

Regards

Élisabeth de Turckheim : évaluer la recherche finalisée

Propos recueillis par Bernard Hubert et Daniel Terrasson

Élisabeth de Turckheim¹, Bernard Hubert², Daniel Terrasson³

¹ Mathématicienne, INRA, Délégation à l'évaluation, 75338 Paris cedex 07, France

² Écologue, INRA, Écodéveloppement, 84914 Avignon cedex 9, France

³ Agronome, CEMAGREF, Direction scientifique, 33612 Cestas cedex, France



Élisabeth de Turckheim, directrice de recherche de classe exceptionnelle de l'Inra, est née en 1946 à Neuilly-sur-Seine. Agrégée de mathématique, elle a d'abord enseigné les statistiques aux étudiants de l'Université Paris-Sud et contribué à la formation en statistique de chercheurs en biologie. Après une thèse d'État en statistique mathématique, elle a poursuivi sa carrière à l'Inra pour y développer des travaux de recherche en lien avec les besoins de modélisation de biologistes. Directrice d'une unité de recherche (1989-1994), puis responsable du département de recherche en mathématique et informatique de l'Inra (1995-2002), elle a confirmé son intérêt pour les travaux de recherche interdisciplinaires au moment de l'émergence de la bio-informatique, puis en animant avec Bernard Hubert un programme de recherche transversal Inra-Cirad sur les démarches, les méthodes et les outils pour la conception et la construction de la décision en agriculture et dans l'espace rural. Elle a ensuite dirigé la délégation de l'évaluation de l'Inra (2003-2011). Elle est actuellement chargée de mission à l'Inra et à l'Observatoire des sciences et techniques (OST).

NSS : D'une façon générale, la revue *Natures Sciences Sociétés* s'intéresse à des formes de production des savoirs qui s'écartent de la production académique traditionnelle, que ce soit parce qu'elles impliquent le croisement entre des disciplines, ou que ce soit parce qu'elles tentent de répondre à des questions de société. Mais ces formes de production posent toujours des problèmes d'évaluation à différents niveaux. C'est le cas lors de l'évaluation des textes qui sont soumis à notre revue. C'est le cas lors de l'évaluation des chercheurs et des équipes qui sont impliqués dans ces formes de production des savoirs.

Élisabeth de Turckheim, vous avez été chargée de l'évaluation dans un grand organisme de recherche, l'Inra ; vous avez participé aux instances scientifiques

d'autres établissements, notamment du Cemagref¹, et vous avez accompagné la création de l'Agence nationale chargée de l'évaluation de la recherche, l'AERES, comme membre de son Conseil. Nous souhaiterions donc que vous nous fassiez part de votre expérience sur la façon dont peut se concevoir l'évaluation d'une recherche qui ne serait pas purement académique, mais aussi sur les conséquences de l'irruption de l'AERES dans le paysage national. Avant d'entrer dans le vif du sujet, pouvez-vous nous rappeler quelles ont été vos dernières fonctions à l'Inra ?

Élisabeth de Turckheim : J'ai été chargée de l'évaluation à l'Inra à partir de 2003, avec le titre de « déléguée à l'évaluation ». C'est une compétence que j'ai construite

Auteur correspondant : B. Hubert,
bernard.hubert@avignon.inra.fr

¹ Le Cemagref est devenu Irstea (Institut national de recherche en sciences et technologies pour l'environnement et l'agriculture) en novembre 2011.

progressivement et qui répondait d'abord au besoin interne de faire fonctionner l'évaluation des chercheurs à l'Inra. Il s'agissait ensuite de repenser nos méthodes d'évaluation pour mieux les adapter à ce qu'on en attend, c'est-à-dire contribuer à l'amélioration de la qualité de la recherche et de sa cohérence avec ce qu'on attend de l'Inra et des organismes finalisés.

NSS : Y a-t-il eu un élément déclenchant qui vous a fait choisir cette voie ?

Élisabeth de Turckheim : Devoir regarder de près l'ensemble de la recherche menée à l'Inra était une mission séduisante. Jusqu'en 2003, j'étais chef du département de mathématiques et d'informatique appliquées et cette expérience a été tout à fait exceptionnelle pour comprendre le fonctionnement de recherches dans d'autres champs que ceux de ma culture, les mathématiques et la bio-informatique. À la fin de mes deux mandats de chef de département, plutôt que reprendre une recherche personnelle spécialisée, j'ai choisi de continuer dans cette voie. D'autre part, en tant que chef de département, j'ai été très impliquée dans l'évaluation, même si je n'avais pas particulièrement de culture sur ce sujet, sauf, comme tout le monde, celle d'avoir monté des dossiers, et celle d'avoir, de temps en temps, été membre d'une commission d'évaluation.

NSS : On peut penser qu'il était intéressant que ce soit un ancien chef de département qui occupe la fonction de délégué à l'évaluation.

Élisabeth de Turckheim : Rappelons que Bernard Hubert et moi-même avons été les deux premiers candidats à une évaluation de nos départements à la fin des années 1990. Nous étions intéressés par cette façon d'analyser nos travaux. Le fait d'être les premiers dans l'exercice nous a conduits à nous poser beaucoup de questions : Que fait-on comme type de recherches ? Quelles en sont les finalités ? Cognitives ? Sociétales ? Comment s'y prend-on pour élaborer les programmes et les mener à bien ? Comment cela est-il évaluable par une commission externe ? Comment présenter dans un document les structures, les objectifs particuliers et communs, les travaux et les résultats, le rôle du département par rapport aux unités de recherche ? Comment présenter les points de vue des partenaires des chercheurs du département dans le dossier ou lors de la visite de la commission ? Pour quoi faire ? À l'Inra, cette réflexion n'existait pas et elle a été très stimulante à mener. Ce n'est donc pas par hasard si j'ai accepté ce poste. Celui-ci s'appelait au départ « secrétariat général à l'évaluation », puis « délégation à l'évaluation », ce qui signifiait que l'Inra voulait lui donner une position plus réflexive, plus stratégique. Cette délégation a été rattachée directement à la présidence et au directeur général délégué aux programmes et à l'évaluation.

NSS : Cette structure est donc directement rattachée à la direction générale, mais elle n'est pas un des services de la DRH comme cela aurait pu l'être.

Élisabeth de Turckheim : Non, la délégation ne dépend pas d'une direction d'appui. En gros, on attendait de moi et de l'équipe que j'ai constituée petit à petit, en premier lieu, de remplir les obligations d'évaluation des chercheurs des instituts de recherche, mais au-delà, de renouveler la réflexion sur l'évaluation, de bien la réintégrer dans le cœur des outils de pilotage.

NSS : Selon les organismes, les pratiques d'évaluation étaient fort différentes à cette période. Pouvez-vous nous dire qui et quelles structures étaient évalués à l'Inra ? Vous nous avez parlé de l'évaluation des départements, était-ce le point de départ ?

Élisabeth de Turckheim : Non, à l'Inra, à l'origine, la mission opérationnelle était d'évaluer les chercheurs en application du décret de 1984² sur l'évaluation individuelle dans le cadre des commissions scientifiques spécialisées (CSS). Au début, la mise en place de cette évaluation a apporté aux chercheurs une vision de leur activité et de leurs résultats indépendante de leur hiérarchie, puis, petit à petit, cette vision externe s'est située pour tenir compte du contexte de leurs unités, qui relevaient, elles, d'une évaluation commanditée par les chefs de département, c'est-à-dire conçue dans un cadre hiérarchique, considérant l'unité de recherche comme une des composantes du département ainsi que de l'Inra. Il s'est ainsi établi une relation triangulaire entre l'évalué, ses responsables scientifiques et la CSS, de façon à ce que les résultats de l'évaluation soient utilisés par les directeurs d'unité et les chefs de départements pour acquérir une vision stratégique des individus et des collectifs au sein du dispositif dont ils étaient responsables, le département de recherche. Mais, en sens inverse, nous devions transmettre aux CSS de l'information pertinente sur le fonctionnement des structures, en particulier des unités dans lesquelles travaillaient les chercheurs dont elles évaluaient les dossiers. Au lieu de mettre en place un processus dichotomique, avec l'avis de la hiérarchie sur les unités, d'un côté, et l'avis d'une commission externe sur les individus, de l'autre, nous avons essayé de faire coopérer l'ensemble.

NSS : En effet, il y avait à l'Inra une évaluation des unités mais qui était réalisée sous la responsabilité des départements. C'était donc un outil de la hiérarchie

² Le décret 83-1260 fixant les dispositions statutaires communes des fonctionnaires des EPST instituait une évaluation tous les deux ans des chercheurs par des instances d'évaluation. Des décrets spécifiques précisaient en particulier la composition de ces commissions scientifiques spécialisées (décret 84-1207 pour l'Inra).

scientifique, organisé par les chefs de départements qui traitaient directement avec les directeurs scientifiques. Mais il n'y avait aucun lien avec le secrétariat général à l'évaluation qui se chargeait de l'évaluation individuelle des chercheurs.

Élisabeth de Turckheim : Le service essayait d'acquiescer a posteriori les informations sur l'évaluation des unités. Mais cette évaluation, conduite par la direction de l'Inra et les chefs de département, n'était pas articulée avec celles des chercheurs. Par ailleurs, un test avait été mené par mes prédécesseurs sur la faisabilité d'une évaluation des ingénieurs par des commissions externes et je devais en tirer les conclusions et mettre en place un processus stabilisé.

NSS : L'évaluation des ingénieurs concernait-elle seulement les ingénieurs de recherche ou également les ingénieurs d'étude ?

Élisabeth de Turckheim : Les deux, et c'est la raison pour laquelle on ne pouvait simplement reproduire la méthode d'évaluation des chercheurs car il fallait une évaluation adaptée aux métiers des ingénieurs et à leurs missions. Il fallait aussi repenser l'articulation entre les évaluations individuelles et les évaluations des unités. Ces dernières n'étaient utilisées que pour le pilotage des structures. En tant que chefs de département, ces évaluations nous permettaient d'analyser la situation des unités pour faire ensuite évoluer le dispositif en fonction des objectifs stratégiques. Lorsque des évolutions profondes étaient demandées aux unités, les commissions d'évaluation des chercheurs et des ingénieurs devaient porter l'attention sur la façon dont ils pouvaient s'adapter et profiter des nouvelles dynamiques collectives. Petit à petit, on a amélioré cette articulation, en commençant par une synchronisation des calendriers. L'évaluation des individus était réalisée tous les 2 ans, puis tous les 4 ans, par ordre alphabétique afin de répartir la population convenablement sur les commissions. L'évaluation des laboratoires se faisait tous les 4 ans et était organisée en fonction de la charge de travail des chefs de département. Pour réaliser cette synchronisation sans alourdir le poids des évaluations, nous l'avons programmée sur 4 ans, tout en nous adaptant au calendrier des vagues universitaires de l'AERES qui a repris en 2007 l'évaluation des unités.

NSS : Pour résumer, votre mission comprenait trois niveaux : évaluation des individus (ingénieurs et chercheurs), évaluation des unités, évaluation des départements.

Élisabeth de Turckheim : Je n'avais pas de responsabilité sur l'évaluation des départements, celle-ci étant pilotée par les directeurs scientifiques et la direction générale. Il y a donc 2 calendriers distincts : celui des départements et celui, quadriennal, des unités et des individus.

NSS : Dans d'autres établissements, l'évaluation a pu connaître des évolutions un peu différentes. Au Cemagref, par exemple, il y avait au départ une évaluation des programmes. Ensuite, il y a eu assez vite une évaluation des agents (ingénieurs et chercheurs) coordonnée avec l'évaluation des programmes, impliquant les mêmes experts mais avec un décalage qui était théoriquement de 6 mois. Jusqu'à la mise en place de l'AERES, au Cemagref, l'évaluation concernait donc les programmes et les individus mais pas les unités de recherche.

Élisabeth de Turckheim : À l'Inra, la démarche de l'intégration des évaluations dans un système d'évaluation a commencé à l'échelle des individus. Après l'évaluation des chercheurs, les missions du service ont inclus celle des ingénieurs. Une vraie réflexion a été menée sur les méthodes à mettre en œuvre et sur les référentiels, appelés assez souvent les critères d'évaluation. Il s'agit d'abord de bien définir les dimensions de l'activité, ainsi que les types de résultats qui doivent être évalués. Nous avons construit ces référentiels progressivement en testant la façon dont ils étaient compris et utilisés par les évalués et par les évaluateurs. Ensuite, nous avons cherché à les rendre cohérents avec ceux de l'évaluation des unités. Ces référentiels ont été construits en tenant compte de toutes les missions de l'Institut, et nous les avons déclinés aux différents niveaux d'organisation (institut et départements, unités, chercheurs et ingénieurs).

S'est ajoutée, au travail interne à l'Inra, une ouverture aux pratiques d'évaluation dans les autres organismes, en Europe et au-delà. Dès 2004, j'ai ainsi participé à un groupe de travail du Comité national d'évaluation de la recherche (CNER), destiné à réfléchir à une mise en commun des systèmes et pratiques d'évaluation de différents organismes de recherche. Cette interaction avec d'autres organismes, comme le Cemagref, le Cirad, s'est révélée intéressante et profitable. Sortir de sa pratique, de ses problèmes opérationnels quotidiens, regarder comment font les autres, est quasiment indispensable dans ce genre de fonction, notamment pour tisser des réseaux avec des collègues d'autres organismes.

Qu'apprendre de l'évaluation dans les pays voisins ?

NSS : Sur ces sujets, les réflexions sont-elles partagées avec des instituts étrangers ?

Élisabeth de Turckheim : Au début, avec Danielle Barret du CNER, qui représentait la France dans le réseau européen EVANET, j'ai participé à un groupe de travail qu'elle avait mis en place pour créer un réseau de responsables de l'évaluation des organismes et établissements de recherche. Nous avons travaillé en particulier sur des documents préparés par les Autrichiens. Ils se

posaient toutes les questions de méthodologie sur l'évaluation d'une structure de recherche pérenne : Qu'est-ce qu'on évalue ? Quelles méthodes ? Nous avons aussi travaillé sur une liste commune des mauvaises et des bonnes pratiques des évaluations de la recherche que nous pratiquions.

Ensuite, j'ai continué au sein d'un forum de l'European Science Foundation (ESF). J'ai tout d'abord noté que la démarche d'évaluations individuelles régulières, hors des procédures de candidatures à des financements, était spécifiquement française. Au niveau des évaluations ex-post, ne sont comparables entre pays que les évaluations des programmes et des structures.

Via ce forum de l'ESF, j'ai croisé de nombreuses personnes, venant pour moitié d'agences de financement, et pour moitié d'organismes de recherche. J'ai aussi complété ma bibliographie. C'est ainsi que j'ai connu les travaux de Jack Spaapen, un chercheur néerlandais qui animait le groupe de recherche ERiC (Evaluation of Research in Context). Ce groupe avait défendu la nécessité d'une évaluation qui prenne en compte le contexte dans lequel sont conduites les recherches et, en particulier, les questions posées par la société, les interactions avec les utilisateurs, les pratiques que cette implication impose. L'évaluation ne peut donc pas se réduire à la mesure des connaissances académiques produites. Nous avons déjà réfléchi à cette question en nous référant, comme le groupe ERiC, aux travaux fondateurs de Bruno Latour, Philippe Larédo et de leurs collaborateurs sur les différentes dimensions de la recherche. Pour construire leur méthode d'évaluation, les Néerlandais avaient travaillé sur des cas. Ils avaient, par exemple, conçu une évaluation pour la recherche agronomique des Pays-Bas, au moment où les instituts de recherche agronomique et l'Université de Wageningen préparaient leur fusion. Ils ont donc inventé une méthode adaptée à l'évaluation d'une recherche qui produit des connaissances et qui répond à des besoins opérationnels des utilisateurs. Leurs arguments et leur méthode pour introduire dans l'évaluation la question de l'utilité sociale des recherches – *a socially robust research* – et l'analyse de l'implication des chercheurs dans leur environnement social était très intéressants³.

Nous avons aussi suivi les travaux des Finlandais qui avaient réalisé en 2006 une étude très complète des méthodes d'évaluation de la recherche et mettaient en route un projet pour concevoir une méthode d'évaluation des impacts de la recherche.

NSS : Les Finlandais ont-ils recadré ces méthodes sachant que le terme « évaluation » n'a pas toujours la même signification ? Les méthodes sont en effet liées à des objectifs, à un cadrage, à une façon de faire de la recherche, à l'appréciation de ce qu'est un impact. Ont-ils également comparé les arrière-pensées de ceux qui appliquent les méthodes ? Les méthodes d'étude et d'analyse d'un objet, quel qu'il soit, sont en effet directement liées aux intentions qu'on a sur l'objet en question.

Élisabeth de Turckheim : oui, c'est vrai, la méthode d'évaluation dépend de ce qu'on veut faire de l'évaluation. En décrivant les différents modèles (dont une analyse fine de 12 protocoles dans différents pays), l'objectif des Finlandais était de définir un outil d'évaluation pour éclairer les décisions d'orientation de la recherche, au niveau du gouvernement ou des directions d'institutions. Mais à ce stade de leurs travaux, il n'y avait pas de proposition nouvelle, en dehors d'une recommandation forte de développer des méthodes d'évaluation de l'impact de la recherche et des mécanismes qui permettent la génération de ces impacts. Ils recommandaient ainsi d'évaluer les processus à quatre stades de leur déroulement : les ressources, les activités, les résultats et les impacts. Depuis cette analyse de 2006, ils ont réalisé un travail très approfondi en développant un cadre d'analyse et des jeux d'indicateurs déclinés selon quatre domaines d'impacts sociétaux. Mais leur méthode est une approche de niveau macro, adaptée au niveau national mais pas à celui des structures de recherche⁴.

Les évolutions du contexte national de l'évaluation

NSS : Après ce rapide panorama de vos activités, pouvons-nous en venir au contexte national ? Pourquoi les questions de l'évaluation qui étaient dans une certaine mesure stabilisées tout en étant critiquées, sont-elles revenues de façon un peu plus forte au goût du jour dans les années 2006-2010 ?

Élisabeth de Turckheim : La raison en est la réforme de la recherche en France qui a été conditionnée par une réforme de l'évaluation. Le discours du président de la République et du ministre de la Recherche était clair : pas de réforme de la recherche, pas de financement pour renforcer la recherche française sans une évaluation digne de ce nom. Cela a parfois été présenté comme « la recherche n'est pas évaluée », ce qui a soulevé un certain nombre de protestations. Mais il est vrai que la recherche

³ Spaapen, J., Dijstelbloem, H., Wamelink, F., 2007. *Evaluating Research in Context: A Method for Comprehensive Assessment*, Consultative Committee for research and development, The Hague, The Netherlands, 152 p., [http://www.nwo.nl/files.nsf/pages/NWOA_73VH8D/\\$file/eric_book_internet.pdf](http://www.nwo.nl/files.nsf/pages/NWOA_73VH8D/$file/eric_book_internet.pdf).

⁴ Luoma, P., Raivio, T., Tommila, P., Lunabba, J., Halme, K., Viljamaa, K., Lahtinen, H., 2011. *Better Results, More Value: A Framework for Analysing the Societal Impact of Research and Innovation*, TEKES review 288/2011.

n'était évaluée ni avec une méthode uniforme, ni avec une même visibilité.

Une évaluation systématique selon une méthode commune a donc été décidée en 2007 avec la création de l'AERES (Agence d'évaluation de la recherche et de l'enseignement supérieur). Elle a été chargée de réaliser elle-même les évaluations, et non pas seulement de les superviser ou de les coordonner. Cette structure opérationnelle est typiquement française et assez unique en Europe. Les Pays-Bas, par exemple, n'ont pas fait le même choix : ils se sont donné un modèle commun, qui est mis à jour régulièrement et que chaque organisme ou université met en œuvre, ainsi qu'une instance qui réalise une sorte de méta-évaluation : elle vérifie que le processus a été correctement conduit et que les décisions sont bien mises en œuvre. Les Britanniques ont, par contre, un système d'évaluation très centralisé et très rigoureux et dont le résultat détermine l'attribution du budget recherche des universités⁵.

Une telle structure, chargée de toutes les évaluations, comporte certes des avantages, en termes d'homogénéité et d'efficacité notamment, mais elle comporte aussi beaucoup de risques. Les méthodes d'évaluation ont un impact sur les pratiques de recherche. Un des risques forts d'une méthode normalisée est de réduire la diversité de la recherche en marginalisant les groupes dont les pratiques ne sont pas celles de la majorité. Un autre risque, découlant de la lourde charge que représentent toutes ces évaluations pour un seul opérateur, est la tentation de réduire l'évaluation à des procédures quantitatives réductrices et normalisantes. Il y a, en effet, une évolution de la gestion des politiques publiques, et de la recherche en particulier, vers une culture du management par performance, appuyé sur l'utilisation d'indicateurs. Une évaluation systématique et normalisée, produisant des notes et des classements, pouvait soutenir la tentation de dériver vers ce modèle de pilotage de la recherche.

L'AERES est organisée en 3 sections.

- La section 1 est celle des établissements. L'objectif était d'organiser l'évaluation des universités ; elle a ensuite été étendue à tous les organismes.
- La section 2, pour les unités de recherche et les autres structures de recherche. S'il était acquis que l'évaluation des organismes devait être externe, cela ne l'était pas pour celle des unités, qui était jusqu'alors pilotée par les organismes. C'était donc un vrai changement.

⁵ Le Higher Education Funding Council of England (HEFCE) réalise toutes les évaluations des départements de recherche des universités. Suite à une révision en profondeur de la précédente méthode (*Research Assessment Exercise*), le nouveau protocole, *Research Excellence Framework*, qui a fait l'objet de travaux et de consultations pendant cinq ans, sera appliqué à partir de 2013 (<http://www.ref.ac.uk>).

- La 3^e section est celle des formations (licences et masters, écoles doctorales...). Leur évaluation était jusqu'alors réalisée au ministère de la Recherche et de l'Enseignement supérieur (MESR).

Confier l'évaluation des individus à l'AERES a été envisagé mais l'idée a finalement été abandonnée au moment de la discussion de la loi.

C'est donc une grosse structure qui s'est mise en place. Même si des interrogations demeuraient sur les procédures qui allaient être mises en œuvre, et si aucune analyse des risques encourus n'avait été faite, le besoin d'une évaluation cohérente avec des méthodes à peu près communes pour l'ensemble de la recherche était reconnu par tous, y compris par les universités dont les équipes n'étaient jusqu'alors pas évaluées.

NSS : Dans la pratique, comment l'AERES s'est-elle mise en état de remplir l'ensemble des missions qui lui ont été confiées ?

Élisabeth de Turckheim : Le besoin d'un certain niveau de qualité dans les évaluations et de méthodes communes se faisait réellement sentir mais les évaluations par l'AERES se sont mises en place trop rapidement. Le conseil de l'AERES, dont j'ai fait partie, a été nommé au mois de mars 2007, et en mai, les évaluations commençaient. L'enjeu a été de tout évaluer tout de suite, avant même de réfléchir aux méthodes. L'AERES a appliqué à tous, sans les adapter, les méthodes de la mission scientifique, technique et pédagogique, instance d'évaluation du MESR et, dans une moindre mesure, celles du Comité national du CNRS. Pour les organismes finalisés, qui ne sont pas dans le modèle standard de production de seules connaissances fondamentales, il y avait un vrai danger. En juillet 2006, au moment où se préparait le décret de création de l'AERES dans les bureaux du ministère de la Recherche, les directions de l'Inra et du Cemagref se sont mobilisées et ont demandé à leurs responsables de l'évaluation de réagir et de proposer une analyse rapide des actions à entreprendre pour que l'évaluation ne soit pas complètement orthogonale aux missions des organismes et, à terme, ne les mette pas en danger.

Un groupe a été constitué et il a produit un premier document sur les enjeux d'une évaluation qui respecte les missions et les pratiques de la recherche finalisée. Le groupe s'est alors rendu compte de la nécessité de poursuivre le travail collectif pour faire des propositions communes de méthodes d'évaluation. Au-delà des organismes du programme 187 de la LOLF⁶ qui en étaient à l'initiative, ce groupe, qui voulait être très transversal et s'est défini comme un « groupe inter-organismes », s'est ouvert à d'autres institutions. Très rapidement, les organismes sous tutelle du ministère de l'Équipement ont

⁶ Le programme 187 de la LOLF est composé des organismes suivants : BRGM, Cemagref, Cirad, Ifremer, Inra, IRD.

rejoint le groupe, puis le CEA, l'Inserm, etc. Ce groupe s'est pérennisé sous l'intitulé « EREFIN » (Évaluation de la recherche finalisée).

Les apports du groupe EREFIN

NSS : Ce groupe a-t-il été reconnu par l'AERES ?

Élisabeth de Turckheim : Très progressivement, l'AERES a signalé la nécessité de tenir compte de la diversité de la recherche, mais sans avoir les outils permettant d'intégrer cette diversité et sans s'investir assez dans la réflexion sur la méthode d'évaluation. Puisque j'étais membre du conseil de l'AERES, représentant l'Inra, le Cirad et le Cemagref, je me sentais investie de la mission de faire la culture des membres de ce conseil, ainsi que de l'AERES dans son ensemble, sur les questions de la recherche finalisée et de l'expertise. Nous avons écrit plusieurs notes sur les méthodes et les critères mais l'AERES n'en retenait que quelques mots ajoutés dans des documents un peu « patchwork ». C'était assez frustrant de réaliser un travail approfondi à une dizaine de personnes, pour n'en voir pioché que quelques bribes, sans compréhension du sens de notre démarche.

Devant cette interaction difficile avec l'AERES, nous avons décidé de poursuivre le travail en direction des unités évaluées, avec la construction d'une méthode leur permettant de présenter leurs résultats de façon cohérente avec leur stratégie. Nous avons ainsi développé l'outil de description du profil de l'unité intitulé « matrice des activités », puis un dictionnaire des produits et enfin des « questions évaluatives » pour permettre de réfléchir à la valeur, à la qualité des résultats décrits. Plutôt que de faire changer d'abord les critères d'évaluation de l'AERES, nous avons donc décalé notre cible et nous avons travaillé en interne dans nos organismes en testant avec les unités les différentes composantes de la méthode, puis en la faisant valider par les directions. Ces dernières l'ont très vite soutenue et ont demandé aux unités de l'utiliser pour préparer leurs rapports d'auto-évaluation pour l'AERES.

NSS : Le paradoxe est le décalage des discours entre la Stratégie nationale de la recherche et de l'innovation (SNRI), qui met pour partie l'accent sur l'innovation, et une mécanique qui reste dans une conception complètement académique de la recherche.

Élisabeth de Turckheim : C'est vrai. Lors du premier cycle, l'AERES n'avait pas investi dans la méthodologie de l'évaluation. La croyance dominante était que de bons chercheurs, reconnus par les communautés de recherche, sauraient évaluer. Or, d'une part, l'évaluation nécessite un recul par rapport aux pratiques habituelles

de la communauté qu'on connaît et, d'autre part, la recherche finalisée ne représente qu'une petite proportion de la recherche française. Elle est donc peu représentée parmi les chercheurs qui reproduisent le modèle de ce qu'ils pensent être une bonne recherche, c'est-à-dire la leur. Oui, il y a eu pendant quatre ans un décalage entre le discours théorique « la recherche doit montrer son utilité, interagir avec des partenaires, avec la société civile » et des comités d'évaluation composés, certes, de très bons scientifiques mais sans recul et sans connaissances des travaux sur les méthodes d'évaluation et leurs impacts (*evaluation studies*).

NSS : Ces questions sont-elles débattues au sein du conseil de l'AERES ?

Élisabeth de Turckheim : Pendant les quatre années où j'ai participé au conseil, très peu. Mais après la conclusion du premier cycle d'évaluation, l'AERES a décidé pendant cette année 2011-2012 de réviser l'ensemble des procédures en reprenant une réflexion de fond sur les principes de l'évaluation, ses objectifs et ses destinataires. Avec le nouveau président de l'AERES, le conseil de l'agence a vu son activité redéfinie de façon plus conforme à ce que doit être le rôle d'un tel conseil. De nouveaux groupes de travail ont été constitués associant des membres du conseil et des délégués scientifiques dont les propositions sont maintenant transmises et discutées au conseil. Dans ce contexte, l'expérience et les travaux d'EREFIN ont été jugés intéressants par la nouvelle équipe de direction et nous avons été associés aux travaux d'un groupe de la section 2 qui a préparé une révision des critères d'évaluation des unités de recherche. Les concepts que nous avons précisés au sein du groupe EREFIN à propos des critères et de la façon de les rendre opérationnels, ainsi que notre expérience du travail interdisciplinaire et inter-établissements ont été utiles au sein de ce groupe. D'une certaine façon, le modèle EREFIN a été accepté.

NSS : Avez-vous diffusé ce modèle ?

Élisabeth de Turckheim : Oui, il est public sur le site internet de l'OST⁷. On y trouve le document final, ainsi qu'en archives, les documents d'étape qui permettent de suivre notre cheminement.

NSS : Pouvez-vous nous donner quelques indications sur la démarche adoptée par le groupe EREFIN ? Par quoi avez-vous commencé ?

Élisabeth de Turckheim : Notre première priorité a été de faire reconnaître, au moment de l'évaluation, l'ensemble des activités conduites par une unité de recherche. Pour cela, nous sommes repartis des idées à la base des représentations classiques que sont la rose des

⁷ <http://www.obs-ost.fr/en/la-cooperative/erefin.html>.

Tableau. La matrice des activités d'une unité de recherche.

<i>Activités</i> <i>Destinataires</i> <i>et collaborateurs</i>	Production de connaissances	Constructions de partenariats, animation de structures et de réseaux	Mise à disposition de connaissances et de ressources (expertise, formation, diffusion)
A. Le monde de la recherche	Production de connaissances génériques	Organisation de communautés scientifiques nationales et internationales, veille scientifique	Développement d'instruments mis à la disposition de communautés scientifiques
B. Les acteurs socio-économiques	Production de connaissances opérationnelles et d'innovations	Construction de partenariats socio-économiques	Expertise-conseil, études, appui technologique et mise à disposition d'équipements ou d'instruments
C. Les pouvoirs publics*	Production de connaissances et de méthodes pour l'appui à la décision et à l'action publiques	Construction de coopérations avec des acteurs publics (hors recherche)	Expertise scientifique collective, expertise et appui méthodologique dans le cadre d'instances réglementaires
D. Les étudiants	Conception d'enseignements, production de documents et de matériels pédagogiques, formation par la recherche	Définition de cursus et de programmes d'enseignement ; animation de réseaux de formation	Enseignements et formations dispensés
E. Les citoyens		Analyse de la demande sociale, veille sociétale, veille industrielle et technologique	Diffusion de connaissances scientifiques, contribution à la culture scientifique et technique, à l'interaction science/société

* Les activités destinées aux instances publiques en charge du pilotage de la recherche (évaluation de projets ou de structures de recherche...) ne sont pas répertoriées dans cette ligne mais dans la ligne A.

vents de la recherche⁸ (Callon *et al.*, 1992) ou les cinq horizons de la recherche⁹ (Latour, 1995). Ces représentations distinguent, dans l'environnement des chercheurs, les différents espaces dans lesquels ils sont impliqués, les groupes sociaux avec lesquels ils collaborent ou auxquels ils destinent les résultats de leurs recherches. Ces groupes ont été classés en 5 types : le monde de la recherche, les acteurs socio-économiques, les pouvoirs publics, les étudiants et les citoyens. Nous avons donc recommandé aux unités qui préparent leurs dossiers d'évaluation de faire l'analyse des destinataires de leurs recherches et d'évaluer comment elles partagent leur temps à développer des produits et des relations avec ces différentes composantes. La répartition de l'investissement pouvait être représentée par une image radar, mais la quantification de cette répartition – les valeurs à porter sur ces 5 axes – n'était pas immédiate : fallait-il mettre

des temps de travail, des indicateurs de résultats, de notoriété ou d'impacts ? Pour décrire les résultats de l'unité, il est aussi apparu nécessaire de distinguer les produits au sens strict (publications, brevets, modèles et simulateurs, prototypes, rapports divers...), de résultats immatériels que sont les relations construites avec différents groupes concernés par la recherche (les *stakeholders* en anglais). L'activité de construction de ces relations partenariales peut être objectivée par des indications comme l'organisation de séminaires ciblés, de sessions de formation pour des professionnels, la participation à des instances d'orientation scientifique des partenaires, etc. Sans être des résultats de recherche au sens strict, ces activités sont essentielles dans la recherche finalisée. Nous avons alors considéré qu'il fallait ajouter une deuxième dimension à celle des destinataires. Trois types d'activités ont été identifiés : la production de connaissances, la construction de partenariats et de réseaux, la transmission de connaissances et la mise à disposition de ressources développées par l'unité auprès de la communauté concernée. Ces trois types d'activité se déclinent pour chacune des cinq composantes de la société, ce qui conduit à une matrice de cinq lignes et trois colonnes (Tab.). Nous avons recommandé aux unités d'analyser

⁸ Callon, M., Larédo, P., Gonard, T., Leray, T., 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.

⁹ Latour, B., 1995. *Le Métier de chercheur : Regard d'un anthropologue*, Paris, Inra.

	Production de connaissances	Construction de partenariats	Mise à disposition de connaissances et de ressources
A. Le monde de la recherche			
B. Les acteurs socio-économiques			
C. Les pouvoirs publics			
D. Les étudiants			
E. Les citoyens			

Fig. 1. Exemple de représentation du profil de l'unité (carte des activités). Les poids relatifs des cellules du tableau, représentés par l'intensité de gris, caractérisent la stratégie de l'unité.

leurs activités avec cette matrice comme fond de carte et, après avoir partagé au sein de l'unité la description des priorités, de colorer les cellules de la matrice en fonction de l'importance de l'activité correspondante dans la stratégie de l'unité, pour en donner une image un peu percutante au comité d'évaluation (Fig. 1). Cette matrice était donc proposée comme un outil de positionnement stratégique.

Pendant ce travail de formulation, nous avons rapproché les vocabulaires utilisés dans les différents organismes et nous avons été attentifs à ne pas utiliser des mots trop réducteurs. Par exemple, nous ne voulions pas employer le terme « valorisation-transfert » pour tout ce qui concerne la recherche orientée par les besoins des utilisateurs. Nous nous sommes ainsi accordés sur l'expression « production de connaissances opérationnelles » pour parler de la production de connaissances contextualisées, répondant à un besoin d'acteurs socio-économiques publics ou privés.

Recherche finalisée et recherche partenariale

NSS : Avez-vous ensuite travaillé sur les produits de la recherche ?

Élisabeth de Turckheim : En effet, il fallait ensuite que les unités puissent rendre compte de leurs résultats et nous nous sommes intéressés aux productions issues des diverses activités identifiées et, comme je vous l'ai dit, il fallait en particulier rendre visibles et évaluables les indications d'activité de la dimension partenariale. Nous avons donc cherché à élargir la notion de « produit » et construit une sorte de dictionnaire des produits ou des faits observables pour chaque cellule de la matrice. La

diversité des membres du groupe EREFIN nous a permis de comprendre et d'intégrer dans ce cadre les spécificités des recherches de chaque organisme, d'approfondir notre compréhension de la diversité des pratiques et des produits de recherche et ainsi d'affiner la définition des cellules de notre matrice. Nous avons proposé une façon de compter les produits par cellule afin de pouvoir représenter la production de l'unité au regard de sa stratégie. Cette représentation par un histogramme avec cinq barres horizontales à trois zones (Fig. 2) permettait à l'unité de réfléchir à la cohérence de sa production et d'attirer l'attention des comités d'évaluation sur les résultats, au-delà de la liste des publications.

Enfin, nous avons travaillé sur la question de la valeur des produits parce que l'évaluation ne se résume pas à un comptage de faits, même s'ils sont observables. Plutôt que de parler de critères de qualité, voire de critères d'excellence, nous avons formulé des « questions évaluatives », notion que par la suite le groupe de travail de l'AERES a reprise sous la dénomination d'« indices de qualité ». La question de la qualité pour des activités partenariales est très peu abordée et encore moins instrumentée. Par exemple, la question des usages des résultats par les groupes sociaux concernés et celle des impacts sur ces groupes sont une dimension importante de la qualité. Pour chaque activité, des questions ont été formulées qui sont reliées à des objectifs et aux intentions qui ont conduit à la mise en place des projets. Ces questions sont destinées à plusieurs acteurs : l'unité qui prépare un bilan, la direction de l'établissement qui précise des missions confiées ou les termes de référence de commissions d'évaluation, les commissions elles-mêmes. Pour nous, le nœud était d'avancer sur l'analyse et l'évaluation de ce qui fait la qualité d'une recherche partenariale, mais plus largement en fait, dans le cadre de la recherche finalisée sous toutes ses formes.

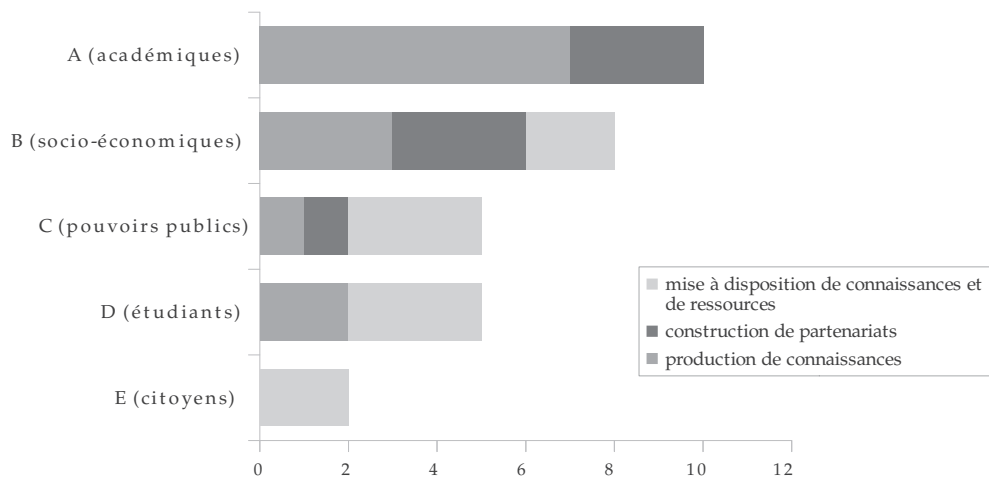


Fig. 2. Exemple d'une représentation synthétique quantitative des différents produits d'une unité en fonction du type de destinataires.

NSS : Le terme de « recherche partenariale » n'est-il pas un peu ambigu ? Pourquoi l'avoir choisi ? Pourquoi n'avez-vous pas repris la référence au concept de recherche de mode 2, développé par Gibbons, Nowotny et leurs collègues¹⁰ ?

Élisabeth de Turckheim : Dans la recherche partenariale, il y a un mode de production de connaissance qui implique très fortement les partenaires à différentes étapes et qui nécessite une pratique de recherche différente et des critères d'évaluation adaptés. En fait, notre objectif était la recherche finalisée, c'est-à-dire une recherche qui répond à des questions posées par la société et qui a en tête l'utilisation des résultats. Mais nous voulions un concept qui couvre plus largement les pratiques que le concept de recherche de mode 2. En effet, par exemple à l'Inra, un dixième des unités seulement peuvent se reconnaître dans celui-ci. Quand les généticiens discutent avec les autres sélectionneurs, où les met-on ? Ils discutent en amont de ce qu'ils vont faire, puis ils réalisent leurs recherches, sans interaction forte avec les partenaires. C'est au moment de la programmation qu'ils interagissent, c'est-à-dire lors de la formulation des questions et de la conception de la problématique. Ce n'est pas tout à fait du mode 2, parce que le travail de recherche se réalise ensuite de façon habituelle, les phases disciplinaires étant souvent prédominantes.

NSS : On a un peu l'impression qu'on n'a pas réussi à trouver une sémantique vraiment claire et partagée.

¹⁰ Cf. Gibbons, M., Nowotny, H., Limoges, C., Schwartzman, S., Scott, P., Trow, M., 1994. *The Next Production of Knowledge. The Dynamics of Science and Research in Contemporary Societies*, London, Sage ; Nowotny, H., Scott, P., Gibbons, M., 2001. *Re-thinking Science: Knowledge and the Public in an Age of Uncertainty*, Cambridge, Polity Press.

Selon les individus ou les institutions, on oscille entre recherche de mode 2, recherche partenariale, recherche transdisciplinaire..., soit de nombreuses appellations, mais pas réellement satisfaisantes.

Élisabeth de Turckheim : J'ai deux points de vue sur cette question. Le premier est celle d'une recherche très différente de la recherche académique, aux modalités opposées ou à contre-pied : c'est le mode 2, qui est un modèle alternatif (ou contre-modèle) au modèle académique dans sa version « pure ». L'important dans le mode 2 est la place du monde réel, qui reste présent tout au long du processus de production de connaissances, et le rôle des partenaires ; ces deux caractéristiques modifient la façon de faire de la recherche. Ces partenaires, par exemple, participent à la construction de la question de recherche, leurs savoirs profanes sont pris en compte, ils sont associés à la conduite des projets de recherche. Cette intrication des deux mondes de la recherche et de la société remet la question de l'appropriation des résultats par les destinataires au cœur du processus de recherche. Mais si cette vision peut paraître un peu « idéale », je la nuance effectivement avec mon deuxième terme ci-dessous.

Ma deuxième vision, en effet, s'élargit vers des démarches qui combinent ce modèle alternatif avec celui de la recherche académique. La direction de l'Inra nous avait confié en 2004 à Bernard Hubert et moi-même une mission dont l'objectif était de réfléchir aux relations recherche-partenariat. Il s'agissait d'étendre, au-delà du département Systèmes agraires et développement (SAD), une vision d'une recherche qui se conduit, au moins partiellement, selon un mode différent, en interaction étroite avec d'autres parties prenantes. Les réflexions conduites au cours de cette mission, telles que nous les avons

caractérisées dans notre rapport remis en 2006¹¹, m'ont été très utiles pour apporter au groupe EREFIN une compréhension de ces pratiques, et nous avons pu nous inspirer des critères de qualité que le groupe sur la recherche partenariale avait déclinés pour les huit étapes d'un projet de recherche partenariale. Mais ces étapes ne sont pas forcément toutes nécessaires. Selon les unités, à l'Inra par exemple, ces caractéristiques sont plus ou moins toutes présentes. Le but de la grille était de pouvoir évaluer à l'Inra des unités qui pratiquent à la fois une recherche de mode 1 et une recherche de mode 2, et, bien souvent, les combinent entre elles. Les apports du travail sur la recherche partenariale devaient être intégrés dans une méthode plus générale ouverte à toutes les pratiques.

À l'Inra, nous avons besoin que la recherche qui répond aux questions de la société soit évaluée par une méthode cohérente avec cette finalité. Cette recherche se construit en amont à partir de questions posées par les acteurs, et, en aval, la question de l'appropriation des résultats et de leurs impacts doit donc forcément être abordée. On ne peut pas se contenter de dire « On a fait un joli papier, on a un joli modèle », la question de son utilité sociale est essentielle. La façon d'interagir avec les partenaires est importante ; elle n'est pas identique à celle mise en œuvre entre chercheurs et il faut en tenir compte dans les évaluations. Ces pratiques de recherche doivent donc être analysées avec des critères qui ne sont pas seulement des critères de « beauté de la science » mais aussi avec des critères de pertinence sociale.

Ensuite, il y a un autre aspect à prendre en compte dans l'évaluation, celui qui concerne le processus de recherche avec tous ces allers et retours entre contextualisation et décontextualisation, entre recherche et intervention, entre recherche fondamentale et appliquée. C'est la remise en cause du « modèle linéaire » dont la première phase serait la recherche fondamentale et la dernière le transfert. Il y a des temps, des moments différents dans les projets. Il y a la question de l'enchaînement de différents types de postures, d'activités, et aussi celle de la gestion de ces enchaînements. Il faut comprendre comment fonctionnent ces processus, quels sont les facteurs de succès. C'est ce que préconisent les chercheurs qui s'intéressent au processus même de l'innovation, qu'ils caractérisent comme un lieu d'apprentissages croisés. Ce sont ces fameux « déplacements » évoqués dans le programme de recherche « Pour et sur le développement régional » (PSDR). La question de départ est modifiée au cours du projet, ainsi que les objectifs, les rôles des uns et

des autres. Il faut savoir gérer ces changements. Il faut expliquer ces dynamiques pour que l'AERES et nous-mêmes dans les instituts apprenions à comprendre ces processus dans une vision globale qui ne serait pas la dichotomie recherche/transfert.

NSS : Mais, à cette époque, les systèmes d'évaluation étaient stabilisés dans les organismes plus finalisés et ces spécificités y étaient déjà prises en compte.

Élisabeth de Turckheim : C'est vrai, les organismes, comme le Cemagref, l'Inra, l'Ifremer, avaient chacun des outils qui étaient à peu près satisfaisants. Mais la seule façon de les faire accepter était de les transformer en un seul outil commun. Ce qui nous a obligés à nous accorder sur une proposition commune qui, de ce fait, devenait plus robuste et justifiait d'autant plus son introduction dans la méthode nationale.

NSS : Cette démarche s'est-elle disséminée dans d'autres organismes de recherche ?

Élisabeth de Turckheim : Au début, les organismes du programme 187 de la LOLF étaient impliqués par une décision des directions mais nous avons vite élargi le groupe à d'autres organismes très finalisés comme le LCPC et l'Inrets (maintenant fusionnés dans l'Institut français des sciences et technologies des transports, de l'aménagement et des réseaux [IFSTTAR]), l'IGN, l'IRSN ; la démarche a aussi beaucoup intéressé l'Inserm, qui voulait soutenir les équipes s'impliquant dans la recherche clinique, et le CEA, qui souhaitait l'ouverture des méthodes d'évaluation aux recherches industrielles.

NSS : Cela étant, même au sein des organismes les plus finalisés, il y a des personnes pour plaider la seule excellence académique. Par ailleurs, l'évaluation et la promotion ne privilégient-elles pas l'aspect « production de connaissances » ?

Élisabeth de Turckheim : Ce sont des aspects qu'il faut faire évoluer, notamment en ayant une vision de l'interdépendance entre les différentes phases de la recherche. Il faut, en particulier, rendre compte et mesurer l'activité dans toutes les phases d'un projet ou d'un programme de recherche finalisée. Il faut prendre en considération le temps du projet et la maîtrise de son déroulement. On sait qu'un projet ne produit pas en permanence des connaissances fondamentales ou des résultats opérationnels mais ces temps différents doivent être expliqués.

Processus de recherche, temporalité et impact

NSS : C'est la prise en compte de la temporalité dans le processus de recherche.

¹¹ Hubert, B., de Turckheim, E., 2006. *L'Évaluation des recherches partenariales à l'Inra*. Rapport pour la direction générale, 25 p ; de Turckheim, E., Hubert, B., Cerf, M., 2009. L'évaluation des recherches partenariales : quelle procédure, quels critères ? in Beguin, P., Cerf, M. (Eds), *Dynamique des savoirs, dynamique des changements*, Octarès, 265-281.

Élisabeth de Turckheim : Oui, c'est reconnaître qu'elle existe. Cela signifie aussi de revenir sur des sujets terminés, ce que nous ne savons pas très bien faire dans nos évaluations quadriennales : revenir sur des impacts quand on commence à les voir, mettre l'accent sur la capitalisation, se poser des questions comme « Avons-nous créé du lien durable ? », « Avons-nous augmenté nos compétences ? ».

NSS : Pourriez-vous développer la question des impacts qui paraît un des points délicats ?

Élisabeth de Turckheim : C'est en effet un point compliqué. En général, on attend au minimum d'une équipe qu'elle précise les impacts qu'elle vise. L'évaluation ne peut pas occulter la question des impacts puisqu'elle doit prendre en compte l'utilité finale des projets. Mais elle ne peut pas reposer seulement sur les preuves de ces impacts, en particulier, on ne peut pas se contenter de demander à les mesurer *in fine*. En effet, plusieurs difficultés rendent l'évaluation de l'impact, entendu comme un changement observable dans la société, difficile. Il y a la question du temps parfois très long entre la production d'un résultat de recherche et son effet, la question de la contribution de nombreux acteurs à la réalisation d'un impact et la difficulté d'attribuer un impact à un groupe de chercheurs particulier, la question de l'asymétrie, soit la réussite surprise (sérendipité) ou l'échec d'un impact dus à beaucoup d'autres facteurs que la qualité et la pertinence du travail de recherche. Une solution est alors de regarder si les conditions qui rendent possible un impact ont été remplies. On peut, en particulier, évaluer la façon dont est conduit le processus de recherche, évaluer la communication avec les destinataires, regarder en amont si la question choisie l'a été avec les bons partenaires, si les conditions de collaboration ont permis des rapports de confiance et équilibrés, si les divers acteurs ont discuté et se sont bien appropriés les produits intermédiaires, etc.

NSS : Quand vous dites « évaluer la façon dont est conduit le processus », voulez-vous renvoyer à la démarche qualité classique ?

Élisabeth de Turckheim : Ce n'est pas aussi normalisable. Il ne s'agit pas seulement de la conduite d'un projet standard, car il peut y avoir des ruptures, des bifurcations, des pauses, etc. Il y a toute l'ingénierie sociale qui se met en œuvre et qui se voit dans les actions qui sont prises. On ne peut pas normaliser ces processus au point de les décrire avec des grilles de contrôle de qualité.

Nos collègues néerlandais ont travaillé sur cette question de l'analyse des interactions entre chercheurs et partenaires. Ils ont proposé le concept d'« interactions productives » qui permet d'anticiper une possibilité d'impact et qui permet de contourner les questions d'identification et de mesure de l'impact ainsi que celle

de l'attribution¹². De son côté, le groupe EREFIN a identifié plusieurs niveaux d'analyse de la qualité des résultats de recherche finalisée de façon à prendre en compte les produits, leur qualité « primaire » – comme le caractère innovant des solutions proposées, l'efficacité de la réponse apportée, la pertinence du mode de diffusion, etc. –, puis leur utilisation qui est une indication de leur utilité, et, enfin, les impacts lorsqu'on peut en observer –effets sur la mise en œuvre de l'action publique, modifications de l'organisation ou des comportements des bénéficiaires, etc.

Enfin, il y a un dernier point qui, selon moi, n'a pas été traité : recueille-t-on des informations, voire un avis, auprès des partenaires ? On le devrait, me semble-t-il, sans pour autant les mettre en position d'évaluateurs. Le dossier d'évaluation pourrait comporter une partie rédigée par les partenaires. Nous insistons sur leur rôle au sein des processus, il est donc logique de ne pas les dispenser complètement d'une contribution au bilan d'un projet partenarial.

NSS : Cela pose-t-il la question de l'évaluation par les pairs ? Au Cemagref, par exemple, il y avait des comités de visite mixtes avec, d'une part, des scientifiques et, d'autre part, des gens qui étaient les utilisateurs de la recherche. Par ailleurs, le mécontentement des partenaires n'est pas forcément le signe d'une action qui n'a pas réussi, mais parfois de résultats qui ne vont pas dans le sens qu'ils souhaitaient.

Élisabeth de Turckheim : Au sein du groupe EREFIN, nous nous sommes demandé si les partenaires devaient avoir un rôle dans l'évaluation. Nous avons plutôt suggéré que les partenaires-utilisateurs ne soient pas directement considérés comme des évaluateurs, mais qu'ils contribuent en participant à une partie du rapport. Par ailleurs, le comité d'évaluation pourrait comprendre des personnalités du monde professionnel qui ont une bonne expérience de la collaboration avec des chercheurs. Les partenaires de l'unité n'évaluent pas le projet, mais, dans un rapport complémentaire au rapport d'évaluation des scientifiques, ils pourraient présenter leur point de vue sur le projet et leur participation. Dans le modèle hollandais, ils répondent à une enquête des évaluateurs. Ainsi, le processus d'évaluation prend du recul par rapport au seul avis des scientifiques sur les collaborations.

L'AERES a identifié ce besoin d'inclure, dans les comités, des experts du monde professionnel, capables de participer à une évaluation, ou des chercheurs ayant une bonne expérience de la collaboration avec des partenaires socio-économiques. Nous avons, certes, un vivier

¹² Projet SIAMPI (Social Impact Assessment Methods for research and funding instruments through the study of Productive Interactions between science and society), 7^e PCRD.

dans les commissions d'ingénieurs, mais autant pour un chercheur, évaluer fait partie de son métier, autant ce n'est pas le cas pour un professionnel. Il peut avoir à négocier avec sa hiérarchie pour prendre le temps de participer à une évaluation d'unité ou de projet de recherche, et cela d'autant plus s'il s'agit d'une unité avec laquelle sa société ne collabore pas directement. On ne peut pas éluder complètement cette question, mais on peut mobiliser des collègues, qui ont bénéficié de formations de base scientifique dans des structures professionnelles, en quelque sorte des personnes intermédiaires entre le monde académique et le monde socio-économique...

Construire une communauté des évaluateurs

Enfin, je voudrais terminer par une dernière idée développée dans le texte de Nicolas Dodier¹³. Fort de son expérience de membre de commissions de spécialistes à l'Inserm, puis à l'Inra, il regrette que la création de l'AERES rende caduques ces commissions qui permettaient la constitution de communautés un peu pérennes où l'on construisait une « métaculture » partagée. Ce type de commission, qui évalue un ensemble d'unités au cours de son mandat, est en effet bien adapté à la construction d'un champ d'analyse des travaux de recherche non académique, même si le succès de ces constructions n'est pas automatique et dépend en particulier de l'ouverture des membres des commissions, de leur intérêt pour découvrir et comprendre des environnements et des pratiques qui ne leur sont pas a priori familiers.

Les échanges, les débats et les négociations dans les commissions sont aussi importants que les avis qui en émanent sous forme écrite. Il s'y construit de la compréhension, une culture commune. Cette culture partagée peut en retour favoriser des collaborations entre spécialités disciplinaires. C'était l'intention de l'Inra lorsque nous avons modifié les contours disciplinaires de certaines CSS pour créer des synergies, par exemple, en

regroupant des écophysiologistes et des chercheurs en biologie moléculaire. Les débats internes qui portaient d'abord sur la comparaison des critères de qualité et sur les modes de publication ont permis des discussions et des points de vue de plus en plus ouverts sur les objectifs des projets de recherche évalués et sur leur conduite.

NSS : Est-ce toujours le cas ?

Élisabeth de Turckheim : Non. La première tentative peut être un échec, et la seconde un succès. Cela dépend des individus et du groupe qu'ils constituent, de l'animation du groupe, mais cela mérite d'être tenté. Dans le fonctionnement de l'AERES avec des comités d'évaluation ad hoc pour chaque unité, ce sont les délégués scientifiques qui pourraient jouer ce rôle. Mais ils sont actuellement renouvelés à 50 % tous les ans et ils ont peu de temps pour du travail de réflexion commun au-delà du pilotage des évaluations des unités dont ils ont la charge. Comment va-t-on créer une métacommunauté où un réel travail collectif est réalisé sur des évaluations de programmes ou d'unités variés ? Pour le champ particulier des recherches partenariales ou transdisciplinaires, on pourrait essayer de constituer, au niveau national, un collectif qui réaliserait quelques évaluations exemplaires sur quelques programmes importants ; une vingtaine de personnes de plusieurs organismes y travailleraient collectivement un peu longuement, ce qui devrait permettre de construire une autre culture de l'évaluation qui pourrait être capitalisée par l'AERES.

NSS : N'est-ce pas ce genre de travail qui est réalisé au sein de revues comme *NSS* ?

Élisabeth de Turckheim : Cela y ressemble en effet. Il faut des personnes suffisamment ouvertes, capables de passer d'un projet transdisciplinaire à un autre, qui connaissent les sujets à analyser, tout en n'ayant pas peur de poser des questions naïves.

Il faudrait que quelques organismes envisagent une opération pilote pour évaluer quelques gros programmes de recherche participative. Ces organismes pourraient ainsi imaginer et tester une méthode qui pourrait éventuellement être étendue. L'AERES pourrait suivre le projet, sans s'engager a priori à transcrire la méthode dans ses propres pratiques. Pourquoi ne pas initier un tel processus dans le cadre des activités de l'association NSS-Dialogues en s'appuyant sur les acquis de ce collectif en matière de programmes transdisciplinaires ?

¹³ *Penser un régime d'évaluation de la recherche scientifique*, intervention de Nicolas Dodier, sociologue, directeur de recherches à l'Inserm et directeur d'études à l'EHESS, présentée le 25 mars 2009 lors d'un débat sur l'AERES. Elle est disponible sur <http://evaluation.hypotheses.org/452>.