), administrative and financial management (management-gearred analysis, balance between resources, linkages with regional authorities and economic partners, investment policy), common facilities (university libraries, job search, but also sport, cultural, health facilities) (CNE, 1993). The term “questionnaire” is rather misleading, since most of the time it also encompasses visits by and extensive dialogue with the secretariat.

18. In addition to the specific issues identified, the terms of reference add transversal issues which the Committee is evaluating. This is currently the case with university libraries, university research policies (the object of an ongoing joint examination with CNER), regional funding (comparative analysis of roles and practices), the handling of non-teaching staff and a departmental evaluation (pharmaceutical studies).

19. The number of experts varies and there have been as many evaluations involving less than five experts as with over 20 experts. However, the Committee has an established policy to limit the number as far as possible. Over an eight-year period, 1 000 evaluations have been carried out. Ninety per cent of the experts are university professors, with only 4 per cent of foreign origin (CNE, 1993).

20. In an average year (1993), the Committee sent 13 000 copies of the reports published during that year and distributed 1 000 more on specific requirements. The reports are free of charge.

21. But return evaluations also indicate cases where the identified weakness of the university government has not evolved over time, and where the same situation is still being faced.

22. For a more in-depth analysis of CNER’s first years, see Larédo and Mustar (1995a).

23. State support being defined by the 22 Ministries which signed the *décret*. The path was also left open for other institutions – private or military – to ask CNER to evaluate them, or for Ministers to ask CNER to evaluate any of the organisations funded by their department.
24. Each year the Ministry in charge of research produces the *État de la recherche et du développement technologique* as an annex to the *projet de loi de finances*. The calculations are based, as suggested in the CNER’s *décret*, on the different operators identified in these “yellow books” from 1989 to 1995.

25. The first cycle saw two presidents, the first having resigned after two years due to his taking up of a new position as chairman of one of the largest French research institutions.

26. The denominations of the three models are the sole responsibility of the author of this paper.

27. CNER, 1996. The citations which follow (unless specific mention is made) are taken from the same report.

28. The CNER’s secretariat underlines, as is also the case for the CNE, the central role of the *chargés de mission* in the actual drafting of the reports.

29. When it is a research institution, both management and workers’ representatives are confronted with experts.

30. The report is available to the public, and the seven first recommendations have been collected in a book published and sold by La Documentation française (1993). A new edition will be published in 1998, presenting the most recent recommendations. All other documents are not endorsed by the CNER and are unofficial, although some have been widely circulated.

31. Respectively 30 000 and 40 000 (Mustar, 1997).

32. In French the official wording is *unité de recherche* while the usual terminology is *laboratoire*, a term which is often misinterpreted in English since, in French, it does not bear any institutional meaning, referring simply to the collective setting in which research activities take place and have been at the core of “recent” science studies. See Latour and Woolgar 1979; Latour 1987.

33. Both situations (recruitment, directorship) correspond to traditional peer-review situations: it is a *concours* (i.e. the number of available positions is known beforehand and only the “best” qualified candidates are selected); the selecting panel is made up of peers, most of whom come from outside INRA: the criterion of excellence (with the number of publications playing major role) is predominant, even though other criteria are increasingly taken into consideration when dealing with promotion to research directors.

34. Following the *loi d’orientation et de programmation pour la recherche et le développement technologique de la France* (LOP) in 1982, a new institutional status – *Établissement public à caractère scientifique et technique* – was created to fit the specific conditions of research activities which, it was thought, might have difficulties developing in the standard framework of administrative establishments. For researchers, this new status had two major consequences: i) they became civil servants (whereas before they were only “under contract” to public research institutes); and ii) it required the creation of a mechanism for their periodic scientific evaluation. This transformation process took some years and ended in rules and configurations specifically tailored to the history and situation of each research institution.

35. They are nominated by the directorate of INRA. In 1996, one-fifth came from teaching institutions linked to agriculture (and often with joint labs with INRA), and the others nearly equally shared between universities and other research institutions (with as usual a significant number of university professors linked to labs associated with the CNRS).

36. Many researchers in such situations seek an “outside” posting, *i.e.* a temporary position outside INRA, usually in a non-scientific position. Those who are unable to stabilise this “managerial” repositioning, face the same problems when they come back to their research positions.
37. The Secrétariat général of the CSS publishes a yearly statement. All data presented here is derived from the *Bilan de la campagne 1996, Préparation de la campagne 1997*, which is widely disseminated within INRA.

38. The statistical analysis of this data has enabled four main configurations to be identified which cross over the institutional situations, showing once more the interest of being in a position to grasp the full range of activities of evaluated researchers.

39. Again their analysis is very revealing. Apart from those explaining the difficulties faced in handling the research programme, three main arguments were developed aiming at directing the evaluation process. Coherence over time was the first argument (“you said that last time and I have started following it this way”). The second argument identified present imbalances, arguing about their temporary (“I had to take a new responsibility, now that things are settling...”) or inherent nature (this is especially true for those in charge of research units). The third attitude links present scientific difficulties to the collective setting (“my position within the research unit”, “the current problems of the unit as a whole”, a “thematic reorientation of the unit which leads to a personal reorientation”, “the difficult relationship with the unit’s hierarchy”).

40. The researcher has the right to refuse the commission’s judgement (there is an official procedure of appeal which only two researchers used in 1996) or to reply to the commission (one researcher in six, 16 per cent, took this opportunity in 1996). This, in particular, helped in solving cases where, due to an incomplete file or the absence of an explanation, the commission gave an undeserved warning.

41. Less than 10 per cent of INSERM personnel work in buildings owned by INSERM.

42. But the balance is different: 60 per cent of members are elected while the other 40 per cent are nominated by the Ministry on the proposal of the general director of INSERM. The directorate of INSERM is thus far less autonomous than that of INRA, where the directorate nominates half of the INRA members and all the outside members.

43. The institution proposes a set of dimensions to complement the originality and quality of the scientific work, which is widely known as the key judgement criterion used by INSERM commissions. These relate to partnerships entered into, to social, medical or economic applications deriving from scientific work, to teaching activities (and especially to PhD training). Other aspects deal with the own dynamics of the unit: its embedment in its geographical setting, the relevance of buildings and equipment and the access to “heavy” equipment, and the handling of individual careers (especially mobility).

44. Practical aspects matter here: the commission designates those (generally two) of the five members who will act as rapporteurs, draft and present the report. The foreign expert is required to write his/her own report.
REFERENCES


CGP (Commisariat Général au Plan) (1991), Outils, pratiques, institutions pour évaluer les politiques publiques, La documentation française, Paris.


CNER (1993a), L’évaluation de la recherche, réflexions et pratiques, La documentation française, Paris.


CNER (1996), L’évaluation de la recherche : un enjeu capital, Rapport d’activité au Président de la République.


MINISTÈRE DE L’ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE (1994), *Consultation nationale sur les grands objectifs de la recherche française*.


