

Document d'analyse de la Commission d'Evaluation de l'INRIA

Que mesurent les indicateurs bibliométriques ?

Préparé par A-M. Kermarrec, E. Faou, J-P. Merlet (rapporteur), P. Robert, L. Segoufin

Validé par la Commission d'Evaluation du 12 septembre 2007

Résumé

La Commission d'Évaluation (CE) de l'INRIA a souhaité mener une réflexion sur les indicateurs bibliométriques de façon à mieux comprendre ce qu'ils mesurent, leur pertinence et leur fiabilité, en particulier dans le contexte de la recherche menée à l'INRIA. Le présent document est le résultat du travail d'un groupe de réflexion animé par Jean-Pierre Merlet et composé également de A-M. Kermarrec, E. Faou, P. Robert et L. Segoufin. Ce document a fait l'objet de discussions lors de deux séances de la Commission d'Evaluation en janvier et septembre 2007 et prend en compte les remarques des membres de la Commission.

Table des matières

| | | |
|----------|--|-----------|
| 1 | Introduction | 4 |
| 2 | Plan du document | 4 |
| 3 | Bibliométrie | 6 |
| 4 | Les sources de citations | 7 |
| 4.1 | Les sources à accès payants | 7 |
| 4.1.1 | Thomson Institute for Scientific Information (ISI) | 7 |
| 4.1.2 | Scopus | 9 |
| 4.1.3 | Problèmes communs aux ressources payantes | 10 |
| 4.2 | Les sources gratuites | 10 |
| 4.2.1 | CiteSeer | 10 |
| 4.2.2 | Google Scholar | 10 |
| 4.2.3 | Citebase | 12 |
| 4.2.4 | Problèmes communs aux ressources gratuites | 12 |
| 4.3 | Comparaison des sources | 13 |
| 4.4 | Les problèmes d'attribution | 15 |
| 5 | Les indicateurs | 15 |
| 5.1 | Les indicateurs pour journaux | 15 |
| 5.1.1 | L'impact factor d'un journal | 16 |
| 5.1.2 | Indice d'immédiateté, immediacy index | 17 |
| 5.1.3 | Demi vie des citations, Cited half-life | 17 |
| 5.2 | Les téléchargements | 18 |
| 5.3 | Les indicateurs quantitatifs | 18 |
| 5.4 | Les indicateurs individuels | 19 |
| 5.4.1 | Le H-index | 19 |
| 5.4.2 | Les variantes du H-index | 20 |
| 5.5 | Les autres indicateurs | 21 |
| 5.6 | Complémentarité et cohérence des indices | 22 |
| 5.7 | Conclusions sur les indicateurs | 22 |
| 6 | Mauvais usages et effets pervers | 23 |
| 6.1 | Les manipulations du JIF | 23 |
| 6.2 | Le "saucissonnage" (<i>salami-slicing</i>) | 24 |
| 6.3 | Les auto-citations | 24 |
| 6.4 | Le frein à la prise de risque | 24 |
| 7 | Le coût des indicateurs et son implication | 24 |
| 8 | La normalisation des adresses | 25 |

| | |
|--|-----------|
| 9 Un exemple d'analyse pour des auteurs de l'INRIA | 26 |
| 9.1 Comparaison entre les sources | 26 |
| 9.2 Recherche institutionnelle | 29 |
| 9.3 H-index | 30 |
| 10 Conclusion | 30 |
| 11 Annexe 1 : les indicateurs de la LOLF | 31 |
| 12 Annexe 2 : L'Observatoire des Sciences et Techniques (OST) | 32 |
| 13 Annexe 3 : glossaire | 33 |

1 Introduction

La Commission d'Évaluation de l'INRIA (CE) a décidé d'entamer à partir de Septembre 2006 une réflexion sur les éléments intervenant dans l'évaluation de l'activité scientifique. Parmi ces éléments figurent les *indicateurs bibliométriques*¹ chiffrés qui sont de plus en plus souvent utilisés pour évaluer l'activité et la qualité scientifique des chercheurs, des équipes, des institutions, voire même d'une nation.

La mention d'*indicateurs* garantit dans le milieu scientifique mondial des réactions émotionnelles de principe qui vont probablement encore augmenter dans les années à venir en raison de l'importance accrue qu'ils ont dans le fonctionnement de la recherche, que ce soit au niveau des structures, des orientations stratégiques ou des financements. C'est aussi le cas en France puisque la LOLF indique de manière relativement précise les indicateurs qui seront utilisés pour évaluer les établissements (voir la liste en Annexe 1); cette utilisation se propage à tous les niveaux comme l'illustre un très récent rapport de l'Inspection Générale de l'Administration de l'Éducation Nationale et de la Recherche (IGAENR)² qui examine dans le détail la nature et l'usage d'indicateurs qui serviront à évaluer et gérer la contractualisation des organismes de recherche avec l'État.

On ne peut cependant nier une influence parfois démesurée de certains de ces indicateurs comme le *journal impact factor* (JIF, section 5) ou le nombre de citations, dont sont tirés, jusque dans la presse nationale, des opinions tranchées sans qu'il y ait examen critique des chiffres, de la méthodologie utilisée pour les produire ni même une parfaite connaissance de leur définition.

On peut aussi constater que ces indicateurs jouent un rôle dans le fonctionnement de l'INRIA, en particulier au niveau de la Commission d'Évaluation (CE). Par exemple certains indicateurs ont été utilisés dans le rapport d'évaluation d'un thème par le panel de rapporteurs, tous extérieurs à l'INRIA. Les membres même de la Commission peuvent faire appel à ces indicateurs lorsqu'ils sont amenés à évaluer des activités qui sont à la marge de leur domaine de compétence (retour d'évaluation, jury de concours par exemple). La Commission d'Évaluation de l'INRIA a donc souhaité faire un point sur ces indicateurs de façon à mieux comprendre ce qu'ils mesurent (et ce qu'ils ne mesurent pas), leur pertinence et leur fiabilité, ceci dans le but premier de faciliter le travail interne de la Commission. Pour cela elle a chargé un groupe de travail de la rédaction d'un document d'analyse qui a fait l'objet de discussions lors de deux séances de la Commission d'Évaluation en janvier et septembre 2007. Le présent document prend en compte les remarques de la Commission qui l'a validé. Les analyses présentées prennent bien sûr en compte les spécificités de l'Institut et ne sauraient donc être transposées à d'autres cas sans une analyse appropriée.

2 Plan du document

Le plan de ce document est le suivant :

- les principes de base de la bibliométrie sont présentés dans la section 3

¹il existe bien sûr d'autres indicateurs, par exemple ceux qui sont utilisés pour évaluer le transfert ou la valorisation, qui ne font pas l'objet de ce document

²Rapport 2007-012, La contractualisation des organismes publics de recherche avec l'État, Avril 2007

- la plupart des indicateurs bibliométriques sont établis à partir de l'analyse des citations. Nous examinons quels sont les organismes, les sociétés et les moteurs de recherche Web qui fournissent ces citations (4), livrons quelques éléments qui permettent d'estimer la fiabilité de la fourniture des citations (4.3) et mentionnons le problème de l'attribution des citations à des institutions (4.4)
- les indicateurs établis à partir des citations sont présentés dans la section (5). Ces indicateurs peuvent être établis pour des publications scientifiques (journaux principalement), pour des institutions ou pour des individus.
- la cohérence des indicateurs utilisés comme mesure de qualité scientifique est examinée en section (5.6) et des remarques de fond sur ces indicateurs sont présentées dans la section (5.7)
- l'utilisation des indicateurs dans la vie scientifique n'est pas neutre car elle expose à des manipulations et plus généralement à des problèmes éthiques qui sont présentés en section (6)
- les coûts d'établissement des indicateurs sont abordés en section (7)
- attribuer des citations à une institution afin de permettre des calculs d'indicateurs pour cette institution est un problème délicat mentionné en section (8)
- des exemples pratiques pour des chercheurs de l'INRIA sont présentés en section (9) : nous examinons pour 4 chercheurs seniors de l'INRIA travaillant dans des domaines différents les résultats de la recherche de citations par les trois moteurs de recherche les plus couramment utilisés et l'influence sur les indicateurs de la disparité de ces résultats. On y montre aussi que les travaux de ces chercheurs ne sont pas uniformément attribués à l'INRIA

En conclusion nous indiquons que si les indicateurs peuvent donner des tendances sur un nombre réduit d'aspects de la vie scientifique, il convient d'être très circonspect dans leur usage en raison de la possibilité d'interprétations erronées, des erreurs de mesure (souvent considérables) et des biais dont ils sont affectés. Un usage abusif des indicateurs est facilité par la nature chiffrée du résultat qui introduit la possibilité d'établir dans l'urgence toutes sortes de statistiques, sans se préoccuper d'en analyser la qualité et le contenu, et en occultant l'examen d'autres éléments de la vie scientifique comme, par exemple, l'innovation et le transfert intellectuel et industriel.

Les constatations et les recommandations les plus importantes sont les suivantes :

1. on pourra constater que les erreurs de mesure dans la recherche de citations sont très importantes et cela est particulièrement vrai pour les travaux effectués à l'INRIA. Nous recommandons donc fortement de multiplier les sources, de faire une analyse critique des résultats (en particulier en faisant appel à des experts des domaines) et de s'attacher, au plus, aux ordres de grandeur des indicateurs
2. nous recommandons d'utiliser la littérature bibliométrique pour cerner la signification réelle des indicateurs et leur biais
3. les indicateurs ne donnent qu'une vue partielle et biaisée de certains aspects de la vie scientifique, sans en recouvrir l'ensemble. Ils doivent donc être complétés, corrigés et commentés par des spécialistes du domaine scientifique et interprétés s'ils sont utilisés dans le cadre d'une évaluation ou d'une prise de décision
4. nous recommandons de ne jamais utiliser des indicateurs pour effectuer des comparaisons

entre domaines scientifiques différents

5. le principe des indicateurs reposant sur une analyse des citations est peu favorable à la prise de risques scientifiques et à l'innovation. Une utilisation abusive (voire pire, automatisée) serait un frein majeur à l'innovation
6. les indicateurs chiffrés sont très facilement manipulables par les individus, les institutions et d'autres acteurs de la vie scientifique (comme les journaux). Le nombre de manipulations augmentent, ce que l'on peut corrélérer à l'effet croissant de l'influence des indicateurs
7. pour ce qui concerne les domaines scientifiques traités par l'INRIA il a été constaté une très mauvaise couverture des sources de citation usuelles (Web of Science, Scopus), pouvant aller jusqu'à l'absurde. Par voie de conséquence les indicateurs construits sur ces sources donnent une information erronée sur le fonctionnement de l'Institut. Une action auprès des sources pourrait éventuellement apporter des améliorations sur l'attribution de citations à l'INRIA mais ne changerait rien au fait que les sources restent insuffisantes pour ce qui concerne l'informatique et les mathématiques appliquées.

3 Bibliométrie

Les indicateurs bibliométriques reposent sur l'analyse des *citations*, c'est-à-dire la partie "Références" d'un article scientifique qui fait mention des travaux effectués par la communauté scientifique (y compris les auteurs de l'article) sur le sujet de l'article. Il faut d'ores et déjà indiquer que les indicateurs ne reposant que sur des citations ne donnent qu'une vision partielle du travail scientifique puisque 90% des papiers publiés dans les journaux scientifiques ne sont jamais cités [9] et qu'il semble difficile de croire que seulement 10% de la production scientifique serait significatif.

La première étape de la construction d'un indicateur est donc le recueil de ces citations : étant donnée l'ampleur de la production scientifique, il s'agit d'un travail colossal qu'il est impossible de mener à bien de manière complète. Les professionnels effectuant ce travail se contentent donc de passer au crible une partie sélectionnée de la littérature mondiale (voir la section 4). La seconde étape de la construction d'un indicateur consiste à en définir la formule mathématique qui, à partir des données brutes des citations, permet d'en obtenir une évaluation chiffrée. Il est aussi possible de construire des méta-indicateurs qui vont combiner les évaluations chiffrées de divers indicateurs pour obtenir d'autres indicateurs (c'est par exemple ce qui est fait pour le classement dit "de Shanghai" des Universités).

Ces indicateurs et la méthodologie utilisée pour les calculer soulèvent les mêmes questions que se pose un physicien face aux résultats d'un instrument de mesure :

1. quelle grandeur est effectivement mesurée par l'instrument ?
2. quel est la marge d'erreur sur cette mesure ?
3. quel est le rapport qualité/coût de la mesure ?

Comme le dit M. Zitt de l'INRA [18] : *At the same time, more precautions are needed, as "desktop" bibliometrics may reinforce a danger already present in less reactive scientometrics :*

the contrast between the highly sensitive nature of evaluation issues and the eagerness of users or clients to elicit a figure or a display and forget crucial warnings about statistical distributions and methodology artifacts. If scientometrics is a mirror of science in action, then scientometricians' particular responsibility is to both polish the mirror and warn against optical illusions.

Les citations étant au cœur de l'établissement des indicateurs, il faut donc examiner quels sont leurs sources, les acteurs du domaine (qui d'ailleurs la plupart du temps ne se contentent pas de recueillir les citations mais proposent aussi les indicateurs) et de faire une analyse critique de la méthodologie employée pour recueillir les citations. En France le principal pourvoyeur d'indicateurs est l'Observatoire des Sciences et des Techniques (OST)³, voir Annexe 2.

4 Les sources de citations

4.1 Les sources à accès payants

Les moteurs de recherche sur le Web permettent deux types de recherche pour ce qui concerne les publications scientifiques :

1. *la récupération d'articles* : des moteurs comme PubMed Central (PMC) pour le domaine médical ou Science Direct permettent de localiser des articles et d'en récupérer le texte, le plus souvent moyennant paiement.
2. *la recherche de citations* : il s'agit principalement de trouver dans une base de données quels sont les articles qui mentionnent dans leur partie "Références" un auteur ou une institution donnée. C'est donc ce type de requête qui est utilisé pour établir la base de citations qui sera ensuite utilisée pour le calcul des indicateurs

Nous nous intéresserons ici seulement aux moteurs du deuxième type.

4.1.1 Thomson Institute for Scientific Information (ISI)

Un acteur incontournable des indicateurs, et le plus ancien, est la société commerciale Thomson ISI fondée en 1960 qui publie annuellement son *Journal Citation Report* (JCR), qui inclut un certain nombre d'indicateurs, dont le *journal impact factor* (JIF), et qui paraît 6 à 7 mois après la fin de l'année.

Le *Web of Science* (WoS) est la source Web des citations utilisées par l'ISI pour le JCR (sa version papier est le *Science Citation Index*, SCI) et contiendrait un total de 38 millions de citations et remonterait jusqu'à 1900 dans sa version étendue. Le WoS inclut l'analyse de 8700 journaux en sciences "dures" (80%) et sciences sociales (20%), dont un petit nombre (190) en accès ouvert, avec une mise à jour chaque semaine. À noter que les domaines sont inégalement couverts : par exemple, les sciences du vivant sont moins bien représentées que dans les bases biomédicales comme *Medline*. Il est aussi reconnu que la couverture des domaines plus proches de la demande sociale ou de l'application (science de l'ingénieur notamment) n'est pas aussi bonne que pour la recherche fondamentale. La couverture de WoS ne représente qu'une faible partie de la littérature scientifique mondiale puisque on estimait en 1999 qu'il y avait un total

³www.obs-ost.fr

de 100 000 revues scientifiques dont 25 000 rien que dans le domaine biomédical⁴. Enfin la plus grosse partie des actes de conférence ne sont pas traités dans le WoS. Il en est de même des articles publiés dans des journaux en accès ouvert, des archives ouvertes ou des pages personnelles.

Le WoS est souvent accusé d'un biais américain. Le WoS de 2005 contient 98.74% d'articles en anglais, 0.234% d'articles en français, 0.205% d'articles en chinois et 0.09155% d'articles en japonais [7]. Il est clair que l'on cite plus facilement des articles de son propre pays⁵ et ce biais a été confirmé pour des cas particuliers (voir la section 4.3).

Il est important de définir ici le sens de *citation* adopté par l'ISI. Tout matériel publié par un journal et cité comme référence compte comme citation pour ce journal. Il en est ainsi bien sûr des articles mais aussi des éditoriaux, des revues de livres, des lettres aux éditeurs. Les indexeurs de l'ISI assignent donc *un type de document* à chaque matériel publié dans un journal, de manière parfois incohérente⁶. Les journaux eux-mêmes sont classés en catégories, certains pouvant être assignés à deux catégories et exceptionnellement à trois. Les indexeurs renseignent le WoS sur les champs suivants, pour tous les auteurs d'une publication donnée :

- Main Organization : un seul champ ;
- Suborganizations : 3 champs distincts
- Ville ;
- Province ;
- Pays ;
- Code postal

Ce choix d'une seule main organisation n'est pas neutre : c'est par exemple ce champ qui sert de base au classement de Shangai.

À l'opposé, le nombre d'*articles publiés* dans une année par un journal donné n'inclut que les articles de recherche. Il est aussi notable que l'ISI ne compte pas comme citation une référence à un *livre*⁷ (par défaut l'interrogation de la base de données ISI pour un auteur donné ne fera donc pas référence aux livres que cet auteur pourra avoir publiés). L'ISI essaie d'éviter de compter les *auto-citations* d'auteurs (c'est-à-dire les citations du ou des auteurs à leur propres papiers), en se contentant toutefois d'éliminer les papiers cités dont le premier auteur est un des auteurs du papier citant (ainsi l'auto-citation entre les membres d'un groupe risque d'être très sous-évalué).

Les citations elles-mêmes sont attribuées à l'ensemble des auteurs. Les adresses sont soumises à une normalisation, qui est cependant loin d'être parfaite.

On doit noter quelques phrases-clé proposées par Thomson à propos du JCR⁸ :

⁴G. Roelants, INFI Newsletter, Décembre 1999

⁵une étude a montré que les auteurs d'articles de biomédecine américains ne citent en moyenne que 12% de références étrangères de moins de 2 ans et 30% de moins de 10 ans (un taux très inférieur au poids réel de la littérature non américaine), *N. Pinhas, Inserm Actualités, 154, Septembre 1997*

⁶*It further complicates the situation that the same genre of item are inconsistently classified*, Peter's Digital Reference Shelf

⁷ISI le justifie par l'existence de livres dit édités où chaque chapitre est écrit par des auteurs différents. Comme seuls les noms des éditeurs sont mentionnés dans les citations les auteurs des chapitres seraient lésés : *For books, the bibliographic information presented in an article's reference list often contains first author only, or a partial list of authors . . . Linking to the source item allows us to expand the reference to include the complete list of authors. Lacking this expanded information would decrease the accuracy of our analysis by under-representing the contribution of additional authors.* On pourrait cependant faire remarquer que tous les livres ne sont pas de type édités.

⁸Peter's Digital Reference Shelf

- *There has been a guide for a long time in the help file about how to use the JCR. Many ignore it and use it in a way it was not meant to be used, such as faculty tenure position, or used alone*
- *There are excellent educators who do not necessarily publish in journals processed by ISI or do not research in a field where citations are profuse*
- *I just wish a better handling of the citable items, and plausibility checks of the scores by experts who know well the journals of a discipline to spot errors at a glance.*

Il faut aussi mentionner la difficulté de l'attribution des citations aux auteurs (en raison des différences de pratique entre les communautés) mais aussi aux institutions. Il est clair que des institutions partenariales comme l'INRIA où les auteurs peuvent se réclamer de plusieurs institutions vont être pénalisés. Une recherche dans le WoS montre ainsi des différences considérables dans l'intitulé de l'institution d'appartenance qui ne facilitent pas le travail des indexeurs. Des articles qui pourraient incontestablement être attribués à l'INRIA vont être attribués au LORIA, à l'IRISA, à l'INRIA Sophia etc. . . Ce problème de la normalisation des adresses pour la France est abordé dans la section 8.

Il faut enfin noter que l'ISI n'est pas simplement un pourvoyeur de citations : cette société est à la source d'une très grande majorité des indicateurs bibliométriques comme le JIF (section 5).

4.1.2 Scopus

Scopus a été lancé à l'automne 2004 par Elsevier et contiendrait 30 millions de citations, qui ne remontent cependant qu'à 1996 (1900 pour le WoS), provenant de 15 000 titres avec review, 12850 journaux dont 1000 en accès ouvert, 700 proceedings, 275 millions de pages Web et 125 séries de livres, mis à jour de manière quotidienne, avec une couverture plus exhaustive que le WoS pour l'ingénierie et apparemment une meilleure ergonomie que le WoS [1, 2]. La distribution des citations serait la suivante [2] :

- santé et sciences de la vie : 25.3 millions de documents
- ingénierie : 8.5
- agriculture et biologie : 3.6
- sciences de la terre et environnement : 1.9
- chimie : 1.3
- physique : 0.59
- sciences sociales : 0.29
- mathématiques : 0.26
- psychologie : 0.23
- économie : 0.22

La provenance géographique des titres est plus variée que pour WoS avec 60% des titres ne provenant pas des USA. Jusqu'en 2003 seule était utilisée l'affiliation du premier auteur mais, à partir de cette date, les affiliations de tous les auteurs sont utilisées. Un effort particulier a été porté pour mieux gérer les homonymies avec une prise en compte des variantes des noms et une utilisation de données comme l'affiliation et le sujet traité pour mieux cerner l'auteur. D'un point de vue bibliométrique, les statistiques proposées sont plus riches que celles du WoS mais il manque la possibilité de faire des analyses par institutions. Scopus est un concurrent

sérieux du WoS.

4.1.3 Problèmes communs aux ressources payantes

Le problème est essentiellement lié au corpus, qui n'est d'ailleurs pas toujours bien connu. On a vu que celui-ci ne représentait qu'une faible proportion de la production scientifique. Le choix des journaux et des actes de conférences est donc important et influencera les indicateurs de citations calculés à partir de ces sources. Cela peut favoriser une sous-thématique au détriment d'une autre, voire fausser complètement le résultat. Par exemple, il semble, au moment où ce rapport est écrit, que le WoS contienne tous les actes publiés dans un LNCS mais aucun acte édité par l'ACM. Dans un domaine où les bonnes conférences ont un label ACM et les mauvaises sont publiées via les LNCS, un indicateur calculé à partir de WoS donne un résultat contestable. Voir la section 4.3.

4.2 Les sources gratuites

La prédominance de WoS et de Scopus et des pratiques qui leur sont associées ont suscité des réactions de la communauté scientifique pour élaborer d'autres modes de recherche de citations et d'évaluation.

4.2.1 CiteSeer

CiteSeer est une base de données spécialisée en informatique qui fournit des indexations sophistiquées des citations mais aussi un classement des conférences, des journaux et des auteurs en informatique.

Une comparaison sur l'année 2003 entre le classement ISI des journaux en informatique reposant sur le JIF et le classement CiteSeer a été établie par la CE. La figure 1 montre le classement chez CiteSeer des vingt meilleurs journaux selon ISI alors que la figure 2 présente le classement ISI des vingt meilleurs journaux selon CiteSeer.

Ces figures montrent des discordances très nettes et surprenantes entre les classements. Nous avons considéré les 20 journaux classés par l'ISI (qui en classe 200) comme les plus importants du domaine et avons examiné leur classement selon CiteSeer (qui classe 1200 journaux et conférences). Le premier journal selon l'ISI, *ACM Computer Surveys*, est 195ème selon CiteSeer, le 2ème selon l'ISI, *Bioinformatics*, n'apparaît pas dans CiteSeer, le 6ème pour l'ISI est 958ème pour CiteSeer... Inversement le premier pour CiteSeer, *ACM Trans. on Computer Systems*, est 26ème pour l'ISI, le 4ème pour CiteSeer, *Computer Networks*, est 122ème pour l'ISI.

4.2.2 Google Scholar

Le but affiché de Google Scholar est d'inclure *peer-reviewed papers, theses, books, preprints, abstracts and technical reports from ...academic publishers, professional societies, preprint repositories and universities ...available across the Web*. Les sources ne sont cependant pas affichées, ni les dates de couverture. Google propose différents types d'indicateurs dérivés des sources Scholar :

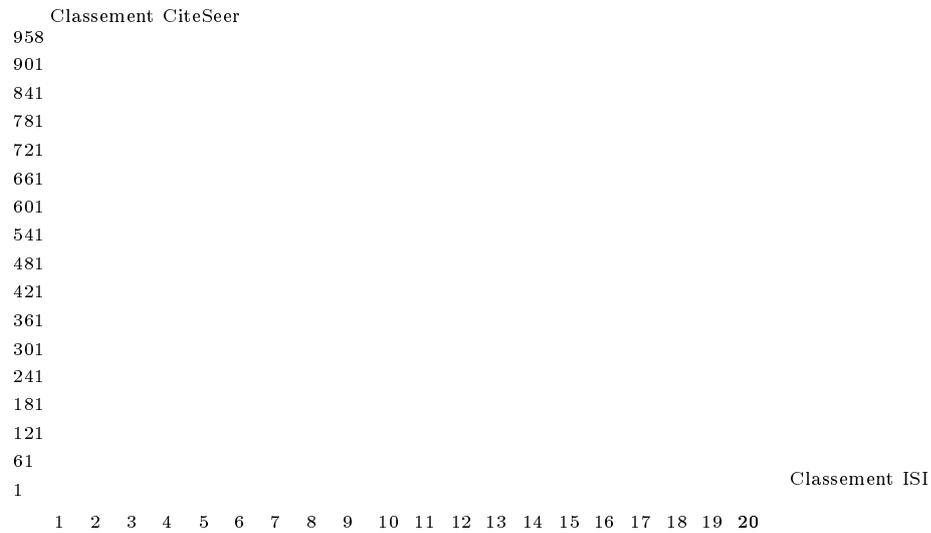


FIG. 1 – En ordonnée le classement CiteSeer des vingt meilleurs journaux en informatique selon ISI. A remarquer que certains journaux du classement ISI n'apparaissent pas dans le classement CiteSeer (et réciproquement).

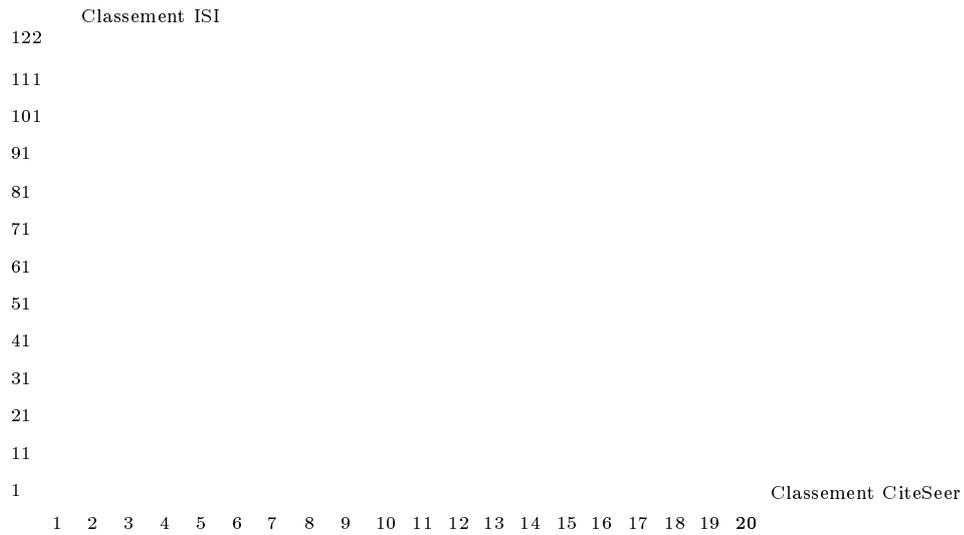


FIG. 2 – En ordonnée le classement ISI des vingt meilleurs journaux en informatique selon CiteSeer.

- documents Web
- citation d'articles
- citation de livres

Les documents Web sont des liens sur des pages Web qui soit décrivent le document soit proposent un accès aux documents. Les citations sont comptées à partir des bibliographies des documents Web.

- Différentes analyses et comparaisons avec d'autres moteurs de recherche ont montré que
- la couverture de Google Scholar est très variable selon les domaines. Certaines études indiquent que les résultats sont bien inférieurs à ceux que l'on pourrait obtenir en utilisant des bases plus spécialisées mais cela n'a pas été le cas pour notre étude décrite en section 9
 - les documents utilisés comme source de citations par Google Scholar sont encore plus mal connus que les sources des bases commerciales
 - les indicateurs obtenus, comme le compte de citations, sont très imprécis
 - il n'y a qu'un format de sortie (par score de citations) alors que WoS propose une analyse plus fine (nombre de publications et de citations par années, élimination des auto-citations, ...)
 - la base de données n'est pas souvent mise à jour (les documents récents ne seront trouvés que dans Google Web)

La mauvaise gestion des homonymies de Google Scholar a motivé Anne-Will Harzing à développer le logiciel **Publish or Perish**⁹ qui récupère les données de Google Scholar mais en fait une analyse plus fine, et propose le calcul d'un nombre important d'indicateurs :

- quantitatifs : nombre total de papiers, de citations, nombre moyen de citations par papier, par auteur, nombre d'auteurs par papier
- individuel : nombre G, nombre H (avec deux variantes), taux de citation pondéré par les dates de publication

4.2.3 Citebase

Dès 1999, Harnad et Brody proposent d'expérimenter un algorithme de citation différent de celui de l'ISI qui se veut plus objectif que les indicateurs de l'ISI qui se concrétise par la création de l'outil **Citebase**¹⁰. Dans ce modèle en plus des citations sont comptabilisés le nombre de téléchargements des articles (toutefois uniquement à partir des données disponibles sur un nombre limité d'archives ouvertes come **arXiv**) ainsi que le temps entre téléchargement et citations pour essayer de mieux cerner la qualité scientifique des articles.

4.2.4 Problèmes communs aux ressources gratuites

Ces ressources gratuites ont des défauts communs liés au fait qu'elles récupèrent (le plus souvent via internet) et traitent de manière automatique leurs données. Nous en citons ici quelques-uns.

1. Identification des personnes. Il est souvent impossible de différencier des personnes ayant le même nom. Les homonymies sont très mal gérées : une recherche de citation sur l'auteur D. Cohen retourne aussi des citations d'auteur du même nom et dont un des prénoms commence par D. À l'opposé, il arrive aussi parfois qu'un seul auteur voit ses papiers et ses citations divisés en deux entrées distinctes¹¹. Les noms accentués ou ne respectant pas

⁹www.harzing.com

¹⁰www.citebase.org

¹¹Par exemple, si votre nom contient "fi" essayez votre vrai nom et votre nom sans le "fi". Cela marche aussi bien avec Google Scholar et CiteSeer. Ceci s'explique par une mauvaise gestion dans certains documents PDF de ces deux lettres lorsqu'elles sont accolées, la lettre "i" disparaissant du document.

le schéma prénom-nom peuvent aussi poser problème puisque ISI stocke les noms d'auteurs exactement comme ils sont trouvés (en permettant toutefois le stockage de variantes, en particulier pour les noms asiatiques). Ainsi l'absence, fréquente, d'un accent dans une citation va conduire à la pénalisation de l'auteur cité.

2. Identification des articles. Ceux-ci étant souvent récupérés automatiquement via Internet, il n'est pas toujours évident d'unifier les diverses versions publiées par tous les coauteurs d'un article sur leur page personnelle. Il n'est donc pas rare de voir un même article donner deux entrées ou plus dans Google Scholar et dans Cite-See, perturbant d'autant le décompte des citations.
3. Identification d'un journal. Il est théoriquement possible de faire des recherches par nom de journal mais c'est au demandeur de donner le nom du journal. Étant donné le nombre important d'abréviations utilisées pour les journaux, il est difficile d'accéder réellement à l'ensemble des volumes d'un journal donné.

Tous ces défauts ont des occurrences très fréquentes. Ils rendent donc peu crédible tout indicateur calculé à partir de ces sources.

4.3 Comparaison des sources

Un article de 2007 [9] étudie les citations obtenues pour 25 chercheurs renommés en science de l'information en utilisant WoS, Scopus et Google Scholar. Scopus propose 35% de citations de plus que le WoS et Google Scholar 160% de plus (ces différences pouvant varier énormément selon les disciplines). Le score de Google devrait être pondéré en raison de l'imprécision des sources et des citations mais il est aussi clair que le fait de rechercher les sources Web permet de mieux prendre en compte les nouvelles habitudes de publication des chercheurs.

Des études récentes se sont intéressées aux citations des 25 auteurs les plus cités en informatique et à leur recoupement dans les bases CiteSeer, Scopus et Web of Science [10]. Il est apparu que seulement 44% des citations étaient communes aux trois bases. Une différence analogue a été observée par Burnham [1] dans le domaine médical¹².

Nisonger (2004, voir l'article de Meho) a examiné ses propres citations dans le Web of Science : cette base contenait 28.8% du total des citations de ses articles, 42.2% des citations de ses articles de journaux, 20.3% de ses citations en dehors des USA et 2.3% de ses citations qui n'étaient pas en anglais.

Meho [10] a publié une comparaison exhaustive du WoS, de Scopus et de Google Scholar en recherchant l'ensemble des citations des 15 membres de son laboratoire (School of Library and Information Science), donc d'un ensemble connu à l'avance. Une première comparaison concerne la facilité d'obtention de données fiables. Clairement la palme revient au WoS dont les données sont plus fiables et les outils plus performants, Scopus requérant deux fois plus de temps et Google Scholar 30 fois plus (avec en prime 12% de citations de son laboratoire non présentes). Une autre constatation a été l'utilité d'avoir la liste des citations connue à l'avance

¹²nombre de citations trouvées pour Arthritis and environment, 1992-date : 405 (WoS), 395 (Scopus), pour Tubercular meningitis, 1992-date : 21 (WoS), 76 (Scopus), pour Barik S*, 1992-date : 93 (WoS), 126 (Scopus) et pour American Journal of Cardiology, 1992-date : 15 619 (WoS), 21 993 (Scopus)

en raison de nombreuses homonymies : ceci permet de corriger manuellement des citations qui avaient été attribuées à d'autres auteurs ayant le même nom ou des noms voisins. Pour les domaines considérés, Scopus donne 13.7% de citations de plus que le WoS et l'ensemble des citations du WoS et de Scopus représentait 35.1% de citations de plus que le WoS seul. Encore plus préoccupant est que ce pourcentage est très variable selon le domaine de recherche, allant de 4.9% à 98.9%. Seules 58.2% des citations étaient communes entre le WoS et Scopus, une des raisons majeures étant que Scopus fournit considérablement plus de citations de conférences que le WoS.

Google Scholar quant à lui trouve 53% de citations de plus que WoS et Scopus combinés, permettant d'accroître le nombre de citations du laboratoire de 93.4%. Même si certaines de ces citations viennent de journaux avec un JIF faible, une grande partie d'entre elles proviennent de conférences (trouvées d'ailleurs principalement sur les pages personnelles des auteurs). Mais là aussi on observe de très fortes variations de l'augmentation du nombre de citations trouvées selon les domaines (allant de 143.6% pour human-computer interaction à 23.4% pour bibliometrics). L'intersection entre Google Scholar et l'union du WoS et de Scopus est très faible (30.8%) et Scholar ne trouve pas 40.4% des citations du WoS et de Scopus. Par contre l'introduction des données de Scholar ne modifie pas sensiblement le classement des chercheurs que l'on obtient en considérant ensemble le WoS et Scopus.

Finalement Meho indique une bien meilleure couverture que WoS ou Scopus de la littérature non anglo-saxonne par Google Scholar avec 6.95% de citations trouvées (respectivement 1.14% et 0.7%).

Pour ce qui concerne l'informatique le Centre for Science and Technology Studies (CWTS) de l'Université de Leiden a constaté que les actes de conférences étaient un élément majeur de dissémination en informatique et a donc entrepris de compléter le WoS en incluant les Lectures Notes in Computer Science (LNCS) de Springer, les conférences d'ACM et les conférences de la Computer Society d'IEEE, augmentant ainsi de 66% la taille de la base de données [12]. Il a été cependant constaté que cet ajout nécessitait un gros travail en particulier pour l'extraction des données des fichiers PDF. Pour mesurer la qualité de la couverture de cette base étendue le CWTS a examiné les citations des articles de la base qui font référence à des articles de la base, ce qui constitue une mesure de la couverture interne : un pourcentage de 100% indiquerait que l'ensemble des articles de la base ne font référence qu'à des articles de la base. Cette couverture interne est de 51% (une augmentation substantielle puisque la couverture interne du WoS en informatique est de 38%, bien moindre qu'en physique ou chimie pour lesquelles elle excède 80%) mais reste encore selon le propos des auteurs "modérée" et qu'il est improbable que l'ajout de conférences pourrait faire monter cette couverture à plus de 80%.

Des responsables de laboratoires hollandais ont réagi en mentionnant des publications qu'ils considéraient comme importantes et qui ne figuraient pas dans la base étendue. Une recommandation du rapport du CWTS est par conséquent qu'une étude bibliométrique en informatique devrait reposer sur une liste de publications qui serait fournie par les chercheurs eux-mêmes ou au moins vérifiée par eux (cette recommandation rejoint d'ailleurs le travail en cours de la CE dont l'objet est d'identifier les journaux et conférences jugées importantes par les chercheurs de l'Institut). Une seconde recommandation est qu'en raison de la diversité des sources de ci-

tation (par exemple en dehors des journaux et des conférences très formatées), il paraît très difficile d’automatiser le comptage des citations et qu’une analyse manuelle est nécessaire : ceci constitue un travail très lourd comme l’atteste la section 9 qui présente les résultats d’une étude préliminaire concernant pourtant uniquement quatre chercheurs de l’INRIA et qui confirme, en pire, les conclusions de Meho.

Cette incohérence des sources jette clairement un doute sur la valeur des indicateurs qui sont publiés à partir d’elles. On peut se poser la question de la validité de publier un JIF avec 3 décimales à partir du moment où l’on est susceptible d’avoir omis plus de 20% des citations des journaux traités (un chiffre pourtant prudent si l’on en croît l’article de Meho), et que cette incertitude va jouer dès la première décimale du JIF.

4.4 Les problèmes d’attribution

- Les indexeurs sont confrontés à deux types de problèmes lorsqu’ils examinent les articles :
- à quel organisme doit on attribuer l’article : le seul élément identifiant est l’adresse des auteurs dont la complexité rend nécessaire un dépouillement manuel (donc peu compatible avec le traitement automatisé des données proposé par les sources de citations). Ce problème est traité dans la section 8
 - pour les articles à plusieurs auteurs comment attribuer le travail entre les auteurs : le plus simple est de répartir la publication à part égale entre les auteurs (ce qui semble être la méthode utilisée par le WoS¹³). Il est cependant très difficile d’obtenir les règles qui régissent les attributions de citations. On peut essayer de plus de prendre en compte les pratiques de publication mais elles sont extrêmement diverses selon les domaines, ce qui en rend le traitement automatique très difficile.

5 Les indicateurs

Il existe une multiplicité d’indicateurs bibliométriques et leur nombre ne fait que croître (probablement un corollaire de leur utilisation accrue)¹⁴. Il n’est donc pas question d’en faire ici un inventaire exhaustif mais plutôt d’examiner les plus utilisés, et qui ont donc été soumis à des analyses poussées, ainsi que de décrire des tendances pour les nouveaux indicateurs.

Tous ces indicateurs reposent sur une source de citation. Ils héritent bien évidemment des problèmes liés à leur source, mentionnés dans la section précédente.

5.1 Les indicateurs pour journaux

L’évaluation de la qualité scientifique d’un article est un problème délicat. Une approche simple consiste à relier la qualité d’un article à la qualité du support dans lequel il a été publié. On substitue ainsi une évaluation du support (le plus souvent un journal scientifique) à une

¹³ *Citations to each article from any other article in the dataset are counted, and each indexed author is credited with the full tally of citations to that article. Thus, an article with ten authors which has been cited 100 times, will attribute 100 citations to each of the ten indexed author names.*

¹⁴ Une littérature abondante sur ce sujet est disponible à indicascience.veille.inist.fr

évaluation individuelle, ce qui évidemment simplifie énormément le travail puisqu'il y a bien moins de supports que d'articles. Cette approche constitue le modèle initial de l'ISI qui a cependant un peu évolué sous la pression des organismes clients.

5.1.1 L'impact factor d'un journal

Le *journal impact factor* (JIF) est un indicateur proposé par l'ISI dans son JCR. Il est calculé à partir du WoS. C'est au départ une notion inventée au début des années soixante par Gene Garfield, le fondateur de l'ISI. Le JIF d'un journal à l'année n est défini comme le rapport entre le nombre de citations dans l'année n d'articles du journal parus dans les années $n - 1$ et $n - 2$ (et uniquement celles-ci), et le nombre total d'articles publiés pendant ces deux années :

$$JIF_n = \frac{C_{n-1} + C_{n-2}}{P_{n-1} + P_{n-2}}$$

La limitation à deux années semble mystérieuse¹⁵. Le JIF est souvent considéré comme un indicateur de la qualité d'un journal et joue un rôle considérable dans le monde scientifique¹⁶. L'ISI indique cependant que le JIF ne doit pas être utilisé pour des domaines différents. En particulier, le JIF dans un domaine impliquant des recherches longues sera automatiquement plus faible que celui des journaux d'un domaine à évolution rapide. Il a été prouvé par exemple que ce fait expliquait qu'en moyenne les journaux de biologie moléculaire (domaine où un article devient rapidement obsolète) avaient un JIF bien plus élevé que la moyenne des journaux de mathématiques [15]. En 1999 le meilleur JIF en mathématiques correspondait à celui du 51ème journal en biologie cellulaire et le papier d'Andrew Wiles sur le théorème de Fermat ne contenait que 4 références sur 84 à des publications qui avaient été publiées dans les deux années précédentes [16]. Des analyses récentes montrent que cette tendance ne s'est pas inversée. Par exemple l'étude du JIF de 181 journaux en mathématiques et de 124 journaux en génétique sur la base du JCR 2005 a montré que si la distribution était comparable, la valeur moyenne du JIF variait d'un facteur 10 entre les deux disciplines, au profit de la génétique [7]¹⁷.

Un autre argument logique contre l'utilisation du JIF pour des comparaisons inter-domaines est que des journaux d'un domaine à faible contenu scientifique peuvent cependant avoir un JIF du même ordre que celui d'un autre domaine avec un haut niveau de recherche puisque le JIF ne dépend que des citations. Réciproquement un domaine très actif peut avoir des journaux avec un JIF faible en raison de pratiques de citations différentes¹⁸ ou une communauté réduite. À ce propos il est souvent indiqué que la taille de la communauté (mesurée en nombre d'articles publiés) influence largement le JIF. Ceci n'est pas exact pour tous les domaines mais de nombreux autres facteurs peuvent influencer de manière arbitraire le JIF : par exemple l'augmentation du nombre d'articles publiés dans un journal où la recherche est très active

¹⁵ISI a proposé récemment le mode d'emploi pour calculer un JIF utilisant une fenêtre de 5 ans

¹⁶*Editors get hired and fired soon after the annual JIF data are published*, Peter's Digital Reference Shelf, August publié chez Thomson Gale qui est une compagnie affiliée à Thomson Institute for Scientific Information (ISI)

¹⁷E. Garfield attire d'ailleurs l'attention sur l'absurdité des comparaisons reposant sur le JIF : *it is absurd to make invidious comparisons between specialist journals and multi-disciplinary general journals like Nature and NEJM. To compare journals you should stick to a particular category as is explained very carefully in the Guide to JCR*, Der Unfallchirurg, 48(2) p.413, June 1998

¹⁸par exemple le taux moyen de citation en 2000 pour la pharmacologie était de 11 alors qu'en génétique il était de 28 [13]

va usuellement conduire à un JIF plus bas car le dénominateur du JIF augmentera alors que les papiers seront la plupart du temps cités en dehors de la fenêtre temporelle utilisée pour calculer le numérateur du JIF [16]. La nature même du matériel publié dans un journal donné peut introduire un biais dans le JIF estimé de 5 à 40%, voir section 6.1.

D'autre part le JIF mesure la *moyenne* des citations d'un journal mais il faut savoir que même pour les journaux à fort JIF les citations proviennent d'au plus 15% des articles publiés [15] : le JIF ne mesure donc pas vraiment la qualité d'un article ou d'un auteur spécifique.

Étant basé sur le WoS, le JIF est souvent accusé d'un biais américain, même dans la communauté anglo-saxonne¹⁹. C'est dû au fait que l'on cite plus facilement des articles de son propre pays²⁰ et ce biais a été confirmé pour des cas particuliers (voir la section 4.3).

À noter qu'en dépit d'une moins bonne couverture par l'ISI des sciences de la vie (voir section 4) celles-ci tiennent une place à part pour le JIF : 12 des 15 journaux avec un JIF supérieur à 10 relèvent de cette discipline. Le caractère relativement généraliste des journaux en science de la vie leur donne un avantage certain par rapport aux bonnes revues spécialisées qui sont très peu citées en dehors de la communauté concernée.

5.1.2 Indice d'immédiateté, immediacy index

Cet indice est aussi proposé par l'ISI dans son JCR. Il est défini comme le rapport entre le nombre de citations d'articles parus à l'année n (et uniquement celle-ci) et le nombre d'articles publiés dans le journal cette même année :

$$II = \frac{C_n}{P_n}$$

Cet indice est souvent considéré comme une mesure de l'impact immédiat d'un journal. Il apparaît cependant que dans bien des cas les journaux présentant des indices d'immédiateté élevés obtiennent ce chiffre en raison d'un nombre important de références à des éditoriaux, qui n'apparaissent pas dans le dénominateur de cet indicateur.

5.1.3 Demi vie des citations, Cited half-life

Cet indice est aussi proposé par l'ISI dans son JCR. Pour l'année n le *cited half-life* est le nombre d'années j tel que 50% des citations du journal de l'année n sont des citations antérieures à l'année $n - j$ et 50% sont ultérieures. Ainsi *Nature Genetics* avait un cited half-life de 4.7 en 2003 car 46.38% des citations de l'année 2003 dataient d'années antérieures à 1999. Cet indicateur fournit donc une information sur la *permanence* des recherches dans un domaine donné.

Les indicateurs comme le JIF ne prenant en compte que des citations relativement récentes, les journaux qui ont des cited half-life plutôt petits vont avoir mécaniquement des JIF plus élevés que ceux avec un cited half-life important.

¹⁹guide "Journal Impact Factors", Juin 2004, The University of Auckland, New-Zealand, www.boulder.nist.gov/div853/Publication%20files/journalimpactfactors.pdf

²⁰une étude a montré que les auteurs d'articles de biomédecine américains ne citent en moyenne que 12% de références étrangères de moins de 2 ans et 30% de moins de 10 ans (un taux très inférieur au poids réel de la littérature non américaine), *N. Pinhas, Inserm Actualités, 154, Septembre 1997*

5.2 Les téléchargements

Le mouvement des éditeurs vers l'accès en ligne permet de mettre en œuvre un autre indicateur : le nombre de télé-chargements. Cette information a l'avantage de pouvoir être obtenue en temps réel et d'après Meho une corrélation a pu être établie entre le nombre de téléchargements et le nombre de citations, quoique ce degré de corrélation varie sensiblement selon les disciplines. Dans ces conditions le nombre de téléchargements fournirait une estimée initiale du nombre de citations à venir des articles.

Des restrictions s'appliquent cependant à cet indicateur :

- il est difficile à établir pour un auteur donné susceptible de publier dans des revues très diverses car il serait très coûteux d'examiner l'ensemble des journaux d'une base de données pour chaque auteur
- il ne prend pas en compte les nouveaux moyens de diffusion utilisés par les chercheurs (pages personnelles, archives ouvertes)
- la pratique de certains éditeurs de mentionner les articles les plus téléchargés favorise mécaniquement ces articles
- la fiabilité est relative. Ces indicateurs sont calculés par les éditeurs eux-mêmes, ce qui constitue un conflit d'intérêt évident

5.3 Les indicateurs quantitatifs

Ce sont les indicateurs les plus faciles à établir à partir des bases de données de citations. Mentionnons comme exemple :

- nombre de publications et de citations pour un groupe de chercheurs
- nombre de publications et de citations par chercheur pour un groupe de chercheurs
- pourcentage de la production mondiale
- nombre de publications intervenant dans les indicateurs ISI
- nombre de publications dans les journaux à fort JIF

Il va de soi que les trois premiers indicateurs ne donnent aucune information sur la qualité du travail scientifique : ils permettent tout au plus d'évaluer si le groupe a une activité de publication "normale", ce qu'il faut placer bien sûr par rapport à la moyenne d'activité des autres groupes travaillant dans le même domaine. Les deux suivants ont une validité corrélée à celle des indicateurs ISI pour le domaine considéré.

Une information intéressante pour l'informatique est fournie dans le rapport du CWTS de l'Université de Leiden [12] : l'évaluation par les pairs de 42 laboratoires en informatique hollandais, tel qu'établie par le Review Committee for Computer Science de la Quality Assurance Netherlands Universities (QANU) a une corrélation certes positive mais faible avec les indicateurs mentionnés ci-dessus. Cette constatation impose donc une réflexion : l'évaluation par les pairs mesure-t-elle mal l'impact des laboratoires ou les indicateurs sont-ils peu appropriés à la mesure d'impact ?

5.4 Les indicateurs individuels

Les organismes utilisant les indicateurs sont demandeurs de mesures qui permettraient une évaluation individualisée des chercheurs, ce qui n'est pas le but des indicateurs de journaux²¹. Dans ce contexte les indicateurs de journaux ne leur permettent qu'une vue indirecte ne conduisant pas aux valeurs chiffrées qu'ils recherchent. Les scientifiques objectent aussi que les analyses en moyenne des indicateurs de journaux ne peuvent pas refléter la qualité d'un élément particulier. Les exemples sont nombreux d'articles publiés dans un journal avec un JIF faible et constituant pourtant une contribution majeure à la science contemporaine et, à l'inverse, d'articles de mauvaise qualité, ou purement polémistes, publiés dans des journaux avec des JIF élevés. En conséquence une tendance actuelle (souvent critiquée d'ailleurs par les bibliométriciens professionnels) est de proposer des indicateurs censés évaluer la qualité scientifique du travail d'un individu.

5.4.1 Le H-index

Le nombre H d'un auteur a été défini par J.E. Hirsch [6] comme le nombre h d'articles de l'auteur qui ont été cités au moins h fois chacun. Il a été proposé comme alternative à d'autres indicateurs (dont les avantages et inconvénients sont repris du papier de Hirsch)

- *nombre total de papier* : qui mesure la productivité mais pas l'impact
- *nombre total de citations* : qui mesure une forme d'impact total mais peut être fortement influencé par le nombre de co-auteurs et par les articles de synthèse
- *citations par papier* : permet des comparaisons entre scientifiques d'âges différents ; difficile à estimer, récompense une productivité faible et pénalise une forte productivité
- *nombre de publications significatives* : nombre de papiers cités plus de y fois, n'a pas les inconvénients des indicateurs précédents mais souffre de l'arbitraire dans le choix de y (on pourrait aussi mentionner la difficulté de la mesure)
- *nombre de citations aux q papiers les plus cités* : n'a pas les inconvénients des précédents indicateurs mais difficile à obtenir et difficulté du choix de q

Hirsch indique que la mesure du nombre H peut être facilement obtenue à partir du Web of Science en utilisant l'ordre *times cited* que propose l'ISI. Mais cela suppose une bonne couverture de l'ensemble des domaines scientifiques (ou au moins une couverture homogène si l'on veut comparer des individus) par le Web of Science, ce qui est loin d'être le cas pour les disciplines de l'INRIA. Hirsch va même plus loin en proposant des valeurs du nombre H pour obtenir la tenure, devenir full professor ou membre de l'Académie des Sciences des USA. Ce n'est qu'à la fin de son papier qu'il aborde brièvement le problème de la valeur du point H selon les disciplines (*h-indices in biological sciences tends to be higher than in physics. . . more research in understanding similarities and differences . . . in different field of science would be of interest*).

Parmi les nombreux problèmes que soulève cet indicateur mentionnons la fiabilité avec laquelle il peut être mesuré. L'ISI a initialement refusé de mettre à disposition cet indicateur, ce

²¹E. Garfield est d'ailleurs très clair sur ce point : *The source of much anxiety about Journal Impact Factors comes from their misuse in evaluating individuals, e.g. during the Habilitation process. In many countries in Europe, I have found that in order to shortcut the work of looking up actual (real) citation counts for investigators the journal impact factor is used as a surrogate to estimate the count. I have always warned against this use.* Der Unfallchirurg, 48(2) p.413, Juin 1998.

qui a conduit à la mise en service d'outils "bricolés" reposant sur les données de Google Scholar, qui outre le fait d'avoir des sources mal identifiées, gère mal les homonymies et les auto-citations. Pour des auteurs ayant des noms relativement communs les résultats obtenus via Google Scholar sont souvent fantaisistes (sans faire une recherche exhaustive on peut ainsi montrer que des auteurs initialement qualifiés d'un H-index de 35, très élevé, voyaient ce nombre ramené à 5 (moyen) dès que l'on examinait plus attentivement les citations trouvées dans Google Scholar). Meho prétend que le nombre H est maintenant facilement calculable à partir du Web of Science (en Octobre 2006 l'ISI a changé de politique et propose le nombre H dans son JCR), Scopus et Google Scholar. Ce n'est pas notre avis : la diversité des réponses dans ces trois bases de données nécessite des recoupements très fins et une analyse manuelle de la validité des données avant de pouvoir obtenir une valeur à peu près correcte de l'indicateur au sens de sa définition (sans présumer de sa validité).

De plus les sources utilisées gèrent très mal les références à des livres ou chapitres de livres, ce qui pénalise fortement les auteurs de livres de référence. On peut ainsi trouver des exemples d'auteurs ayant écrit très peu d'articles mais dont les livres sont reconnus et qui ont un nombre H très faible : seul l'indicateur *max citations* qui est fourni en plus du nombre H peut alors indiquer la discordance entre le nombre H et l'influence réelle de l'auteur, ce qui nécessite alors un examen très attentif des citations elle mêmes.

Une analyse critique du nombre H a été publiée par H.L. Roediger [14] qui a relevé les caractéristiques suivantes :

- le nombre h est corrélé avec l'âge
- retard : le nombre h peut substantiellement augmenter même si le chercheur n'est plus actif depuis longtemps
- le nombre h est sous-estimé pour les chercheurs ayant publié des livres
- le nombre h ne met pas en valeur les contributions très importantes d'un auteur
- le fait que les citations soient attribuées à l'ensemble des auteurs ne permet pas de prendre en compte les pratiques des domaines, qui sont très variées. L'ordre des auteurs reflète souvent l'importance des contributions. Mais, par exemple Roediger, psychologue, indique que dans son domaine les contributions du premier et dernier auteurs sont les plus importantes. Dans d'autres domaines, l'ordre alphabétique est la règle, ou bien le premier auteur est celui qui a apporté la contribution la plus importante.

On peut mentionner d'autres inconvénients majeurs du nombre H : les citations négatives (c'est-à-dire très critiques vis-à-vis d'un papier) ne sont pas prises en compte, il ne prend pas en compte les travaux très cités et il ignore le nombre total de citations.

5.4.2 Les variantes du H-index

Pour combler les lacunes du H-index, des variantes ont été proposées

- *a-index* : le nombre moyen de citations pour les articles retenus dans le calcul du nombre H
- *g-index* : proposé par Egghe [3], c'est le nombre g d'articles dont la somme des nombres de citations est au moins g^2 (un g -index de 10 indique que l'auteur a écrit 10 papiers dont la somme des citations est au moins de 100)

Hirsch a d'ailleurs encouragé le développement de ces indicateurs alternatifs. On peut toutefois estimer que la multiplication des indicateurs sans un minimum d'analyse critique fait planer un large doute sur l'effectivité de la relation entre qualité scientifique et indicateurs.

5.5 Les autres indicateurs

Des indicateurs plus sophistiqués, reposant toujours sur les citations, ont été proposés pour prendre en compte trois biais possibles :

- année de publication : les publications plus anciennes sont davantage citées
- le type de document : le nombre de citations varie considérablement selon ce type, par exemple les articles de “review” sont en général plus cités que les articles scientifiques
- le domaine : les pratiques de publication diffèrent considérablement selon les domaines scientifiques.

Les indicateurs avancés reposent sur un processus de normalisation qui tente de corriger ces biais. Les indicateurs avancés les plus connus sont le *field normalized citation score* dont une forme est le *crown indicator* qui compare le nombre moyen de citations attribué à une “unité” (un chercheur, un laboratoire) au nombre moyen de citations dans les publications internationales de la même année, dans le même domaine et sur le même type de document. Un crown indicator de 0.9 indique que les publications analysées sont citées 10% moins que la moyenne mondiale.

Un autre indicateur avancé est le *Top 5%* qui calcule, pour un groupe d'auteurs, la part qui est dans les 5% des articles les plus cités dans le monde la même année, dans le même domaine et pour le même type de document. Une valeur supérieure à 1 indique que le groupe a plus de publications dans le groupe des 5% des publications les plus citées que la moyenne mondiale.

La volonté louable de prendre davantage en compte les spécificités des domaines scientifiques pour produire ces indicateurs avancés se heurte à plusieurs difficultés :

- la définition du domaine fait appel à beaucoup de subjectivité (à quel niveau de granularité est-il nécessaire de descendre pour réellement refléter un domaine ?)
- le traitement des données ne peut qu'être manuel (par exemple un journal peut couvrir plusieurs domaines, il est donc nécessaire d'en trier les citations)
- l'imprécision importante des sources de citation intervient à deux niveaux, sur le calcul de la moyenne mondiale et sur les citations du groupe considéré.

L'utilisation unique des citations pour évaluer la qualité scientifique d'un article et son impact suscite beaucoup d'interrogations dans la communauté scientifique, celle qui est évidemment le plus au courant des biais que cette méthodologie peut créer. Cette interrogation a suscité des initiatives pour proposer d'autres modèles. Par exemple en biologie et en médecine dans le service *Faculty of 1000*²² proposé par *BioMed Central* l'analyse d'un article repose sur la lecture coopérative des articles par un groupe de spécialistes coopté dans un domaine précis.

²²www.facultyof1000.com

5.6 Complémentarité et cohérence des indices

Une étude [8] (relativement ancienne cependant) des indicateurs du JCR sur 19 ans a montré que les indicateurs des journaux pouvaient être distribués en quatre groupes : bas, central, haut, extrême (journaux avec des indicateurs très différents des autres). Les groupes haut et extrême ne représentent que 10% du total alors que 50% sont dans le groupe central. Ainsi, en 1994, 6.7% des journaux avaient un JIF supérieur à 3, 27% un JIF entre 1 et 3, le reste ayant un JIF inférieur à 1. À noter que les journaux dans lesquels publient les chercheurs de l'INRIA sont quasiment uniquement des journaux des 2ème et 3ème groupes, donc ayant un JIF entre 1 et 3. Il faut effectivement porter très peu d'attention aux décimales indiquées dans les indicateurs comme le JIF : d'un strict point de vue méthodologique seul le premier chiffre peut avoir un sens vis-à-vis des erreurs de mesure.

Pour ce qui concerne la cohérence on peut noter des discordances entre les opinions des experts et les classements que l'on pourrait établir à partir des indicateurs ISI. Par exemple les équipes de robotique de l'INRIA ont proposé en 2005 à la CE un classement des journaux du domaine²³. Ce classement est assez sensiblement différent de celui qui pourrait être établi à partir des indicateurs fournis par le JCR. Une étude de 1995 [5] a comparé le classement construit par 50 chercheurs des NIH et le JIF de l'ISI. Les 10 premiers journaux pour les experts avaient un rang de classement variant entre 3 et 160 dans le classement JIF, ce qui reflète bien que le JIF caractérise plutôt la visibilité des travaux que la qualité de la production scientifique. Une discordance similaire entre classement ISI et CiteSeer pour l'informatique a été présentée dans la section 4.2.1.

Des chercheurs de renom ont même proposé un moratoire sur l'utilisation des indicateurs dans le processus d'évaluation en attendant une révision et une réforme profonde [4].

5.7 Conclusions sur les indicateurs

Pour chaque indicateur, deux problèmes se posent. Le premier concerne le sens de l'indicateur lui-même et comment il doit être interprété. Le second concerne la façon dont il est calculé et si le nombre obtenu reflète bien le nombre attendu.

Répondre à la première question nécessiterait une recherche plus poussée que celle effectuée dans le cadre de ce rapport. On peut se demander pourquoi le JIF se contente des citations des deux années qui suivent la publication d'un article (ISI indique d'ailleurs une procédure permettant le calcul sur une fenêtre de 5 ans). Pourquoi 2 est plus pertinent que 1,3 ou tout autre nombre ? Le H-index est une fonction qui peut paraître ad hoc mais pourquoi cette fonction serait-elle meilleure qu'une autre ? On peut par exemple calculer l'importance d'un papier en fonction de l'importance des papiers qui le citent en faisant un point fixe, et obtenir l'importance d'une personne en fonction de l'importance des papiers qu'il/elle écrit. On peut ainsi imaginer des milliers de fonctions et seule une étude approfondie permettrait, peut être, d'en faire émerger les plus pertinentes.

²³Ce classement prend bien sûr en compte les spécificités des recherches menées à l'INRIA et l'on pourrait très bien concevoir qu'un autre classement soit établi dans un cadre différent

Enfin nous tenons à dire ici que tous les indicateurs majeurs, en particulier tous ceux présentés ici, sont des fonctions sur le nombre de citations d'un papier. Ils mesurent donc tous un facteur d'impact, qui vaut ce qu'il vaut et qui doit être utilisé au regard des réserves mentionnées dans ce rapport, mais qui reste un facteur d'impact et en aucun cas un facteur d'excellence scientifique.

Le point le plus inquiétant concerne le calcul de ces indicateurs. On l'a vu, les bases à partir desquelles il est calculé ont des limitations importantes aussi bien sur le corpus que sur la pertinence des résultats, en particulier pour Google Scholar et tout autre logiciel faisant du traitement automatique à partir de documents récupérés via Internet. Les nombres obtenus sont donc des approximations, souvent très grossières, du nombre attendu. Les indicateurs sont donc des nombres à la fiabilité douteuse.

6 Mauvais usages et effets pervers

L'importance des indicateurs a comme effet pervers que l'on peut être tenté de manipuler les indicateurs pour les faire croître sans pour autant améliorer en quoi que ce soit la qualité de la recherche. Nous mentionnons ici quelques manipulations faciles et qui ont été observées, sans prétendre à une quelconque exhaustivité.

6.1 Les manipulations du JIF

Une pratique courante dans certains domaines, et qui commence à se propager aux journaux qui publient des articles de chercheurs de l'INRIA, consiste à demander aux auteurs d'articles acceptés dans un journal d'inclure des références d'articles de ce même journal (ce qui constitue une forme *d'auto-citation*), sans qu'il existe une relation vraiment claire entre les références et le sujet de l'article²⁴. Une analyse exhaustive des références du WoS a montré un taux de référence erroné (références n'ayant pas de lien avec le sujet du papier) d'environ 7%, ce taux étant beaucoup plus élevé dans certaines situations spécifiques [11]. L'ISI est conscient de ce problème puisqu'il propose pour les journaux le *Self-Citing rate* qui est le rapport entre les citations publiées dans le journal et le total des citations du journal : on se trouve alors dans la situation paradoxale où un journal devrait interdire aux auteurs de citer des articles qu'il publie . . .

Une autre tactique classique consiste à jouer sur la définition même du JIF en publiant du matériel, comme des éditoriaux²⁵ qui pourront être comptabilisés dans le numérateur du JIF, mais qui ne le seront pas dans le dénominateur²⁶.

²⁴*Some authors need no arm-twisting, they volunteer to include reference to articles in the target journal even though they may not be relevant to the submitted papers*, Peter's Digital Reference Shelf, August publié chez Thomson Gale

²⁵qui peuvent eux-mêmes faire référence à des articles du journal : (*à propos d'un éditorial de Diagnostica*) *I must warn you that the text is about a half page and the rest is the listing of all articles published in Diagnostica in the previous two years and usually nothing else*, Peter's Digital Reference Shelf, August publié chez Thomson Gale

²⁶*take for example the paper about the accessibility of information on the Web (labeled as editorial material by ISI (note du groupe de travail : donc non compté dans le dénominateur du JIF)) from Nature. It was already cited by 325 article*, Peter's Digital Reference Shelf, August publié chez Thomson Gale

6.2 Le "saucissonnage" (*salami-slicing*)

Face aux définitions des indicateurs courants et à l'importance qu'ils prennent, les chercheurs pourraient être tentés de diviser leurs publications pour une recherche donnée en plusieurs articles contenant un élément particulier de leur recherche, seul l'ensemble des articles en question pouvant refléter l'importance du travail mené, assurant ainsi un nombre important de citations²⁷.

6.3 Les auto-citations

L'auto-citation fait référence au fait qu'un auteur peut faire référence à ses propres articles. Cette pratique n'est évidemment pas condamnable a priori mais pourrait influencer les indicateurs si elle devenait malicieuse. Différentes études ont montré que l'auto-citation n'apportait pas de changement majeur sur la valeur des indicateurs si l'on regardait un domaine assez vaste et sur une période assez longue, à part dans des domaines particuliers²⁸. On peut toutefois craindre que la pression de la gouvernance par les indicateurs conduise à des comportements qui viseraient à l'optimisation des indicateurs : il serait par exemple facile de mettre en place une stratégie de citations au sein d'un groupe, difficilement détectable, qui permettrait d'augmenter sensiblement le point h de chacun des membres du groupe.

6.4 Le frein à la prise de risque

Une prise en compte exagérée des indicateurs pour évaluer les recherches individuelles peut pousser les jeunes chercheurs à aller vers des domaines où il est possible d'obtenir des résultats rapidement au détriment de recherches de plus longue haleine²⁹. De même les indicateurs peuvent être un frein à l'innovation puisque souvent les innovations majeures sont le fait d'une communauté réduite et que, de plus, les indicateurs sont peu appropriés pour détecter rapidement des recherches innovantes³⁰ (par exemple un nouveau journal ne pourra avoir un JIF qu'au mieux trois ans après sa création).

7 Le coût des indicateurs et son implication

L'INRIA a acquis l'accès au WoS dans le cadre d'un groupement de commande (GC) national avec une cinquantaine d'établissements (EPST, Universités, ANR, OST...) sur 3 ans :

²⁷ *While the practice of "salami slicing" is frowned upon, it is widely practised by those who think that the length of their curriculum vitae measures its strength, moreover it is inevitable in reporting major studies...*, R G Newcombe and J Stebbing, Postgraduate Medical Journal 2006 ;82 :355-356

²⁸ *Self-citations that appear in prestigious high-impact economics journals have a statistically positive, but numerically small, effect on a subsequent article's total citation count and on the quality of the citing journal.*, M. H. Medoff, Scientometrics, 2006, vol. 69, no1, pp. 69-84

²⁹ *les moyens vont souvent chez les majoritaires ce qui favorise le conservatisme... Il est difficile de donner sa chance à un jeune talent, la tendance étant plutôt à l'immersion dans l'existant*, M. Vert, Lettres des Sciences Chimiques, 64, Janvier 1998

³⁰ *Une conséquence déplorable... pourrait être que les scientifiques, de manière à améliorer leur statut et/ou leur financement, soient tentés de négliger ou d'abandonner les thèmes de recherche moins populaires au profit des thèmes plus rentables. Cela mènerait à un appauvrissement de la connaissance scientifique dans bien des domaines virtuellement intéressants*, G. Roelants, INFI Newsletter, Décembre 1999

notre périmètre d'accès est limité essentiellement au Science Citation Index Expanded avec une profondeur d'archives limitée à 1991 ; en 2007 le périmètre a été élargi au Journal of Citation Report.

De plus, une deuxième convention a été passée entre le CNRS (opérateur du groupement de commande) et l'INRIA pour prendre à sa charge la moitié de notre facture : le GC nous permet cependant de payer ces ressources à 30% de leur tarif institutionnel.

Le double effet de cette économie et de la prise en charge du CNRS ramène à environ 6.000 euros par an la facture de l'INRIA pour le WoS (au lieu de 40.200 USD). Il faut depuis 2007 ajouter environ 1.000 euros par an pour le JCR, soit un total de 7.000 euros par an.

L'INRIA n'est pas actuellement abonné à Scopus et il est donc difficile d'estimer le coût d'un abonnement. Il est cependant généralement estimé que ce coût s'établit à environ 85 à 95 % de celui du WoS [2].

L'accès aux sources payantes est donc relativement onéreux alors que nous avons vu que le taux de couverture imparfait de ces outils devrait amener à multiplier les sources pour établir de manière raisonnable les indicateurs demandés et/ou pour contrôler les biais méthodologiques des calculs des indicateurs. De plus pour le cas particulier de l'INRIA l'étude préliminaire présentée en section 9 montre que les résultats obtenus via les sources payantes sont bien plus médiocres que ceux obtenus à l'aide des outils gratuits.

Dans un autre domaine il serait a fortiori pernicieux que les indicateurs soient utilisés pour guider les choix des abonnements dans les centres de documentation.

8 La normalisation des adresses

Une mesure qui permettrait aussi de rendre plus justes les indicateurs fournis par l'ISI serait de signaler à cette société le caractère spécifique de l'Institut et de normaliser les adresses des chercheurs de l'INRIA de façon à pouvoir compter comme publications INRIA les publications des équipes projets INRIA du LORIA ou de l'IRISA. Le CNRS a procédé de cette manière en faisant entrer dans les abréviations choisies par l'ISI le terme CNRS. Il conviendrait de vérifier si l'on ne pourrait pas faire de même pour l'INRIA.

Ce problème est identique pour le CNRS qui a défini pour ses besoins propres la notion de *publication des laboratoires du CNRS* comme toute publication dont l'un des signataires a donné comme adresse une unité soutenue ou évaluée par le CNRS, que le signataire soit salarié du CNRS ou non³¹. Cette notion ne résout cependant en rien le problème des indicateurs fournis par des organismes extérieurs comme l'ISI.

L'OST a d'ailleurs lancé le programme *Normaddresses*³² dont le but est de proposer à l'ISI une nomenclature des laboratoires français. Nous citons le site de cette initiative :

L'initiative de l'OST a pour objectif d'améliorer la qualité des traitements appliqués aux adresses françaises du Web of Science par Thomson Scientific. Elle doit s'accompagner de deux types d'action : en amont, un travail de normalisation des adresses des manuscrits qui sont envoyés par les chercheurs aux éditeurs de revues et, en aval, une familiarisation des utilis-

³¹ *Les publications des laboratoires du CNRS et leur impact*, Unité d'indicateurs de politique scientifique, Mars 1999

³² voir www.obs-ost.fr/e-docs/00/00/00/A0/document_actu.phtml

teurs d'indicateurs avec les différents modes de calcul utilisés par les producteurs. Tout d'abord, les chercheurs doivent être informés de l'importance d'un libellé complet des adresses dans les manuscrits qu'ils soumettent aux éditeurs : c'est une étape primordiale, car ce sont ces informations, nécessaires à un bon repérage de l'article, de ses auteurs et de leurs institutions de rattachement, qui seront transcrites dans l'article publié. L'attention des chercheurs doit cependant être attirée sur le fait que les éditeurs limitent souvent le nombre de caractères autorisés pour libeller l'adresse. L'étape suivante est celle de la création de la notice bibliographique. Lors de la mise en base des articles dans le Web of Science, Thomson Scientific n'élimine a priori aucune des institutions apparaissant dans les adresses, mais il en modifie parfois le libellé par souci de standardisation. De même, il arrive que l'éditeur change l'ordre des informations : entre institutions, et entre structures d'une même institution (institution, département, laboratoire, etc..). Le traitement des adresses "mixtes" (unités associées) pose des problèmes particuliers, pas toujours bien résolus par l'éditeur. En aval, la production d'indicateurs bibliométriques nécessite des méthodes rigoureuses. Ainsi, l'OST travaille ligne à ligne, avec chaque institution, au repérage de ses articles, et il tient compte des multi-affiliations d'unités mixtes, qu'il différencie des co-signatures. Cependant, de nombreuses études dans le monde reposent sur des choix de traitements automatiques qui ne repèrent que certaines institutions ou certaines occurrences de noms, ou qui limitent leurs repérages à la première institution mentionnée dans chaque adresse. Selon l'objectif poursuivi, ces choix sont ou non adaptés mais il est important, lorsqu'on utilise des indicateurs, de bien connaître les caractéristiques des traitements qui ont été employés pour leur calcul, afin de permettre une bonne interprétation.

Au vu de la complémentarité des sources il semblerait nécessaire de ne pas se limiter au WoS et proposer la même procédure à Scopus et à Google Scholar.

9 Un exemple d'analyse pour des auteurs de l'INRIA

9.1 Comparaison entre les sources

Nous avons effectué un travail préliminaire de comparaison de recherche des citations sur WoS, Scopus et Google Scholar (GS) pour quatre chercheurs de l'INRIA (dans l'ordre, en vision par ordinateur, en réseaux, en mathématiques financières et en informatique pure). Les données brutes ont été corrigées manuellement pour éliminer les doublons et corriger les titres (un article donné a fréquemment un titre orthographié de manière différente selon la base de données). Ce travail consiste à récupérer les données de citations, à les sauver dans des fichiers et à leur faire subir un premier traitement manuel (pour uniformiser les titres des articles) pour pouvoir effectuer ensuite un traitement automatisé. Ce traitement est lourd (environ 1/2 journée pour 2 chercheurs) mais nécessaire car les données brutes varient considérablement d'une base de données à l'autre (non seulement pour un article donné mais aussi sur les résultats comme nous le verrons). Le *Nombre total d'articles trouvés* indique le nombre total de papiers différents trouvés dans l'union des trois bases de données, *Articles trouvées dans* le nombre d'article trouvés dans une des bases de données et *Nombre de citations* le nombre total de citations

référéncées dans une base donnée. On obtient les résultats donnés dans la Table 1³³.

| Chercheur | 1 | 2 | 3 | 4 |
|--|------|-----|-----|------|
| Nombre total d'articles trouvés | 152 | 91 | 86 | 128 |
| Nombre total d'articles trouvés (avec au moins une citation dans GS) | 137 | 69 | 75 | 114 |
| Articles trouvées dans WoS | 31 | 13 | 16 | 13 |
| Articles trouvées dans Scopus | 52 | 23 | 27 | 4 |
| Articles trouvées dans Google | 140 | 89 | 77 | 126 |
| articles $\in WoS, \notin Scopus$ | 6 | 3 | 2 | 12 |
| articles $\notin WoS, \in Scopus$ | 27 | 12 | 13 | 3 |
| articles $\in WoS \cap Scopus$ | 25 | 10 | 14 | 1 |
| articles $\in WoS \cup Scopus$ | 58 | 25 | 29 | 16 |
| articles $\in WoS, \notin Google$ | 2 | 0 | 3 | 2 |
| articles $\notin WoS, \in Google$ | 111 | 76 | 64 | 115 |
| articles $\in WoS \cap Google$ | 29 | 13 | 13 | 11 |
| articles $\in Scopus, \notin Google$ | 11 | 2 | 9 | 0 |
| articles $\notin Scopus, \in Google$ | 99 | 69 | 59 | 122 |
| articles $\in Scopus \cap Google$ | 41 | 20 | 18 | 4 |
| Nombre de citations WoS | 237 | 163 | 104 | 78 |
| Nombre de citations Scopus | 652 | 334 | 122 | 52 |
| Nombre de citations Google | 1859 | 981 | 565 | 2324 |

TAB. 1 – Articles et citations trouvées dans WoS, Scopus et Google Scholar et dans différentes combinaisons de ces bases, à comparer au nombre total d'articles trouvés dans l'union des 3 bases (1ère et 2ème lignes).

À titre indicatif nous avons aussi consulté des bases de données plus spécialisées en informatique (et gratuites) : DBLP³⁴ (qui ne permet pas de trouver les citations) et CiteSeer. C'est

| Chercheur | 1 | 2 | 3 | 4 |
|---|-----|-----|---|-----|
| Articles trouvés dans DBLP | 54 | 19 | 1 | 39 |
| Articles trouvés dans CiteSeer | 58 | 49 | 1 | 42 |
| Nombre citations trouvées dans CiteSeer | 134 | 103 | 0 | 818 |

TAB. 2 – Articles et citations trouvées dans les bases de données DBLP et CiteSeer pour les 4 chercheurs.

évidemment pour le chercheur en informatique que les nombres d'articles trouvés par DBLP et CiteSeer sont les meilleurs (et très supérieurs aux résultats de WoS ou Scopus). Il reste cependant très en dessous de ceux de GS. Par contre pour les autres les résultats vont du moyen (chercheur 1 et 2) au très mauvais (chercheur 3).

Le pourcentage des articles indexés par les différents moteurs de recherche parmi l'ensemble des papiers indexables est donné en Table 3.

Il apparaît donc sur ces quatre exemples que l'utilisation conjointe de WoS et Scopus ne permet de repérer qu'environ 1/3 des articles susceptibles d'être cités et que c'est en informatique pure que le nombre est le plus faible (12.5%). On remarque aussi que pour les papiers trouvés avec une utilisation conjointe de WoS et Scopus on obtient une intersection relativement faible

³³Dans les tables le symbole \cup indique l'union, \cap l'intersection, \in l'appartenance, \notin la non appartenance

³⁴<http://www.informatik.uni-trier.de/~ley/db/>

| | WoS | Scopus | Google | WoS \cup Scopus | WoS \cap Scopus |
|-------------|--------|--------|--------|-------------------|-------------------|
| Chercheur 1 | 20.4% | 34.2% | 92.1% | 38.16% | 16.45% |
| Chercheur 2 | 14.3% | 25.3% | 98.9% | 27.47% | 11% |
| Chercheur 3 | 18.6% | 31.4% | 89.5% | 33.72% | 16.28% |
| Chercheur 4 | 10.15% | 3.12% | 98.44% | 12.5% | 0.78% |

TAB. 3 – Nombres de papiers trouvés dans chacun des moteurs de recherche rapportés à l’ensemble trouvé en combinant les résultats de WoS, Scopus et GS

entre les deux moteurs (typiquement 50% ou moins). Enfin, des papiers effectivement très cités peuvent ne pas apparaître dans une base.

On peut aussi remarquer que le nombre de citations trouvées varie considérablement selon le moteur de recherche. Il est clair que le balayage systématique du Web par GS va permettre de trouver un ensemble bien plus large de citations pertinentes que celui trouvé par les autres moteurs. Toutefois cette recherche souffre de nombreux défauts :

- pas de traitement des auto-citations
- tous les documents ont la même valeur : une citation dans un rapport de recherche de master est mis sur le même plan qu’un article de revue
- les documents comptant comme citations sont de natures très variables : un examen au hasard des citations fournies par GS nous a permis de trouver des articles non publiés, des rapports de contrats industriels et même des CV

Même si l’on exclut Google Scholar, on remarque des différences notables entre WoS et Scopus. Pour l’illustrer on prend pour le chercheur 4 les 10 papiers (les 4 pour Scopus) les plus cités dans une base de données et on examine le nombre de citations trouvées dans les différentes bases de données pour obtenir les tableaux suivants. On remarquera l’incohérence des résultats que ce soit dans les papiers indexés (papier très cité dans une base de données mais absent dans une autre) ou dans l’ordre en terme de nombre de citations. Par exemple, si Scopus et GS sont cohérents et classent premier le même papier (avec respectivement 44 et 208 citations, Tables 5,6) alors que ce papier n’apparaît pas dans le WoS et qu’il ne recueille que 2 citations pour CiteSeer (ce qui le situe bien au delà de la 10ème place selon cette base). Inversement le meilleur papier selon WoS (Table 4) n’apparaît même pas dans les 10 premiers selon GS et est absent chez Scopus.

| Numéro papier par rang de citation dans WoS | 1 | 2 | 3 | 4 | 5 | 76 | 7 | 8 | 9 | 10 |
|---|----|----|----|----|----|----|----|---|----|----|
| Nombre de citations WoS | 22 | 13 | 12 | 11 | 7 | 5 | 4 | 2 | 1 | 1 |
| Nombre de citations Scopus | - | - | 2 | - | - | - | - | - | - | - |
| Nombre de citations GS | 67 | 62 | 67 | 67 | 85 | 38 | 47 | 8 | 14 | 47 |
| Nombre de citations CiteSeer | 22 | 10 | 23 | 6 | 26 | 31 | - | 0 | 2 | - |

TAB. 4 – Nombre de citations selon WoS pour les 10 papiers les plus cités du chercheur 4 selon cette base (2ème ligne) et nombre de citations de ces papiers selon les autres bases (lignes suivantes). Un "-" indique un papier ne figurant pas dans une base de données.

| Numéro papier par rang de citation dans GS | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|--|-----|-----|-----|-----|----|----|----|----|----|----|
| Nombre de citations GS | 208 | 160 | 111 | 101 | 95 | 85 | 83 | 74 | 73 | 70 |
| Nombre de citations WoS | - | - | - | - | - | 7 | - | - | - | 22 |
| Nombre de citations Scopus | 44 | - | - | - | - | - | - | - | - | - |
| Nombre de citations CiteSeer | 2 | 152 | 68 | 74 | 32 | 26 | 60 | - | 43 | 22 |

TAB. 5 – Nombre de citations selon GS pour les 10 papiers les plus cités du chercheur 4 selon cette base (2ème ligne). Un "-" indique un papier ne figurant pas dans une base de données.

| Numéro papier par rang de citation dans Scopus | 1 | 2 | 3 | 4 |
|--|-----|----|----|---|
| Nombre de citations Scopus | 44 | 5 | 2 | 1 |
| Nombre de citations WoS | - | - | 12 | - |
| Nombre de citations GS | 208 | 37 | 67 | 6 |
| Nombre de citations CiteSeer | 2 | - | 23 | - |

TAB. 6 – Nombre de citations selon Scopus pour les 4 papiers les plus cités du chercheur 4 dans cette base (2ème ligne). Un "-" indique un papier ne figurant pas dans une base de données.

Pour les 4 papiers de ce chercheur les plus cités respectivement dans les 4 bases de données, on obtient le nombre de citations indiqué dans la Table 7.

| Classement du papier | 1 | 2 | 3 | 4 |
|------------------------------|-----|-----|-----|-----|
| Nombre de citations WoS | 22 | 13 | 12 | 11 |
| Nombre de citations Scopus | 44 | 5 | 2 | 1 |
| Nombre de citations GS | 208 | 160 | 111 | 101 |
| Nombre de citations CiteSeer | 152 | 74 | 68 | 60 |

TAB. 7 – Nombre de citations du chercheur 4 pour les 4 papiers les plus cités selon chacune des 4 bases de données. Dans une colonne donnée, le papier considéré n'est pas forcément le même selon la base de données.

9.2 Recherche institutionnelle

La recherche institutionnelle ne devrait pas non plus donner de bons résultats puisque pour les 4 chercheurs concernés nous avons trouvé dans le WoS les 9 affiliations suivantes :

INFIA ROCQUENCOURT (sic!)
 INRIA
 INRIA RENNES
 INRIA RHONE ALPES
 INRIA ROCQUENCOURT
 INRIA SOPHIA ANTIPOLIS
 INST NATL RECH INFORMAT & AUTOMAT
 INST NATL RECH INFORMAT & AUTOMAT ROCQUEN COURT (sic!)

NAT RES INST COMP SCI & CONTROL

Scopus fait légèrement mieux avec "seulement" 6 affiliations :

INRIA

Inst Natl de Recherche en (sic!)

I.N.R.I.A.

Inst. Natl. Rech. Info. et Automat.

LORIA

LIFIA-IMAG and Iuria Rhone-Alpes (sic!)

Il est impossible de faire sur GS une recherche par affiliation pour ces auteurs.

9.3 H-index

Il est enfin intéressant de présenter le h-index de ces quatre chercheurs, calculé (après un tri manuel) sur chacun des trois moteurs de recherche, Table 8. La disparité dans les valeurs

| Chercheur | 1 | 2 | 3 | 4 |
|----------------|----|----|----|----|
| h-index WoS | 8 | 5 | 6 | 5 |
| h-index Scopus | 14 | 6 | 6 | 2 |
| h-index GS | 21 | 12 | 12 | 25 |

TAB. 8 – H-index des 4 chercheurs INRIA sélectionnés selon les différentes base de données

de cet indicateur reflète évidemment la faiblesse marquée dans la recherche de citations et tout indicateur reposant uniquement sur les citations sera soumis aux mêmes types d'aléa de grande ampleur.

10 Conclusion

Les indicateurs sont des instruments de mesure importants mais qui doivent être utilisés de manière éclairée en respectant des règles de base :

- Les indicateurs actuels sont essentiellement des facteurs d'impact ce qui n'est pas synonyme de qualité.
- Il faut simplement s'attacher aux ordres de grandeur des indicateurs car même l'utilisation de sources diverses ne permet pas d'obtenir une grande précision.
- Il est nécessaire de corriger les indicateurs par l'avis d'experts.
- Il est nécessaire d'utiliser plusieurs indicateurs³⁵. Par exemple pour un domaine donné on ne peut dissocier le JIF du cited half-life.
- Il est nécessaire de recouper des sources diverses pour obtenir une information pertinente.

³⁵selon Meho *relying exclusively on Web of Science and a single citation measure will, in many cases, no longer be an option for making accurate impact assessments*

- Plutôt que de mesurer la qualité des journaux par des indicateurs chiffrés, il conviendrait de les classer en groupes sur la base de critères qualitatifs recueillis auprès des chercheurs eux-mêmes. Un travail dans ce sens est en cours : la Commission d'Évaluation de l'INRIA cherche à dresser un guide des publications de référence en faisant appel à l'expertise des projets.
- Ne jamais faire de comparaison inter-domaines.

Ces recommandations rejoignent en partie celles proposées par M. Zitt et G. Filliatreau (directrice de l'OST) à propos des biais et limites des indicateurs qui [19] :

- ne rendent compte que d'une partie du spectre d'activité
- doivent être calibrés pour "comparer des choses comparables"
- sont peu adaptés à l'observation des émergences
- requièrent une diversité d'angles d'attaques et de niveaux d'observation
- doivent être complétés, dans le cadre d'une évaluation, par d'autres éléments, en particulier l'avis de pairs.

Pour ce qui concerne l'INRIA on ne peut que prôner une meilleure utilisation des références du rapport d'activité. L'INRIA a la chance de disposer chaque année de la quasi intégralité de ses publications dans un document relativement structuré. P. Robert a montré que l'on pouvait aisément utiliser cette base pour établir des statistiques correctes sur les publications de l'INRIA. Un effort supplémentaire permettrait de corriger certains indicateurs.

En conclusion, on peut retenir la position saine donnée par Vinay et Baverel [17] : *il faut critiquer les [indicateurs], en même temps qu'on les considère ; il faut surtout rejeter la tentation qu'ils véhiculent d'un automatisme dans l'évaluation. . .*

11 Annexe 1 : les indicateurs de la LOLF

Voir le site www.obs-ost.fr/projet_u.phtml pour une description et les méthodes de comptage. On notera que les indicateurs bibliométriques sont très peu nombreux et résultent des travaux de la Mission Interministérielle sur la Recherche et l'Enseignement Supérieur (MIREES). Manifestement la Direction Générale de la Recherche et de l'Innovation (DGRI) désirait plus d'indicateurs³⁶ mais ce souhait s'est heurté à la crainte de ne pas pouvoir les renseigner.

Publications

- Part des publications de référence internationale dans la production scientifique de la France, de l'UE25, du monde (total et par discipline)
- indice de citation à deux ans (indice d'impact relatif), total et par discipline

Brevet

- nombre de brevets déposés respectivement à l'INPI et à l'OEB

³⁶Le rapport de l'IGAENR indique : *Observation de la DGRI : la réalité d'un opérateur et la mesure de son efficacité s'apprécient par un ensemble d'indicateurs, qui, regardés ensemble, permettent une expertise qualitative de l'organisme. Ex : aucun indicateur ne permet à lui seul de "mesurer" le niveau d'excellence d'un opérateur dans le champ de la recherche fondamentale. Toute une série d'indicateurs sont nécessaires : publications, citations, évaluations externes et internes, chercheurs accueillis (dont étrangers), nombre de chercheurs disposant de la HDR, nombre d'invitations dans des conférences internationales, distinctions, . . . De plus ces indicateurs devraient être mesurés quantitativement mais aussi qualitativement*

PCRD

- taux de participation dans les projets financés par les programmes cadre de l'UE (total et par finalité)
- taux de coordination dans les projets financés par les programmes cadre de l'UE (total et par finalité)

12 Annexe 2 : L'Observatoire des Sciences et Techniques (OST)

L'OST est un groupe d'intérêt public créé en 1990 et renouvelé en 2002 dont la mission est *de concevoir et produire des indicateurs relatifs aux activités scientifiques, technologiques et d'innovation et permettre leur interprétation en termes de position de la France dans l'Europe et dans le monde.*

Selon le site web de l'OST ses actions s'articulent autour de trois axes :

- la production d'indicateurs standards, de tableaux de bord, et d'un rapport biennal intitulé "Indicateurs de sciences et de technologies",
- la réalisation d'études portant sur des domaines de recherche, des secteurs technologiques, l'activité de R&D d'institutions, de pays, de régions...
- des travaux de recherche et développement sur les indicateurs, les méthodologies d'évaluation stratégique et les outils de pilotage.

L'OST a en particulier été chargé par le Ministère de produire les indicateurs de production scientifique et technologique des opérateurs du programme 150 "Formations supérieures et recherche universitaire"³⁷. Il est d'ailleurs intéressant d'examiner les indicateurs bibliométriques retenus par l'OST pour ce programme et la méthodologie utilisée. En premier lieu la base de données est unique (ISI) et la première phase du programme a consisté à repérer les opérateurs dans les listes institutionnelles fournies par ISI, manifestement de manière manuelle. Pour l'attribution il existe deux types d'attributions :

- comptage de présence : une attribution pour chaque opérateur apparaissant dans un article
- comptage fractionnaire : crédit au prorata des contributions des opérateurs, le prorata étant lié au nombre d'adresses des publiants et du nombre de disciplines dans lesquelles l'article est classé

Les indicateurs utilisés sont alors

- l'indice d'immédiateté à deux ans mais après répartition des articles dans 8 domaines (mathématiques, physique, médical, sciences pour l'ingénieur, chimie, biologie fondamentale, science de l'Univers, biologie appliquée). Il s'agit alors du rapport entre la part de citations de l'opérateur divisé par sa part de publications
- degré de visibilité : après une normalisation par discipline, spécialités et journaux, on définit des classes de visibilité (premier 5% d'articles les plus cités, 5% suivants, 10% suivants...). On définit alors l'indice d'activité par classe de visibilité en faisant le rapport entre le pourcentage d'articles de l'opérateur dans une classe par le pourcentage définissant la classe (par exemple si l'opérateur a 7.5% d'articles dans la classe des 5% d'articles les plus cités son indice d'activité pour cette classe sera de $7.5/5 = 1.5$)

³⁷www.obs-ost.fr/projet_u.phtml

L'OST indique cependant des réserves quant à l'utilisation de ces indicateurs pour une comparaison inter établissements : stabilisation des données, analyse des "usages" de publications et de citations au niveau des spécialités, analyse des données dans le contexte de l'établissement.

13 Annexe 3 : glossaire

- *article* : publication d'un texte scientifique dans un journal ou dans les actes d'une conférence. La notion d'article peut varier lorsque l'on recherche les *citations* selon la source qui établit les citations. Ainsi certains éléments de journaux comme les éditoriaux sont pris en compte pour l'énumération des citations mais pas en tant qu'articles publiés par le journal (ce qui intervient dans le dénominateur du JIF)
- *auto-citation* : pour un journal ce sont les citations dans le journal de travaux publiés dans le même journal. Pour un chercheur ce sont les citations d'un auteur à ses propres travaux
- *citation* : mention dans un *article* des travaux effectués par la communauté scientifique sur un sujet donné. Traditionnellement cette mention se fait en fin de l'article avec des renvois dans le corps du texte. Une citation peut être *positive* (les auteurs indiquent la qualité de la recherche) ou *négative* (les auteurs ne mentionnent l'article que pour en souligner la piètre qualité, voire pire)
- *indicateur* : élément (usuellement chiffré) censé fournir une indication pour l'évaluation d'une activité scientifique

Références

- [1] Burnham J.F. Scopus database : a review, Mars 2006. www.bio-diglib.com/content/3/1/1.
- [2] Dess H.M. Database reviews and report : Scopus, Hiver 2006. www.istl.org/06-winter/databases4.html.
- [3] Egghe L. Theory and practice of the g-index. *Scientometrics*, 69(1) :131–152, 2006.
- [4] Ergma E. et al. Science, morality and quality, Juillet 2002. ALLEA Biennial Yearbook.
- [5] Foster W.R. Impact factor as the best operational measure of medical journals. *Lancet*, 346(8985), 11 Novembre 1995.
- [6] Hirsch J.E. An index to quantify an individual's scientific research output, 29 Septembre 2005. arXiv :physics.
- [7] Leydesdorff L. Caveats for the use of citation indicators in research and journal evaluations, Avril 2007. Preprint, SIGMETRICS.
- [8] Magri M-H. et Solari A. SCI JCR : a tool for studying journals. *Scientometrics*, 35 :93–117, 1996.
- [9] Meho L.I. The rise and rise of citation analysis. *Physics World*, Janvier 2007.
- [10] Meho L.I. et Yang K. A new era in citation and bibliometric analyses : Web of Science, Scopus, and Google Scholar. *Journal of the American Society for Information Science and Technology*, 2006. a paraître, <http://arxiv.org/abs/cs/0612132v1>.

- [11] Moed H.F. The impact-factor debate : The ISI's uses and limits. *Nature*, 415(6873) :731–732, 2006.
- [12] Moed H.F., Visser M.S. Developing bibliometric indicators of research performance in computer science : an exploratory study *CWTS report 2007-1*, Leiden University, Février 2007
- [13] Rehn C. et al. Bibliometric handbook for Karolinska Institutet, 27 Novembre 2006.
- [14] Roediger H.L. The h index in science : a new measure of scholarly contribution. *The Academic Observer*, 19(4), Avril 2006.
- [15] Seglen P.O. Why the impact factor of journals should not be used for evaluating research. *British Medical Journal*, 314(7079) :458–502, 1997.
- [16] Sutherland W.J. What do impact factor tell us. *TREE*, 14(10) :382–384, 1999.
- [17] Vinay P. et Baverel G. Au fur et à mesure : à propos de l'évaluation numérisée de la production scientifique. *Médecine/Science*, 10 :701–703, 1994.
- [18] Zitt M. Facing diversity of science : A challenge for bibliometric indicators. *Measurement*, 3(1) :38–49, 2005.
- [19] Zitt M. et Filliatreau G. Bibliométrie et indicateurs : rôle de l'OST. In *Rencontres 2005 des professionnels de l'IST*, Nancy, 20-22/06/2005.