

"Les évaluations de la recherche"

Compte rendu du colloque du Centre D'Alembert, 13 & 14 mai 2009, Orsay

Jean-Claude Vial

Dans son allocution d'ouverture, G. Couarraze, Président de l'Université Paris-Sud (UPSud), a loué les aspects « échange et liberté de pensée » de ce genre de colloque. Il a rappelé l'actualité du sujet et la nécessité d'une réflexion épistémologique. Il comprend « EvaluationS » comme multiple par ses acteurs, ses outils, ses finalités. De la même façon, il faut entendre « la recherche » par les diverses disciplines et aussi les systèmes de recherche (telle UPSud) dont l'efficacité doit être mise en regard des moyens mis en œuvre.

J-F. Pome, au nom du Pôle de Recherche et d'Enseignement Supérieur (PRES – UniverSud) a présenté UniverSud, ses objectifs (notamment un rapprochement Grandes Ecoles-Universités), ses moyens (budget de recherche de l'ordre de 2,4 Meuros) et ses outils d'évaluation.

A. Revcolevschi, Vice-doyen, Directeur de la Division de la Recherche à la Faculté des Sciences d'Orsay, a donné une idée des débats au Conseil de la Recherche et rappelé que UPSud a été évaluée par l'AERES. Il attire l'attention sur trois aspects de l'évaluation :

1/ l'excès de classement et la tentation de se limiter aux divers indicateurs, notamment bibliométriques. Il donne quelques contre-exemples : ceux d'Einstein et du mathématicien français (orcéen) qui n'avait pas publié pendant des années, avant d'obtenir la médaille Fields.

2/ il rappelle que le débat n'est pas formel : il va falloir déterminer les crédits pour le prochain quadriennal et il demande à tous quelle équation utiliser ?

3/ il insiste sur une juste évaluation des doctorants tenant compte de la curiosité et de la prise de risque.

J-M. Lefour, Directeur du Centre d'Alembert a enfin remercié tous les organismes qui soutiennent le colloque et son comité d'organisation.

La première partie était consacrée à l'histoire comparée de l'évaluation dans la gestion de la recherche publique et privée, fondamentale ou finalisée

Sont intervenus : B. Bellon, Professeur émérite en sciences économiques, UPSud, («Evaluation scientifique et évaluation managériale»), P. Laredo, Directeur de recherche à l'Université Paris-Est (École des Ponts, Laboratoire Territoires, Techniques, Société) et Professeur à l'Université de Manchester, («L'évaluation par les pairs, quel rôle des organisations dans son déroulement et ses résultats») et P. Lazar, ancien Directeur général de l'INSERM et ancien Président de l'IRD («L'évaluation, un processus de décision nécessairement diversifié, sélectif et engagé»).

Quelques caractéristiques de la recherche sont rappelées: accumulation temporelle des découvertes/connaissances, existence de frontières non pas nationales mais cognitives et enfin mondialisation des connaissances combinée à une certaine raréfaction des finances publiques. Evaluer est une exigence à la fois politique, économique (qui concerne le payeur, l'acteur, l'utilisateur), éthique (cf. génétique, atome, ..). Plusieurs schémas sont proposés : une triple hélice constituée de l'Université (ses trois composantes), la Société (gouvernants et contribuables), l'Industrie à quoi s'ajoutent les media ; ou encore le triangle : gouvernement, opérateurs, acteurs. L'évaluation comporte aussi des risques : elle crée une image qui va devenir « réalité » dont la complexité ne va pas être saisie par tous (journalistes). Une fois posé le processus en termes d'input (hommes, argent, instruments, immobilier) et output (articles, brevets,

livres, ...), tout tient dans la « boîte noire » située entre les deux et qui contient: la nature des processus de recherche, la gestion des institutions, les relations internationales, la gestion quotidienne. La grande diversité de processus d'évaluation est soulignée: aux USA, l'évaluation est proche du « client » (étudiant, ..) ; en Angleterre, elle est très hiérarchique ; enfin au Canada, l'évaluation se fait par des entreprises extérieures. Une des questions posées dans cette boîte noire est : l'analyse des processus d'évaluation par les pairs est-elle dissociée des contextes organisationnels dans lesquels ils s'insèrent ? A partir de dispositifs d'évaluation très voisins, le CNRS et l'INSERM ont abouti à des résultats différents. Se pose aussi la question de la science « hétérodoxe » et de son financement : que va faire par exemple, l'European Research Council (ERC) où quelques Geuros sont mis en place ? A cause de la multiplicité des Agences, la recherche hétérodoxe est sous-financée en Europe, au profit de la recherche « mainstream ». De la même façon, dans les programmes blancs de l'ANR, tous les biais sont possibles (conformisme, prétention à manager la recherche ..). L'initiative de l'AERES de classer les labos selon quatre lettres est vue comme une réduction scandaleuse de la complexité des évaluations. En ce qui concerne la connaissance (l'un des deux déterminants dans notre société, avec l'argent), 3 domaines sont structurants : l'accroissement des connaissances, le transfert des savoirs, le transfert opérationnel des connaissances (mobilisation des savoirs en vue des applications) ; ils sont fortement hiérarchisés : « faire de la recherche » avant tout, alors même que l'enseignement à temps plein devient une nécessité. Ce dernier est si complexe qu'il a besoin d'une évaluation spécifique. Il est préférable d'avoir des chercheurs de qualité dans des domaines « partout denses ». Quant aux procédures d'évaluation, on pourrait voir les priorités suivantes : 1/ Gestion du temps : il convient que le temps d'évaluation ne dépasse pas le temps de l'action ou qu'on ne tombe pas dans le « tout quantitatif » ; il faut installer un principe de subsidiarité au niveau européen; rôle fondamental de l'Unité de Recherche avec possibilité de fermer –au bout de douze ans- un laboratoire, ce qui a été réalisé à l'INSERM ; recruter plus jeune avec des passerelles entre les 3 modes de transfert des connaissances, 2/ Nature des évaluateurs. L'évaluation est un choix : la mise sur pied de comités d'experts est une régression car l'existence d'élus permet un contrôle déontologique et l'acceptation par la base. 3/ « Evaluer, c'est choisir » (P. Mendès-France). Enfin la multiplication de primes individuelles attribuées aux chercheurs, risque de renforcer l'individualisme et non l'esprit d'équipe. La discussion a permis de relever que la France a la chance d'avoir du soutien de base et qu'il y a plus de financements industriels dans l'Université européenne qu'aux USA. Le labo joue un rôle central, y compris pour « cacher » la recherche hétérodoxe !

Les acteurs de l'évaluation

Sont intervenus : J-F. Mela, Mathématicien, Professeur émérite à l'Université Paris XIII, ancien chef de la Mission scientifique universitaire (MSU) (« Réévaluation de l'évaluation. Qui évalue quoi et pourquoi ? »), Y. Langevin, Directeur de l'Institut d'Astrophysique Spatiale, ancien Président de la Conférence des présidents du Comité national, (« La pratique de l'évaluation dans la section Système Solaire et Univers Lointain du Comité national »), F-A. Wollman, Directeur de l'Institut de Biologie Physico-Chimique, Université Pierre et Marie Curie (Paris 6) («L'évaluation scientifique des unités de recherche et des chercheurs : l'AERES a-t-elle amélioré le système ?»), J-P. Finance, Professeur d'informatique, Président de l'Université Henri Poincaré (Nancy I), ancien Président de la Conférence des Présidents d'Université (CPU) (« Stratégies des universités et Evaluation de la Recherche ») et B. Castro, Directeur Scientifique Affaires Industrielles à Sanofi-Aventis, («L'évaluation de la R&D dans l'Industrie Pharmaceutique»). Le paysage a changé depuis 25 ans. La recherche ne se limite plus aux organismes. La

composante universitaire a cru (démographie, politiques d'association et de contractualisation). Le pilotage par l'Etat a eu pour conséquence la contractualisation et donc l'évaluation de toutes les équipes, sur tous les campus. En 2006, a été créée l'AERES en tant qu' « autorité administrative indépendante » (du même type que la CNIL, CSA, ..) à mission « universelle », une structure qui n'existe dans aucun autre pays. Les « Etats Généraux de la Recherche » (2004) étaient plutôt favorables à la création d'une « Instance nationale d'évaluation scientifique » (INES) agissant pour le compte de l'ensemble des Etablissements. Son fonctionnement devait s'inspirer de celui du Comité national de la recherche scientifique du CNRS (CoCNRS). La présence d'élus était sujet à débats (sont-ce des experts ?). La création de l'AERES, constituée seulement d'experts –nommés-, et surtout sa pratique ont fait l'objet de beaucoup de critiques (abus d'indicateurs quantitatifs, ..), notamment en SHS. Une des justifications avancées était que l'évaluation collégiale et contradictoire du CoCNRS coûte cher (un super CoCNRS coûterait 7 Meuros par an et constituerait une charge de travail multipliée par 2,6 en comparaison avec le CoCNRS). Par ailleurs, la création de l'AERES obéit à une logique budgétaire : l'Etat distribue l'argent en fonction des résultats des opérations inscrites dans la LOLF, cette Loi qui affiche déjà des indicateurs tels que nombre de publications, indice de citation, brevets, Aujourd'hui, l'évaluation est reportée vers les Universités qui pratiquent alors l'auto-évaluation, une situation diamétralement opposée à celle des USA où ce sont la NSF et les autres agences qui procèdent à l'évaluation des projets, évaluations sur lesquelles s'appuient les Universités pour former leurs propres jugements. Quant aux Britanniques, ils sont en train de transformer leur RAE (Research Assessment Exercise qui examine les départements universitaires tous les 4 à 6 ans) en REF (Research Excellence Framework) avec comme principal indicateur les financements que ce projet est capable d'attirer ! Ce report d'évaluation vers les Universités nécessite une aide opérationnelle pour faire un diagnostic personnalisé à partir des indicateurs et des notes de l'évaluation nationale. L'analyse comparée des fonctionnements du CoCNRS (à partir d'une de ses sections) et de l'AERES a été très révélatrice. Les tâches d'évaluation des sections du CoCNRS portent sur les chercheurs (recrutements, promotions) et unités de recherche (jusqu'à l'arrivée de l'AERES). Cette évaluation par les pairs, collégiale et transparente, est internationalement reconnue. Les critères sont clairement affichés, qu'il s'agisse des recrutements (à travers le « coloriage » des postes, la section 17 est à même de mener une politique scientifique). L'évaluation des chercheurs tous les deux ans permet de déceler des problèmes potentiels. Lors de l'examen des dossiers de promotions DR2 - DR1, puis DR1 - DRCE, un poids de plus en plus important est attribué aux critères concernant la reconnaissance internationale et les responsabilités dans la discipline. La bibliométrie est donc considérée comme l'UN des critères. Elle peut pénaliser les thématiques en émergence, favoriser les effets de mode mais elle est de plus en plus significative avec du recul. Au moment des Etats Généraux, face à la grande diversité des systèmes d'évaluation (CEA, CNRS, INRA, et les Universités..) le souci d'évaluer à la fois chercheurs et unités de recherche avec la plus grande cohérence s'était manifesté. Ce sont des comités de suivi qui ont proposé une instance unique (février 2005) dont les résultats d'évaluation des unités seraient confiés aux instances concernées, seules responsables des évaluations individuelles. Pourquoi l'AERES n'a-t-elle pas répondu aux attentes ? Quelques raisons sont données : les rapports, étant publics, ne peuvent indiquer des noms de personnes ; les rapports des comités de visite peuvent être modifiés sans le consentement du Comité ! En physique, les comités ayant refusé de noter, ce sont les délégués de l'AERES qui ont effectué, depuis leurs bureaux, ce classement. Devant ces aberrations, l'AERES a lancé en 2008 une démarche « qualité » (exigences / satisfaction « client »). Cette démarche n'avait-elle pas pour objectif de

confirmer la qualité de l'AERES ? Si l'AERES est simplement une Agence de notation n'est-elle pas illégitime ? Un point de vue bien différent a été porté sur l'AERES. Le contexte français étant caractérisé par une triple dualité historique (Universités – Écoles, Accès non sélectif – Accès sélectif, Universités – Organismes), les Universités ont une place à trouver entre gestions nationale et facultaire. L'intervention de l'Etat n'est pas assez stratégique. On constate un poids hypertrophié des disciplines et peu de culture européenne et internationale. Au cours des trente dernières années, le nombre d'étudiants a été croissant ; l'immobilier a suivi et les effectifs d'EC ont suivi une croissance épisodique. Cette massification a conduit à un changement qualitatif des métiers. On ne peut nier des aspects positifs récents : quadriennal (qui a induit une politique d'établissement), unités mixtes ; création des écoles doctorales. On note aussi la prise en compte d'une dimension territoriale (relations avec collectivités), la mise en place du processus de Bologne (LMD), la mise en place de la LOLF et d'instances d'évaluation. Quelques questions fondamentales sont soulevées : comment prendre en compte la diversité du paysage universitaire ? passer du mythe de l'égalitarisme à la recherche de l'équité ? comment améliorer la Gestion des Relations Humaines (GRH) ? L'augmentation des moyens en personnels n'a pas conduit à un meilleur fonctionnement. La loi d'avril 2006 sur la recherche a notamment abouti à la création des PRES (une bonne échelle d'organisation) et au développement de politiques de sites. La loi 10 août 2007, réclamée par la CPU, est une bonne loi de décentralisation (autonomie et responsabilité des Universités). Mais la question des moyens est une condition *sine qua non* du succès: il faudra voir ce qu'il en est réellement des + 5 Meuros en 5ans et + 4 Meuros pour la recherche annoncés. La première étape consiste à construire une stratégie d'établissement qui combine bottom-up et top-down. Les modes d'action des établissements passent par une allocation de moyens, les recrutements et la GRH. Les partenariats ont un grand rôle à jouer. Il faut reconnaître différentes formes d'excellence, ce qui conduit à une stratégie propre à chaque établissement. Dans ces conditions, il faut adapter les instruments d'évaluation à la mesure objective de ces diverses formes. L'évaluation est un outil de pilotage et de mise en œuvre opérationnelle. Il faut une démarche « qualité » qui porte sur les processus (de production) et sa relation avec l'auto-évaluation (certaines universités ont été accréditées ISO9001). Il faut séparer les pouvoirs d'évaluation de ceux de décision, la transparence, l'aspect contradictoire et le besoin d'une instance de recours. La question de la GRH est centrale : comme les universités deviennent des employeurs, il faut être rigoureux sur les comités de sélection, la gestion des primes et des carrières. C'est donc une « révolution » qui est induite par le triangle Autonomie <-> Responsabilité <-> Qualité & Evaluation .

Dans le privé, l'évaluation est essentiellement hiérarchique. Pour l'embauche externe , il y a contacts entre chef de service et « candidat » (même dans le cas d'une candidature spontanée). Le tri initial par CV est fait par le chef de service (CS), puis le chef du chef. Les entretiens d'embauche se font plutôt en tête à tête : avec le CS puis le DRH. La décision est collégiale avec primauté au CS ; le DRH fixe le salaire. La procédure est simplifiée dans le cas d'un CDD ou d'un ex-thésard. Quant à l'avancement, elle passe par un entretien annuel avec le supérieur hiérarchique immédiat (« cet entretien est la pierre angulaire de l'évaluation des personnels »). Les promotions de grade et changement de secteur sont réalisés dans l'enveloppe budgétaire de façon collégiale (négociées avec les syndicats). Quant à l'évaluation des programmes de recherche, il s'agit de « viser la mise sur le marché de nouveaux produits pharmaceutiques apportant un service de santé ». C'est donc une recherche finalisée sur programmes qu'il s'agit d'évaluer à toutes les étapes : depuis l'évaluation initiale, l'évaluation en cours et enfin l'évaluation des projets de développement. Le DG peut décider alors du passage en phase 3 clinique après avis extérieurs (scientifiques,

autorités de santé, agences de brevet). L'évaluation de la compagnie pharmaceutique est faite par les clients (agences, prescripteurs, malades) et les investisseurs. Pour conclure, l'orateur insiste sur l'ineptie de la programmation des découvertes, donnant l'exemple du Taxotère trouvé par Pierre Potier mais non programmé. La « recherche à risques » se fait dans « des jardins secrets », mais c'est de plus en plus difficile à cause d'un contrôle de gestion trop tatillon. On retiendra parmi quelques fortes formules « On ne met pas les chercheurs du privé sous tension. »

La métrique de l'évaluation

Sont intervenus : M. Blanc, Directeur Général adjoint chargé de la Recherche, École Polytechnique (« Métriques, classements et politique scientifique des Etablissements »), Y. Gingras, Professeur, Chaire de recherche du Canada en histoire et sociologie des sciences à UQAM (« Du bon usage de la bibliométrie »), J-M. Schlenker, Professeur, Institut de mathématiques, Université Paul Sabatier (Toulouse III) (« Utilité et limites des indicateurs bibliométriques : l'exemple des mathématiques »), M. Beaudouin-Lafon, Professeur à UPSud, Directeur du Laboratoire de Recherche en Informatique (« Indices et classements : à qui profitent les chiffres ? »), Les divers classements (Shanghaï ; WRWU, Times Higher Education, Ecole des Mines) sont caractérisés par surabondance et anarchie. Shanghaï donne la primauté à la qualité de l'enseignement, et de l'institution (nombre de Nobel, de chercheurs les plus cités), au nombre de publications (notamment dans Nature et Science) et enfin à la « performance académique » au regard de la taille de l'institution. En 2008, Paris 6 est 42^{ème}, Paris 11 : 49^{ème}, l'X : 119^{ème}. Le Webometrics Ranking of World Universities (WRWU) met l'accent sur la visibilité et l'impact des publications de l'établissement sur le Web ainsi que le nombre de consultations du site. Avec WRWU, Paris 6 est 13^{ème}, l'X : 333^{ème}. Times Higher Education tient notamment compte du taux d'encadrement, du nombre d'enseignants et étudiants étrangers et celui d'employeurs intéressés. Dans les cent premiers, on trouve l'ENS Ulm au 28^{ième} rang et l'X au 34^{ième} rang ; les autres Etablissements français ne sont pas classés. Le classement de l'Ecole des Mines tient compte des trajectoires professionnelles (nombre de postes de direction) ; ainsi l'ESCP-EAP est au 10^{ième} rang (elle a fourni le patron de Total) et l'Ecole Polytechnique au 47^{ième} rang (directeur de Nissan). La corrélation Shanghaï-Times est faible, signe que les critères influencent fortement les classements. En résumé, on ne peut que noter de multiples biais dans ces classements : effet de taille, effets de thématiques scientifiques (santé, environnement, .. surévalués). Par ailleurs, dans ces classements les objets de la métrique sont très divers : devenir professionnel, rayonnement académique, ouverture internationale, volume et impact de la recherche, notoriété. Ce qui est oublié, ce sont les inventions & brevets, la promotion sociale des étudiants, la promotion de la culture scientifique, la qualité des relations au sein de l'Etablissement (la notion de projet partagé). La granularité de la recherche se déclinant en: chercheur/ita (ingénieur-technicien-administratif), équipe, labo, faculté, établissement, les indicateurs les plus visibles et les plus « médiatisés » se concentrent sur chercheurs et établissements alors que les niveaux intermédiaires sont très importants. L'évaluation doit porter non sur l'établissement en soi mais sur son projet. Une fois défini le projet, alors la définition d'une métrique adaptée peut être une aide à la politique scientifique de l'Etablissement.

L'évaluation de la recherche n'est qu'un des champs d'application de la bibliométrie : histoire, sociologie des sciences, économie de la recherche, politique scientifique. Elle porte sur différentes échelles : micro (individus : la bibliométrie est à éviter), méso (labos ..), macro (pays). On note la croissance exponentielle des publications (facteur

10 sur 40 ans) et des références. L'âge médian des références augmente depuis 1960. Dans les sciences naturelles, le chercheur individuel n'existe plus ; la moitié des articles sont en collaboration internationale. Dans les sciences sociales, le travail se fait de plus en plus en équipe ; mais il reste des objets « locaux ». L'économie prend le statut d'une science. La bibliométrie est essentielle pour cartographier l'ensemble des pratiques et sortir des perceptions locales et anecdotiques ; elle permet d'identifier les tendances (régionales, nationales, mondiales) ; les pratiques de publication, de références et de collaboration étant différentes dans les diverses disciplines, les indicateurs doivent tenir compte de ces pratiques.

En mathématiques, comme dans les autres sciences, les journaux à comité de lecture sont la clé de voûte du système académique ; cela permet de sous-traiter l'évaluation scientifique à des experts compétents, aux éditeurs et referees des revues spécialisées. Mais en maths, les revues sont fortement hiérarchisées. La publication des articles est lente (3 ans) et le travail de vérification des referees peut prendre des semaines. L'accumulation des citations est aussi très longue (20 ans). Les directeurs de thèses ne cosignent pas avec leurs doctorants. Les collaborations sont relativement rares, il y a peu de co-auteurs. La bibliométrie est donc surtout utile pour les institutions, les labos/départements, certaines études statistiques (valables si le nombre de chercheurs.an est bien supérieur à cent). Pour les évaluations individuelles, les maths utilisent le concept de publications majeures (cf. l'exemple de la médaille Fields 2002). Avec l'autonomie des Universités, la bibliométrie peut permettre un diagnostic des points forts de l'institution et une calibration des évaluations inter-disciplines. Elle peut être un indicateur de succès d'une politique de recherche. En conclusion, l'utilisation d'outils bibliométriques est irréversible mais les domaines et modalités d'applications doivent être contrôlés. Les indicateurs doivent être adaptés aux spécificités disciplinaires. « Les évaluations doivent *in fine* reposer sur des experts scientifiquement compétents, même s'ils utilisent des indicateurs quantitatifs ».

Les deux indices bibliométriques courants sont analysés. L'indice H : il existe des Nobel avec $H=15$ et Engelbart (inventeur de la souris) a un $H=11$. Comme on a des lois de puissance reliant qualité des revues et articles avec les taux de citations, les outils statistiques naïfs (tels la moyenne) ne veulent rien dire. Le facteur d'impact : il est défini, lui, comme le rapport entre le nombre de citations l'année N d'articles publiés les années $N-1$, $N-2$ et le nombre total d'articles publiés l'année N . Mais il faut veiller à la bonne couverture des publications par ISI (60% en physique, 15% en informatique). La discussion a porté sur le manque de propositions de bons indices, l'effet des prépublications qui conduit à la fin de la hiérarchisation et sur la hiérarchisation des revues par l'AERES, notamment en économie. Pour Y. Gingras : « pour construire, il faut d'abord détruire », c'est-à-dire abolir les indicateurs dangereux et choisir des indicateurs adéquats.

Usages de l'évaluation et conséquences sur les pratiques de la recherche

Sont intervenus : J-Y. Merindol, Administrateur de l'ENS-Cachan, Président honoraire de l'Université Louis Pasteur (Strasbourg I) (« L'évaluation, pour qui ? ») et C. Paradeise, Sociologue, Professeure à l'Université, Paris Est-LATTS et Présidente de l'IFRIS (« Dispositifs de jugement et appropriation de l'évaluation »). Il est rappelé que les évaluations sont multiples : agences de notations financières, sociales ou environnementales, ONG, classements internationaux, normes, audits, inspections. Les objets à évaluer le sont aussi, comme on l'a vu (personnes, laboratoires et établissements).

Les dispositifs de jugement sont passés d'une évaluation « sociale » à une évaluation « experte », qu'il s'agisse du jugement, des outils, de l'impact ou de la logique. Le lien entre évaluation et allocation s'est renforcé à cause de la

multitude des classements, systématisé dans les pays d'Etat-Providence mais avec de fortes variations des usages par pays et disciplines. Cette mutation serait due en partie à la perte de crédibilité du jugement des pairs et à l'évolution générale des modalités de l'action publique (sur fonds de contraintes budgétaires). L'évaluation « experte » a créé une émotion collective en France. Parmi les craintes, on notera celle de la segmentation par tâches des « knowledge workers » (enseignement, recherche), de la mise sous tutelle des professionnels par les managers, la non-reconnaissance de la diversité des missions académiques (« l'enseignement comme punition »). Est également posée la question de la légitimité du pilotage bureaucratique (choix démocratiques vs. choix technocratiques). Au niveau professionnel, se pose la question de la dualisation des professions académiques avec les «stars» contre professionnels « ordinaires» et les «permanents» contre les «précaires».

En clôture, J-C. Vial, Directeur de Recherche à l'Université Paris-Sud (Paris XI), rappelle les objectifs de départ du Colloque : 1/ apporter aux scientifiques évaluateurs des éléments d'information sur les outils, les procédures d'évaluation, et aussi sur les conséquences des évaluations, 2/ apporter à tous les scientifiques des éléments d'information sur les finalités et modalités des évaluations dont ils sont l'objet. 3/ apporter à toutes les personnes impliquées dans la réflexion sur la recherche et dans la gestion de la recherche, des éclairages variés sur ce passage obligé de l'activité de recherche qu'est l'évaluation. Le colloque est intervenu à un moment on ne peut plus opportun (AERES, ANR ..) où nombreux sont les questionnements sur les modalités à mettre en oeuvre aujourd'hui (cf. question du vice-doyen de la Faculté d'Orsay : « quelle équation dois-je utiliser dans l'attribution des crédits ? »). Une réelle inquiétude est apparue dans la communauté scientifique sur «l'évaluation» de notre recherche par le pouvoir politique (et l'opinion publique), alors qu'elle aspire à une cohérence à l'intérieur du système hexagonal et en Europe. Sur la question de l'évaluation : quoi, pour quoi, est apparu un consensus sur l'exigence politique, économique, éthique de l'évaluation, qu'il s'agisse de « la » recherche (thématiques, disciplines), des structures, des acteurs. A la question de l'évaluation : pour qui ? le colloque s'est demandé qui passe la commande ? qui en tire le bénéfice ? La réponse globale étant : les fournisseurs de crédits (très divers en France, Europe), les labos eux-mêmes, l'établissement (cf. ci-dessous), les acteurs eux-mêmes, ... L'autonomie des Universités conduit à l'autoévaluation, ce qui pose le problème de cette évaluation dans le contexte plus large (national, international) des divers disciplines et thématiques. A la question de l'évaluation : qui ? le colloque a bénéficié de l'exemple de fonctionnement d'une section du CoCNRS qui satisfait aux critères de collégialité de transparence et où l'évaluation se fait à partir d'une réflexion prospective. L'expérience de l'INSERM est également enrichissante. On peut donc parler de « modèles » qui marchent mais ... coûteux (?). Restent les « autres » (par exemple les. Enseignants-Chercheurs). Des critiques sévères du fonctionnement AERES ont été formulées au cours du colloque, allant de la non-prise en compte des problèmes individuels jusqu'à la participation à l'allocation de moyens. Des raisons ont été avancées : opacité dans la constitution de l'AERES, absence de relais communauté-experts (élus), critères utilisés. La question a été posée de l'attitude à prendre : refuser ou infléchir ? Le colloque a reçu de précieuses informations sur ce qui se passe dans le privé qui procède lui aussi à l'évaluation des personnels, des programmes de recherche, des projets de développement et des organisations. Sauf celle des acteurs, l'évaluation est relativement collégiale mais reste hiérarchique. On retiendra la remarque « Il n'y a pas de plan pour trouver »... Le colloque est revenu en détail sur la question des métriques : les divers classements des établissements, les divers biais possibles (sur les critères et les résultats), les effets de taille, de thématiques scientifiques, les oublis (inventions & brevets, promotion sociale des

étudiants, promotion de la culture scientifique, qualité des relations au sein de l'Etablissement). Et puis les situations sont très différentes selon les domaines, même si dans les sciences naturelles & sociales, on constate l'importance croissante du travail en équipe, y compris international. La situation des humanités reste différente (malheureusement aucun exposé ne leur a été consacré). La situation est aussi très particulière en maths où la plupart des indicateurs n'ont aucun sens et où les situations peuvent être différentes entre sous-disciplines. Parmi les questions en suspens, on notera l'impact de l'arrivée des pré-publications, le positionnement de la bibliométrie (affaire des scientifiques ou des bibliométristes ?), le besoin de bons indices (lesquels ?), l'utilisation des indicateurs par les media (la société). Parmi les oublis et pour ce qui pourrait être la suite de notre réflexion, on trouve l'évaluation des ITA (pourtant des acteurs importants de la recherche, dans les sciences dures notamment), le besoin d'une analyse plus approfondie des modes de fonctionnement des autres organismes et leurs point forts/faibles –une remarque faite aussi par les étudiants assistant au Colloque et qui sont un peu restés sur leur faim par manque d'exemples concrets-, et enfin l'évaluation des très grands équipements.

Les présentations et débats sont disponibles sur le site <http://www.centre-dalembert.u-psud.fr/> (onglet : "conférences en ligne"). Il suffit de demander un login.