

# Indicateurs de suivi de l'activité scientifique de l'Inria

Préparé par Luce Brotcorne, Anne Canteaut, Aline Carneiro Viana, Céline Grandmont (coordinatrice), Benjamin Guedj, Stéphane Huot, Valérie Issarny, Guillaume Pallez, Valérie Perrier, Vivien Quema, Jean-Baptiste Pomet, Xavier Rival, Sylvain Salvati, Emmanuel Thomé (coordinateur)

La Commission d'Évaluation de l'Inria a mis en place début 2020, à la demande de la Direction Générale, un groupe de travail visant à réfléchir à des indicateurs pour l'analyse qualitative de l'activité scientifique de l'Institut. Ce document est le fruit des discussions de ce groupe, enrichies par les échanges avec la Direction Générale d'Inria et avec l'ensemble de la Commission d'Évaluation. Au-delà de cette « commande », il nous a semblé important de saisir cette opportunité pour mener une réflexion collective, qui dépasse le cadre de notre Institut, sur les indicateurs, et sur ce qu'ils peuvent ou ne peuvent pas dire de nos pratiques scientifiques.

## 1 Commande et objectif de ce rapport

### 1.1 Commande

L'objectif est de ce travail est de répondre au premier jalon du Contrat Objectifs-Performance (COP) de l'Inria<sup>1</sup> :

*Inria précisera le contenu du rapport d'analyse qualitative de son impact scientifique qu'il produira chaque année.*

Le rapport dont il est question est produit, notamment, à la destination des tutelles de l'Institut. Des précisions sur les indicateurs sur lesquels reposera cette analyse qualitative sont apportées page 49:

*Inria produira un document annuel présentant une analyse qualitative de l'impact scientifique par domaine de recherche, reposant notamment sur le nombre de publications dans une liste choisie de journaux et conférences avec un auteur membre d'une équipe-projet, la participation à des comités éditoriaux de revues phares, les bourses ERC, les prix internationaux, et de manière générale tout élément permettant d'apprécier qualitativement son impact scientifique. Même si elle n'est pas associée à un indicateur particulier, cette action associée au Jalon 1 du COP est structurante pour l'appréciation de l'impact scientifique d'Inria.*

Bien que l'objectif affiché soit clairement celui d'une analyse qualitative de l'activité scientifique de l'Institut, tous les indicateurs mentionnés en exemples dans ce paragraphe sont des indicateurs quantitatifs. En effet, il est ressorti de nos discussions avec la Direction Générale qu'une présentation de la production scientifique s'appuyant par exemple sur une sélection de faits marquants n'était pas souhaitée, et que cette « analyse qualitative » devait pouvoir faire l'objet d'une comparaison entre les valeurs présentées à l'année  $N$  et à l'année  $N + 1$ . Des éléments qui évaluent une sorte d'impact à 10, 15, 50 ans (tout ce qui est temporel) n'ont pas non plus vocation à rentrer dans le périmètre de cette étude.

Par ailleurs, des éléments comme le nombre de doctorants, le nombre de thèses soutenues, l'attractivité des postes ou le temps moyen passé en postes par les chercheurs fournissent naturellement des informations importantes sur l'activité scientifique d'un organisme de recherche.

1. Contrat Objectifs-Performance « Ambition 2023 » de l'Inria, signé le 18 février 2020. [lien](#).

Toutefois, de tels indicateurs relèveront du volet RH de la présentation de l’institut et n’ont donc pas vocation à entrer dans le périmètre de cette étude. Il en est de même pour certaines dimensions de l’activité scientifique, comme la valorisation ou la médiation scientifique, qui sont suivies par d’autres indicateurs du COP.

## 1.2 Objectif poursuivi par le groupe de travail

Jusqu’à présent l’indicateur utilisé pour mesurer l’activité scientifique de l’Inria se réduit au nombre total de publications de l’institut dans les journaux et dans les conférences. Dans ce contexte, l’objectif que s’est assigné le groupe de travail CE est donc de réfléchir à la possibilité de définir des indicateurs permettant de mieux rendre compte de la diversité de l’activité scientifique de l’Inria et de son impact, étant entendu que ces indicateurs chiffrés sont forcément réducteurs et *ne peuvent qu’être des éléments permettant d’illustrer/enrichir/étayer une analyse qualitative effective et argumentée que la direction de l’institut doit continuer de privilégier dans le dialogue avec nos tutelles.*

En effet des indicateurs chiffrés ne couvrent pas toutes les dimensions de la recherche scientifique et ne sont pas à même de mesurer, à eux seuls, l’impact, la qualité, la profondeur, l’originalité, la prise de risque de ces recherches. Le véritable impact d’un travail de recherche ne peut se réduire à un chiffre (comme un nombre de citations) ni à une série de chiffres, et il ne peut se mesurer qu’après un certain nombre d’années [WVS17]. Il est généralement difficile à prévoir à travers une analyse de la production scientifique de l’année en cours.

Conformément au COP, l’objectif de ces indicateurs est d’aider à analyser la production scientifique de l’institut (et son évolution au fil des années) et non de l’évaluer. Nous reviendrons plus en détail sur cette distinction importante dans la section 3.

Au delà du cadre de la « commande » du COP, il nous semble utile de **saisir cette opportunité pour mener une réflexion collective sur les indicateurs, et sur ce qu’ils peuvent ou ne peuvent pas dire de nos pratiques scientifiques**, à la suite du précédent rapport de la CE sur le sujet [KFM+07]. Un institut de recherche promeut des valeurs communes quant à l’évolution de l’activité scientifique. Sans que notre regard se limite à ce cas, mentionnons que c’est particulièrement le cas concernant la façon de publier les travaux scientifiques, qui est en pleine transformation. Dans ce document, nous avons également souhaité mettre en avant des indicateurs qui promeuvent ces valeurs communes, lorsque cela nous a semblé possible.

## 2 Utilisation et biais des indicateurs

Toute réflexion à propos de la définition d’indicateurs quantitatifs amène à s’interroger sur le cadre de leur utilisation, les biais qui y sont associés, et les effets pervers qu’ils produisent. Nous avons essayé de synthétiser ici quelques éléments de réflexion autour de ces trois problématiques, étudiées notamment dans [Gin16].

Nous commençons par présenter les principes qui devraient être respectés lors de leur utilisation pour l’évaluation d’un institut, d’un projet, d’un chercheur (Section 2.1). De tels principes sont également mis en avant par différents textes et manifestes en vue, que nous mentionnons également en annexe de ce document. Après quoi nous essayons de synthétiser quelques éléments concernant les biais connus de la bibliométrie et des indicateurs quantitatifs en général (Section 2.2).

### 2.1 Principes de base à propos de la production et l’utilisation d’indicateurs

Pour chaque indicateur quantitatif les réponses aux questions suivantes doivent clairement être énoncées

- *Que cherche-t-on à mesurer?* Un indicateur doit répondre à une question, cette question doit être précise.
- *Comment l’indicateur a-t-il été construit?*

- *L'indicateur mesure-t-il bien ce que l'on souhaite mesurer?* On sait aujourd'hui que le H-number, censé mesurer à la fois la production scientifique d'un chercheur et sa qualité, est fortement corrélé au nombre de publications [Gin20]. C'est donc essentiellement la *production* scientifique qu'il mesure.
- *Quels sont les biais associés et les écueils qu'ils peuvent induire?* Nous abordons de tels biais en 2.2
- *Quel est son cadre d'utilisation?* On peut par exemple citer le facteur d'impact des revues, créé en 1975 initialement pour éclairer les bibliothèques sur les choix à opérer en matière d'acquisition de revues. Son utilisation pour évaluer les articles ou les chercheurs est donc un détournement de son objectif d'origine.

Le dernier point est particulièrement important à nos yeux. Pour le travail qui nous concerne ici, les indicateurs sont à destination des tutelles de l'institut. **Ces indicateurs ne sauraient être déclinés au niveau de l'évaluation des équipes de recherche ou encore moins des chercheurs.**

Par ailleurs, **tout indicateur doit être accompagné d'un argumentaire qualitatif.** En effet, si on n'y prend garde les indicateurs de « productivité », auxquels on prête souvent la qualité d'objectivité, tels que le nombre d'articles dans des revues ou conférences, le nombre de citations, le nombre de séjours à l'étranger, peuvent être amenés à peser plus lourd que les réelles contributions à la connaissance. À titre d'exemple, la Commission d'Évaluation s'attache à une analyse qualitative et demande aux candidats aux concours de recrutement ou à une promotion de mettre en avant quelques articles (typiquement 3) marquants et/ou représentatifs de leur activité de recherche. La quantité n'est pas du tout un gage de succès!

De plus, les indicateurs construits afin de quantifier la qualité de la production, l'impact scientifique (sorties) devraient pouvoir être mis en regard d'indicateurs permettant d'analyser les moyens mis à disposition pour mettre en œuvre la recherche scientifique qui a permis cette production (entrées).

Enfin, ici, afin de prendre en compte la diversité des disciplines présentes à l'institut, la variété de nos missions de recherche, il nous apparaît indispensable de **privilégier un ensemble grand d'indicateurs.**

## 2.2 Biais des indicateurs

La plupart des indicateurs mesurent mal la qualité de la recherche, car ils sont biaisés. De nombreux travaux pointent ces biais, par exemple [LTZ<sup>+</sup>16, OM17, ALW19]. Nous mentionnons également un ensemble de biais qu'il nous semble particulièrement pertinent d'avoir à l'esprit dans le contexte de l'Inria.

Tout d'abord l'utilisation d'un indicateur unique (ou d'un petit nombre d'indicateurs) ne prend en compte que certaines formes de production scientifique (généralement les publications), et n'accorde aucune importance aux autres (par exemple les logiciels). Il n'est pas uniquement question ici de « trous dans la raquette ». La Commission d'Évaluation place haut dans ses priorités le fait de reconnaître la diversité des aspects du travail scientifique. Un mauvais ensemble d'indicateurs aurait un effet clairement adverse. On peut citer en exemple des activités dont l'institut est fier, et qui passent sous le radar des évaluateurs quantitatifs (exemple: **STEEP, cycle comprendre et agir**). Avec des œillères constituées par des indicateurs restrictifs, l'institut prend le risque de perdre ces thématiques, à terme. Mentionnons ensuite les biais thématiques:

- Si certains indicateurs peuvent être pertinents pour une discipline, ils ne le seront pas pour une autre: la publication dans des conférences est la norme en informatique tandis que ce n'est pas le cas dans la majorité des disciplines mathématiques où les recherches sont principalement publiées dans des revues. Compter les conférences en mathématiques n'a donc que peu de sens.
- Le choix de certains indicateurs conduit à la mise en avant d'une thématique plutôt qu'une

autre, même à l'intérieur d'une sous-discipline (typiquement, par le choix de revues ou conférences « de référence »). Les exemples sont nombreux.

- En cryptographie, la conférence de référence dédiée aux aspects théoriques (*Theory of Cryptography Conference - TCC*) est classée A par CORE, alors que celles dédiées à la cryptographie symétrique (*Fast Software Encryption - FSE*) ou à la cryptographie à clef publique (*Practice and Theory in Public Key Cryptography - PKC*) sont classées B.
- En Calcul Haute-Performance (HPC), la conférence dédiée au logiciel MPI, EuroMPI, l'un des rendez-vous importants de cette communauté, est classée C par CORE. Similairement, la conférence majeure d'un point de vue algorithmique (IPDPS, rang A) est aussi beaucoup moins importante pour la communauté Architecture (mais obtient le même rang). Il ne semble pas qu'il existe en HPC de conférence de rang A\* selon CORE.
- En machine learning, les conférences phares (NeurIPS, ICML) ont connu une croissance impressionnante (NeurIPS 2019: plus de 11000 soumissions, plus de 1600 papiers acceptés – environ 10 fois les chiffres du milieu des années 2000). D'autres conférences parmi les plus prestigieuses du domaine comme COLT rassemblent environ 100 articles (pour 400 soumissions). Les H-nombres de ces conférences sont très différents et s'il est clair dans la communauté ML qu'un article à COLT est prestigieux, ça ne se voit probablement pas beaucoup en dehors, où NeurIPS et ICML écrasent tout et sont les seuls vraiment visibles. D'autres domaines comme l'interaction homme-machine connaissent une évolution similaire.

Par ailleurs les indicateurs mesurent difficilement l'impact réel des recherches ou leur qualité intrinsèque

- les journaux et conférences considérés comme les meilleurs ne sont pas forcément ceux qui publient les travaux ayant le plus grand impact à long terme. Par exemple, les travaux interdisciplinaires souffrent clairement d'une reconnaissance tardive, sont peu cités à court terme, et les journaux les plus prestigieux sont souvent très réticents à les publier [WVS17]. Les « meilleures » revues ont plutôt tendance à un certain conformisme et à rejeter les articles qui proposent des idées nouvelles et des recherches considérées comme « risquées » [WVS17], comportement souvent également reproché aux agences de financement et résumé de manière humoristique dans [Pet12].
- Le nombre de citations d'un article n'est pas une mesure d'une qualité intrinsèque de l'article puisque des expériences ont montré qu'un même article publié dans des revues différentes n'obtenait pas le même nombre de citations.

Nous reviendrons dans la Section 5 sur des exemples plus concrets de biais pour des indicateurs précis.

## 3 Indicateurs qui évaluent ou indicateurs qui mesurent

### 3.1 Deux types d'indicateurs

Les indicateurs peuvent être utilisés essentiellement dans deux contextes distincts: afin d'éclairer une prise de décision en ce qui concerne le financement d'un projet, d'un laboratoire, la promotion d'un chercheur ou d'un enseignant-chercheur, il s'agit alors d'évaluer, ou bien afin d'aider une équipe, un laboratoire à suivre son activité, à définir des objectifs structurants, il s'agit alors d'un « audit » [LTZ<sup>+</sup>16, p. 48].

Nous entendons donc que deux lectures peuvent être faites du rôle d'un ensemble d'indicateurs.

- Des indicateurs qui évaluent (un institut, un site, etc). Nous revenons abondamment sur ce contexte, qui est largement discuté dans la littérature.
- Des indicateurs qui mesurent, pas plus. Ce terme correspond mieux à la notion d'« analyse qualitative » mise en avant dans le COP.

Bien entendu, tout indicateur, dès lors qu'il est chiffré, peut être compris comme une mesure, d'une chose plus ou moins bien définie, mais qu'on mesure par un nombre. Toutefois une telle

mesure n'est pas nécessairement associée à un jugement de valeur. C'est lorsqu'un tel jugement de valeur est absent que nous parlerons spécifiquement d'indicateurs de mesure.

Quelques exemples sont utiles. Le nombre de publications de l'Inria dans des revues ou conférences en biologie est un indicateur qui mesure mais qui peut difficilement être utilisé pour évaluer car son interprétation est complexe: le fait que ce nombre soit élevé montre que les travaux de l'institut ont un impact dans d'autres disciplines, mais cela peut aussi être le signe que l'institut a moins d'impact dans les domaines qui sont son cœur de métier. Au contraire, le nombre de publications dans les revues ou conférences prédatrices est incontestablement un indicateur qui évalue car il juge de la qualité des publications. Il est associé à un jugement de valeur: publier dans des revues prédatrices n'est pas souhaité. Comme tous les indicateurs qui évaluent, son utilisation a un effet prescripteur, mais qui est unanimement considéré comme positif. A l'inverse, il existe de multiples indicateurs qui évaluent et ont un effet prescripteur néfaste.

Le contexte de la « commande » du COP est clairement celui d'un « audit », d'une « analyse qualitative » de l'activité scientifique et son impact. Le second type d'indicateurs, qui simplement mesurent, est donc parfaitement approprié. Toutefois, il est légitime de se demander si des indicateurs classiques d'évaluation, porteurs d'un effet prescripteur, pourraient être employés dans ce contexte dans la mesure où ils ne seraient pas utilisés pour les évaluations d'équipes ou des chercheurs.

### 3.2 Effets pervers pour la recherche

Même pour les indicateurs qui sont présentés comme un instrument de mesure, les études montrent que, dès lors qu'ils sont affichés et qu'ils induisent un jugement de valeur, ils deviennent prescripteurs et modifient rapidement les comportements collectifs et individuels. Au cours des 20 dernières années, de nombreuses études ont mis en évidence une modification importante des pratiques des chercheurs dans les pays où le financement de la recherche s'appuie sur des indicateurs, notamment bibliométriques, e.g. [JCDLCRPF02, Bor11, Wei05]. Mais de tels changements sont également observés dans les pays où ces indicateurs n'ont pas de répercussions directes sur les financements [MS14, Soc17]. La Royal Society se fait par exemple l'écho d'une pression ressentie par les jeunes chercheurs à « publier vite et souvent » pour favoriser une promotion [Soc17, Page 4]. De manière peut-être plus surprenante, elle note aussi que, même si l'utilisation de certains indicateurs comme les facteurs d'impact des journaux est clairement déconseillée au Royaume-Uni, les chercheurs sont persuadés du contraire. **Même s'ils ne sont pas déclinés sur le plan individuel, la simple définition d'indicateurs de production scientifique affecte fortement les pratiques scientifiques.** De nombreuses études se sont alors attachées à identifier les conséquences néfastes de cet impact sur la qualité de la recherche. Nous en détaillons ici quelques exemples.

- Citons en premier lieu l'Italie qui a introduit l'utilisation d'indicateurs quantitatifs tels que le nombre de publications et le nombre de citations en 2011 dans l'évaluation des chercheurs et de la recherche. L'étude [BA19], récemment publiée, montre une augmentation des auto-citations (bien que l'Italie ait un faible taux de collaborations internationales): la communauté scientifique dans son ensemble a ainsi adapté sa stratégie pour maximiser un des indicateurs utilisés et devenir la championne de l'auto-citation.
- Donner une plus grande importance à une liste de journaux ou conférences a un effet prescripteur qui oriente les thématiques de recherche (il y a des études en SHS où les chercheurs ont cessé de s'intéresser à des sujets à dimension nationale pour publier plus dans des revues internationales [LTZ+16, p.69, note 1]). Mais c'est aussi le cas dans nos domaines, comme le montre l'exemple de la « hiérarchie » des conférences de cryptographie en fonction des thématiques.
- Compter un nombre de publications ou d'un autre type de réalisation encourage la recherche superficielle, et dissuade les chercheurs de travailler sur les sujets difficiles ou risqués, qui peuvent prendre plus de temps. Ce qui motive le sous-titre de l'article [Par07] de David Parnas: *Counting papers slows the rate of scientific progress*. Dans la communauté Systèmes,

la production d'un article d'excellente qualité (comme ceux présentés à SOSP, OSDI) requiert plusieurs années de travail (3 hommes-an en moyenne, selon une étude réalisée il y a quelques années), alors qu'il est possible de produire des articles de rang A en quelques mois. Compter un nombre de publications n'encourage clairement pas les articles profonds et originaux. En interaction homme-machine, la méthodologie de validation des résultats la plus solide requiert souvent de conduire des expérimentations en laboratoire et sur le terrain, avec un grand nombre d'utilisateurs. Cela peut prendre plusieurs mois voire années, et devoir publier plus et plus rapidement tend au mieux à influencer le choix des sujets de recherche, au pire à faire des sacrifices méthodologiques acceptés par la communauté (par exemple, se contenter d'études en laboratoire avec un faible nombre de participants là où des études de terrain donneraient potentiellement des résultats plus solides).

- Les établissements de recherche tentent de maximiser à tout prix leur nombre de publications/citations. Les universités d'Arabie Saoudite offrent à des collègues des postes à temps partiel à l'unique condition qu'ils mentionnent l'université dans leurs affiliations<sup>2</sup>. Il y a des exemples similaires dans des universités européennes qui ont les moyens. Des journaux adoptent également de telles pratiques pour truquer leur facteur d'impact<sup>3</sup>.
- Le saucissonnage des papiers: beaucoup de chercheurs sont tentés de diviser les résultats d'un travail afin d'en tirer le plus grand nombre d'articles possibles (on parle de *smallest publishable unit*).
- La multiplication des journaux et conférences sur des sujets étroits, qui satisfait à la fois les chercheurs qui peuvent afficher un comité éditorial supplémentaire sur leur CV, et les éditeurs commerciaux qui augmentent la facture de leurs bouquets de revues et conférences.
- La multiplication des auteurs fantômes pour maximiser le nombre de publications au niveau individuel. Il y a pléthore de travaux sur le sujet, par exemple [WFFD11]. On peut lire aussi l'édifiante étude sur les scientifiques les plus prolifiques [IKB18] intitulée *Thousands of scientists publish a paper every five days*, ou également l'enquête de *Science* sur le marché noir des auteurs [Hvi13]<sup>4</sup>.
- L'amplification de l'aspect normatif: les évaluateurs qui utilisent ces indicateurs sont souvent choisis parmi ceux qui les maximisent et donc ne sont pas enclin à les questionner et à les remettre en cause.

Si ces excès sont souvent associés à la recherche médicale, ou à divers domaines des sciences de la vie, nos disciplines sont également concernées (de manière sans doute moins délirante mais tout aussi préoccupante) et les comportements mentionnés précédemment sont dénoncés par divers collègues [Par07]. La Commission d'Évaluation s'est toujours refusée à utiliser les indicateurs bibliométriques [KFM<sup>+</sup>07]. Pour autant, malgré cette résistance, on peut se demander si ceux-ci n'ont pas fait insidieusement leur œuvre normative, et conduit inconsciemment l'institut à accentuer le poids des publications dans les évaluations individuelles, au détriment de certaines disciplines (par exemple le « système »).

Au vu de ces nombreux effets pervers, nous trouverions **particulièrement dangereux** que l'Inria emprunte la voie suivie par l'Inserm [Ins16, p. 2]:

*La cellule Mesures, Indicateurs, Bibliométrie fournit aux commissions scientifiques spécialisées, aux commissions ad hoc ou au conseil scientifique de l'Inserm des synthèses bibliométriques pour soutenir leur expertise lors des évaluations individuelles (chercheurs candidats au concours DR ou à la promotion DR1, DRE) ou lors des processus de création d'équipes ou d'unités de recherche.*

S'il ne nous semble pas acceptable d'évaluer à l'aide d'indicateurs, la simple utilisation d'indicateurs d'évaluation dans un but d'audit est quant à elle une ligne de crête difficile à suivre, eu égard à l'influence, documentée, qu'a leur utilisation sur les pratiques scientifiques. C'est un choix d'ordre politique.

2. <https://science.sciencemag.org/content/334/6061/1344/tab-pdf>

3. <https://scholarlykitchen.sspnet.org/2011/10/17/gaming-the-impact-factor-puts-journal-in-time-out/>, et un exemple récent sur [cette page](#).

4. Voir aussi <https://retractionwatch.com/2016/10/24/seven-signs-a-paper-was-for-sale/>.

## 4 Indicateurs quantitatifs pour suivre l'activité scientifique de l'Inria

Dans cette section nous avons essayé de proposer des indicateurs prenant en compte la diversité présente à l'institut, diversité qui en fait sa richesse. **Ces indicateurs ne seront utilisés par la CE, ni pour l'évaluation des équipes ni pour celle des individus.** En particulier car nous leur reconnaissons de multiples biais (par exemple autour des publications et des projets qui ne représentent pas la seule dimension de la recherche scientifique faite au sein de l'institut). Ils nous paraissent néanmoins intéressants car ils permettent d'avoir une vision d'ensemble des activités de l'Inria et de leurs évolutions et peuvent ainsi servir de base à une discussion avec les tutelles.

La méthodologie que nous proposons pour les étudier, en accord avec la section 2, est:

- d'une part de présenter ce que nous cherchons à suivre,
- puis de présenter les indicateurs que nous proposons,
- enfin de présenter les possibles biais et risques de ces indicateurs.

Nous agrémentons cette liste d'éléments de réflexion sur notre compréhension, dans le contexte Inria, de la difficulté à rassembler les données nécessaires aux indicateurs que nous proposons.

Le COP (Section 3.1, p. 49) propose six indicateurs pour rendre compte des aspects suivants:

1. La vitalité du modèle organisationnel de l'Inria
2. L'excellence
3. La prise de risque scientifique
4. L'interdisciplinarité
5. Le développement de technologies logicielles
6. La formation par et à la recherche

Les indicateurs qui suivent ont vocation à apporter une dimension qualitative, que nous proposons de décliner en regard de ces six points. Nous complétons l'objectif de mesure de l'interdisciplinarité en y adjoignant un regard sur le panorama disciplinaire au sein de l'institut.

### 4.1 La vitalité du modèle organisationnel de l'Inria

Le COP propose de suivre la durée d'instruction pour la création d'une équipe-projet.

Il nous semble que le modèle d'équipe-projet Inria a pour but de pousser la collaboration autour d'un thème commun. Nous souhaitons donc répondre à la question suivante: l'activité scientifique de l'Inria atteint-elle son but de mise en valeur de travaux collectifs? L'aspect fédérateur du modèle d'équipe-projet est-il une vérité, ou est-il une façade qui cache une simple somme d'aventures de recherche individuelles? Nous proposons, pour répondre à cette question, de suivre les valeurs suivantes:

#### Proposition d'indicateur(s)

- Répartition du nombre de permanents d'une même équipe Inria par publications (camembert avec des proportions pour 0, 1, 2, 3 etc).
- Pour chaque projet de nature collaborative financé auquel participe un chercheur ou une équipe Inria, nombre de permanents d'une équipe Inria et nombre d'équipes Inria différentes impliquées.

**Difficulté de construction** À partir de la base HAL, le critère « chercheur permanent » n'est pas facile à obtenir: d'une part parce que l'accès à cette information qui relève des bases RH n'est pas une évidence, d'autre part parce qu'il n'existe pas d'identité unique des auteurs dans HAL (beaucoup d'auteurs, même parmi les personnels Inria, n'ont pas d'identifiant HAL).

La synthèse du second indicateur proposé nécessiterait de mobiliser plusieurs directions fonctionnelles, car on peut penser que la donnée n'existe pas actuellement sous une forme qui puisse être agrégée.

**Biais et risques** Pour le premier indicateur il est clair que la situation peut être très différente en fonction de la taille des équipes. Il ne s'agit en aucune façon d'inciter à des collaborations fictives. Il est clair que, dans certains domaines, les collaborations se font dans d'autres dimensions que la publication, que dans d'autres on encourage les doctorants à publier seuls. Cet indicateur n'a pas vocation à être utilisé pour évaluer les équipes-projets, mais bien pour suivre les évolutions de pratiques scientifiques. Un des effets néfastes de l'affichage de cet indicateur serait, par exemple et pour caricaturer, que le REP co-signe systématiquement toutes les publications des membres de son équipe.

Le risque du second indicateur est que 100% des projets soient collaboratifs. C'est pourquoi nous étudions la répartition en fonction du nombre de chercheurs impliqués et/ou du nombre d'équipes Inria impliquées. À nouveau, les principaux risques que nous envisageons sont ceux liés à une mauvaise interprétation/usage de l'indicateur (par exemple: ce n'est pas « maximiser le nombre de projets incluant uniquement une équipe Inria »).

## 4.2 « L'excellence »

Pour suivre le critère « excellence », le COP propose de suivre le nombre de bourses ERC obtenues. Nous détaillerons dans la Section 5 pourquoi nous avons choisi de ne pas retenir ce critère.

Suivre « l'excellence » de l'institut est sans aucun doute la question la plus difficile à poser pour des indicateurs chiffrés étant donné le manque de définition « d'excellence ». Le choix des indicateurs qui définissent cette excellence est un choix politique. Ici clairement la direction demande que les ERC en fassent partie.

Ainsi plutôt que de suivre cette question mal définie, nous avons décidé de suivre sur ces différents points:

- L'internationalité;
- Les pratiques scientifiques;
- Et enfin nous rajoutons un indicateur autour des ERC.

En soi, ces indicateurs ne décrivent pas l'excellence d'une recherche. Mais ils suivent des éléments qui paraissent importants pour décrire l'institut.

### 4.2.1 Suivre l'internationalité de l'Inria

Pour introduire ces indicateurs, notons qu'ils décrivent probablement plus la visibilité de l'Inria que son excellence.

#### Proposition d'indicateur(s)

- Impact international: proportion de publications avec un coauteur en poste à l'étranger parmi le nombre total de publications.
- Évolution géographique des principaux partenaires des projets de recherche collaboratifs internationaux (projets européens, équipes associées, projets industriels, ...).

**Difficulté de construction** Le premier indicateur semble facile à construire, avec une attention à garder sur les *fausses* collaborations internationales (par exemple un ancien doctorant parti en postdoc à l'étranger et qui publierait sous sa nouvelle affiliation).

Le second peut nécessiter une pondération assez complexe à établir (volume financier du projet, part du partenaire Inria, centre de gravité, ...). Compter les projets comme unités a certes des limites que l'on reconnaît, mais a aussi le mérite de la simplicité.



**Biais et risques** Comme pour tous les indicateurs de suivi, une mauvaise interprétation (et en particulier y associer un objectif chiffré ou de maximisation) est un risque. Donner « l'internationalité » comme critère d'excellence amène le risque de passer à côté de collaborations locales qui pourraient d'une part être plus fructueuses, et d'autre part avoir plus de sens, si on prête attention aux discussions sur l'impact énergétique de nos recherches.

#### 4.2.2 Présenter un panorama de pratiques

L'excellence est liée à certaines pratiques et certaines responsabilités, aussi nous proposons de suivre également comme mesure de l'excellence le fait de mettre en avant des modèles de publication vertueux, ainsi que la *slow science*.

##### Proposition d'indicateur(s)

- La proportion de publications sur des modèles ouverts (« *green* » ou « *diamond open access* »).
- Montant total des APC<sup>5</sup> payés par l'Inria (déjà suivi, normalement).
- Proportion de publications disponibles en texte complet sur HAL.
- Nombre de publications dans des journaux prédateurs.
- Nombre de badges ACM « *reproducible research* » (science ouverte et reproductible), ou équivalents (des initiatives similaires émergent dans divers domaines).
- Nombre de reproductions d'expériences effectuées par des scientifiques Inria.
- Nombre de mises à disposition de jeux de données collectées.

**Difficulté de construction** S'il est envisageable de croiser les publications sur HAL avec des bases telles que DOAJ, il faut veiller à écarter de la sélection les journaux s'appuyant sur un modèle auteur-payeur, et pratiquant des APC prohibitifs.

Concernant les journaux prédateurs, des listes existent, mais elles semblent rarement mises à jour. De plus, certains éditeurs de journaux prédateurs vont jusqu'à harceler les auteurs de telles listes et faire pression sur leur université d'appartenance pour être retirés de ces listes [Bea17].

Les deux derniers indicateurs sont probablement difficiles à construire en l'état actuel, tant des évolutions dans cette direction voient le jour dans de multiples domaines.

**Biais et risques** En voulant apporter un éclairage vers certaines évolutions de pratiques que nous jugeons vertueuses, nous provoquons une distorsion entre les domaines thématiques: certains domaines sont en pointe sur les questions de révolution des modèles de publication et/ou sur le sujet de la science reproductible, tandis que d'autres ont plus de difficulté à s'échapper des modes de fonctionnement actuels. On peut également ajouter que la notion de reproduction d'expérience, en elle-même, a un sens très variable entre les domaines thématiques, jusqu'à être sans objet pour les travaux les plus théoriques, qu'il n'est bien entendu pas question d'oublier.

Enfin, un des points importants relevés par Beall [Bea17] lors de la construction de sa liste de journaux prédateurs a été l'opposition majeure des promoteurs de l'open access à tout prix. Un journal prédateur, sans valeur scientifique, peut tout à fait relever du « *gold open access* », en pratiquant des APC indus, et en ouvrant certes l'accès aux travaux ensuite.

5. APC=« *article processing charges* ».

### 4.2.3 Indicateur autour des ERC

#### Proposition d'indicateur(s)

Vecteurs d'indicateurs déclinés autour de

- Proportion des propositions individuelles ERC, et des propositions d'ANR JCJC, financées.
- Proportion des propositions ERC Synergy et ANR collaboratives financées.

**Biais et risques** Le principal biais de ces indicateurs est que rien n'indique qu'ils permettent de suivre vraiment un critère d'excellence. Si la qualité des propositions émanant de l'institut est bien sûr un facteur déterminant, on peut voir également de multiples facteurs qui influencent la valeur que prend cet indicateur: la capacité de l'institut à répondre aux appels à projets, le degré d'auto-censure, la qualité du « *coaching* » des soumissions, etc. Par ailleurs, même si l'on admettait que le nombre d'ERC mesurait un type d'excellence de l'institut, un autre biais est qu'il ne reflète quand même pas *toute* l'excellence de l'institut dans sa diversité.

Le COP mentionne explicitement un indicateur lié aux propositions ERC, et cette mesure est souvent mise en avant comme critère d'excellence (comme le nombre de prix Nobel a un poids déterminant dans le classement de Shanghai). Néanmoins, il nous semble trop restrictif de limiter aux ERC, ce pourquoi nous proposons d'inclure d'autres types de succès à des appels à projets sélectifs. Il est de toute façon évident que même à l'échelle de l'institut, les nombres dont il est question pour les appels ERC, à plus forte raison si on les observe discipline par discipline, sont trop limités pour qu'on puisse en tirer une information pertinente.

Comme pour les autres indicateurs, un risque serait que ce nombre soit vu comme un but à maximiser. Il existe deux manières de maximiser ce nombre: maximiser le nombre de projets acceptés, ou réduire le nombre de projets soumis. Les risques liés à la maximisation du nombre de projets acceptés nous semblent déjà existants (et donc non liés à cet indicateur). Ils n'en sont pas moins préoccupants: certains sujets sont plus appréciés dans le contexte de l'ERC que d'autres, et l'impact que cela peut avoir sur l'orientation des thématiques de recherche au sein de l'institut mérite d'être surveillé. Quant au risque d'une censure à la soumission des projets (« à moins d'être extrêmement sûr du fait que mon projet ne passe je n'ai pas le droit de le soumettre »), il nous paraît moindre étant donné les considérations politiques actuelles. Il conviendra de réévaluer cela et ses conséquences si elles changeaient.

### 4.3 La prise de risque scientifique

Le COP propose de suivre le ratio entre le flux annuel de projets encourageant la prise de risques (actions exploratoires) et le nombre d'équipes-projets. L'un des principaux problèmes tel que présenté dans le COP est le chiffrage de ce critère (« 10% dès 2020 ») et sa signification. Est-ce un but? Est-ce que ce 10% correspond au pourcentage de recherche à *risque* souhaité à l'Inria? Comment a été obtenu ce nombre de 10%?

En accord avec le COP et la littérature [Jac16], il nous semble que le financement sur des contrats industriels (thèse Cifre par exemple), ou plus généralement tout financement qui est subordonné à des politiques d'orientation scientifique qui sont à la fois assez dirigées, et externes à l'institut, encourage moins la prise de risque. À ce titre, nous proposons de suivre l'évolution des sources de financement de thèses dans le cadre de l'analyse de l'activité scientifique de l'institut. Il s'agit bien entendu d'un « indicateur d'entrée »: au vu de l'information que l'on en extrait, peut-on dire que l'Inria a les moyens de sa prise de risque scientifique?

#### Proposition d'indicateur(s)

- Répartition des sources de financement de thèse.

**Difficulté de construction** Au moins à gros grain, cet indicateur semble facile à construire.

**Biais et risques** Il convient de rester conscient que la conjoncture actuelle en terme de postes académiques n’encourage possiblement pas la prise de risque sur des thèses. Pour protéger les jeunes docteurs l’encadrant peut pousser sur des sujets plus « sûrs ». Toutefois à nouveau en tant qu’indicateur de suivi ces répartitions nous semblent intéressantes.

#### 4.4 L’interdisciplinarité et le panorama disciplinaire

Le COP propose de suivre le nombre (ou ratio) d’équipes-projets actives dans l’interdisciplinarité, en précisant la définition de cette notion d’activité par: « qui publie simultanément dans des journaux/conférences de référence de disciplines différentes, à la fois numérique (informatique-mathématiques) et autres disciplines (biologie, santé, agronomie, SHS, etc.) ». Cet indicateur mérite d’être plus discuté et clarifié, ce que nous faisons ci-dessous. Nous proposons d’ajouter un indicateur complémentaire, qui se place au niveau des publications elles-mêmes, plutôt qu’au niveau des équipes.

Avant l’interdisciplinarité, nous proposons de suivre le panorama thématique présent à l’Inria et son évolution au cours du temps, en suivant les deux premiers indicateurs ci-dessous. En tant que Commission d’Évaluation, nous apprécierions de disposer de ce panorama. Nous ajoutons un dernier indicateur qui relève de l’interdisciplinarité.

##### Proposition d’indicateur(s)

- La proportion de publications par discipline (mathématiques, informatique, biologie, santé, physique, etc).
- La proportion d’ERC par domaine.
- Nombre (ou ratio) d’équipes-projets « actives dans l’interdisciplinarité », au sens du COP.
- Proportion de publications avec un ou plusieurs co-auteurs d’une autre discipline.

**Difficulté de construction** Dans la base HAL, la catégorisation des publications dans quelques grands domaines (une douzaine) est possible, mais elle est faite de façon assez négligente par les auteurs. C’est donc une donnée difficile à utiliser. Pour les articles, il est possible de pallier en partie cette difficulté, en croisant cette information avec celle des domaines scientifiques des journaux, qui sont référencés dans des bases externes telles Scimago. Toutefois, il serait préférable que l’information *princeps* soit de qualité. À cet effet, on pourrait suggérer que le rapport d’activité annuel des équipes-projets soit assorti d’une invitation à un meilleur étiquetage des publications. La catégorisation des auteurs dans des thématiques scientifiques est probablement encore plus difficile.

La proportion d’ERC par domaines, au moins au grain des panels ERC, est facile à collecter (et probablement déjà suivie par ailleurs).

**Biais et risques** Ces indicateurs n’ont pas de portée évaluative. Tout au plus peut-on craindre une volonté d’affichage d’une interdisciplinarité de façade, mais ce risque nous semble être faible.

Au niveau des biais, les deux indicateurs qui tentent de mesurer l’interdisciplinarité sont sujets à caution. D’une part, la mesure au niveau des équipes peut mal repérer le fait que certaines équipes peuvent avoir un spectre thématique extrêmement large sans qu’on puisse réellement dire qu’il y a du liant entre deux points extrêmes de ce spectre. D’autre part, la mesure au niveau des publications peut surpondérer des publications où la contribution de la discipline extérieure est anecdotique.

## 4.5 Le développement de technologies logicielles

Le COP propose de suivre le nombre d'équipes-projets ayant un ingénieur de développement affecté dans la durée, du fait de leur intensité technologique. C'est un indicateur qu'on pourrait qualifier d'« indicateur d'entrée », au même titre que l'indicateur que nous proposons pour la prise de risque scientifique.

Y adjoindre une autre proposition nous a semblé difficile.

### Proposition d'indicateur(s)

- Nombre de projets logiciels *open-source* développés par l'Inria, faisant l'objet d'un développement actif depuis plus de deux ans et avec une communauté d'utilisateurs externes existante.

**Difficulté de construction** La source première peut être la base BIL, mais elle ne permet pas de qualifier l'existence d'une communauté d'utilisateurs, et le degré d'activité du développement peut également être difficile à déterminer.

**Biais et risques** Nous reconnaissons sans difficulté que cet indicateur est très parcellaire. Tout le développement logiciel de l'institut n'est pas *open-source*, et le simple nombre, qui met sur un pied d'égalité des développements logiciels très différents, est très insatisfaisant.

Par ailleurs, il nous a semblé souhaitable mais très ardu de proposer des indicateurs pertinents pour rendre compte des développements matériels et de l'activité d'expérimentation de l'Institut, qui sont des aspects qui ne sont pas mentionnés dans le COP.

## 4.6 La formation par et à la recherche

Le COP propose de suivre le ratio du nombre de doctorants sur le nombre de chercheurs permanents. Cet indicateur est intéressant, et sa déclinaison par thèmes également.

### Proposition d'indicateur(s)

- Ratio du nombre de doctorants sur nombre de chercheurs permanents titulaires d'une HDR, décliné également par grand domaines scientifiques et/ou par centre Inria.
- Ratio de doctorants étrangers.

**Difficulté de construction** L'indicateur du ratio du nombre de doctorants par permanent HDR nous semble, en vérité, assez difficile à collecter dans de nombreux cas. Qualifier l'information « doctorant » dans une équipe n'est pas trivial, et il en va de même pour l'information « permanent HDR ». C'est encore plus délicat lorsqu'il s'agit de personnels non rémunérés par l'Inria. Donc dans ce ratio, ni le numérateur ni le dénominateur ne sont des données faciles à collecter. On peut y ajouter le fait que certaines équipes ont des permanents qui, de par leur fonction (ils peuvent avoir d'autres employeurs), peuvent avoir un degré d'implication assez modeste dans la formation doctorale.

Les mêmes réserves valent pour les ratio des doctorants étrangers.

**Biais et risques** Le biais du ratio global, collecté au niveau de l'institut, est patent: les pratiques varient entre les domaines, et également entre les écoles doctorales (dont beaucoup imposent des limites de taux d'encadrement, qui peuvent être *très* variables), si bel et si bien que cet indicateur peut ne pas mesurer grand-chose d'interprétable. Le décliner par discipline fera certainement apparaître des différences. L'hétérogénéité restera de mise, tant les pratiques, et l'investissement des encadrants dans l'encadrement, peuvent varier considérablement. Il est également patent que

les bassins de recrutement des doctorants sont très divers en fonction des sites Inria: on s'attend à ce que l'environnement joue un rôle important dans la mesure.

Les ratios de doctorants étrangers peuvent être significativement affectés par des éléments externes à l'institut (appels spécifiques par exemple), ce qui est susceptible d'introduire un biais.

En matière de risque, comme dans la plupart des cas, la maximisation est porteuse de travers évidents. Pousser au nombre réduit la qualité de l'encadrement. Vouloir maximiser le ratio (en minimisant le dénominateur) est également susceptible de constituer un frein au passage de l'HDR, ou à l'intégration dans une équipe de permanents HDR qui, pour des raisons qui peuvent être légitimes, ne souhaitent pas encadrer.

Porter le regard sur ces valeurs peut être intéressant. Promouvoir leur augmentation sans se soucier de l'effet très délétère de certaines stratégies de maximisation pourraient induire serait véritablement dangereux.

## 5 Indicateurs que nous avons choisi d'écartier

### 5.1 Les « classements » sont sources de problèmes

Dans cette section nous discutons certains indicateurs proposés initialement et pourquoi nous avons choisi de ne pas les considérer. Le premier concerne le **nombre de publications dans les « meilleures » conférences et revues**. La difficulté ici est de définir un référentiel de ces revues et conférences. Outre le fait que les habitudes de publications diffèrent très fortement d'une discipline à l'autre, que ce référentiel est, par essence, mouvant, la définition même et l'affichage de ce référentiel a un effet prescripteur qui peut être néfaste. A titre d'exemple, la communauté mathématique française considère qu'aucun consensus n'existe sur cette question (cf. rapport HCéres<sup>6</sup>):

*Il existe un grand nombre de revues répertoriées comme des revues de mathématiques (environ 700 d'après Thomson Reuters). Il n'existe pas de consensus de la communauté sur un classement de ces revues.*

En effet, quand on les examine de près, les « bugs » dans les classements de journaux et conférences, et les biais qu'ils induisent, sont tellement fréquents qu'on se demande s'ils ne sont pas la norme. C'est particulièrement criant concernant les conférences. Nous ne donnons ici que deux exemples par souhait de concision, mais nous pourrions les multiplier.

- La conférence « *Static Analysis Symposium (SAS)* » classée A par CORE et unanimement reconnue comme telle par la communauté a été soudainement rétrogradée C par CORE suite à une confusion d'acronyme. Il a fallu plus d'un an pour faire corriger cela.
- En HPC, la conférence HiPC était de rang A jusqu'en 2017, où elle a été déclassée en rang C (elle était inscrite sous un autre nom en rang C). Après demande d'évaluation du dossier par les *program chairs*, elle a été reclassée en rang B en 2017 (soit trois rangs différents la même année). Elle vient de passer en « conférence nationale » sans argument supplémentaire autre que « elle a toujours lieu en Inde » (même si Core admet la qualité de la conférence et des taux d'acceptation, ainsi que la réputation et l'internationalité de l'équipe d'organisation).<sup>7</sup>

Un nombre croissant d'instances reviennent sur l'utilisation de tels référentiels. Cela n'a rien d'étonnant. Si un institut de recherche comme l'Inria choisissait d'afficher un référentiel pour une communauté scientifique pour laquelle le consensus est qu'il n'y a justement *pas* de consensus, cet affichage serait mauvais pour l'institut et ses personnels.

En outre, on peut craindre que les classements de journaux introduisent un biais en faveur des modèles de publication les moins ouverts, et des études confirment cette crainte [SPF18]. Ils

6. <https://www.hceres.fr/sites/default/files/media/downloads/ST1%202021.11.2017.pdf>

7. On liste sans peine plusieurs conférences internationales qui ont lieu systématiquement dans un pays donné ou une zone géographique donnée: Crypto (USA, A\*), SC (USA, A), Eurocrypt (Europe, A), Euro-Par (Europe, A), PRDC (Pacifique, B). L'internationalité de ces conférences n'est pas remise en question. La diversité des classements associés reflète (au mieux) une forme d'arbitraire.

contribuent donc à une perpétuation d'un modèle de diffusion scientifique dont, par ailleurs, un institut de recherche comme l'Inria promeut la remise en question.

## 5.2 Mieux vaut lire les articles !

Plus généralement, il est impossible de mesurer la qualité et l'impact de travaux de recherche par le niveau présupposé des journaux ou conférences dans lesquels ils sont publiés.

- La simple notion de « publication », dans certains domaines, ne veut plus dire grand-chose. Dans certains domaines de l'informatique, on distingue les publications qui donnent lieu à un poster, celles qui donnent lieu à un poster avec un article écrit associé, celles qui donnent lieu à un poster avec une présentation orale, etc. Clairement, compter cela avec un nombre entier n'est pas approprié.
- Tous les référentiels de journaux ou de conférences, ou les indicateurs (généralement fondés sur le nombre de citations) sont très biaisés, par exemple en faveur des « gros » journaux généralistes (ainsi le taux de citation des publications d'un journal est fortement corrélé à son nombre de pages [MS14]), alors que bon nombre d'articles importants sont publiés dans des journaux ou conférences spécialisés.
- Des « publications » à fort impact peuvent être issues de blogs, magazines, pas seulement de journaux ou conférences « prestigieuses ». Par exemple, en IHM, l'article « *The design and implementation of pie menus* » de Don Hopkins<sup>8</sup> dans *Dr.Dobb's journal* en 1991. Dans le domaine de la correction d'erreurs, la découverte des turbo-codes par C. Berrou et A. Glavieux, incontestablement un des résultats les plus importants de ces 30 dernières années qui a valu par la suite de multiples prix à leurs auteurs, a été présentée à la conférence *IEEE International Conference on Communications*, classée B par ERA, et qui n'est considérée comme conférence-phare du domaine dans aucun classement.
- De tels classements pris pour de l'évaluation encouragent à publier dans les journaux ou conférences les plus faibles de chaque catégorie [Moo16].

Ainsi, le nombre de publications dans des conférences de rang A/A\* ou dans les « meilleures » revues cherche à évaluer une sorte de mesure d'impact sans toutefois y parvenir et l'on se rend bien compte qu'à moins de lire les travaux il est impossible de juger de manière automatique de l'impact d'un papier (et même alors, parfois il faudrait une vision à plus long terme).

## 5.3 Des indicateurs numériques peu convaincants

Le deuxième indicateur concerne le **nombre d'ERC**, évoqué dans le COP pour effectuer un suivi scientifique de l'Inria. Nous avons choisi de l'écartier car il nous semble être un indicateur qui cible un type très particulier d'appel à projet et donc peu à même de couvrir la diversité existante à l'institut. Par ailleurs, le risque que nous voyons à afficher ce nombre unique est qu'il serait interprété comme un objectif à maximiser par les individus et un passage obligé permettant d'assurer une progression de carrière. S'il est besoin de le rappeler, ce n'est *pas* le cas ! Si l'on peut comprendre que l'institut encourage les soumissions, cela ne peut être fait sans frein. Trop soumettre induirait une considérable perte de temps et d'énergie. C'est pour cette raison que le choix d'un ratio nous a paru préférable.

Le troisième indicateur concerne le **nombre de comités de programme et éditoriaux**. Nous avons de multiples exemples de dérives associées à ces comités :

- On se contentera de citer ce que signifie le fait d'être l'un des 724 *associate editors* pour le journal *J. Alzheimer's disease*: « *Once an article is accepted, the corresponding author is invited to become an Associate Editor of JAD with a one-year term* »<sup>9</sup>.

8. <https://www.drdoobs.com/database/the-design-and-implementation-of-pie-men/184408667?pgno=5>

9. <https://www.j-alz.com/edittorrole>

- D’une façon similaire, dans les grosses conférences de vision ou d’IA, être membre du comité de programme signifie faire un rapport.
- Les responsabilités importantes dans les comités de programme pour évaluer les chercheurs poussent au *copinage*. Il arrive notamment que, dans certaines conférences internationales, la plupart des rôles scientifiques qui, au contraire des tâches d’organisation, ne nécessitent pas de présence physique sur les lieux de la conférence, soit néanmoins attribuée à des chercheurs de l’institution ou du pays organisateur (la conférence ICPP 2019<sup>10</sup> illustre ce biais dans le domaine du HPC, mais on pourrait trouver des exemples similaires dans d’autres domaines).
- Même sans évoquer des biais flagrants comme ci-dessus, le niveau de reconnaissance que signifie le fait de participer à un comité de programme ou à un comité éditorial peut être sans aucune comparaison possible d’une discipline à une autre. Assez souvent, le nombre de comités de programmes ou de comités éditoriaux est une mesure du réseau des individus, ou des équipes.

Afficher un tel indicateur promeut des pratiques de plus en plus fréquentes et peu vertueuses, et n’apporte pas d’information utile.

## 6 Annexe

Nous rappelons ici les recommandations en matière d’utilisation d’indicateurs scientométriques du Manifeste de Leiden<sup>11, 12</sup> publié en 2015 qui visent à améliorer les pratiques d’évaluation :

1. La description quantitative doit servir d’appui à une évaluation qualitative par des experts
2. Mesurer la performance au regard des missions de recherche de l’institution, du groupe ou du chercheur.
3. Protéger l’excellence dans les domaines de recherche importants à l’échelle locale.
4. Maintenir une collecte de données et des processus d’analyse ouverts, transparents et simples.
5. Permettre aux chercheurs évalués de vérifier les données et analyses.
6. Tenir compte des différences entre disciplines en matière de publication et de citation.
7. Baser les évaluations des chercheurs sur un jugement qualitatif de leurs travaux.
8. Éviter les simplifications abusives et les fausses précisions.
9. Reconnaître les impacts systémiques des évaluations et des indicateurs (toujours privilégier plusieurs indicateurs).
10. Réévaluer régulièrement et faire évoluer les indicateurs.

Par ailleurs, le Hcéres, qui évalue des structures de recherche et non des individus, se reconnaît dans ces principes<sup>13</sup>, et affiche des recommandations qui peuvent être synthétisées comme suit :

1. L’évaluation par les pairs est le fondement des pratiques évaluatives et des processus de publication scientifique.
2. Le Hcéres fait le choix d’une évaluation multicritérisée.
3. La méthodologie employée met l’accent sur la dimension qualitative de l’évaluation des structures de recherche. Autoévaluation: mise en avant de 20% des résultats les plus importants.
4. L’évaluation doit prendre en compte la spécificité disciplinaire.
5. Il est recommandé de privilégier la portée des résultats, sans recourir nécessairement et exclusivement à des indicateurs bibliométriques. Indicateurs : outils de travail complémentaires des jugements évaluatifs qualitatifs.
6. Le département Observatoire des Sciences et Techniques du Hcéres produit des rapports d’indicateurs bibliométriques utilisés dans le cadre de l’évaluation des coordinations territoriales et des organismes de recherche.

10. <https://www.hpcs.cs.tsukuba.ac.jp/icpp2019/committees/index.html>

11. <https://www.ouvrirlascience.fr/le-manifeste-de-leiden-pour-la-mesure-de-la-recherche/>

12. <https://www.ouvrirlascience.fr/wp-content/uploads/2018/09/leidenmanifesto-hceres-ost.pdf>

13. <https://www.hceres.fr/fr/decoder/evaluation-de-la-recherche-criteres-et-indicateurs>

## Références

- [ALW19] Dag W. Aksnes, Liv Langfeldt, and Paul Wouters. Citations, citation indicators, and research quality: An overview of basic concepts and theories. *SAGE Open*, 9(1), 2019. [10.1177/2158244019829575](https://doi.org/10.1177/2158244019829575).
- [BA19] Petrovich E Baccini A, De Nicolao G. Citation gaming induced by bibliometric evaluation: A country-level comparative analysis. *PLoS ONE*, 14, 2019. [10.1371/journal.pone.0221212](https://doi.org/10.1371/journal.pone.0221212).
- [Bea17] Jeffrey Beall. What I learned from predatory publishers. *Biochemia medica*, 27(2):273–278, 2017. [10.11613/BM.2017.029](https://doi.org/10.11613/BM.2017.029).
- [Bor11] Lutz Bornmann. Mimicry in science? *Scientometrics*, 86:173–177, 2011. [10.1007/s11192-010-0222-8](https://doi.org/10.1007/s11192-010-0222-8).
- [Gin16] Yves Gingras. *Du Bon usage des indicateurs: Critères de validité, échelles et disciplines*, pages 67–73. Volume 258 of Leclercq et al. [LTZ<sup>+</sup>16], juin 2016. [lien](#); [lien2](#).
- [Gin20] Yves Gingras. Why the h-index is a bogus measure of academic impact. *The Conversation*, 2020. 2020/07/08. <https://theconversation.com/why-the-h-index-is-a-bogus-measure-of-academic-impact-141684>.
- [Hvi13] Mara Hvistendahl. China’s publication bazaar. *Science*, 342:1035–1039, November 2013. [10.1126/science.342.6162.1035](https://doi.org/10.1126/science.342.6162.1035); [lien2](#).
- [IKB18] John P. A. Ioannidis, Richard Klavans, and Kevin W. Boyack. Thousands of scientists publish a paper every five days. *Nature*, 561:167–169, 2018. [10.1038/d41586-018-06185-8](https://doi.org/10.1038/d41586-018-06185-8).
- [Ins16] Inserm, Cellule MIB. Les indicateurs bibliométriques à l’Inserm. Brochure d’information, Inserm, 2016. <https://www.inserm.fr/professionnels-recherche/bibliometrie>.
- [Jac16] Jacq, Annick. Les nouveaux modes de pilotage de la recherche: quels effets? *Nat. Sci. Soc.*, 24(4):371–378, 2016. Séminaire 2014-2015 du Centre d’Alembert de l’Université Paris-Sud.
- [JCDLCRPF02] E. Jiménez-Contreras, E. Delgado López-Cózar, R. Ruiz-Pérez, and V.M. Fernandez. Impact-factor rewards affect Spanish research. *Nature*, 417:6892, 2002. [10.1038/417898b](https://doi.org/10.1038/417898b).
- [KFM<sup>+</sup>07] Anne-Marie Kermarrec, Erwan Faou, Jean-Pierre Merlet, Philippe Robert, and Luc Segoufin. Que mesurent les indicateurs bibliométriques? Rapport de la commission d’Évaluation de l’Inria, INRIA, 2007. <https://hal.inria.fr/hal-02989342>.
- [LTZ<sup>+</sup>16] Bruno Leclercq, Benoît Timmermans, Edwin Zaccai, Barbara Clerbaux, Marek Hudon, and Hugues Bersin, editors. *L’évaluation de la recherche en question(s)*, volume 258 of *Mémoires*. Académie royale de Belgique, juin 2016. [lien](#); [lien2](#).
- [Moo16] Imad A. Moosa. A critique of the bucket classification of journals: The ABDC list as an example. *Economic Record*, 92(298):448–463, 2016.
- [MS14] Carolin Michels and Ulrich Schmoch. Impact of bibliometrics studies on the publication behaviour of authors. *Scientometrics*, 98:369–385, 2014. [10.1007/s11192-013-1015-7](https://doi.org/10.1007/s11192-013-1015-7).
- [OM17] Kai A. Olsen and Alessio Malizia. Counting research → directing research. the hazard of using simple metrics to evaluate scientific contributions. an EU experience. *Journal of Electronic Publishing*, 20(1), 2017. [10.3998/3336451.0020.102](https://doi.org/10.3998/3336451.0020.102).
- [Par07] David Lorge Parnas. Stop the numbers game. *Communications of the ACM*, 50(11):19–21, 2007. [10.1145/1297797.1297815](https://doi.org/10.1145/1297797.1297815).
- [Pet12] Gregory A. Petsko. Goodbye, Columbus. *Genome Biology*, 13, 2012. [10.1186/gb-2012-13-5-155](https://doi.org/10.1186/gb-2012-13-5-155).



- [Soc17] Royal Society. Research culture embedding inclusive excellence. <https://royalsociety.org/topics-policy/publications/2018/research-culture-embedding-inclusive-excellence/>, 2017.
- [SPF18] Joachim Schöpfel, Hélène Prost, and Amel Fraisse. Plus ou moins open : les revues de rang A en sciences de l’information et de la communication. In Philippe Bonfils, editor, *Information scientifique et diffusion des savoirs : entre fragmentations et intermédiaires*, volume 15 of *Revue Française des Sciences de l’Information et de la Communication*. OpenEdition journals, 2018. [10.4000/rfsic.4706](https://doi.org/10.4000/rfsic.4706).
- [Wei05] Peter Weingart. Impact of bibliometrics upon the science system: Inadvertent consequences? *Scientometrics*, 62:117–131, 2005. [10.1007/s11192-005-0007-7](https://doi.org/10.1007/s11192-005-0007-7).
- [WFFD11] Joseph S Wislar, Annette Flanagin, Phil B Fontanarosa, and Catherine D DeAngelis. Honorary and ghost authorship in high impact biomedical journals: a cross sectional survey. *BMJ*, 2011. [10.1136/bmj.d6128](https://doi.org/10.1136/bmj.d6128).
- [WVS17] Jian Wang, Reinhilde Veugelers, and Paula Stephan. Bias against novelty in science: A cautionary tale for users of bibliometric indicator. *Research Policy*, 46(8):1416 – 1436, 2017. [10.1016/j.respol.2017.06.006](https://doi.org/10.1016/j.respol.2017.06.006).