

Research evaluation of the Italian CNR institutes: a missed opportunity?

Stefania Giuffrida (*), Alberto Silvani(°), Giorgio Sirilli (*)

Consiglio Nazionale delle Ricerche (*)
Università di Milano (°)

Abstract

The paper deals with the institutional evaluation of the research infrastructure of the Italian National Research Council (CNR). The evaluation was commissioned by the agency itself and was carried out by a panel of experts. The paper analyses the procedure adopted and the results obtained with the aim of evaluating to what extent the exercise was appropriately designed and performed, identifying the strengths and weaknesses of the methodology adopted, suggesting relevant methodological changes, evaluating the impact of the evaluation within and outside the agency. The exercise is compared with similar evaluations carried out by other European agencies. From a science policy viewpoint it is argued that the evaluation of the CNR institutes is a missed opportunity if its outcome is not used for policy making – and after nine months from the release of the evaluation report it has not been so used.

Introduction

The increasing importance of knowledge as a key factor in the development of society and the growing complexity of innovation due to the wider application of science and technology, requires a thorough reconsideration of R&D policies.

The “soft side” of knowledge creation and diffusion also acquires greater importance: human resources, with their managerial and organizational skills, play an increasing role in the performance of R&D activities.

Moreover, a number of different links between the international, national and local levels of innovation systems lead to a greater complexity in the relationships between the various stakeholders and to the creation of a marked interdependence between the different levels and the ensuing increase in secondary effects that this interdependence gives rise to.

The dynamics of these ongoing processes brings about changes, arouses interests, causes different options to develop: all these factors result in an increasing request for evaluation that is directed towards the actors involved in the decision-making process leading to the definition of R&D, education and innovation policies as well as towards the scientific institutions. These institutions differ from many other social systems both in their type of ultimate goals, which are basically the pursuit of knowledge and not necessarily the achievement of socio-economic objectives, and in their effort to achieve higher quality by means of comparative analysis and peer review. In other words, the peculiarity of the subject and the real difficulties that are connected with R&D evaluation have so far led to a distinction being made between an “internal” evaluation, based on rules and procedures created within the scientific community, and an “external” evaluation, linked to the need for justifying the use of public money by the scientific community and assessing the possible impact of their activities (Lyall et al., 2004; A. Baccini, 2010).

The demand of evaluation exhibits new features. On the one hand, there is an increasing social demand to re-negotiate the “social contract” between taxpayers and scientists according to which R&D yields, by definition, positive results in terms of new knowledge and its application (products, processes and findings): concerns are raised in connection with issues like the ethical implications of biotechnologies or the impact of new discoveries and new technologies on the physical environment. On the other hand, from the evaluation perspective it is difficult to look at all scientific developments in a harmonious and coherent framework, and often contradictions are inherent in the exercise – how to reconcile different views on scientific excellence, policies, goals, impacts. How common criteria could be standardized and used to evaluate all research in a comparative way? And how to offset costs against benefits? In this context the outcome of evaluation exercises must be aimed at fostering a greater awareness of the social and economic benefits of policy-making (Airaghi et al., 1999); measurement tools and standards could help but should be properly used (Fahrenkrog et al., 2002).

Generally speaking, it may be argued that the conceptual setting of research evaluation tends to focus on the identification of requirements - including evaluation criteria and their implementation – and on the definition of questions which cannot be answered by means of the traditional methods of assessment. These methods are *de facto* often limited to ranking projects or institutions in order of scientific merit. The increasing complexity of the R&D system means that the notion of the evaluation process being merely a way of assessing results is now obsolete and that it should now be thought of as an essential part of the feedback for designing R&D policies (Georghiou, 1998; Georghiou, Larédo, 2006).

New evaluation demands, wide-reaching impact analyses, oriented assessment activities and data gathering also affect the evaluation divide between national evaluation systems. This has

important consequences in terms of successful models to be adopted (CNR, 2003; Silvani, Sirilli, 2001).

Over the last few years the evaluation of research and higher education has become a key issue in the debate on science and technology policy in Italy. This debate is no longer restricted to the inner circle of people and organizations directly involved in the system, but has become part of the political discourse in the mass media, centered mostly on the quality and performance of universities and research organisations. In this context the governance and management of scientific institutions and the overall functioning of the research system as such still remain neglected. Nevertheless the need to address the issue of the efficiency and effectiveness of the public research and higher education system has led to the institutionalization of evaluation with the recent creation of ANVUR, the National agency for the evaluation of the university and research. The agency is expected, in the mind of some policy makers, to play a “saving” role of a heavily criticized public infrastructure: the government argues that a pre-condition for pouring money into the public research and education system is the “check” of the “quality” of each individual component (Silvani et al., 2005).

The most relevant experience in Italy was the evaluation of universities, research agencies (including the National Research Council – CNR) and some private organizations financed by the public sector carried out almost ten years ago by CIVR (Research evaluation committee) set up by the Ministry for universities and research (MIUR). The Committee delivered a report (CIVR, 2006) drafted by a panel of experts and external referees using the simplified model of the English Research Assessment Exercise (RAE, 2008). The results of this exercise were used by the minister to allocate a part of the “additional fund” of the yearly budget (7 per cent of the total) to the scientific institutions supervised by MIUR on the basis of their scientific merit.

After the CIVR evaluation, and before its repetition due to be made in 2011, CNR decided to carry out a self evaluation with the aim of identifying the strengths and weaknesses of its research institutes, giving visibility to their excellence, attaching priority to the most promising research lines amenable to attract external funding. This was a timely initiative, given that CNR is being reorganized, and its new statute is going to be soon adopted by MIUR, its supervising and financing governmental body.

This paper analyses the evaluation procedure adopted by CNR with the aim of evaluating to what extent the exercise was appropriately designed and carried out, identifying the strengths and weaknesses of the methodology adopted, suggesting relevant methodological changes, evaluating the impact of the evaluation within and outside the agency. The final section deals with implications in terms of science policy.

Methods and procedures

The evaluation of the CNR institutes was set forth in the CNR the guidelines “*Obiettivi, modalità e criteri per la valutazione degli Istituti del CNR*”, adopted in 2007 by the CNR board of directors (CNR, 2007). In the mandate, actors, procedures and time schedule of the evaluation exercise were described.

The objective of the exercise was to:

- evaluate the past activity of the research institutes in order to: verify the matching between their original mission and the present research lines; measure the quality of the scientific results in the international context; identify the weaknesses of the system and the ways to tackle the problems;

- identify the perspectives of the institutes, in order to: implement the necessary organizational changes; increase the value of existing potentialities; develop the most promising research lines able to attract more financial resources in a context of budget restraints.

The main bodies involved in the evaluation process were:

- the General Panel (GP), composed of 16 distinguished scientists and technologists belonging to the Italian scientific community (one for each of the macro-areas in which CNR operates), responsible for the overall management and coordination of the process and for the drafting of the final report to be delivered to the CNR board of directors. It may be assumed that the decision to select only Italian scientists was due to the fact that the GP was meant to play basically a role of supervision, monitoring and dialogue with the various actors, and that the members' profile should reflect their recognised scientific standing, knowledge of the national S&T system, and the capacity to synthesize a large amount of information;
- 26 Thematic Panels (TP) composed of 156 scientists (40% foreigners or Italian expatriates) with different and complementary competences, able to evaluate the heterogeneous activities of the CNR institutes.

During the process, GP and TPs were assisted by a task force composed of technical and administrative staff, responsible to provide the necessary support, materials, information and assistance to the experts.

In the Guidelines it was specified that the evaluation would start in 2007, but due to various reasons, including a significant change in the composition of the responsibilities of top management, it started in April 2009 and was completed in March 2010.

At the outset of the exercise the GP elaborated a questionnaire to be completed by the institute directors to collect information on seven indicators for the period 2003-2007:

- personnel;
- number of publications (type, number of citations, collaboration with other organizations);
- capability to promote and disseminate research results;
- intellectual property rights (patents and copyrights);
- teaching activity and scientific exchange;
- participation in scientific projects and budget acquired by third party funding;
- management of infrastructure.

The questionnaire to be filled out by the institutes in the area of social sciences and humanities was slightly modified in order to take into account their own specificities (e.g. the number of citations or translations of books for social sciences and humanities has a quite different meaning than in the case of natural sciences and engineering).

The questionnaire was complemented by a short description of the mission of the institute and the main results achieved in the five year period, i.e. the most important publications, citations, national and international awards, spin-off enterprises, collaborative agreements with firms.

The number of citations was calculated using different data bases: in order to evaluate the quality of the data provided, institutes were requested by TPs to specify the source of information (ISI Web of Science, Scopus, Google Scholar). Furthermore, institutes were requested to include only the publications which received more than 20 citation with no time restrictions ("old" publications were considered worth attention) and to exclude self-citations. This information base was meant to provide the GP with a sound "objective" knowledge of scientific production, personnel, financial resources, as well as "subjective" and qualitative information on the infrastructure, quality of the scientific and technical output, the capability of innovation of the institutes, the social and economic spillovers of the scientific activities.

One of the aims of the evaluation was to give a total score to each institute in a comparative way. The GP developed a weighting algorithm whereby the total of the scores assigned to the seven indicators should be equal to maximum a 100. In order to take into account the specific characteristics of the various sectors, TPs were allowed to define their own maximum scores for each individual indicator. The only constraint was that the first indicator, publications, should receive a minimum score of 40 for the natural sciences and engineering and of 50 for social sciences and humanities. The maximum scores for each TP are reported in Table 1. It should be noted that some TPs, namely Physics, Civil, Industrial and Computer engineering lumped together some dimensions.

Table 1. Maximum scores assigned by Thematic Panels

Panel	Publications	Promotion and dissemination	Editorial activity	Patents	Training	Projects and contracts	Management facilities and infrastructures	Total
A.1 - Mathematics	50	10	10	0	10	15	5	100
A.2 - Computer sciences	40	20	5	5	10	10	10	100
B.1 - Physics	55	5	5		5	10	20	100
C.1 - Chemistry	45	5	5	10	5	25	5	100
D.1 - Material sciences and technologies	50	5	5	10	5	10	15	100
E.1 - Earth sciences	50	5	5	5	5	20	10	100
E.2 - Environmental sciences	50	5	5	5	5	20	10	100
F.1 - Biological, biochemical and pharma	55	5	5	10	5	15	5	100
F.2 - Biotechnologies	50	5	5	5	5	25	5	100
G.1 - Neuroscience	70	2	2	2	2	20	2	100
G.2 - Medical sciences	60	5	5	5	5	15	5	100
H.1 - Agricult. sciences, agrofood and veter.	50	8	5	12	5	12	8	100
I.1 - Civil engineering and Architecture	50	10	10		30			100
L.1 - Industrial engineering*	40	10	25		5	15	5	100
L.2 - Computer engineering	40	10	30		20			100
M.1 - Diagnostics, restor. and conserv. of Cultural Heritage	50	10	4	4	8	16	8	100
M.2 - Assessment, exploitation Cultural Heritage	50	10	10	0	10	10	10	100
N.1 - Antiquity sciences	50	10	10	0	10	10	10	100
N.2 - Philological-literary sciences	68	6	6	0	4	6	10	100
N.3 - Historical-artistic sciences	65	10	5	0	5	10	5	100
O.1 - Historical-geographical sciences	50	10	5	5	10	10	10	100
O.2 - Philosophical sciences	50	10	10	0	10	10	10	100
O.3 - Pedagogical and psychological sciences	65	6	5	0	6	13	5	100
P - Legal sciences	50	15	5	0	10	15	5	100

Q - Economical and statistical sciences	41	11	13	0	10	20	5	100
R - Political and social sciences	50	10	10	0	10	10	10	100

In the period March 2009 and January 2010 the TPs visited the institutes and drafted a report on the basis of the questionnaires, of the meetings with directors, principal investigators, junior researchers and, in some cases, technicians and administrative staff, and of a visit to the facility. This report contains the score for each indicator and a qualitative assessment of the institute's performance and potential, as well as suggestions and recommendations for future developments. Table 2 shows the average scores assigned by each TP.

Each institute, due to its heterogeneous research lines, was visited by two or three TPs. In some particular cases TPs visited only a limited number of institutes.

Amongst the panels belonging to the area of natural sciences and technology, Panel Agricultural and veterinarian sciences appeared to be rather "parsimonious" (the average score was 56.22), and Panel Mathematics as the most "generous" (average of 98.00) (Table 2). Amongst TPs assessing institutes in the social sciences and humanities the more "parsimonious" was TP Political and social sciences (average of 60.20), and the most "generous" was TP Philosophical sciences (average of 97.50). Comparing the scores attributed on average by each panel, the GP noted a general tendency of panels assessing a larger number of institutes to give lower average scores. This could be due to a learning process, as members of panels visiting several institutes had more opportunities to meet each other many times, setting the operative standards, calibrating the indicators.

Table 2. Average scores assigned by Thematic Panels and number of institutes evaluated

Panel	Average score	Number of institutes evaluated
A.1 - Mathematics	98.00	2
A.2 - Computer sciences	85.63	8
B.1 - Physics	63.77	18
C.1 - Chemistry	74.96	24
D.1 - Material sciences and technologies	75.30	25
E.1 - Earth sciences	73.00	11
E.2 - Environmental sciences	64.23	14
F.1 - Biological, biochemical and pharmaceuticals	74.98	26
F.2 - Biotechnologies	63.79	14
G.1 - Neuroscience	83.00	4
G.2 - Medical sciences	66.33	9
H.1 - Agricultural sciences, agrofood and veterinary	56.22	14
I.1 - Civil engineering and Architecture	73.46	7
L.1 - Industrial engineering	75.28	9
L.2 - Computer engineering	73.58	12
M.1 - Diagnostics, restoration and conservation of cultural heritage	79.50	8
M.2 - Assessment, exploitation cultural heritage	82.50	3
N.1 - Antiquity sciences	90.75	4
N.2 - Philological-literary sciences	94.00	7

N.3 - Historical-artistic sciences	77.50	2
O.1 - Historical-geographical sciences	83.00	2
O.2 - Philosophical sciences	97.50	2
O.3 - Pedagogical and psychological sciences	93.00	2
P - Legal sciences	78.25	5
Q - Economical and statistical sciences	65.01	5
R - Political and social sciences	60.20	5

The scores assigned to different sections of the same institute by the various (two or three) TPs in some cases were rather close to each other, while in others the difference was much higher.

In order to assess if these differences were due to systematic different approaches amongst TPs or to real structural differences, the GP calculated the “severity index” for each TP (the ratio between the average of evaluations of the whole 26 TPs and the TP’s average evaluation). The figures are shown in Table 3.

Table 3. Severity index

Panel	Severity index
A.1 - Mathematics	0.73
A.2 - Computer sciences	0.83
B.1 - Physics	1.12
C.1 - Chemistry	0.95
D.1 - Material sciences and technologies	0.95
E.1 - Earth sciences	0.98
E.2 - Environmental sciences	1.11
F.1 - Biological, biochemical and pharmaceuticals	0.95
F.2 - Biotechnologies	1.12
G.1 - Neuroscience	0.86
G.2 - Medical sciences	1.07
H.1 - Agricultural sciences, agrofood and veterinary	1.27
I.1 - Civil engineering and Architecture	0.97
L.1 - Industrial engineering	0.95
L.2 - Computer engineering	0.97
M.1 - Diagnostics, restoration and conservation of Cultural Heritage	1.01
M.2 - Assessment, exploitation Cultural Heritage	0.97
N.1 - Antiquity sciences	0.89
N.2 - Philological-literary sciences	0.86
N.3 - Historical-artistic sciences	1.04
O.1 - Historical-geographical sciences	0.97
O.2 - Philosophical sciences	0.83
O.3 - Pedagogical and psychological sciences	0.87

P - Legal sciences	1.03
Q - Economical and statistical sciences	1.24
R - Political and social sciences	1.34

In cases where the difference between the lower and the higher score assigned by the two or three TPs to each institute was lower than 20%, the final score was calculated as the average of the two.

In cases of larger discrepancies the final score was assigned by GP through an iterative process where the TPs coordinators were consulted, the severity index was taken into account and a thorough analysis was made of the TP's report. An example of the latter case is shown in Table 4.

Table 4. An example of final evaluation from the GP

Name of institute	Panels involved	Score Panel 1	Score Panel 2	Score Panel 3	Final score from GP	Notes
ICIB	B.1, D.1, F.1	51	67	84.5	68	The institute was visited by Panels B1, D1 and F1 which assigned respectively the scores 51, 67 and 84.5, which are quite diverse amongst them. The GP notes that the cybernetic component of the institute was reduced in size, while the biophysics component, which is evaluated very positively, was significantly expanded. This may explain the large differences between the scores. After consultation with the Panel coordinators, and taking into account the "severity" index, the institute assigned the score of 68.

The General Panel report

In the final report the GP concluded that "The performance of the CNR institutes was quite good: the average score of institutes in the areas of natural sciences and engineering was 73 (on a the scale between 0 and 100), while the score of institutes in social sciences and humanities was 82. The difference was deemed to be non significant, being mostly due to differences in measurement standards." (CNR, 2010). This assessment is consistent with a recent analysis of the participation of European research institutions to the VII Framework Programme (EC, 2010).

The GP stressed the fact that the CNR institutes are characterized by a very high heterogeneity from the point of view of differences in fields of science, methodologies, technological development, multidisciplinary, financial resources, applicability potential and socio-economic impact of results.

From a methodological point of view, the GP recommended that in the future the two groups of scientific areas (natural sciences and engineering, social science and humanities) remain separated. Experience showed in fact that differences in objectives, research methodologies, evaluation criteria suggest to be deal with them separately.

The GP also recommended that TPs avoid visit too few institutes: experience showed that TPs which visited two or three institutes had scant general vision of the role of their specific area within CNR and had the tendency to assign very high scores.

The General Panel delivered the following general conclusions. “Several institutes are in a difficult financial situation, which inevitably influences their performance. In some cases the research lines were found too fragmented and not well coordinated among different institutes or departments. The average age of researchers was found to be, in general, too high. Poor promotion of research results has a negative impact both on the exploitation of the outcomes at the social and economic level and on the external image of the institution as a whole. In general the CNR scientific infrastructure is quite good in attracting resources from third parties.”

Fragmentation and too little coordination between different laboratories/units of the same institute is mostly due to the fact that in the recent past the government pushed CNR to merge its 300 laboratories and institutes, considered to be too many, into a smaller number. This merge led to placing under the same roof significantly diverse scientific laboratories which continue to live their own life with little common ground.

The pros and cons of the exercise

All in all, it may be argued that the pros of the evaluation are the following:

- this was the first internal evaluation carried out by CNR,
- the process was a credible one, given the presence of international experts,
- the evaluation process was welcomed by researchers,
- the results were deemed to be potentially used for improving the scientific network and for promoting the carrier of institutes’ researchers,
- the evaluation was a good opportunity to start collaboration between those evaluated and the evaluators.

The cons are the following:

- the emphasis of the exercise was placed on the scientific dimension, while little attention was paid to the management of the institutes,
- the report put too much emphasis on the quantitative dimension producing basically a “league table” of the institutes,
- the time between the period of reference of the information supplied to the TPs (2003-2007) and the site visit was too long: in some cases the situation had changed considerably,
- institutes were not asked to give a thorough long term strategic analysis,
- the institutes’ staff had no chance to comment on the results in an iterative process,
- the periodicity of the exercise was not specified,
- the GP’s final report gave little guidance to the agency’s governing body in terms of proposals for future restructuring of the scientific network.

A comparison with European organisations

A relevant methodological question is how the CNR exercise compares with similar evaluations carried out by other research institutions. Table 5 shows a comparison with the Spanish CSIC, the German Max Plank, and the French CNRS (CSIC, 2009; Max Plank, 2010; CNRS, 2010). The following parameters are analyzed: coverage, self-evaluation of institutes, scope of the evaluation, choice of peers, sources of information, site visits, involvement of the institutes evaluated, emphasis of the evaluation in the final report, expected role of evaluation in the decision making process, follow-up of results.

Table 4. A comparison between evaluations in some European countries

Parameter	CNR	CSIC	Max Plank	CNRS
Periodicity	Occasional	Every 4 years (with annual monitoring)	Continuous. The evaluation is based on the results for the past 2 years	Every 2 years evaluation of researchers' activity
Coverage	Only scientific aspects	Both scientific and managerial aspects	Both scientific and managerial aspects	Both scientific and managerial aspects
Self-evaluation of institutes	Mostly quantitative information collected through questionnaire	Strategic plans prepared by each research line and submitted in a report on past activity and perspectives	Report on the positioning of the institute; quantitative and qualitative information	No self-evaluation of institutes; self-evaluation of individual researchers
Overall scope of the evaluation	Evaluation of institutes, not of individuals	Assessment of fulfillment of objectives of the agency's mission; allocation of budget.	Evaluation of institute, including research groups and programs, appointments of directors.	Policy, programs, monitoring of research units and researchers' activity, advancement of career of researchers
Peers	Nominated by the Board. No check of conflict of interests. Receive remuneration for the service	External experts, mostly foreigners integrated with agency's experts acting as supervisors; receive remuneration	Rotated every 6 years amongst the most distinguished international scientists suggested by institutes. Conflict of interests avoided or declared. Receive no remuneration	Members of the CNRS community; other members nominated by the agency and the Ministry
Sources of information	Questionnaires, institutes' internal data bases and site visits	Reports prepared in the context of strategic planning plus meeting with directors	Status reports and site visits	Internal reports, self-evaluation of researchers validated by managers
Site visit	Usually one day per institute, self organised by panel	No site visit, but meeting with directors	Two-three days, with the participation of agency's senior executives and administrative manager	No site visit
Involvement of evaluated institutes	Only during the site visit. No chance to discuss outcome of the evaluation	No information	Involved in the selection of peers and in the discussion of results of the evaluation	No information
Results of evaluation in final report	Mostly quantitative evaluation, in some cases adjusted by the general panel	Qualitative report identifying strengths and weaknesses of agency, suggestions for management in general	Qualitative evaluation	Quantitative and qualitative evaluation
Expected role of evaluation in the decision making process	Exercise not clearly inserted in the agency' decision process	Changes to the agency's strategic plan	Decisions about financing of programs, creation/closure/reorientation of institutes, dismissal of directors/group leaders,	Closure of research units, advancement of career of researchers, changes in strategy
Follow-up of	No follow-up at	Each 4 years with the	Continuous	Continuous

results	the moment	new action plan		
----------------	------------	-----------------	--	--

For most of the parameters the Italian exercise looks different from the others which, in turn, look pretty similar amongst them. The main features of the CNR evaluation, as compared with the other European agencies, are the following:

- occasional exercise,
- focus on the scientific dimension,
- no involvement of researchers in the selection of peers and no chance for them to discuss the outcome of the evaluation,
- heavy emphasis on the quantitative dimension,
- no strategic recommendations formulated to be adopted by the agency's governing board,
- exercise not formally inserted in the agency's strategy formulation.

The similarities regard:

- the use of a panel of experts and site visits,
- the gathering of basic information collected for the purpose in various ways.

Discussion and policy implications

The GP Report overemphasizes the quantitative dimension of the evaluation: this leads the reader to direct his attention to the ranking of the institutes, while the recommendations to be eventually implemented by the CNR board of directors are quite general and not amenable to be operationalised. As a matter of fact, one of the appendixes to the Report displays the fact-sheets drafted by TPs for each institute setting forth the qualitative evaluation including recommendations for action. This “gold mine” of information could be “exploited” by the CNR Departments which supervise the institutes in order to elaborate specific scientific and organizational recommendations. All in all, the Report does not make any management and policy recommendations for the restructuring of the scientific network, nor suggestions on how to use the available information for this purpose.

At the time of writing this paper – nine months after the release of the Report - the outcome of the evaluation has not yet been used for policy purposes. A first opportunity could have been the allocation of posts of researchers to the various CNR institutes which was decided by the board of directors by the time of the release of the Report. A second opportunity could have been the drafting of the CNR budget for the year 2011 in which the channeling of the resources could have been linked to explicit criteria of scientific quality and strategic priorities.

The lack of impact of the evaluation of the CNR institutes can be linked to various reasons:

- the evaluation was very much centered on the scientific and technological performance and little on the organizational-managerial dimension,
- the evaluation exercise is not formally included in the CNR decision making process – in other words it is not specified to what extent the recommendations of the GP would be used in implementing policies regarding institutes, groups, individual researchers,
- CNR is a public agency embedded in the large Italian bureaucratic system which allows little scope for change in terms of restructuring organizations, moving people around, closing or deeply reshuffling institutes,
- in the present juncture the financial situation of the agency is close to the survival level: given the fact that all institutes were deemed to be viable – even at different levels of performance – it is extremely difficult to subtract resources to somebody in order to promote

somebody else. The use of the outcome of the evaluation would have been more effective in an expansion phase where additional money could be channeled to the more promising avenues,

- over the last two decades the agency has been subjected to continuous reforms and at the moment a new statute is going to be adopted by the government: the general attitude of “wait and see” does not encourage changes in the organization of the institutes.

Overall, it may be argued that, from the point of view of assessing the scientific and technological potential of the CNR network, the exercise may be considered a success even though a “gold mine” of qualitative information and recommendations remain to be “dug”. However, the GP report cannot be used as an instrument for policy making. In terms of science policy the CNR evaluation was therefore a missed opportunity. Even though it was the first attempt to carry out a systematic assessment of the agency’s whole research infrastructure, and it was efficiently carried out, it might have been expected that the investment of human and financial resources (hundreds of people and a direct cost of 1.8 million euro, excluding the opportunity costs) would have yielded more effective results. One has the impression that the evaluation was meant to be addressed mostly to external actors – government, general public, parliament, the press – in order to legitimize the social role of the agency, rather than to be a management tool to be used for internal decision making.

Looking ahead, the exercise could be transformed from a missed opportunity into an investment for the future if further evaluations are carried out with the proviso that they take on board the suggested improvements in methodology, cover the whole range of dimensions (scientific, organizational, managerial), and that evaluation is institutionally inserted in the agency’s decision making process.

References

- Airaghi A., et al. (1999). Options and limits for assessing the socio-economic impact of European RTD Programmes, Report by the Independent Reflection Group of the European Technology Assessment Network (ETAN) to the European Commission DG XII, Evaluation Unit, Brussels. January, 36 pp.
- Baccini A., (2010) *Valutare la ricerca scientifica*, Il Mulino, Bologna.
- CIVR - Comitato di Indirizzo per la Valutazione della Ricerca, (2006). *Risultati dell’esercizio di valutazione triennale*, Ministero dell’Istruzione, dell’Università e della Ricerca, Roma, CIVR, <http://www.vtr2006.cineca.it/>.
- CNR (2003). *CNR Report 2003. Research Results 2002*, Rome.
- CNR (2007). *Obiettivi, modalità e criteri per la valutazione degli Istituti del CNR*”, delibera CdA n. 101/2007, 11 giugno.
- CNR (2010). *Valutazione degli istituti del CNR. Relazione finale del Panel generale*, CNR, Roma, marzo, <http://www.cnr.it/sitocnr/IICNR/Attivita/ValutazioneIstituti.html>.
- CNRS (2010). Comité national de la recherche scientifique www.cnrs.fr/comitenational.
- CSIC (2009). *Action Plan 2010-2013 Executive Summary*, October www.csic.es.
- Fahrenkrog G., Polt W., Rojo J., Tübke A., Zinöcker K., (eds), S., (2002). *RTD Evaluation Toolbox*, IPTS, European Commission and Joanneum Research, Vienna, June.
- EC (2010). *Interim Evaluation Report of the Seventh Framework Programme*, Report of the Expert Group, Brussels, 12 November.
- Georghiou L., (1998). Issues in the evaluation of innovation and technology policy, *Evaluation*, vol.4, n. 1, 37-52.
- Georghiou L and Larédo P. (2006). "Evaluation of Publicly Funded Research: Recent Trends and Perspectives." In *OECD Science, Technology and Industry Outlook*, ed. OECD, 177-199. Paris: OECD.

- Lyall, C., Bruce A., Firm J., Firm M., Tait J. (2004). Assessing end-use relevance of public sector research organizations, *Research Policy*, 33, 73–87.
- Max Plank Society (2010). *Evaluation: the procedures of the Max Planck Society*, June.
- RAE (2008). Research Assessment Exercise 2008, <http://www.rae.ac.uk/>.
- Silvani A. & Sirilli G. (2001). A review of the Italian S&T policy and system of innovation, in P. Larédo and P. Mustar (eds), *Research and Innovation Policies in the New Global Economy*, Elgar, Abingdon, UK.
- Silvani A., Sirilli G., Tuzi F. (2005). R&D evaluation in Italy: more needs to be done, *Research Evaluation*, vol. 14, n. 3, December, pp. 207-215.