

EREFIN

Groupe de travail inter-établissements sur
l'évaluation de la recherche finalisée

Un jeu de descripteurs quantitatifs pour une approche intégrative de la production et des résultats d'une unité de recherche

Table des matières

1 Contexte et objectifs du projet	2
1.1 L'évaluation de la recherche impose de prendre en compte la réalisation de l'ensemble de ses missions	2
1.2 Une typologie commune pour composer un profil personnalisé pour chaque unité	3
1.3 Le besoin et l'usage de descripteurs quantitatifs	4
2 Construction de deux jeux de descripteurs	5
2.1 Les trois types de descripteurs	5
2.2 La liste des productions scientifiques de l'unité.....	6
2.3 Les descripteurs de niveau 1	7
2.4 Quelques pistes pour construire des descripteurs de niveau 2	8
2.5 Pour aller plus loin : des questions de méthode	8
2.6 Conclusion très provisoire.....	9
3 Références	10

Note sur le titre : Nous aurions pu dire une approche *intégrée*, mais le mot *intégrative* nous paraît mieux traduire le rôle de l'évaluation dans une dynamique d'amélioration

1 Contexte et objectifs du projet

1.1 L'évaluation de la recherche impose de prendre en compte la réalisation de l'ensemble de ses missions

La contribution à l'innovation et au développement d'une économie de la connaissance est au cœur des objectifs déclarés par les gouvernements européens depuis la définition de la stratégie de Lisbonne en 2000. Cependant, la question de la mesure de l'efficacité de la recherche dans les processus d'innovation et plus largement de la mesure des impacts sociétaux de la recherche est une question difficile. C'est en particulier ce que montre une étude très exhaustive, conduite à la demande du gouvernement de Finlande. Les auteurs du rapport, qui ont passé en revue les méthodes d'environ une centaine d'agences de financement ou d'organismes de recherche et analysé de façon approfondie douze procédures, soulignent l'absence de méthodes de référence pour évaluer les impacts culturels et sociétaux de la recherche (Kanninen et Lemola, 2006). Les diverses procédures d'évaluation abordent ainsi 4 volets : l'allocation des moyens, les impacts sur les communautés scientifiques, les collaborations avec les utilisateurs et les transferts de connaissances, l'utilisation des résultats de la recherche et leurs impacts sociétaux. Il ressort que la valeur d'un dispositif de recherche dépend de la réalisation d'une large gamme d'activités. Parmi ces activités, celles dont doivent répondre les collectifs de recherche vont bien au-delà de la seule production de nouvelles connaissances.

La diversité des activités de la recherche est donc essentielle à la réalisation de ses missions envers la société. L'attention à accorder à des missions comme l'expertise, l'appui aux politiques publiques ou la contribution à la culture scientifique de la société est régulièrement réaffirmée. La nécessité de prendre en compte l'ensemble des missions de la recherche lors des évaluations s'impose dans les principes. C'est ainsi que la Loi de Programmation de la Recherche¹ ainsi que le décret relatif au fonctionnement de l'Agence de l'évaluation de la recherche et de l'enseignement supérieur (Aéres) sont explicites sur l'obligation de prendre en compte dans l'évaluation l'ensemble des missions et des activités de recherche des organismes, respectant ainsi la « diversité de nature et de mission des entités évaluées² ». Ce décret précise en particulier que la *notation* des unités sera « établie sur la base d'une analyse multicritère³ ».

La mise en pratique de cette recommandation est essentielle pour tous les établissements et organismes de recherche. Cela l'est de façon plus essentielle encore pour les organismes de recherche finalisée (ou spécialisée) qui investissent une part importante de leurs moyens dans des activités différentes de la recherche fondamentale. Une évaluation qui ne prendrait en compte que la seule production de connaissances génériques (ou académiques) rendrait rapidement difficile la prise en charge par les équipes de recherche des activités de communication et de collaboration avec des acteurs sociaux hors de la recherche et, à terme, pourrait remettre en question l'existence même de ces organismes, de même que l'abandon de la mission de production de connaissances fondamentales remettrait en cause leur existence comme organisme de recherche. Mais le besoin d'une évaluation globale⁴, intégrant tous les aspects de l'activité des équipes de recherche ne concerne pas les seuls organismes finalisés ou spécialisés. Comme le montrent des recherches en sociologie de la science, la recherche, aussi fondamentale soit-elle, est caractérisée par des liens forts avec différents groupes sociaux, même celle qui se présente comme la plus fondamentale (Latour 1994 ; Laredo et Mustar, 2000).

La réorganisation du système de recherche français conduit les établissements et organismes à repenser leur rôle dans le dispositif national et à consolider leur méthodes de pilotage scientifique et leur gouvernance. Les nouvelles structures qui se mettent en place ont le même besoin d'explicitation leur stratégie et leurs objectifs ainsi que les outils qu'elles se donnent pour en suivre la réalisation. L'évaluation est un outil important pour la conduite d'un établissement ou d'un institut. La cohérence du système d'évaluation avec les missions et les objectifs stratégiques doit être bien perçue au sein de chaque collectif, mais elle doit être aussi très lisible depuis l'extérieur, en particulier pour les membres des commissions d'évaluation appelées à évaluer leurs différentes entités. La mise en place en mars 2007 d'une agence nationale d'évaluation⁵ renforce ce besoin de lisibilité de la stratégie de chaque entité évaluée par l'Agence et celui de méthodes d'évaluation partagées qui respectent les missions et les spécificités de chacune.

À l'initiative des organismes relevant du programme 187⁶ de la LOLF, un groupe de travail, nommé

1 Loi de programme n° 2006-450 du 18 avril 2006

2 Décret 2006-1334 du 3 /11/06, Art. 1er

3 Décret 2006-1334 du 3 /11/06, Art. 12

4 nos collègues anglo-saxons disent « a comprehensive assessment »

5 l'Aéres : agence d'évaluation de la recherche et de l'enseignement supérieur

6 programme intitulé « Recherche dans le domaine de la gestion des milieux et des ressources »

EREFIN (Évaluation de la REcherche FINalisée) réunit depuis deux ans des personnes en charge de l'évaluation dans une dizaine d'organismes pour travailler en commun sur des outils permettant d'explicitier une stratégie d'organisme aux différents niveaux d'organisation et de façon plus précise au niveau des unités de recherche.

1.2 Une typologie commune pour composer un profil personnalisé pour chaque unité

Le groupe EREFIN s'est d'abord attaché à construire une **typologie des activités** des unités de recherche et de leurs produits. Une première version de cette typologie se présente sous la forme d'un tableau décrivant succinctement une dizaine d'activités-types, leurs produits les plus caractéristiques et une première liste de critères pour en évaluer la qualité.

Dans cette typologie, les activités sont distinguées selon les différentes composantes sociales avec lesquelles l'unité interagit : le monde de la recherche, les acteurs socio-économiques, les pouvoirs publics et l'ensemble de la société. Ce point de vue reprend les idées à la base des représentations classiques que sont la rose des vents de la recherche (Callon *et al.*, 1994) ou les cinq horizons de la recherche (Latour, 1994). En dehors des activités de production, il est apparu souhaitable d'explicitier des activités de type *organisationnel*, qui s'exercent dans le monde de la recherche et dans le monde socio-économique.

En effet, le rôle important des relations partenariales dans les processus d'innovation et la nécessité d'explicitier puis d'évaluer l'activité de construction de partenariats conduit à considérer cette activité en soi et à en mesurer les résultats intermédiaires, traduisant la confiance acquise par les chercheurs auprès des professionnels et la réalité de projets communs. Ainsi les partenariats ne sont pas seulement vus comme un moyen de produire des connaissances mais comme aussi comme un objectif de l'activité. Par exemple, pour l'activité de recherche fondamentale, l'animation de communautés scientifiques, le rôle structurant que peut jouer une unité dans un telle communauté est le résultat non seulement de la publication de très bons articles mais aussi de la construction de projets et programmes nationaux et internationaux, de l'organisation de congrès et séminaires, de l'édition de journaux et revues, résultats dont il est utile de rendre compte.

Ainsi ont été décrites séparément la production de connaissances fondamentales et l'organisation de communautés scientifiques ou aussi la production de connaissances opérationnelles (ou connaissances contextualisées) et la construction de partenariats socio-économiques. Cette distinction permet aux unités de rendre compte de leurs activités à l'échelle de temps des évaluations (quadriennales). En effet, si les interactions de l'unité avec son environnement sont essentielles pour la réalisation de ses missions, le temps pour que ces interactions produisent les résultats escomptés est souvent long : des partenariats durables sont nécessaires pour co-produire des innovations et il est indispensable de pouvoir analyser les processus de coopération sans attendre de disposer de tous leurs effets⁷.

La typologie ainsi construite, intitulée « grille APC » pour Activités-Produits-Critères, comporte ainsi une dizaine d'activités-types (EREFIN, 2007b). Elle est proposée aux unités pour analyser leurs activités et pour définir leur projet pour la période suivante. En effet, la qualité du dossier d'évaluation est un facteur essentiel de l'efficacité du processus d'évaluation et sa contribution à l'amélioration de la qualité de la recherche. Un dossier qui traduit la capacité de l'unité à analyser elle-même sa production puis à réfléchir sa stratégie est un élément essentiel, voire le plus important, de ce qui est produit dans un processus d'évaluation. Des outils d'évaluation bien pensés sont donc au moins aussi utiles pour l'auto-évaluation que pour une évaluation externe. La typologie APC des activités a donc été testée avec quelques unités au cours de l'année 2007 : dans plusieurs cas « à froid » c'est-à-dire hors du contexte d'une évaluation imminente, mais aussi pendant la préparation d'un rapport d'évaluation. Sans qu'une analyse approfondie ait encore été faite de l'usage de cette grille, il est apparu que les intitulés sont compréhensibles et que les 11 activités-types couvrent bien toute la gamme nécessaire⁸.

Cette grille a aussi été utilisée par l'Inra et le Cirad pour tester la faisabilité et l'intérêt de la

7 Plusieurs auteurs en « innovation studies » décrivent le caractère itératif des processus d'innovation vus comme le résultat de nombreuses interactions entre des réseaux sociaux qui constituent des communautés plus ou moins stables de parties prenantes. L'environnement social des chercheurs contribue à la sélection des options possibles des projets et les apprentissages croisés ont un rôle important. Des dispositifs de collaboration et d'interaction au sein desquels ces processus se développent constituent donc déjà un résultat. Enfin, une exploitation des évaluations successives, dans une dynamique d'accompagnement de l'amélioration de la qualité de la recherche, permettrait aussi de revenir sur l'histoire et de faire voir le rôle de ces processus et de ces dispositifs de coopération lorsque des impacts de recherches antérieures sont identifiables.

8 ... étant bien entendu qu'il n'est pas attendu qu'une unité émerge à tous les types, mais bien que toute unité retrouve toutes ses activités dans la liste. L'usage de cette typologie doit inciter à préciser des priorités, une stratégie.

rédaction d'un « contrat d'unité » entre une unité et la direction de l'organisme. Ce contrat explicite en début de période quadriennale, pour chaque type d'activité, des objectifs « opposables » lors de l'évaluation suivante, sous la forme des produits attendus. Dans cette expérience aussi, la typologie des activités et la liste générique des produits décrite dans la grille APC a été jugée adaptée et utilisée.

1.3 Le besoin et l'usage de descripteurs quantitatifs

L'usage du référentiel d'activités est un premier cadre pour analyser et décrire la stratégie de l'unité et plus particulièrement le choix de son profil, c'est-à-dire la répartition de ses moyens selon différentes missions et objectifs. En effet, une unité ne peut pas être impliquée fortement dans chaque type d'activité. Ce n'est ni la réalité, ni un objectif de management cohérent de considérer qu'une « bonne » unité doit être performante sur tous les types de produits de la recherche. C'est au contraire un choix stratégique de définir quelles sont les activités prioritaires d'une unité. Par ailleurs, c'est seulement au niveau d'un département de recherche ou d'un organisme que l'ensemble des missions de l'organisme doivent être remplies.

L'usage de ce référentiel est ainsi une première étape dans l'instrumentation d'une évaluation globale, intégrant l'ensemble des activités. Pour résumer la production et les résultats d'une unité et en suivre l'évolution, des indicateurs sont nécessaires. Des indicateurs pertinents et équilibrés sont alors utiles pour commencer une évaluation : ils donnent aux experts une première image de la production de l'unité et de ses relations avec son environnement social. Cette image et les questions que posent ces indicateurs, doivent les conduire à réaliser une analyse qualitative des résultats replacés dans le contexte de l'unité : son environnement scientifique et institutionnel, sa dynamique.

Une évaluation *intégrative* devrait par ailleurs comprendre plusieurs niveaux. Nous en distinguons trois : i) l'évaluation de la production et des résultats de premier niveau des activités partenariales, prenant en compte tous les produits et résultats, ii) l'analyse de la stratégie mise en oeuvre et la cohérence de la production avec la stratégie de l'unité (équilibre, interactions entre les différentes activités...), iii) les impacts des résultats et de l'activité.

Des indicateurs pour expliciter une stratégie : Pour l'unité qui prépare son évaluation, le calcul de descripteurs quantitatifs ou d'indicateurs selon une règle commune bien explicitée et utilisable pour toutes les unités est une information sur la manière dont des observateurs externes peuvent percevoir leur activité et leur production. Cette confrontation des chiffres avec la propre perception de l'unité de son activité doit susciter un retour sur la stratégie retenue au début de la période évaluée et, si besoin, une inflexion des priorités et un effort de précision des objectifs pour la période à venir. L'utilisation des indicateurs ou de descripteurs quantitatifs, suivis dans le temps, permet aussi de montrer la dynamique de l'unité. Ces descripteurs peuvent donc aussi être utilisés, après la prise en compte de l'évaluation, pour préciser les objectifs convenus entre l'unité et la direction de l'organisme, tels qu'ils pourraient être consignés dans la lettre de mission au Directeur d'unité. Qu'il soit utilisé pour faire un bilan ou pour préciser une stratégie, l'usage d'un jeu complet de descripteurs quantitatifs, reconnu par l'Agence et compris par les experts, serait une garantie que l'évaluation sera cohérente avec les missions et objectifs de l'unité.

Le groupe EREFIN s'est donc défini un objectif de court terme qui consiste à construire un jeu de descripteurs de l'activité et de la production d'une unité. Ce jeu doit être considéré comme une mesure multidimensionnelle et devra donc être utilisé en bloc. Dans une deuxième étape, qui serait réalisée au sein d'un collectif plus large que le groupe⁹, les principaux descripteurs (choisis pour conserver au mieux la multi-dimensionnalité) pourraient être consolidés et normalisés pour être transformés en indicateurs et utilisés aussi à l'échelle de départements de recherche, voire des organismes.

Cet objectif est semblable à la définition d'un *REPP (Research Embedment and Performance Profile)*, qui a été développée dans le projet ERiC¹⁰ par Spaapen et al. (2007). Leur méthode comporte 4 étapes : l'analyse des missions et du profil de l'unité (auto-analyse de l'unité ou analyse par les membres du projet ERiC), l'utilisation de descripteurs et la construction du REPP, le recueil des avis des principaux partenaires ou utilisateurs (stakeholder analysis), un retour de l'unité sur l'analyse des missions et leur réalisation. L'expérience particulièrement bien documentée et analysée du projet ERiC et les bases théoriques qui fondent leur démarche conforte le groupe EREFIN dans son projet de définir des descripteurs, puis une représentation simplifiée du profil de production de chaque unité de recherche. Les deux applications rapportées par Spaapen et al. concernent les sciences agronomiques (Wageningen University and Research Centre) et les sciences pharmaceutiques (Universités d'Utrecht et de Groningen).

Une première représentation du profil de l'unité est celle de l'utilisation de son temps de travail. Les unités qui ont testé la grille APC en 2007 ont jugé utile et réaliste d'estimer le pourcentage de temps total des membres de l'unité selon 6 axes issus des types d'activités, obtenant ainsi une

⁹ le projet est de développer cet objectif au sein de *La Coopérative*

¹⁰ Evaluating Research in Context : A method for comprehensive assessment

première image radar du temps investi par l'unité sur ses différentes missions. D'autres unités ont préféré représenter la répartition de leurs moyens par un histogramme¹¹.

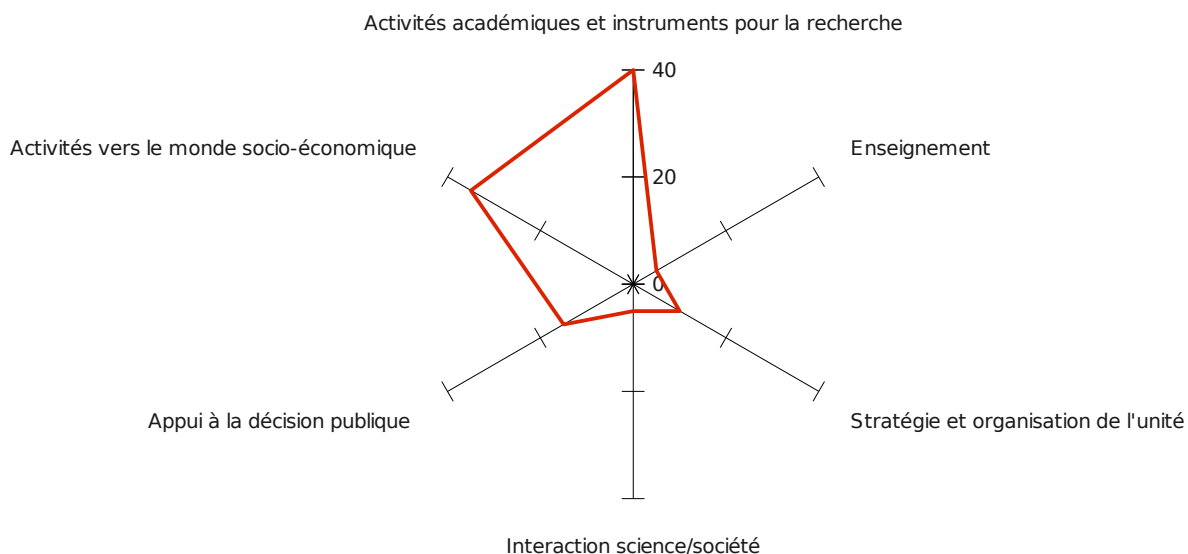


Figure 1 : Une représentation possible de la répartition du temps de travail d'une unité (% par activité)

Une telle représentation peut aussi être utilisée pour montrer les évolutions que l'unité propose au cours de la période suivante. Une deuxième « étoile » peut alors être dessinée sur les mêmes axes.

2 Construction de deux jeux de descripteurs

2.1 Les trois types de descripteurs

Les indicateurs ou descripteurs proposés doivent être une aide à la représentation, puis à l'évaluation de la production et de sa qualité. Un des critères de qualité de la recherche est l'impact des travaux. Lorsqu'il s'agit de l'impact sur les communautés scientifiques et sur le développement de la science, les méthodes de mesure existent et leurs qualités de même que leurs biais sont largement étudiés. Par contre, il est plus difficile de trouver des grandeurs corrélées de façon satisfaisante avec les impacts sur l'innovation et le développement social. Dans les processus d'innovation par exemple, les rôles respectifs des équipes de recherche et des partenaires et acteurs du terrain sont difficiles à démêler. En effet, les équipes de recherche ne sont pas entièrement maîtresses des processus de transfert et d'innovation et l'observation de succès ne rend pas compte du travail réalisé et de compétences d'ingénierie sociale mobilisées pour la conduite de projets au sein desquels doivent collaborer des acteurs ayant des objectifs, des contraintes et des cultures différents. Enfin, l'échelle de temps pour évaluer ces impacts est en général plus longue que ce qui serait nécessaire au pilotage à court et moyen terme. L'importance des dispositifs de coopération dans ces processus conduit à porter une partie de l'analyse sur la nature des interactions et les activités des chercheurs pour construire des partenariats durables avec des acteurs socio-économiques ou institutionnels.

Aux descripteurs de premier niveau que sont les dénombrements de produits : nouvelles connaissances véhiculées par des textes (publications et brevets), des logiciels, des outils d'aide à la décision..., synthèses et mises en perspectives de connaissances (comme les rapports d'expertises collectives) et qu'on qualifiera de *descripteurs de production*, le groupe tenté d'ajouter

¹¹ Dans son étude des unités de recherche en sciences pharmaceutiques, J. Spaapen utilise des représentations plus simples (en « camemberts ») en demandant à l'unité son investissement dans les 3 grands domaines académique, industriel et sociétal (incluant dans ce dernier l'appui à la décision publique et la formation). Il compare ensuite le temps consacré à chaque domaine, l'auto-estimation par l'unité de son influence dans chaque domaine avec son analyse du nombre et de l'importance des partenaires de l'unité.

des *descripteurs d'interaction* et des *descripteurs d'impact*. Plus précisément,

- les **descripteurs de la production** mesurent la quantité (ou l'existence) de produits de différents types : les produits comptés dans ces indicateurs sont ceux qui apparaissent dans la liste des produits du rapport de l'unité ;
- les **descripteurs des interactions**¹² permettent d'approcher l'investissement de l'unité dans des relations partenariales, dans la construction de réseaux scientifiques ou avec des professionnels,
- les **descripteurs d'impact** permettent d'aborder la question de la qualité des résultats au regard d'objectifs de niveau plus général liés aux missions des établissements et des organismes. Une des questions est de mesurer l'utilisation des résultats scientifiques par les destinataires. Il est sans doute difficile pour l'unité d'apporter toutes les preuves de cette utilisation dans un pas de temps compatible avec l'évaluation d'un bilan quadriennal. Mais tenter de renseigner certains indices constitue néanmoins un moyen pour l'unité de s'interroger sur l'impact réel et potentiel de ses travaux. Cette analyse devrait ensuite contribuer à la définition de sa stratégie et à celle de l'organisme.

Un premier jeu de descripteurs, dénommés *descripteurs de niveau 1*, est un résumé quantitatif des productions de l'unité. Ces descripteurs de production ont donc été construits à partir de la liste des produits associés à chaque type d'activité. Ils seront résumés par 5 à 11 nombres (selon la diversité des partenaires de l'unité) qui donnent une première image simple du profil d'activités, construite à partir des résultats facilement identifiables. Une deuxième jeu de descripteurs, dits *descripteurs de niveau 2*, comprendra d'autres descripteurs d'interaction ainsi que des descripteurs d'impact.

2.2 La liste des productions scientifiques de l'unité

La liste des « productions scientifiques » de l'unité demandée dans le rapport d'évaluation doit être plus complète et structurée. En cohérence avec la typologie APC, cette liste serait organisée en 11 rubriques comme suit :

Liste des produits et résultats

1. Connaissances fondamentales : Publications scientifiques et autres vecteurs de diffusion
2. Partenariats scientifiques, animation des communautés scientifiques nationales et internationales
3. Développement de grands instruments pour la recherche
4. Partenariats avec des acteurs socio-économiques
5. Connaissances opérationnelles et produits pour des acteurs socio-économiques
6. Expertises et études pour des acteurs socio-économiques
- 4'. Partenariats avec des acteurs publics (hors recherche)
- 5'. Connaissances opérationnelles et produits pour des acteurs publics
- 6'. Expertises et études pour des acteurs publics
7. Enseignement et formation
8. Veille sociétale et diffusion de la culture scientifique

Ce qui devrait apparaître dans chacune des rubriques est précisé dans un document annexe. En fonction de son profil, **l'unité utilisera les rubriques pertinentes au regard de ses productions** et structurera chacune de ces rubriques en autant de sous-rubriques qu'elle le jugera utile.

¹² le mot anglais est souvent *embedment*

2.3 Les descripteurs de niveau 1

Le jeu de descripteurs en cours de définition par EREFIN vise à pouvoir s'appliquer à une grande diversité d'unités de recherche évaluées par l'Aéres. Ce jeu est un résumé quantitatif des produits présentés dans la *Liste des produits et résultats*. Leur comptage est donc relativement rapide à partir de cette liste. À partir de ce tableau présenté dans le dossier de l'unité, les experts pourront moduler ces comptages des produits en fonction de leur qualité et définir une nouvelle valeur de ces descripteurs. Cette analyse représentera une première étape de leur évaluation.

Ces comptages sont résumés par type de destinataires sous la forme de sommes pondérées. Les coefficients de pondération ont été déterminés avec l'expertise du groupe¹³. Ils seront revus après un premier test sur une vingtaine ou une trentaine d'unités des différents organismes du groupe.

En aucun cas, ces descripteurs ou leur résumé ne doivent être compris comme des critères d'évaluation : chacun des nombres doit s'analyser par rapport à des objectifs et l'optimum pour une unité ne peut être d'avoir le chiffre maximal sur chaque dimension. Une colonne *Niveau d'atteinte des objectifs* permet à l'unité de rapporter ces résultats aux objectifs fixés et l'incite ainsi à une analyse de ces résultats et à une critique constructive de l'usage de ces descripteurs.

Tableau 1 : Résumé des descripteurs de niveau 1 (voir en annexe le Tableau 2 de ces descripteurs)

	Type de résultats	Descripteur synthétique (de niveau 1)	Niveau d'atteinte des objectifs
	Résultats vers le monde académique	SA	
1	Connaissances fondamentales : publications scientifiques et autres vecteurs de leur diffusion	S1	
2	Partenariats scientifiques, animation de communautés scientifiques nationales et internationales	S2	
3	Développement de grands instruments pour la recherche	S3	
	Résultats vers le monde socio-économique	SB	
4	Partenariats avec des acteurs socio-économiques	S4	
5	Connaissances opérationnelles et produits pour des acteurs socio-économiques	S5	
6	Expertises et études pour des acteurs socio-économiques	S6	
	Résultats pour les pouvoirs publics	SC	
4'	Partenariat avec des acteurs publics (hors recherche)	S'4	
5'	Connaissances opérationnelles et produits pour des acteurs publics (hors recherche)	S'5	
6'	Expertises et études pour les acteurs publics	S'6	
	Résultats pour l'enseignement et la formation	SD	
7	Enseignement et formation	S7	
	Résultats relatifs à l'interaction avec la société	SE	
8	Veille sociétale, veille industrielle et technologique (nationale et mondiale) ; diffusion de la culture scientifique et technique	S8	

Les descripteurs synthétiques permettent de construire une image simplifiée sous forme d'un histogramme en barres. Nous préférons en effet cette représentation à une image radar dans laquelle l'ordre des axes influe sur la surface qu'on peut être tenté d'interpréter, à tort, comme représentant la production totale¹⁴.

¹³ ...alors que Spaepen les calcule à partir de la distribution de leurs valeurs sur la « population » des unités de chaque étude.

¹⁴ En particulier si la mesure sur certains axes est nulle, la surface peut se trouver très réduite même si d'autres contributions sont très importantes.

Produits et résultats par type de destinataires

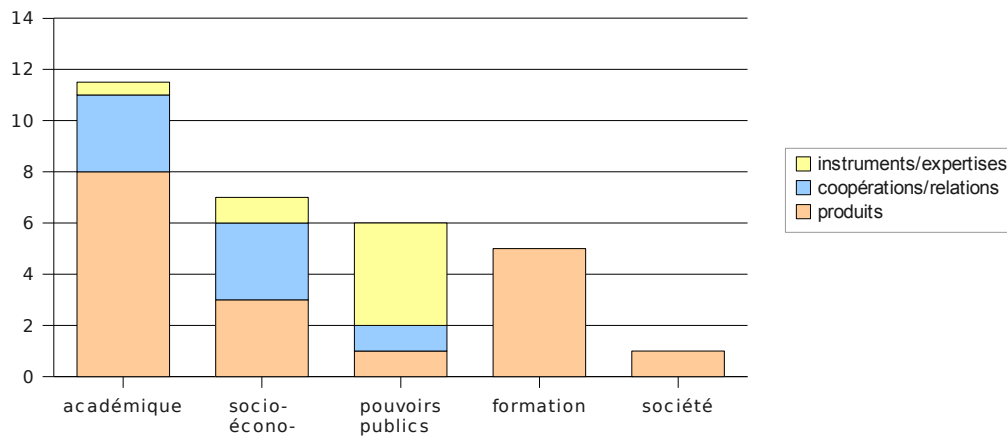


Figure 2 : Représentation de la répartition des produits et des résultats de l'unité, par type de destinataires

2.4 Quelques pistes pour construire des descripteurs de niveau 2

Le jeu de descripteurs de niveau 2 a deux objectifs : compléter les descripteurs de niveau 1 pour l'analyse des interactions de l'unité avec ses différents partenaires et réfléchir sur les impacts à plus long terme. C'est ici qu'on peut regarder les vecteurs des interactions que sont, comme le disent Callon et ses co-auteurs (1992), les textes, les personnes, les artefacts et l'argent. Ces descripteurs sont d'autre part une incitation à analyser les impacts de la recherche, son influence sur l'action des partenaires et des destinataires finaux des recherches conduites dans l'unité.

Ces descripteurs de niveau 2 seraient utilisés par les unités en choisissant ceux qui sont pertinents en fonction du contexte et de l'histoire de l'unité. Ils n'ont pas un objectif de synthèse d'un profil mais d'aide à l'analyse des résultats à plus long terme.

Ils sont encore moins utilisables comme une mesure de la qualité que les descripteurs de niveau 1, parce que d'une part ils sont difficiles à obtenir et à valider (ce sont seulement des indicateurs au sens courant du terme c'est-à-dire des indices) et parce que les effets ou les impacts dépendent en grande partie des partenaires et des destinataires.

La construction d'un jeu de descripteurs de niveau 2 est en cours. Ce travail fera lui aussi l'objet de test auprès d'un panel d'unités, mais aussi de collaborations avec des spécialistes de la construction d'indicateurs et des chercheurs dont c'est le domaine de recherche.

2.5 Pour aller plus loin : des questions de méthode

Plusieurs questions de méthode sont importantes pour la construction d'indicateurs. Les descripteurs proposés ici étant de possibles précurseurs d'indicateurs, il convient de les aborder dès maintenant.

- **la validation des données** : où sont-elles disponibles ? qui les valide ? cette question est fondamentale pour la définition d'indicateurs dignes de ce nom. Pour ce qui concerne les données servant à calculer les descripteurs de production, les produits devront tous être dans la liste des produits du dossier d'évaluation. Les références pour vérifier leur existence devraient être, dans la mesure du possible, apportées par l'unité dans le dossier et les documents disponibles sur place pour le comité de visite. Lors de l'évaluation, les experts seraient invités à prendre connaissance des descripteurs calculés par l'unité : ils valideraient ou corrigeraient les descripteurs de production au vu des références des produits (dans le dossier) ou de leur examen lors de la visite. Les données qui interviennent dans les descripteurs d'interaction sont en partie des données administratives (flux de personnes, recettes) et sont en général déjà présentes, sous une forme éventuellement différente, dans la partie administrative du dossier de l'unité. Ces données devraient pouvoir être vérifiées dans les systèmes d'information des organismes. Enfin, pour les indicateurs d'impact comme par exemple les citations des travaux de l'unité dans des textes réglementaires, la vérification de chaque fait rapporté est quasi

impossible dans une procédure d'évaluation d'unité sauf si les experts peuvent recouper les affirmations de l'unité avec des informations qu'ils auraient par ailleurs. Les indicateurs d'impact sont donc beaucoup moins fiables et seront analysés avec circonspection ;

- **la normalisation des comptages** : Si on souhaite effectuer des comparaisons entre unités, il faut prendre en compte l'effet taille de l'unité et par exemple rapporter chaque indicateur quantitatif à une grandeur de référence qui prend en compte cette taille. La grandeur de référence la plus naturelle est le nombre des personnels de l'unité ou le budget à coûts complets de l'unité pour les données financières. Mais il faut éviter de donner une impression injustifiée de précision à de tels quotients/rapports. Dans une première approche, nous proposons de ne pas faire de normalisation, mais de mettre à disposition des experts des tableaux résumant les moyens de l'unité, tableau à regarder en parallèle de celui des descripteurs.
- **la période de temps** sur laquelle les données sont enregistrées : la date de l'évaluation, très anticipée par rapport à la fin de la période quadriennale contractuelle, pose la question de l'intervalle de temps sur lequel sont recueillies les données. Or cette durée doit être très explicite. Selon l'histoire de l'unité et le descripteur calculé, la durée pertinente est soit les 4 années¹⁵ précédant le dépôt du dossier, soit dans le cas d'une création récente d'unité, la période écoulée depuis le début du contrat quadriennal. L'unité qui remplit le tableau précisera cette durée en commentaire en début du tableau des valeurs des descripteurs ;
- **le grain de la structure sur laquelle sont calculés ces descripteurs** : il n'est déjà pas habituel de calculer des indicateurs à l'échelle des laboratoires et les problèmes que pose l'interprétation de petits nombres doivent conduire à regarder ces descripteurs avec beaucoup de précaution dans le cas de petites structures, voire à ne pas les utiliser ;
- **la question de la couverture** et de l'équilibre du nombre de descripteurs par type d'activité de ce jeu de descripteurs reste à approfondir au fur et à mesure des mises en test et retours d'analyse : les descripteurs de production permettent-ils d'approcher les questions qu'on doit se poser sur la production ? que manque-t-il aux descripteurs d'interaction pour mieux appréhender l'implication de l'unité dans son environnement ? les descripteurs d'impact sont-ils utilisables ou à quelle conditions et quels autres aspects des impacts devraient être dépistés ?

Tableau 3 : Grandeurs de référence pour interpréter (ou pondérer) les descripteurs dans le calcul des descripteurs de niveau 1

Notation	Intitulé de la grandeur de référence	Valeur
	Période retenue :	
E1	Nombre de chercheurs, enseignants-chercheurs, ingénieurs de recherche et équivalents (personnels titulaires et non titulaires) Total compté en années	
E11	Nombre d' enseignants-chercheurs	
E12	Nombre de directeurs de recherche, professeurs et membres de l'unité habilités à diriger des recherches (dont les CR1 avec HDR), à la date de dépôt du dossier	
E2	Nombre des personnels non comptés dans E1	
F1	Budget annuel moyen (2 dernières années) de l'unité à coûts complets	
F2	Budget annuel moyen (2 dernières années) hors salaires des personnels permanents	

2.6 Conclusion très provisoire...

Au cours de l'année 2009, le groupe EREFIN va donc poursuivre ce travail de construction de ces jeux de descripteurs en travaillant avec des unités afin de confronter cette proposition aux réalités du « terrain ». Cette étape conduira peut-être à une évolution importante de cette première version.

Dès maintenant, nous invitons tout établissement ou toute structure intéressé par cette approche à analyser ce travail et à s'en saisir, en retenant cependant les trois principes qui l'ont fondé :

¹⁵ à cause des calendriers de l'évaluation, il ne s'agit donc pas de 4 années *civiles* et cela pose un problème pour la comparaison (ou l'extraction) de ces données avec ce qui est enregistré dans les bases de données

- cet outil est destiné aux établissements, aux unités et aux autres collectifs de recherche afin de les **aider à définir leur stratégie** puis à suivre dans le temps les résultats obtenus dans une démarche d'auto-évaluation ;
- il est conçu pour une **gamme très large d'unités**, y compris au-delà des unités des organismes de recherche finalisée, car il est flexible et permet de décrire chaque construction particulière, associant de façon originale différentes activités au cours d'une histoire. En général, toutes les activités et tous les types de résultats ne seront pas présents pour une unité à une évaluation donnée. Même des unités très spécialisées peuvent utiliser ces descripteurs, en ne gardant que ceux qui sont pertinents par rapport à leurs missions et leur stratégie. Ces unités seront par contre légitimes de compléter par des descripteurs plus détaillés de leur « rubrique » ;
- cet outil permet une description grossière d'un profil et est une introduction à l'analyse d'un bilan mais il n'est **pas une grille de notation**. Une évaluation digne de ce nom, et la notation qui éventuellement la résume, n'a rien à voir avec un algorithme qui transformerait ces descripteurs en notes, et ceci même si les objectifs sont définis sous la forme de pondération entre les différentes rubriques.

Le groupe met donc cet outil à la disposition de tous les acteurs de la recherche, mais il en critique par avance toute utilisation contradictoire avec ces principes.

3 Références

Callon M., Larédo P., Mustar P. (1994) Panorama de la science française, *La Recherche* 25, 264, 378-384

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

Latour B. (1994, 2ème édition 2006) *Le métier de chercheur ; regard d'un anthropologue*, QUAE Éditions, Série Sciences en questions

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

Dejours C. (2003) L'Évaluation du travail à l'épreuve du réel. Critique des fondements de l'évaluation, *INRA Éditions, Sciences en questions*

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

Jack Spaapen, Huub Dijstelbloem and Frank Wamelink (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_73VHUD_Eng

EREFIN (2007a) *L'évaluation des unités de recherche : propositions pour la prise en compte de l'ensemble des missions et des activités de la recherche finalisée*

EREFIN (2007b) *Les activités élémentaires des unités de recherche : la grille APC*

EREFIN (2008) *Évaluation des unités : proposition d'une grille d'évaluation pour les experts*. Note du groupe EREFIN destinée à l'Aéres (6p)