

Towards a comprehensive evaluation: a common framework to evaluate research groups with different missions

EREFIN Group¹

Summary

This paper presents the work of a group named Erefin associating evaluation officers from a number of French research institutions involved in mission-oriented research. The work relies on practical experience of evaluation in these different institutions. It has the operational objectives to provide common tools to help the institution, the divisions and the research groups to describe their scientific profiles and strategies, to achieve self-evaluation and to prepare evaluation reports for the assessment by external visiting committees. The provided tools are both general and versatile so that they are applicable to a large range of disciplines and research practices and respect each specific research type.

Key words

Research evaluation, mission-oriented research, research profile, quantitative descriptors, evaluative questions

1 The need for a comprehensive evaluation

The essential contribution expected from science and technology to the development of a knowledge-based society is enhancing the attention to the assessment of scientific research. This issue of extending the measure of quality of research to include the assessment of its economic, societal and cultural impacts is clearly understood and addressed. For instance, the study directed by the Academy of Finland (Kaninen & Lemola, 2006) has surveyed the different approaches to measure the impacts of basic research. New projects as SIAMPI have been launched to design a methodology and relevant sets of indicators to measure the impact of research on society.

In mission-oriented research institutions, the research is driven by social needs. They achieve targeted research that aims at solving problems issued from the actors practice that deserves investment in public research. This implies a strong interaction with stakeholders from the institute management as well as from the research groups. The policy design and research processes are much more socially interactive and the relationship with society goes beyond a preliminary consultation of stakeholders before defining research programmes or their cooperation to transfer scientific results to end users. Stakeholders and end users may be involved in the different steps of the research groups activity, inferring on their scientific agenda, intervening in the follow-up or management of research programmes or contributing to knowledge production. As described by Michael Gibbons and co-authors (Gibbons et al. 1994), such interaction between science and society leads to a different practice of research.

Mission-oriented research institutions combine fundamental and applied research. Fundamental research is necessary to fill the gaps not covered by disciplinary knowledge that are revealed while translating societal issues into scientific questions. These institutions aim at developing efficient interaction between this two approaches. They are therefore strongly concerned by the issue of relevant evaluation methods which take into account the whole range of their activities and evaluation criteria that are clearly related to the final objectives of the research.

The need for a comprehensive evaluation is not only relevant for mission-oriented or specialized research groups and institutions. As widely explained in political and strategic statements, public

1 Names and Institutions of the contributors are given at the end of the paper. *Corresponding author:* Élisabeth de Turckheim, Institut National de la Recherche Agronomique, France
elisabeth.deturckheim@paris.inra.fr

research is not only expected to increase scientific knowledge but other objectives are assigned as the contribution to innovation, to scientific expertise and decision support to policy makers or the contribution to societal debates and scientific literacy of the citizens. Whatever the weights these activities have in the agenda of the research groups, there is a need to extend the tools commonly used to assess the quality of academic research with other information, indicators and impact assessment methods related to their specific missions and social environments. The ability of the evaluation tools to cope with these aspects and their versatility to fit any type of research is a condition for the legitimacy and acceptability of a general evaluation procedure.

2 Objectives of the Erefin group

In France, a national agency of evaluation (*Agence d'évaluation de la recherche et de l'enseignement supérieur Aeres*) has been created in March 2007 and is now in charge of the evaluation of the whole national public research system. Both curiosity-driven and mission-oriented institutions and units are to be evaluated by the agency which had to define the evaluation processes and to develop evaluation tools to achieve its mission.

Before the creation of the agency, mission-oriented institutions had their own evaluation systems, combining scientific assessment by external committees and strategic decision processes involving stakeholders. None of these systems could be applied at the national level. Therefore, there was a need to improve the basic common framework for scientific assessment in order to properly evaluate research groups and institutions working in different contexts.

Hence, several mission oriented research institutions launched a working group – the Erefin² group – aimed at discussing and comparing the coherence of evaluation methods and criteria with each institution strategic objectives. As a great similarity was observed between the requirements of the members of the group with respect to relevant evaluation tools, the group decided to design a common set of such tools to support the evaluation of the “research units”.

The Erefin group, did not develop any theory but relied on literature and on the previous studies or operational tools used by the participating institutions. Concerning the literature, the ERiC project was found a very useful support to confirm the relevance of Erefin's approach (Spaapen *et al.* 2007). This project is based on smart research in science policy and provides a very elaborated method that has been used in different fields. Erefin did not achieve such a professional work but its objectives and its method are very similar.

The “tool kit” developed by the Erefin group has three pieces:

- A. **a typology of activities** displayed in the *activity table* which describes the whole set of possible elementary activities. It is a support to define the profile of a given research group as a combination of chosen elementary activities;
- B. a check list of possible outputs and short-term results for each elementary activity with a set of **quantitative descriptors** to summarize the whole set of outputs in a synthetic histogram showing the level of diversity of the production;
- C. a set of **evaluative questions** to analyse mid and long term results and foster evaluation on qualitative aspects that take into account the research group objectives and its social environment.

3 The Erefin tool kit

A typology of elementary activities

The set of elementary activities is organized as a two-entry table and is designed to cover the whole range of fundamental, applied, collaborative research and any combination of these different approaches of research production. The first entry distinguishes the different social groups with

² Erefin stands for *Évaluation de la Recherche Finalisée*. A dozen of research institutions have joined the group in the period 2007 - 2009. The group was initiated by research institutes in environment, agriculture, land use and natural resources management. Other mission-oriented institutes soon joined the group as Inserm (biological, medical and public health research) and CEA (defence and security, energy, health and information technology) or graduate institutes of science and technology as *AgroParisTech* and *École des Mines de Paris*.

which the unit cooperates. This point of view is usual in science studies as illustrated by the compass card of research (Callon et al. 1992, Larédo & Mustar, 2000). It is also the starting point of the methodology of the sci_Quest method (EriC project). The second entry aims at identifying knowledge production from relational activities between the unit and its social environment and from communication and transfer of knowledge and resources to different users. This second entry of the table does not suggest to classify each research project in one single cell but to identify how the different phases of a research programme combine contextual problem solving and production of generic knowledge, cooperation with scientists and interaction with stakeholders. This is particularly necessary for collaborative projects operated in a context of application.

A particular research group is not expected to be seriously involved in each type of activity, whereas a larger division or a whole institute may have to achieve an overall balance to meet its strategic objectives. Organizing the missions and profiles of the different units into a coherent system is therefore necessary and the negotiation of their strategic objectives with the units could be formulated with the elementary activities of the table.

Table 1: The “activity table”

<i>Social actors concerned (collaborators and users)</i>	1. Knowledge production	2. Partnership and network management	3. Transfer and communication
A. Academic actors	Producing scientific certified knowledge	Organizing scientific communities, surveying scientific trends	Developing infrastructures and resources available for scientific communities
B. Socio-economic actors	Producing knowledge in a context of application (transferable, embodied knowledge)	Designing and managing economic partnership	Consulting and contractual studies for firms
C. Public policy makers	Producing knowledge for decision support of governmental bodies	Designing and managing cooperation with public actors (outside research)	Expertise work for governmental bodies , commitment in safety agencies or other public bodies
D. Students	Training young scientists; designing education programmes and methods	Managing education programmes and networks	Teaching
E. Citizens	³	Surveying industrial, technological and societal trends and needs	Communicating scientific and technical knowledge, contributing to scientific literacy and societal debates; consulting for societal organizations
F. The unit	Designing the scientific strategy of the unit	Management of skills and resources of the unit; internal cooperation	

Different graphical representations are possible to show the activity profile. Radar graphs are widely used but they rely on quantitative data as time score attributed to the different activities. A proposed qualitative representation is a map based on an 18-cell grid, where the intensity of the colour of each cell of table 1 is related to the strategic value of the activity, like heights on a geographic map.

³ The activity of knowledge production intended to citizens is difficult to isolate: in a sense any scientific knowledge is aimed at the whole society. But new scientific results are necessarily communicated and debated first with a scientific community or a public stakeholder that represents the citizens. Therefore this cell is generally empty.

Figure 1: An example of *strategic map* of a research group

	Konowledge production	Partnership and network management	Transfer and communication
A. Academic actors			
B. Socio-economic actors			
C. Public policy makers			
D. Students			
E. Citizens			
F. The unit			

Quantitative descriptors and check list for outputs

The description of outputs or short term results has to cover any type of product in order to show the consistency of a proclaimed strategy with actual results.

Outputs related to knowledge production are fairly accountable but interaction activities can also be followed through some kind of outputs. For instance, indication about membership in stakeholders advisory boards, involvement in collaborative research projects, or financial support for projects from stakeholders (including PhD scholarships) are relevant facts to account for this activity. The hypothesis is that these indications are related with other informal tasks contributing to the management of the cooperation with stakeholders which are not easily accountable.

Erefin established a check list of outputs associated to each cell of the table of activity and was careful that groups in each field could find all relevant items in the list, be they carrying out medical, agricultural or environmental research and engineering or highly specific technological research. This led to write a long list of possible outputs. In a second step, a shorter list of about 50 main outputs were selected for their traceability and representativeness.

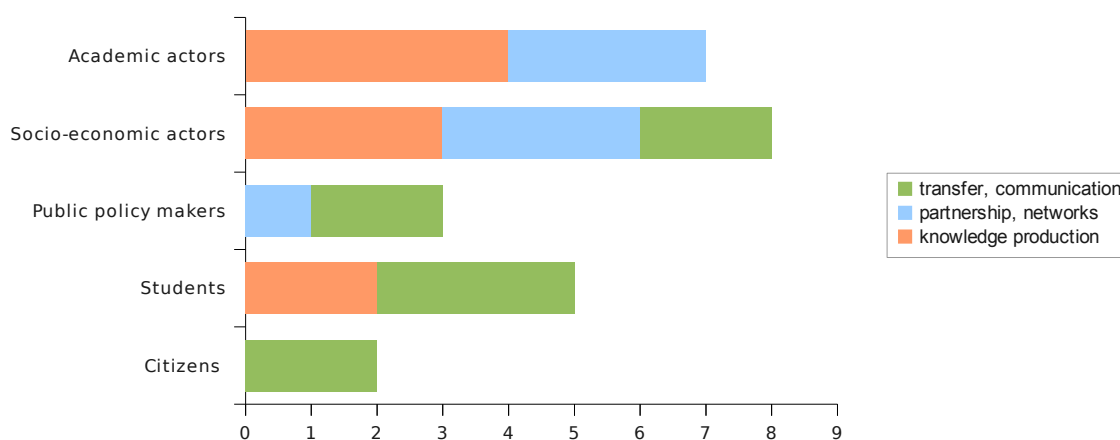
In a third step, a quantitative approach was attempted because the groups need to summarize their production with a small set figures and evaluation committees are expecting them. Erefin proposed a set of *quantitative descriptors* based on weighted sums of the counts of the main outputs of the second list. Weights could be chosen in relation to a specific strategy and discussions rose in the group showing differences in the institution strategies and a compromise was to be found. They could not be based on statistics because, except for a few outputs, no statistical data is available at the national level⁴. The proposed weights (see appendix 1) have been finally based on a rough estimate of amount of necessary working time of scientists to produce the output.

Using these descriptors a synthetic histogram can be drawn showing the performance profile of the unit. Such a multidimensional image could incite the group to a reflexive work and prevent the evaluation committees to forget some facet of the group activity.

The words *quantitative descriptors* have been preferred to *indicators* because designing indicators deserves more care, especially considering the issue of normalization, in particular for benchmarking uses. The group will consider its future involvement in common work with indicator producers to transform some of these descriptors into proper indicators which can be aggregated and analysed at a larger scale.

4 On the contrary, the REPP (Research Embedment and Performance Profile) of the sci_Quest method used in the ERiC project is based on the distribution of the number of each type of product in the considered institution or set of institutions

Figure 2: An example of a group production histogram



Evaluative questions

The set of quantitative descriptors is only capturing outputs that are observable in the short term. But the evaluation cannot be restricted to the count of outputs. There is a need for qualitative approach where outputs are related to the objectives of the group and to the missions of the research institution or division. These objectives are mid and long term objectives and the short term outputs are only steps towards their achievement. For instance, if an objective of the research group is to support decision of some actors, the actual use of the decision support tools is to be questioned. Also the impact of the use of new methods or devices produced by the research work on the users' activity or organization is to be considered.

The mid and long term objectives of the research are also dependant on the context. The research processes involve other groups and actors and the evaluation has to make clear how the evaluated group contributes to the whole process. Among the different factors, those depending on the research group activities such as the quality of the cooperation management, the relevance of communication supports are criteria that evaluative questions should address.

Therefore, different point of views about the quality of a research work are to be debated. To invite to such debate and help to explore the different dimension of quality, Erefin decided to formulate *evaluative questions* for each activity. These questions are for different actors: the group preparing its evaluation report, the institution managers defining missions to the groups or writing the terms of reference for evaluation committees and evaluation committees.

Finally, a special mention is to be made for the activities related to the group scientific strategy and governance (line F). No quantitative descriptors have been defined to summarize the outputs of these activities. As the descriptors and evaluative questions for the other cells do not grasp the coherence of the different activities into a comprehensive research strategy, this important issue is approached with evaluative questions associated to the last line of the activity table.

4 Erefin interactive working process

The table of activities and the descriptors have been designed through an interactive process that associated an increasing set of institutions, so that the experience in agriculture and environment of the first participants was fruitfully enriched by other domains (medical, engineering, public safety...) leading to improved definitions, new representations and common understanding of the framework. These tools are therefore the result of a learning process associating different actors.

At the different steps of the design, chosen research groups have tested the tools and managers have been interviewed. An interactive phase was performed through the experiment that was implemented with the support of the Inra Deputy Director General in 2007. The aim was to explore the feasibility of a contract between a research unit and the division management where long term goals as well as explicit outputs and accountable results to reach within four years were stated. This

procedure aimed at helping the units to define their objectives and to negotiate them with the division management. The experiment was performed through a close interaction between a small number of unit directors and the two members of the Erefin group and several discussion meetings inside each unit. This allowed a better understanding of the preliminary version of the tools and led to further improvement.

The unit directors involved considered the Erefin tools were very useful for internal communication and discussion and, for those units which have been evaluated soon after, it has been an efficient tool for the preparation of the evaluation report. They also pointed that the right level of description of the expected outputs was difficult to find and they pointed that the next evaluation could not be a simple comparison between the expected outputs described in the “contract” and the achieved ones. This is because research units have to be able to adapt their goals to unexpected changes in their environment. Therefore the ex post evaluation of the contract has to take account of the arguments explaining the changes in the targeted outputs. At the end, the frame to describe objectives and outputs was considered as more attractive than the preparation of a precise contract. Another outcome of the experiment was the consciousness of the division management that such a process, in a simplified version, could provide a powerful tool to design the division scientific policy. The division managers found Erefin tools helpful to write the engagement letters to the unit directors at the beginning of their commitment.

Another important interaction for the group was with the staff of the *Observatoire des Sciences et des Techniques*. OST experience reinforced the understanding that developing operational tools is a relevant approach to sustain policy objectives. This helped the group to understand that the ability to describe and measure the variety of missions and activities is essential to preserve the diversity of the research system in the long term. This collaboration also comforted the group in its choice to develop the framework before considering the design of proper indicators.

5 Conclusion

The work presented in this paper has been launched to help in promoting a comprehensive evaluation of research. Taking advantage of previous work on this important issue, the Erefin group was able to build a common framework to describe the scientific strategy of mission-oriented research groups and to report on their results.

The implementation of the Erefin tools has not achieved the refinement of the empirical work of the two studies carried out by sci_Quest research team in the ERiC project, but the aim was that the tools are understood and used at different levels by the different actors. The framework has been approved and endorsed by the institute managers and is gradually appropriated by research groups involved in strategic research as they are evaluated by the agency in the four year rhythm of the French assessment and contractual protocol.

This was possible because the priority of the project management was focused on the cooperation between the institutions and the tools were designed for a flexible use by the research groups rather than in a top down process. In fact, the main expected impact of Erefin's work is an improvement of the self evaluation step which has to be achieved by the lab units when they prepare their evaluation by the agency.

Erefin's framework is too sophisticated to be performed by the French evaluation agency since the agency has to evaluate about 800 various lab units each year. However, the agency agrees with the need for adapting the general evaluation tools to specific situations and the agency managers are supporting the declination of the general guidelines to one specific field, engineering, promoting this adaptation as a “cas d'école” (Dhainaut et al. 2009).

Further steps will consider a change of scale and a focus on some specific activity because there is need to improve or to design indicators for activities which are not commonly described. Presently on the table is the issue of collecting precise information on the expertise activities achieved by the labs and institutions to support public action and decision. Related to this need of the institutions, there is a need to design indicators at the national scale to help the follow-up of the work of public research organizations in support public policies.

Lastly, our work illustrates the new production mode recommended by Lepori et al (2008) to design and produce innovative indicators on a robust basis. In this respect, Erefin plays a role in representing the users in a networks where indicators designers and data producers cooperate to meet the demand of new indicators and their consolidated production.

References

Callon M., Larédo P., Gonard T., Leray T., and Rabeharisoa V. (1992) *Defining the Strategic Profile of Research Labs: the Research Compass Card Method* in Raan A.F.J. v.,(ed) *Science and Technology in a Policy Context*, Leiden: DSWO Press

Dhainaut J.-F., Cormier M., Glorieux P., Menand A., Allano S., Carraro L., Chabbal R., Clerc J.-F., Dubuisson B., Lermينياux C., Sagaut P., Schmitt M. (2009) *Évaluation de la recherche en ingénierie : un cas d'école* www.aeres-evaluation.fr/Evaluation-de-la-recherche-en

Gibbons M., Nowotny H., Limoges C., Schwartzman S., Scott P., Trow M., (1994). *The New Production of Knowledge: The Dynamics of Science and Research in Contemporary Societies*. Sage, London

Kanninen S., Lemola T. (2006) *Methods for Evaluating the Impact of Basic Research Funding : An analysis of Recent International Evaluation Activity*, Academy of Finland www.aka.fi/publications

Larédo P., Mustar P. (2000) *Laboratory Activities Profile : An Exploratory Approach*, *Scientometrics*, 47-3, 515-539

Lepori B., Barré R., Filiatreau G. (2008) *New perspectives and challenges for the design and production of S&T indicators*, *Research Evaluation* 17(1), 33-44
Spaapen J., Dijkstra H. and Wamelink F. (2007)

SIAMPI (*Social Assessment Methods for research and funding instruments through the study of Productive Interaction between science and society*) is a European consortium funded by European commission's Seventh Programme (FP7) Science in Society (SiS)
www.nwo.nl/nwohome.nsf/pages/NWOA_7RZC9Q

Spaapen J., Dijkstra H. and Wamelink F. (2007) *Evaluating Research in Context*, Consultative Committee of Sector Councils for Research and Development (COS) The Netherlands
www.eric-project.nl/nwohome.nsf/pages/NWOA_6TZJ28_Eng

Appendix 1: the Erefin quantitative descriptors

In the following table, we show the descriptors for the cells B1, B2 and B3 of the activity table.

Table 2: Quantitative descriptors for activities with socio-economic actors (B1- B2 - B3)

	Output	Count	Weight (indicative)	Weighted count
Producing knowledge in a context of application	Papers in technical or professional journals with a large audience in the targeted community		1	
	Technical/methodological guides, survey books for professionals		4	
	Decision support or negotiation tools and models delivered to users; registered software available to users, patents		2	
	Trademarks, plant breeders rights		3	
	Clinical trials, epidemiology studies		2	
	Prototypes, standardized procedures, pilots, demonstrators delivered to the user		3	
	Licences related to patents filed by the group		3	
	Firms created by members of the group		4	
Descriptor				b1
Designing and managing economic partnership	Research contracts or common research projects with stakeholders with either financial support from stakeholders greater than 50 000€ or involving 0.5 year or more FTEs of the research group personnel		1	
	PhD projects elaborated with stakeholders and supported by them (ex. co-financing of PhD grants)		2	
	Membership of scientists in stakeholder strategic councils		0.25	
Descriptor				b2
Consulting and contractual studies for firms	Expert missions and studies achieved for partners or clients		1	
	Recommendations for protocols (ex. clinical protocols)		1	
	Personnel FTEs (years) devoted to the development of a large instrument used by socio-economic actors		2	
Descriptor				b3

Appendix 2: Evaluative questions

We propose here evaluative questions for the activities of the lines C and D of the activity table.

Not all questions are relevant for a given research group and some others are certainly missing. These questions are not evaluation standards but an incitement to raise questions about the quality of the research that are relevant for the assessed group.

Quantitative answers are generally not appropriated and the straightforward conclusion that “more is better” is often wrong. These evaluative questions neither deserve a *yes or no* answer but the right balance between two ideal options (or extreme types) is depending on the context.

Examples of facts which may indicate that the objectives are reached are shown with arrows. More indications should be found by the group to provide evidence of the impact of its activity.

Activities for public decision support

C1-C3 Knowledge production and expertise work for public policy support

- Quality of the products and relevance of their diffusion
 - innovativeness of the proposed solutions, originality of the methods
 - efficiency of the proposed solutions, relevance of the transfer process, of the communication support (ex. choice between ad hoc report or generic paper)
- Level of involvement in knowledge production for governmental bodies and its transfer
 - intervention mode: either in reaction to a demand or in a proactive mode or proactive (alert duty...)
 - traceability of the command, of the transmission of the answer
- Use of the research results in the design or the implementation of a public policy
 - explicit references, other proofs of actual use (ex. teaching the personnel of public bodies how to use the provided tools)
 - contribution to norms and rules, to monitoring plans
 - use of the provided tools in negotiation processes

C2 Partnership with public stakeholders

- Existence of institutionalized relationship
 - duration of common projects
 - involvement of the public stakeholder in the production of the output
- Positioning of the stakeholder
 - scope and range of action of the stakeholder (geographic, thematic ...)
- Feedback for the research group
 - interactions and cooperation (with other research groups, inside the group itself...) as a consequence of the commitment in answering a request
 - emergence of new research questions derived from the stakeholder request

Activities for education and training

D1-D3 : Design and implementation of training programmes

- Impact on the teaching contents related to scientific progress in the field of the group
 - existence of organized opportunities to discuss about the scientific progress to insert in training programmes
 - existence of one or more persons responsible for the interaction between science and education

→ change or creation of new training modules as a result of such a strategy

- Involvement in changes of training programmes related to societal needs
 - survey of the new skills needed by socio-economic actors
- Feedback on the research group: recruitment of PhD students, partnerships...
- Quality and relevance of research training
 - PhD quality: published papers, other published and transferred products,
 - professionals skills of trained students
 - employment of graduate students and doctors
 - reflexive thinking on the training strategy (choices of contents, of PhD subjects..)

D2 : Partnership in training networks

- Involvement in international training networks
 - co-tutoring students, mobility of students in these networks
 - contribution to the design on international training courses, Summer schools, tutorials in conferences
 - participation to *Erasmus mundus* masters et PhDs

Appendix 3: the Erefin Group in December 2009

Agrocampus Ouest, *Pierre-Guy Marnet*

AgroParisTech, *Marianne Lefort, Gilles Trystram*

Bureau de recherches géologiques et minières (BRGM), *Françoise Aye*

Centre de coopération international en recherche agronomique (Cirad), *Anne-Yvonne Le Dain*

Commissariat à l'énergie atomique (CEA), *Annick Nguyen*

Institut de recherche en sciences et technologies de l'environnement (Cemagref), *Luc-André Leclerc*

Institut français du pétrole (IFP), *Xavier Montagne*

Institut français de recherche pour l'exploitation de la mer (Ifremer), *Jean-Luc Devenon*

Institut géographique national (IGN), *Jean-Philippe Lagrange*

Institut national de la recherche agronomique (Inra), *Édith Legouy, Élisabeth de Turckheim*

Institut national de la santé et de recherche médicale (Inserm), *Isabelle Henry*

Institut national de recherche sur les transports et leur sécurité (INRETS), *Jean-Baptiste Lesort, Dominique Mignot*

Institut de radioprotection et de sureté nucléaire (IRSN), *François Bréchnignac*

Institut de recherche pour le développement (IRD), *Benoit Lootvoet*

Laboratoire central des ponts et chaussées (LCPC), *François de Larrard, Rémi Pochat*

Mines ParisTech (ENSMP), *Emmanuel Roussel, Michel Schmitt*

Ministère de l'écologie, de l'énergie, du développement durable et de la mer- Commissariat général au développement durable - Direction de la recherche et de l'innovation- (MEEDDM/CGDD/DRI), *Brigitte Hébrard*

Observatoire des sciences et des techniques (OST), *Ghislaine Filliatreau*

SupAgro Montpellier, *Jacques Maillet*