

Ce qu'il y a de bien (ou de mal) dans l'évaluation bibliométrique telle que pratiquée par l'AERES

NICOLAS HOLZSCHUCH

Institut national de recherche en informatique et en automatique (Inria)
nicolas.holzschuch@inria.fr

Depuis 1997, **Nicolas Holzschuch** est chargé de recherche à l'Institut national de recherche en informatique et en automatique (Inria) Rhône-Alpes. Il est membre de la commission d'évaluation de l'Inria. Il est aussi l'auteur du blog « Une odyssée à Ithaca », <http://n.holzschuch.free.fr/dotclear/>

* Extrait de *Full Metal Jacket*, de Stanley Kubrick, 1987.

- Marine, what is that button on your body armor?
- A peace symbol, sir.
- Where'd you get it?
- I don't remember, sir.
- What is that you've got written on your helmet?
- « Born to kill », sir.
- You write « born to kill » on your helmet and you wear a peace button. What's that supposed to be, some kind of sick joke? [...]
- I think I was trying to suggest something about the duality of man, sir.
- The what?
- The duality of man. The jungian thing, sir.

- Soldat, c'est quoi ce bouton sur votre gilet pare-balles ?
- Un symbole de paix, chef.
- Où l'avez-vous eu ?
- Je ne m'en souviens pas, chef.
- C'est écrit quoi sur votre casque ?
- « Né pour tuer », chef.
- Vous écrivez « né pour tuer » sur votre casque et vous portez un badge de la paix. Ça veut dire quoi ? C'est une blague de malade ? [...]
- Je pense que j'essayais de suggérer quelque chose à propos de la dualité de l'homme, chef.
- La quoi ?
- La dualité de l'homme. Le truc de Jung, chef*.

Bref rappel des faits

L'AERES¹, c'est l'agence chargée d'évaluer le système universitaire français (voir encadré page ci-contre) : laboratoires, universités et formations. L'évaluation d'un laboratoire (et, par extension, de l'université qui l'héberge) se fait par la visite d'un comité d'experts. Et l'une des choses que regardent les experts, c'est la proportion de chercheurs « publiants » dans le laboratoire. C'est facile, on fait remplir une fiche à chaque chercheur ; on compte le nombre de publications ; on regarde si ça dépasse un certain seuil ; si oui, on décide qu'il est publiant, sinon, eh bien non ; et l'on divise par le nombre total de chercheurs dans le labo. Les labos sont notés en fonc-

tion de leur taux de chercheurs non-publiants, et leurs crédits peuvent en dépendre (avec 99 % d'entre eux qui ont publié, vous avez plus de chances de voir la dotation du labo augmenter qu'avec 50 %, CQFD...).

Les critères sont publics² et simples. Au moins en apparence... Il faut avoir au moins « n » publications de bonne qualité au cours des quatre dernières années. Sachant que « n » est variable suivant la discipline et le statut (chercheur ou enseignant-chercheur). En mathématiques, par exemple, il faut avoir publié deux articles au cours des quatre dernières années, quel que soit votre statut. Pour un chercheur en

1. Agence d'évaluation de la recherche et de l'enseignement supérieur.

2. Cf. AERES, Section des unités de recherche, critères d'identification des chercheurs et enseignants-chercheurs « produisant en recherche et valorisation », www.aeres-evaluation.fr/IMG/pdf/Criteres_Identification_Enspts-Chercheurs.pdf

chimie, le taux passe à quatre articles au cours des quatre dernières années (et seulement deux en quatre ans pour un enseignant-chercheur en chimie).

De la définition d'une « bonne » revue

Là où tout cela se corse, évidemment, c'est pour définir ce qu'est une publication « de bonne qualité ». Là encore, au départ, c'est simple : est considérée comme une publication de bonne qualité, un article dans un journal de bonne qualité. Et l'on fournit la liste des « journaux de bonne qualité », pour chaque discipline. Les spécificités de chaque discipline sont aujourd'hui prises en compte : ainsi, pour les sciences humaines et sociales, les ouvrages sont comptabilisés ; de même, entrent en compte les conférences en sciences et technologies de l'information et de la communication.

Les critères sont simples, et pourtant la polémique a été assourdissante. Je ne vais pas ici vous rejouer l'histoire de l'opposition aux critères de l'AERES, mais disons qu'il y a eu discussion... et qu'elle continue. Aujourd'hui, je voudrais donner mon opinion sur le fait même de conduire une évaluation sur ces critères.

Ce qu'il y a de mal dans la bibliométrie pratiquée par l'AERES

De la quantité... et la qualité ?

Il faut se rappeler que les chercheurs sont des êtres humains, et les êtres humains, comme le rappellent mes économistes favoris, réagissent aux mesures et aux critères. Si l'on dit que les labos sont évalués uniquement en fonction du nombre de publications, et non de leur qualité, alors les labos vont faire du chiffre. De la quantité, uniquement de la quantité, au détriment de la qualité individuelle de chaque publication.

D'autre part, les critères étaient, au départ, mal définis. Dans certaines disciplines, comme l'informatique, une conférence peut être considérée

Le rôle de l'AERES

L'AERES est l'Agence d'évaluation de la recherche et de l'enseignement supérieur. Depuis 2007, l'agence est chargée d'évaluer l'ensemble des composantes de la recherche et de l'enseignement supérieur en France : les structures de recherche (laboratoires), les enseignements (diplômes) et les établissements (universités, écoles). Chaque évaluation est effectuée par un comité d'experts, nommés par l'AERES. L'évaluation des unités de recherche repose sur un dossier portant sur l'activité au cours des quatre dernières années, et sur une visite réalisée par le comité d'experts. Le rapport final prend en compte de nombreuses données, qualitatives et quantitatives.

Parmi les données qualitatives, on note le caractère ambitieux et original de la recherche entreprise au sein du laboratoire, le projet du laboratoire, la gouvernance.

Parmi les données quantitatives, on note le nombre de publications, le nombre de thèses soutenues, le nombre de brevets ou de contrats. Et également le nombre de chercheurs « publiants » ou « produisant » ; c'est la définition de ce terme qui est l'objet du présent article.

comme meilleure que la meilleure des revues dans ce domaine. Dans d'autres, comme les sciences humaines et sociales, il vaut mieux écrire un livre qu'un article. Avec les critères tels que publiés initialement, on aurait pu avoir d'excellents chercheurs considérés comme non-publiants. Depuis, les critères ont été révisés pour inclure les conférences en sciences et technologies de l'information et de la communication et les ouvrages en sciences humaines et sociales.

Le vrai problème, c'est que le statut (publiant/non-publiant) est binaire, et que le travail à fournir pour passer la limite est connu d'avance. Si j'ai publié le bon nombre d'articles la première année, je peux me permettre de me reposer pendant trois ans, sans que cela change mon statut (je reste publiant). Inversement, si j'ai plusieurs articles refusés, une fois dépassé le temps

moyen de réaction d'une revue dans mon domaine – souvent plus d'un an –, ce n'est plus la peine de travailler, mon statut ne changera pas, je resterai non-publiant. C'est tout le problème d'un facteur binaire avec une limite connue à l'avance³.

Hors de la liste, point de salut

En outre, seules comptent les publications dans des revues qui sont incluses dans la fameuse liste. Si j'ai publié vingt articles dans la *Revue alsacienne de biologie moléculaire*⁴ (qui n'est justement pas intégrée à la liste), ils ne sont pas pris en compte, et je suis donc classé comme non-publiant. Pour les revues qui ne font pas partie de la liste, c'est une catastrophe, et elles ont d'ailleurs enregistré une grosse baisse du nombre d'articles soumis. Pour les chercheurs qui avaient pris l'habitude de publier dans certaines revues, c'est aussi une catastrophe. Se retrouver non-publiant alors qu'on en a écrit et publié une vingtaine, cela fait « mal au cœur »...

Ce problème est démultiplié pour les chercheurs (et les laboratoires) qui sont à la frontière de plusieurs disciplines. Un chercheur en histoire de l'économie, qui publie dans des revues d'histoire, peut être jugé comme non-publiant s'il est évalué par des économistes, puisqu'il n'a écrit aucun article dans des revues d'économie.

On peut ajouter que l'Agence ne regarde que les revues, et non pas l'article individuellement. Un article fondateur, s'il est publié dans une revue qui n'est pas classée A, ne compte pas pour l'AERES. Un article très mauvais dans une bonne revue, par contre, compte bien comme une publication.

Se pose, par ailleurs, le problème de la constitution des « listes de revues ». Ce travail a été confié à des comités, qui ont eux-mêmes consulté la communauté. Néanmoins, en analysant leur travail, plusieurs biais sont

3. On pourra consulter à ce sujet : Maya Beauvallet, *Les stratégies absurdes : comment faire pire en croyant faire mieux*, Éditions du Seuil, 2009.

4. C'est un exemple imaginaire...

visibles dans les premières versions des listes de revues :

- Dans certaines matières, on ne trouve que les revues où publient les membres du comité. Un biais de sélection classique – on juge mieux ce que l'on connaît – mais qui aboutit à l'exclusion de revues de très bonne qualité, y compris certaines des meilleures revues internationales !

- Dans d'autres matières, on constate que la position des revues dépend, en grande partie, de leur facteur d'impact (voir encadré). Or, le facteur d'impact est mesuré à partir des bases de données bibliographiques (Thomson ISI, Scopus)... qui sont notoirement incomplètes. Outre le fait que, jusqu'à récemment, ces bases de données ne comptaient pas les conférences (donc un article fondateur publié à Siggraph⁵ et cité dix fois l'année suivante dans des conférences aurait eu un impact zéro), elles oublient tous simplement plusieurs articles, et parfois des années entières de certains journaux. Le taux de couverture⁶ est très variable selon les disciplines et les bases de données. Les grandes disciplines historiques, telles que la physique ou la biologie, ont l'avantage d'être bien indexées. Pour des disciplines plus récentes, comme l'informatique, le taux de couverture tombe autour de 60 %⁷. Et nous sommes encore dans

Le facteur d'impact

Le facteur d'impact est l'un des moyens les plus employés pour mesurer la qualité moyenne d'une revue scientifique. Il est calculé chaque année par les grandes bases de données bibliographiques (ISI, Scopus, Scimago...).

Pour chaque journal, on compte le nombre de citations qu'il a reçues, et on divise par le nombre total d'articles publiés. On en déduit un nombre moyen de citations par article. Le comptage des citations et des articles est fait sur une fenêtre temporelle limitée, allant de 2 à 5 ans.

C'est une évaluation de la qualité de la revue (une bonne revue publie souvent des articles qui sont très cités), et par extension des articles qui sont publiés dans cette revue. Comme toute moyenne, elle peut cacher de grandes disparités : un article très moyen peut être publié dans une bonne revue (mais il ne sera pas cité pour autant), et inversement.

C'est aussi une prophétie auto-réalisatrice : une revue avec un facteur d'impact élevé sera plus réputée, donc elle sera plus lue. Ses articles auront alors plus de chance d'être cités.

les sciences dures. Les sciences humaines et sociales sont, elles, encore moins bien reconnues et indexées par les bases de publications. Dans ces conditions, le facteur d'impact d'une revue n'est pas forcément une donnée fiable.

- Enfin, presque toutes les revues classées A⁸ par l'AERES sont des revues de langue anglaise. Quoi de plus normal, me direz-vous, si vous venez des sciences dures, ou si vous avez toujours publié en anglais. Ou si vous avez parcouru le vaste monde, et que vous vous êtes rendu compte que la *Revue alsacienne de biologie moléculaire*, en dehors de l'université qui la publie, personne n'en a jamais entendu parler. Oui, mais, pour quelqu'un qui n'a jamais publié en anglais, c'est très, très difficile la première fois. Les

8. Les revues sont classées A, B ou C mais seules les revues A comptent pour déterminer le statut publiant/non-publiant.

articles ne suivent pas forcément le même plan, il y existe des codes à respecter, des références à citer que vous ne connaissez pas, etc. Et s'il n'y a personne pour vous apprendre les codes, eh bien, vous voilà coincés. Il y a une grosse *learning curve*⁹... et vous êtes tout en bas.

Ce qu'il y a de bien dans la bibliométrie pratiquée par l'AERES

Une pression pour publier

Que nous soyons enseignant-chercheur ou chercheur, à l'université, nous sommes soumis à de nombreuses pressions et charges : faire des cours, remplir des rapports, monter des formations, écrire des articles... Et certaines de ces pressions sont fortes : si vous ne faites pas vos cours, on va s'en rendre compte ! Si vous ne rendez pas votre rapport à temps à l'administration, elle va vous tomber dessus. Si vous ne remplissez pas votre demande de projet à l'Agence nationale de la recherche (ANR) à temps pour la *deadline*, vous n'aurez plus de sous pour votre recherche. La seule des charges des enseignants-chercheurs qui n'était pas soumise à une pression quelconque jusqu'ici concernait les publications. Ces dernières pouvaient donc devenir le maillon faible... L'AERES vient ici rétablir l'équilibre des pressions en faveur des publications.

L'effet boule de neige

D'autre part, l'Agence mène une évaluation collective : l'ensemble du laboratoire est menacé (de perdre son statut de A+, par exemple, et les financements qui vont de pair) s'il y a trop de chercheurs non-publiants. Et ce système est très efficace ! C'est ce qu'on appelle le problème des démenageurs de pianos. Si l'un des démenageurs ne fait pas sa part du boulot, les autres vont très vite s'en rendre compte, et lui expliquer délicatement

9. Courbe d'apprentissage.

5. La plus grande conférence en informatique graphique (c'est-à-dire à la fois la plus prestigieuse et la plus sélective).

6. C'est-à-dire la proportion des publications qui sont indexées par la base de données bibliographiques.

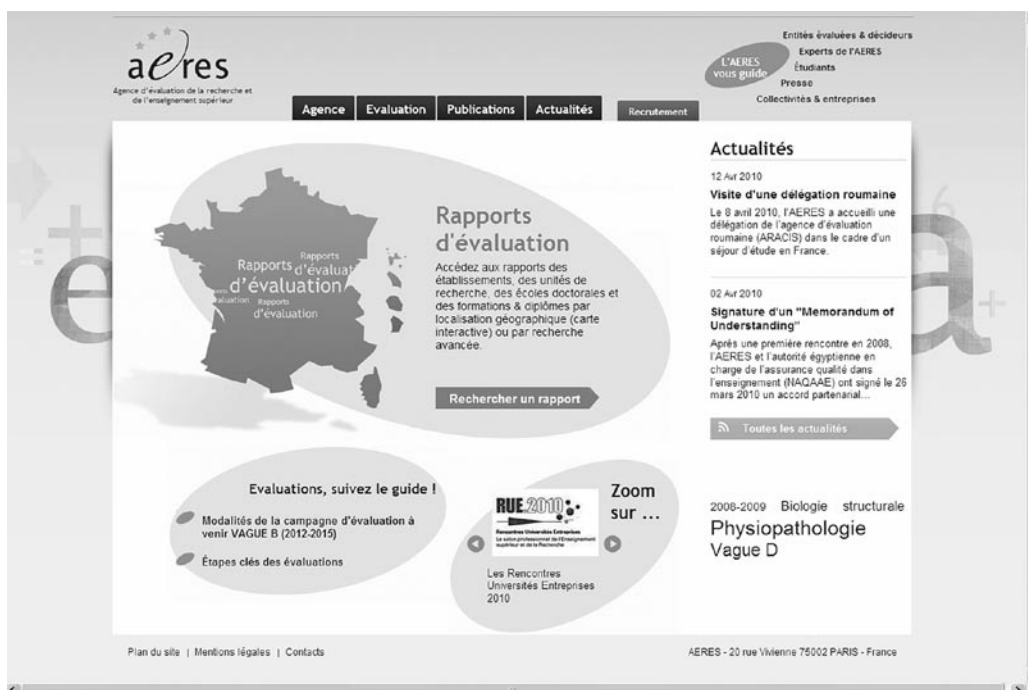
7. Lire sur ce sujet :

– Anne-Marie Kermauer, Erwan Faou, Jean-Pierre Merlet (rapporteur), Philippe Robert et Luc Segoufin, *Que mesurent les indicateurs bibliométriques ?* Document d'analyse de la Commission d'évaluation de l'Inria, www.inria.fr/inria/organigramme/documents/ce_indicateurs.pdf

– Massimo Franceschet, « A comparison of bibliometric indicators for computer science scholars and journals on Web of Science and Google Scholar », *Scientometrics* (2010).

– Henk F. Moed et Martijn S. Visser, *Developing Bibliometric Indicators of Research Performance in Computer Science: An Exploratory Study, Research Report to the Council for Physical Sciences of the Netherlands Organisation for Scientific Research (NWO)*, CWTS Report 2007-01, février 2007, www.cwts.nl/pdf/NWO_Inf_Final_Report_V_210207.pdf

→ Ce qu'il y a de bien (ou de mal) dans l'évaluation bibliométrique telle que pratiquée par l'AERES :



Page d'accueil du site de l'AERES

que ça serait bien qu'il travaille un peu pour l'équipe... Pareil avec les non-publiants ! Et depuis que les critères de l'AERES ont été rendus public (vers 2007), les effets se sont bien fait sentir : plusieurs chercheurs ont ressorti des projets d'articles qui dormaient, les ont soumis, les publiés, etc. Le nombre total de publications augmente. L'effet s'est également fait sentir au sein des revues : plusieurs d'entre elles ont modifié leurs pratiques, augmenté leur sélectivité et leur travail d'édition, se sont mises à publier en anglais... Pour pouvoir passer en « rang A ». Et tout ça, c'est plutôt bon pour la recherche.

Un talent qui s'entretient

Par ailleurs, publier, ce n'est pas comme le vélo : c'est un talent qui s'entretient. Plus on publie, plus on arrive à publier facilement. Mon observation (empirique, mais pluridisciplinaire et multicontinentale) est qu'il y a des effets de seuil : en dessous d'un certain nombre de publications par an, on n'arrive plus à se tenir au courant de ce qui se passe dans son domaine. Au-dessus d'un certain nombre, on sait

rapidement ce qu'il faut mettre dans le résumé, l'introduction, le plan, et on libère ainsi du temps pour se focaliser sur le contenu scientifique. Cette observation est, m'a-t-on dit, confirmée par les bibliométriciens. On peut la résumer ainsi : plus un chercheur a de publications à son actif, moins il doit fournir d'efforts pour la prochaine publication¹⁰.

D'autre part, l'AERES ne se focalise pas sur le nombre total de publications, mais sur le nombre total de publications dans des revues de rang A. Ce qui limite l'effet qualité/quantité dont je parlais au début et dont tout le monde a peur. Il va être dur de faire de la quantité avec des articles publiés dans *Nature* (ou à *Siggraph*).

Enfin, les critères de l'AERES ne font que souligner une évidence : un article fondateur, s'il est publié dans la *Revue alsacienne de biologie moléculaire* (encore elle !), ne sera jamais lu. Il l'aurait peut-être été il y a cinquante ans, ou même trente ans, quand il

¹⁰. Voir, par exemple, Thierry Lafouge, Christine Michel, *Entropie et distributions bibliométriques*, Actes du Congrès SFBA, Île Rousse, septembre 1999, http://archivesic.ccsd.cnrs.fr/sic_00000340/en

y avait moins de revues et moins de publications, mais de nos jours, il ne sera jamais lu. Donc, le publier dans la *Revue alsacienne de biologie moléculaire*, c'est du gâchis. Et en donnant ce petit coup de pouce, l'AERES va pousser à ce que des articles fondateurs soient publiés dans des revues où ils seront lus, et augmenter l'impact scientifique, le vrai, de la recherche française.

On a coutume de dire que toute évaluation quantitative reposant sur un indicateur chiffré aboutit à une perversion du système, où les acteurs ne raisonnent plus qu'en fonction de l'indicateur et non en fonction des autres objectifs. En ce sens, l'évaluation effectuée par l'AERES a nécessairement des inconvénients et aura des effets pervers. En même temps, le système dans lequel s'insère l'évaluation n'est pas forcément parfait, il est simplement soumis à différentes forces, qui le maintenaient plus ou moins en équilibre. L'un des effets de cette évaluation sera de modifier l'équilibre. Nous verrons si cette modification se fera en bien ou en mal. ●

Avril 2010

* Cet article est issu d'un billet paru sur le blog de l'auteur.