

Les dérives de l'évaluation de la recherche, du bon usage de la bibliométrie

Quelques repères autour de la conférence d'Yves Gingras.

I - Autour de la bibliométrie

Traitements basés sur l'utilisation de méthodes statistiques et mathématiques, appliqués à des corpus de documents, plus particulièrement des articles de revues ou des brevets. Elle permet d'alimenter des indicateurs bibliométriques.

ADBS (Association des professionnelles de l'information et de la documentation - www.adbs.fr)

L'ensemble des procédures qui contribuent à l'évaluation de la production scientifique d'un chercheur (ou d'un ensemble de chercheurs) à partir du nombre de ses publications, du prestige des revues dans lesquelles elles ont été faites, et des citations auxquelles elles ont donné lieu. En tout état de cause, la bibliométrie ne mesure pas la qualité d'un chercheur mais seulement ses citations, sans préjuger des raisons qui l'ont amené à être cité.

Du bon usage de la bibliométrie pour l'évaluation individuelle des chercheurs.

Rapport de l'Académie des sciences, janvier 2011 (p.9).

« La bibliométrie tend aujourd'hui à saturer l'évaluation ». Pansu P., Nicole Dubois N., Dubois J.L. *Dis-moi qui te cite, et je saurai ce que tu vauds ; que mesure vraiment la bibliométrie ?*

II - Une remise en question généralisée de l'évaluation de la recherche

- Du bon usage de la bibliométrie pour l'évaluation individuelle des chercheurs

Bach J.F., Jérôme D., 2011. *Synthèse du rapport de l'Académie des sciences.*

Texte original en ligne sur [http : //www.academie-sciences.fr](http://www.academie-sciences.fr)

L'évaluation bibliométrique est fondée sur l'utilisation des citations des articles scientifiques afin d'estimer leur influence et dans l'idée de donner une mesure de leur "qualité". Elle est rendue possible par l'existence de bases de données informatiques qui permettent, pour un chercheur donné, un calcul d'indicateurs à partir des citations de tous ses articles publiés.

La bibliométrie est largement utilisée dans de nombreuses disciplines pour l'évaluation des individus ou des institutions. Néanmoins, cette utilisation fait débat avec des avis très tranchés, positifs ou négatifs selon les disciplines considérées. C'est dans ce contexte que l'Académie des sciences a souhaité étudier le problème en tenant compte le plus précisément possible des différences culturelles et des pratiques de chacune des disciplines sur le sujet.

L'outil bibliométrique doit son succès à son apparente simplicité d'utilisation, contrastant avec la lourdeur et les limites de l'évaluation qualitative par les pairs. Il associe des mesures quantitatives à chaque chercheur, une façon attrayante de fournir une première évaluation rapide fondée sur des données factuelles et objectives. De fait, la bibliométrie peut être utile pour une instance d'évaluation afin de réaliser un premier tri et se faire une opinion immédiate.

Néanmoins, l'outil bibliométrique, utilisé de façon inadéquate, expose à des biais sérieux. En premier lieu, cet outil ne remplace en rien l'évaluation qualitative par les pairs, mais représente surtout un instrument intéressant pour les

évaluateurs. Seule l'évaluation qualitative peut apprécier l'originalité et l'esprit d'innovation d'un chercheur. Cette réserve majeure implique que la bibliométrie ne doit être utilisée que par les pairs, et en aucun cas par des évaluateurs sans connaissance de cet outil, ou n'ayant pas examiné les dossiers scientifiques dans leur ensemble et avec le temps nécessaire à ce travail.

Sur un plan technique, l'outil bibliométrique n'a de valeur que si certaines conditions d'utilisation sont respectées. Tout d'abord, les calculs des indices bibliométriques doivent avoir été correctement réalisés. De plus, dans les disciplines où l'ordre des auteurs n'est pas alphabétique mais tient compte de la contribution de chacun d'entre eux, la place de chaque auteur dans un article doit être prise en considération, ce qui n'est habituellement pas comptabilisé. Surtout, et ce point est crucial, il n'est pas possible de comparer les indices bibliométriques de chercheurs appartenant à des disciplines ou même des sous-disciplines différentes. Les valeurs des indices bibliométriques varient en effet beaucoup selon les disciplines, au point qu'il doit être fortement recommandé de se référer, pour chaque discipline, à la distribution des indices autour de la moyenne, en tenant compte de l'âge des chercheurs.

- Revues et évaluation, un monde de classements

Pontille D. et Torny D., 2011. *Carnets de bord*, n°17, Ivry, Inra, UR Ritme.

Plusieurs listes et classements de revues ont progressivement vu le jour dès les années 1970 (...) Souvent centrés sur des enjeux académiques, ces classements de revues n'étaient pas exempts de préoccupations gestionnaires, à l'image du Science citation index (Wouters, 1999) ou des incitations financières associées à ceux des écoles de commerce. (...) Le caractère inédit des classements de revues contemporains réside davantage dans leur prolifération en tant qu'instrument de politique publique.

- Slow Science et désexcellence

Olivier Gosselain, professeur à l'université libre de Bruxelles formation d'archéologue, spécialisé en ethnographie des techniques

<http://www.sisyphes.upmc.fr/~lemoine/docs/La%20Desexcellence.pdf>

Le mot « excellence », qui revient sans arrêt, n'est absolument pas propre au monde académique. Il relève d'une méthode de management qu'on a vu émerger dans le privé au début des années 1980, mais qui a été transférée progressivement dans les services publics d'une façon générale et à l'université en particulier. Elle repose sur une notion d'accroissement constant et illimité de la qualité (dans un contexte d'hyper-compétition), avec comme corollaire, une évaluation permanente et un souci de gouvernance et de transparence, au nom de la nécessité de rendre des comptes au public – justification numéro un.

Le hasard a fait que nous avons été soumis quelques temps plus tard à une évaluation de laboratoire. L'outil destiné à mesurer nos performances était un formulaire standard, sorte de canif suisse du coaching en entreprise, avec son inévitable analyse SWOT¹. C'était déconcertant de naïveté et presque embarrassant à remplir. Mais à une question au moins, portant sur notre conception de la recherche, nous avons apporté une réponse sincère : plaisir et créativité. Ce fut le point de rupture pour les duettistes en costume sombre qui pilotaient le groupe d'évaluation. De tels termes, nous affirmaient-ils, étaient tout simplement inacceptables. Indignes de figurer sur un formulaire d'évaluation et preuves évidentes de notre manque de sérieux.

- « Le problème est la prétendue objectivité du chiffre »

Barbara Cassin B. Entretien paru dans *Le monde des livres*, 18/04/2014.

Je ne suis pas contre l'évaluation, mais contre un chiffrage généralisé (...) Ce qui me paraît important, c'est de savoir qui donne les critères, pourquoi, à quel moment. Le mouvement est à toujours plus d'évaluation appliquée à des choses toujours moins évaluables.

¹ « Strengths, Weaknesses, Opportunities, Threats ». C'est sur cette base que s'élaborent les plans stratégiques dans les milieux d'affaire et, depuis quelques années, dans les universités.

III. Qu'est-ce qu'un bon chercheur ?

Blog d'Antoine Blanchard, sociologue des sciences

<http://www.enroweb.com/blogsociences/index.php?2009/05/15/395-qu-est-ce-qu-un-bon-chercheur>

« Le bon chercheur il publie mais le mauvais chercheur il publie aussi ». Olivier Le Deuff

- Hypothèse 0 : le bon chercheur c'est celui dont les pairs disent qu'il est un bon chercheur

Ca, c'est le schéma classique, le raisonnement pré-bibliométrique. On part du principe que seuls les pairs peuvent évaluer une recherche, dire si un chercheur est bon ou pas et si ce qu'il fait a 5 ans d'avance ou 10 ans de retard. Sauf qu'il faut pour cela des pairs bien informés sur l'état actuel de la recherche, si possible mondiale. Heureusement, on en trouve. Ensuite, il faut qu'ils puissent juger et rendre un verdict le plus objectif possible, sans être "parasités" par des considérations extérieures. C'est plus facile si l'on fait appel à des tiers neutres. Mais alors il faut qu'ils puissent se projeter dans la recherche qu'ils évaluent, qu'ils aient le temps d'en lire les articles et d'en saisir toutes les dimensions. Appliquez ça notamment aux SHS où la diversité des questions de recherche, des cas étudiés ("terrains") et des cadres théoriques fait que chaque chercheur travaille à peu près tout seul. C'est très difficile et imparfait. Mais prenez aussi les sciences dures où, comme l'expliquait Pierre Joliot, ce qu'un chercheur considère comme ses meilleurs articles sont souvent les plus originaux, les plus novateurs. Un pair évaluateur qui passerait en revue sa bibliographie s'arrêterait-il sur cette poignée d'articles encore incompris ou bien considérerait-il que les autres sont les plus marquants ? Bien souvent, la recherche innovante et fertile est inévaluable au moment où elle se fait...

- Hypothèse 1 : le bon chercheur c'est celui qui publie beaucoup

Comme l'ont montré Latour et Woolgar dans *La vie de laboratoire*, la publication d'articles est au cœur de l'activité du chercheur : les acteurs reconnaissent que la production d'articles est le but essentiel de leur activité. La réalisation de cet objectif nécessite une chaîne d'opérations d'écriture qui vont d'un premier résultat griffonné sur un bout de papier et communiqué avec enthousiasme aux collègues, au classement de l'article publié dans les archives du laboratoire. Les nombreux stades intermédiaires (conférences avec projection, diffusion de tirés-à-part, etc.) ont tous un rapport sous une forme ou sous une autre avec la production littéraire.

Le bon chercheur, ce serait donc celui qui produit des résultats et qui arrive à publier beaucoup. Sauf que le paysage des revues scientifiques est un peu le monde des Bisounours et toute recherche (y compris de mauvaise qualité) peut se publier, même en passant par le filtre des "rapporteurs" — d'où la citation d'Olivier Le Deuff reproduite en-tête. Qui plus est, il est souvent facile de saucissonner son travail en un maximum d'articles, d'avoir quelques signatures de complaisance ou de participer à un programme de recherche en physique des hautes énergies qui vous assure une présence au firmament des auteurs.

- Hypothèse 2 : le bon chercheur c'est celui qui est cité

Finalement, et c'est le principe de base de l'analyse des citations, un chercheur qui cite un autre chercheur donne une accolade qui prouve que l'article a été remarqué, qu'il a eu une vie après la publication. C'est la seule chose qu'on puisse affirmer avec certitude, mais la citation est ce qui se rapproche le plus d'une monnaie d'échange du capital scientifique et par extension de la qualité d'un chercheur. On peut donc penser que le bon chercheur c'est celui qui est cité. Mais que penser des articles frauduleux ou rétractés qui continuent d'être cités, des auto-citations qui permettent d'augmenter son score tout seul ou des citations qui viennent d'articles de seconde zone ? C'est pour ces raisons que les analyses de citation s'appuient essentiellement sur les données de Thomson Reuters (Science Citation Index), qui a des critères stricts d'inclusion des revues et de calcul des scores de citation. Mais cette base de données a un fort biais vers les revues anglo-saxonnes et ses critères de scientificité ne sont pas forcément partagés par tout le monde.

- Hypothèse 3 : le bon chercheur c'est celui qui publie beaucoup et qui est cité

Que se passe-t-il si l'on combine deux qualités que devrait posséder un bon chercheur : publier beaucoup et être cité ? On obtient un indicateur composite, qu'Yves Gingras qualifie d'hétérogène, comme le nombre moyen de citations par article ou l'indice h. Avec cet indice, on peut dire qu'un chercheur A qui a publié trois articles cités soixante fois (indice $h = 3$) est moins bon qu'un chercheur B ayant publié dix articles cités onze fois (indice $h = 10$). Mais est-ce que cela traduit bien la réalité ? Yves Gingras, dans sa note sur "La fièvre de l'évaluation de la recherche" qui vient d'être reprise dans le numéro de mai de La Recherche, écrit que non. Que le chercheur A n'est pas, en réalité, moins bon que B.

C'est ce point particulier que je voudrais analyser plus en détail. Vaut-il mieux favoriser celui qui a publié beaucoup et qui a réussi que chacun de ses articles soit tout de même remarqué ou celui qui a peu publié et qui a été très remarqué ? La réponse n'est pas évidente mais j'entends, au fond de la salle, que le chercheur parcimonieux A doit être préféré. C'est en effet la réponse classique, pas tant parce que son total (180) est supérieure à celui du chercheur B (110) que parce que sa fulgurance et sa brillance nous impressionnent. Mais la bibliométrie a mis en évidence l'effet Matthieu selon lequel on donne plus à ceux qui ont déjà. Et donc qu'il est plus facile de recevoir sa 60e citation quand on est déjà cité 59 fois que de recevoir sa 11e citation quand on peine à se faire remarquer. (...) Mais je voulais montrer par cet exemple que les outils de la bibliométrie offrent des pistes de réflexion et des débuts de réponse, qui peuvent être contre-intuitifs.

- Hypothèse 4 : le bon chercheur c'est celui qui ne fait pas comme les autres

Cette dernière hypothèse est de moi. Elle se veut un peu provocatrice mais n'est sans doute pas si loin de la réalité. Déjà, elle voit le chercheur au-delà de son activité de publication et inclut son rôle de passeur, de communicateur... Et surtout, c'est un plaidoyer pour la diversité en science. (...) Je ne dis pas que tous les chercheurs se valent (...) mais il faut aussi accepter que tous les chercheurs ne se ressemblent pas et qu'ils ne soient pas facilement "benchmarkables".

IV - Que fait l'Inra ?

- L'Inra a participé au groupe EREFIN (groupe de travail inter-établissements sur l'évaluation de la recherche finalisée)

Extrait du rapport « Evaluation des collectifs de recherche : un cadre qui intègre l'ensemble des activités », 2010.

La nécessité d'une évaluation contextualisée

Pour un établissement de recherche finalisée la mission est double : faire progresser la connaissance et répondre à des questions scientifiques suscitées par des enjeux de société et leurs prolongements opérationnels. Les recherches conduites visent à répondre à des problèmes issus de la pratique de différents acteurs socio-économiques. Cet objectif, qui impose le plus souvent un détour par de la recherche fondamentale avec une production de connaissances génériques, implique pour l'institution ou le collectif de recherche de construire des interactions fortes à la fois avec des partenaires scientifiques et avec des partenaires socio-économiques ou des utilisateurs des résultats de la recherche. Ces interactions interviennent dans les différentes étapes de définition d'une stratégie de recherche, de production des connaissances, de transfert et d'utilisation pour l'action et l'innovation. L'évaluation doit donc prendre en compte et analyser les processus mis en œuvre dans le développement de ces différents partenariats.

Plus généralement, l'évaluation d'un collectif de recherche doit être conduite en cohérence avec l'ensemble de ses missions et de l'environnement dans lequel elles s'exercent ainsi que du mode d'organisation et de collaboration choisi pour la réalisation des objectifs. Ce besoin d'une évaluation globale (*comprehensive assessment*) n'est pas restreint aux seuls établissements de recherche finalisée ou spécialisée. Comme l'ont montré des recherches en sociologie des sciences, les liens forts avec différents groupes sociaux caractérisent la recherche dans son ensemble (Latour 1994 ; Laredo et Mustar, 2000). Enfin, la loi de programmation de la recherche de 2006 est explicite sur la diversité des activités de la recherche qui concourent à la réalisation de ses missions envers la société. En conséquence, quel que soit l'équilibre entre des activités d'expertise, d'aide à la décision, de formation, de contribution à l'innovation, les outils

utilisés pour l'évaluation d'un collectif de recherche doivent permettre d'examiner la réalité et la qualité de chacune de ces composantes.

Conclusion

Alors que les indicateurs de production de connaissances sont utilisés depuis longtemps (et que l'analyse de leurs usages et mésusages est bien documentée), le caractère très partiel ou très expérimental des indicateurs existants pour les autres activités de recherche incite à poursuivre les travaux qui permettront de concevoir des indicateurs de qualité sur ces activités.

-- **NORIA** (<https://intranet.noria.inra.fr>)

NORIA est un outil d'aide à la définition d'une stratégie de publication et à l'évaluation des productions scientifiques créé par l'Inra.

Bibliographie

- Hervé Rostaing H., 1996. *La bibliométrie et ses techniques*. Paris, Sciences de la Société/CRRM.
- Stengers I., 2011. *Another science is possible! A plea for slow science*. Bruxelles, ULB, faculté de philosophie et lettres.
http://three rotten potatoes.files.wordpress.com/2011/06/stengers2011_pleaslowscience.pdf
- *Déclaration de San Francisco sur l'évaluation de la recherche*, 2013. <http://am.ascb.org/dora/>
- Moati A., 2013. Du rififi dans l'édition scientifique. *Tangente*, n° 153, juillet-août, pp. 38-39.
- Pansu P., Dubois N., Dubois J.L., 2013. *Dis-moi qui te cite, et je saurai ce que tu vauds : que mesure vraiment la bibliométrie ?* Grenoble, Presses universitaires de Grenoble.
- Volland-Nail P., 2013. *Stratégies de publication scientifique*. Quae (publication électronique).
- Gingras Y. (entretien). *Faire de la sociologie des sciences avec un marteau ? Science et éthique en action*. Partie 1 : <http://zilsel.hypotheses.org/246> (2013) et partie 2 (2014). <http://zilsel.hypotheses.org/284>
- Pumain D., Dardel F., 2014. *L'évaluation de la recherche et de l'enseignement supérieur. Rapport à madame la ministre de l'Enseignement supérieur et de la Recherche*.
- Cassin B., 2014. *Derrière les grilles, sortons du tout évaluation*. Paris, Édition Mille et une nuits.
- Gingras Y., 2014. *Les dérives de l'évaluation de la recherche et développement au bon usage de la bibliométrie*. Paris, éditions Raisons d'agir.
- *Rapport 2014 de l'OST*.
http://cache.media.enseignementsup-recherche.gouv.fr/file/Actus/98/8/Rapport_Pumain_Dardel_295988.pdf

Ce document est proposé par le groupe Sciences en questions comme éclairage autour du sujet qui sera développé par Yves Gingras lors de ses conférences.

Coordination : Michelle Cussenot.