

Les classements à l'international des revues en SHS

David Pontille, Didier Torny

► **To cite this version:**

David Pontille, Didier Torny. Les classements à l'international des revues en SHS. Christine Barats; Julie Bouchard; Arielle Haakenstad. Faire et dire l'évaluation dans l'enseignement supérieur et la recherche. L'enseignement supérieur et la recherche conquis par la performance , Presses des Mines, pp.137-153, 2018, 978-2356714855. halshs-01697116

HAL Id: halshs-01697116

<https://halshs.archives-ouvertes.fr/halshs-01697116>

Submitted on 31 Jan 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Les classements à l'international des revues en SHS

in C. Barats, J. Bouchard, A. Haakenstad (eds.), 2018,
Faire et dire l'évaluation dans l'enseignement supérieur et la recherche,
Paris, Presses des Mines, Collection Sciences Sociales, p. 137-153.

David Pontille et Didier Torny

Centre de Sociologie de l'Innovation - Mines ParisTech
Institut Interdisciplinaire de l'Innovation (CNRS UMR 9217)
PSL Research University

david.pontille@mines-paristech.fr

didier.torny@mines-paristech.fr

Depuis les années 1980, les gouvernements et les agences publiques de nombreux pays portent une attention particulière, accentuée par la compétition internationale, aux modalités d'évaluation des productions académiques [Auranen et Nieminen, 2010]. Un tel engouement s'est traduit dans une batterie d'instruments grandissante [Hicks, 2012], parmi laquelle figure une technologie élémentaire : un classement de revues, conçu comme une aide à l'évaluation de la quantité et de la qualité des productions scientifiques de diverses institutions (e.g. universités, organismes de recherche, ou laboratoires). Bien que plusieurs classements de revues aient été élaborés dès les années 1970 [Glenn, 1971 ; Moore, 1972 ; Giles et Wright, 1975], le caractère inédit de ceux qui ont émergé au cours des années 2000 réside dans leur statut d'instrument de politique publique. C'est le cas de l'Australie, du Brésil, de la France, de la Flandre, de la Norvège, et des Pays-Bas où cette technologie d'évaluation est en vigueur pour certains domaines – notamment en sciences humaines et sociales (SHS).

Cette profusion de classements de revues interroge d'autant plus qu'elle a généralement été envisagée comme une extension des méthodes du *new public management* au monde académique [Bruneau et Savage 2002 ; Bruno, 2008]. Dans cette perspective, les producteurs de classements de revues sont campés en bureaucrates ou gestionnaires dont l'unique enjeu est d'imposer à l'enseignement supérieur et la recherche (ESR) divers critères de rentabilité, d'efficacité, ou de productivité à court terme. Or ce point de vue néglige les formes de participation des personnes, collectifs et institutions qui sont l'objet même de cette technologie d'évaluation. Il raisonne sur la base de deux mondes étanches, l'ESR d'un côté et les instances gestionnaires de l'autre. Une telle dichotomie est pourtant loin d'être évidente.

Promu à la fois par des instances de gestion de la recherche et par des chercheurs, le classement de revues se voit souvent assigné plusieurs motifs en tension. Le premier est de distinguer les « meilleurs » lieux de publication et de miser ainsi sur une incitation dont le modèle est tourné vers la reconnaissance internationale – c'est-à-dire le plus souvent les exigences et standards en vigueur dans une partie du monde anglophone – quelles que soient les disciplines, SHS comprises. Le deuxième objectif est de combler les lacunes des bases de citations telles le Web of Science¹ ou Scopus, maintes fois épinglés pour la faible représentativité des revues non anglophones en SHS [Hicks, 1999 ; Katz, 1999 ; Pons-Novell et Tirado-Fabregat, 2010]. Ces deux aspects sont dorénavant bien connus, largement discutés et critiqués [Lemercier, 2010]. Un troisième motif, beaucoup moins repéré et aux conséquences non négligeables sur les formes d'évaluation, est cependant à l'œuvre : très souvent, les promoteurs de classements de revues souhaitent explicitement développer une technologie d'évaluation qui soit ajustée aux SHS par sa position intermédiaire entre les indices bibliométriques et des systèmes d'évaluation par les pairs. Non seulement les classements de revues ne consistent pas en un outil de mesure fondé sur une formule plus ou moins complexe (à l'inverse du facteur d'impact des revues par exemple), mais les revues qui y figurent, et leur rang, font également l'objet d'une approbation par des

¹ Pour les SHS, le Web of Science est composé du Social Science Citation Index et du Arts and Humanities Citation Index, deux outils produits par l'Institute for Scientific Information respectivement à partir de 1972 et 1978.

spécialistes du domaine en question. Présentant un étalonnage des revues les unes par rapport aux autres, le classement est considéré comme un équipement du jugement qui est ouvert au dénombrement ordonné sans se fonder sur le comptage de citations, et qui simultanément donne des indices de qualité sans nécessairement passer par la lecture des articles.

Prenant au sérieux cette position, nous explorerons ici les modes d'existence de cette technologie d'évaluation spécifique. Bien que la formule générique du « classement de revues » se propage au plan international², différentes versions se développent parallèlement : leurs modalités de production, les valeurs défendues par leurs promoteurs et leurs usagers, aussi bien que leurs formes concrètes sont extrêmement variées. Cette variation s'articule à un degré de technicité qui va de la simple liste alphabétique, sans hiérarchisation interne et fondée sur un critère d'inclusion unique (e.g. « revue à comité de lecture »), au classement sophistiqué dans lequel les revues sont ordonnées en différents rangs, à partir soit d'une mesure numérique continue, soit de notes attribuées à dire d'experts. Entre la liste et le classement, la catégorisation regroupe les revues dans des classes homogènes identifiées par une lettre, un nombre, ou encore des étoiles. Bien qu'isolables sur le plan analytique, ces dénominations constituent un espace de possibles mobilisable dans la fabrique, la critique et les usages. Par exemple, en France l'Agence d'Evaluation de la Recherche et de l'Enseignement Supérieur (AERES) considérait avoir produit une « liste » en 2008, alors que ses critiques évoquaient quasi systématiquement un « classement » [Pontille et Torny, 2012].

Cette plasticité des formes invite à rendre compte des variations d'un mouvement d'ampleur à partir d'une sélection raisonnée. Nous étudions ici les cas de deux pays « marginaux » (Australie et Norvège) qui cherchent explicitement à mesurer leur importance sur le plan mondial en développant un outil pour l'ensemble des disciplines. En 2004, le gouvernement norvégien et la Norwegian Association of Higher Education Institutions (UHR) codéfinissaient un modèle d'évaluation de la recherche destiné à répartir les financements publics dans lequel étaient érigées des listes de revues et d'éditeurs de livres. Et en juin 2008, l'Australian Research Council (ARC) présentait ses « draft rankings », résultat d'un vaste exercice de consultation destiné à produire un classement des revues. En contrepoint, un troisième cas nous permettra d'illustrer une démarche centrée sur une partie des SHS, sur une base transnationale et ayant pour objectif explicite de remplacer l'Arts and Humanities Citation Index comme référence pour les Humanités, en incluant notamment des revues dans des langues minoritaires. Six ans après le lancement du projet, l'European Science Foundation (ESF) publiait progressivement en 2007 et 2008 les « initials lists » de l'European Reference Index for the Humanities (ERIH). À l'étude détaillée de ces trois classements, nous adjoindrons quelques éléments illustratifs issus d'autres cas, situés en position intermédiaire entre valorisation de la marge et volonté d'internationalisation.

Afin de rendre compte des écarts entre les classements de revues, notre analyse sera volontairement restreinte à leurs conditions de fabrique³, sans revenir sur les fortes critiques

² « Journal rankings » ou « league tables » en anglais.

³ L'analyse s'appuie sur un large corpus de textes produits par ces trois institutions sous la forme de documents officiels (rapports, guidelines) et d'informations disponibles sur leur site web, ainsi que de nombreuses prises de positions publiques (éditoriaux, pétitions, points de vue, articles...) à propos de ces classements de revues.

qu'ils ont suscité de la part d'universitaires et de chercheurs qui, d'un cas à l'autre, partagent nombres d'arguments [Pontille et Torny, 2010]. Nous mettrons l'accent sur quatre opérations fondamentales communes à ces différents processus afin de saisir leur variation : la désignation des instances chargées d'élaborer les classements de revues, le découpage des domaines disciplinaires, la sélection des revues pertinentes et la qualification pour les hiérarchiser. L'étude des trois classements de revues proposée ici permettra ainsi de discuter les enjeux qui travaillent cette technologie d'évaluation spécifique.

La production des listes : entre désignation des experts et consultation des pairs

Un dispositif récurrent caractérise l'élaboration des classements de revues : loin de résulter du travail d'une seule personne, ceux-ci émergent toujours d'un processus de délégation à un ou plusieurs groupes réunis et mandatés pour juger de la qualité des revues candidates [Hicks et Wang, 2011]. Si l'appel au jugement d'autres personnes reconnues compétentes caractérise également les trois cas analysés ici, les formes concrètes de cette délégation sont spécifiques à chacun.

En Norvège, le *Research National Council*, créé en 1993 par la fusion d'organismes plus spécialisés, a mené des évaluations en mobilisant le jugement d'experts internationaux, qui ont été ensuite comparées avec des mesures de citation [Aksnes et Taxt 2004]. Réalisées sous l'égide du gouvernement norvégien, ces évaluations visaient à améliorer la qualité des productions scientifiques et à favoriser l'internationalisation de la recherche. Cet objectif était également partagé par l'UHR, qui s'est vu confier en 2003 la réalisation de bases bibliographiques exhaustives [Sivertsen, 2008]. En dépit des variations importantes existant entre les SHS et la plupart des autres disciplines, l'UHR a proposé un système unifié fondé sur le recueil des ISSN (revues) et ISBN (ouvrages et chapitres d'ouvrage) considérés comme académiques [UHR, 2005]. Ce recueil exhaustif des lieux de production effective des chercheurs, confié à un institut de recherche en sciences sociales pendant deux ans, a permis à l'UHR d'élaborer une première base bibliographique, améliorée avec l'aide de la *National Library of Norway*. Un « comité académique », regroupant les représentants disciplinaires de l'UHR, a ensuite produit des critères pour classer les supports de cette base en deux « niveaux » (level 2 et level 1) et proposé un système de points prenant en compte le type de support, le « niveau » et les cosignatures. Cet outil a été approuvé par le gouvernement norvégien et a commencé à être utilisé en 2006, les bases bibliographiques étant mises à jour chaque année par consultation publique sur le site web de l'UHR.

En Australie, la production en recherche a été utilisée à partir de 1995 pour répartir le financement de base des universités et, en conséquence, la production totale, mesurée à partir du nombre d'articles référencés dans le Web of Science, a cru fortement [Butler, 2004]. Face aux critiques d'un découplage de la qualité, le gouvernement australien a adopté une procédure d'évaluation qualitative proche du Research Assessment Exercise britannique. Une nouvelle institution, le *Research Quality Framework* a été créée sous la tutelle de l'ARC pour proposer d'autres critères. Deux types d'indices ont été substitués aux mesures classiques (e.g. facteur d'impact ou nombre d'articles) : les citations brutes et des classements des lieux de production, que ce soit des articles, ouvrages ou conférences. Un vaste exercice de classement a ainsi été réalisé à l'automne 2007, sur la base de listes de

revues préétablies par le *Research Evaluation and Policy Project*, un centre de recherche spécialisé qui a développé une expertise pour les SHS [Donovan, 2005]. Des groupes disciplinaires, mis en place par ce centre – et le *Department of Education Science and Training* a insisté sur leur indépendance – étaient invités à opérer une large consultation des communautés scientifiques pour produire des classements organisés en quatre catégories (A*, A, B, C).

Créée en 1974, l'ESF est un regroupement de soixante-dix-neuf organisations, issues de trente pays européens, visant à promouvoir la science de haute qualité et les formes de coopération scientifique à travers le financement de programmes de recherche. En son sein, le comité de pilotage pour les Humanités a été chargé en 2001 de compiler des listes de bonnes revues académiques. Entre 2003 et 2004, les communautés scientifiques ont alors été sollicitées afin de fournir, à travers un processus de consultation nationale, la matière première du projet ERIH. Après l'élaboration d'une première liste de revues, le comité de pilotage a nommé 15 groupes d'experts (composés chacun de quatre à sept personnes) pour effectuer des éliminations, additions et corrections, ainsi que pour classer les revues retenues en trois catégories (A, B, C), le tout bénéficiant d'un financement européen pour le développement d'infrastructures. L'objectif était d'établir des premières listes à l'automne 2006, qui ont ensuite fait l'objet d'une large consultation européenne, auprès des organisations membres de l'ESF, des communautés scientifiques nationales, des associations professionnelles et des centres de recherche. Début 2007, les listes ont été révisées par les groupes d'experts à la lumière des 4 818 commentaires formulés lors du processus consultatif. Après une accréditation par un comité de l'ERIH, les listes ont été progressivement publiées en 2007 et 2008.

Bien qu'elles élaboraient un outil d'évaluation relativement similaire, les trois organisations n'ont pas développé le même mode opératoire. D'une part, la production d'un premier classement de revues s'appuyait sur des collectifs différents. En Norvège, l'instrument a résulté du travail d'un comité académique unique, regroupant des représentants disciplinaires, qui a produit des critères pour classer les revues et les éditeurs identifiés préalablement dans une base bibliographique par un organisme spécialisé. En Australie, l'élaboration de l'outil a été confiée, après la constitution préalable de listes de revues, à des groupes disciplinaires placés sous l'égide d'un centre de recherche spécialisé dans l'évaluation académique. Le classement de l'ERIH a quant à lui découlé de l'activité de quinze groupes d'experts internationaux, nommés par le comité de pilotage pour les Humanités, qui ont corrigé et complété des listes disciplinaires préétablies par des institutions des organisations membres de l'ESF. D'un cas à l'autre, les collectifs qui ont produit la technologie d'évaluation « classement de revues » étaient différents tant par leur composition (un comité unique, des groupes disciplinaires, des experts internationaux) que par leur principal mandat (élaborer des critères de classement d'une base bibliographique, organiser une large consultation, amender et valider des listes existantes).

D'autre part, la production d'un premier classement de revues résultait de processus très variables de consultation des communautés scientifiques. Les groupes disciplinaires australiens étaient chargés de pratiquer une large consultation, en particulier des sociétés savantes, alors que les groupes d'experts de l'ERIH ne consultaient que par la voie institutionnelle des membres de l'ESF. De son côté, le comité académique norvégien de l'UHR s'est d'abord appuyé sur des données agrégées, puis a également consulté les

membres universitaires pour l'établissement des premières listes. Dans chaque cas, les groupes n'opéraient donc pas seuls, ils intervenaient à l'interface d'une instance, gouvernementale ou européenne, et des professionnels invités à s'exprimer sur la première ébauche du classement. Leur action prenait cependant place dans un agencement particulier : le comité académique norvégien est intervenu après la constitution d'une base bibliographique par un organisme spécialisé et son amélioration par les membres de la bibliothèque nationale ; les groupes disciplinaires australiens ont consulté les communautés scientifiques une fois les listes érigées par un centre de recherche ; les groupes d'experts de l'ERIH ont d'abord corrigé la première ébauche des listes, puis les ont révisées à partir des commentaires formulés lors d'une phase de consultation.

La délimitation de domaines disciplinaires

La plasticité de la technologie d'évaluation est également manifeste dans la définition de son périmètre. Chaque classement de revues concerne des disciplines qui ne représentent qu'une partie des domaines scientifiques dont s'occupent les organisations : alors que l'ESF se concentre sur les Humanités pour l'ERIH, l'ARC considère des indices de qualité tout spécialement adaptés aux sciences sociales, et l'UHR distingue trois groupes de disciplines, les SHS se répartissant dans deux (B et C), à l'exception de la psychologie regroupée avec des disciplines ne publiant quasiment que dans des revues.

Tout d'abord, la délimitation des ensembles disciplinaires était sensiblement différente d'une organisation à l'autre (cf. tableau 1). Avec sa forte orientation en sciences de la documentation, l'UHR a distingué 25 disciplines qui ne reprenaient pas nécessairement des découpages usuels, mais s'appuyaient sur les catégories du Web of Science. L'ERIH a différencié depuis le début 15 disciplines dans les Humanités, qui incluent des ensembles universitaires classiques (e.g. Philosophie, Anthropologie) et des champs de recherche plus contemporains (e.g. Gender Studies). L'ARC a reconnu 10 ensembles disciplinaires pour les SHS, allant de « Education » à « Law » en passant par les « Studies in Human Society ». Ces vastes catégories comprennent entre quatre et neuf « champs de recherche » (Fields of Research – FoR), par exemple pour Economics : Economic Theory, Applied Economics, Econometrics, et Other Economics. Le cas de la psychologie illustre bien à lui seul l'importante flexibilité des rattachements des disciplines à diverses catégories : relevant des « humanités » pour l'ERIH, classée dans le même groupe que les sciences du vivant à l'UHR, identifiée comme une « science humaine et sociale » par l'AERES, la psychologie est également rattachée à la section 27 du CNRS en France, « comportement, cognition, cerveau », intégrée à l'Institut des sciences biologiques.

Tableau 1. La gamme disciplinaire dans trois classements de revues

UHR 2006	ERIH 2008	ARC 2008
Anthropology and ethnology	Anthropology	—
Archaeology	Archaeology	—
Architecture	Art, architectural design history	—
Art history	—	—
Business administration, management	finance, —	Commerce, Management
Classical studies	Classical studies	—
Criminology	—	—
Drama	—	—
—	—	Economics
Education	Pedagogical and educational research	Education
Geography, demographics and regional development	—	—
—	Gender studies	—
History	History	History and Archaeology
—	History and philosophy of science	—
Humanistic media research	—	—
Law	—	Law and Legal Studies
Linguistics	Linguistics	Language, Communication and Culture
Library and information science	—	—
Labour research	—	—
Literary research	Literature	—
Media and communication	—	—
Music	Music and Musicology	—
—	Oriental and African studies	—
Philosophy	Philosophy	Philosophy and Religious Studies
Political science	—	—
Psychology	Psychology	Psychology and Cognitive Sciences
—	Religious studies and theology	—
Social economics	—	—
Social work	—	—
Sociology	—	—
—	—	Studies in Creative Arts and Writing
—	—	Studies in Human Society
—	—	Tourism and Services

Ensuite, la gamme des disciplines couvertes était également très hétérogène. Si des disciplines traditionnelles se retrouvaient dans les nomenclatures produites par les trois organisations, elles n'étaient pas toujours appariées de manière identique. Par exemple l'histoire figurait en « History » à l'UHR et à l'ERIH, mais en « History and Archaeology » à l'ARC. De son côté, la sociologie était isolée à l'UHR, alors qu'elle ne relevait pas des humanités à l'ERIH, et qu'elle était un champ des « Studies in Human Society » à l'ARC. Par ailleurs, certains domaines de recherche n'étaient identifiés que dans un seul classement de revues. Ainsi, les « Gender Studies » n'étaient présentes qu'à l'ERIH, le domaine « History and Philosophy of Specific Fields » figurant uniquement à l'ARC, la « Criminology » et « Drama » seulement à l'URH. La mise en regard avec d'autres classements de revues nationaux renforce le constat de l'existence de singularités : ainsi au Brésil⁴ et en France, on trouve respectivement « Éducation physique » et « STAPS » qui relèvent du même champ disciplinaire, absent des trois classements étudiés ici. Les Brésiliens considèrent également des disciplines qui leur sont propres, comme « Sciences sociales appliquées » ou « Santé publique ». Ces écarts sont bien entendu le résultat d'histoires institutionnelles nationales, mais marquent également l'état de rapports de forces entre disciplines : en France, les « Études de genre » n'ont pas été reconnues comme domaine de recherche autonome, alors que les « Sciences de l'information et de la communication » ont fini par obtenir l'officialisation de leur propre liste de revues [Pontille et Torny, 2012].

Enfin, de tels découpages posaient très directement le problème des revues interdisciplinaires et transdisciplinaires, à propos desquelles les positions des organisations divergeaient. L'ERIH considérait que ces revues pouvaient relever de plusieurs groupes d'experts : « Journals covering overlapping fields may be found in the lists produced by two or more panels (possibly even with different gradings) » (ERIH, 2007 : 1). Inversement, l'ARC soutenait une vision strictement disciplinaire, en demandant aux groupes disciplinaires de ne pas classer les revues relevant majoritairement d'autres « champs de recherche » (FoR) et en créant une catégorie spécifique de « multidisciplinary journals »⁵. C'était également l'option retenue par l'UHR qui a défini deux champs, « cross-disciplinary social science research » et « cross-disciplinary humanities research », qui regroupaient les laissés-pour-compte des choix disciplinaires, comme les « women studies » [UHR, 2005 : section 5.4] ; le plus important pour l'UHR étant que ces recherches relevaient du groupe C avec ses règles propres.

La variété de ces découpages montre donc à quel point des catégories qui sont, à distance, évidentes et consensuelles comme « les humanités » ou « les sciences sociales », donnent lieu à des arrangements bien spécifiques d'une organisation à l'autre. Une fois les domaines disciplinaires clairement délimités et les groupes de personnes réunis autour d'un mandat précis, l'élaboration proprement dite des listes de revues pouvait commencer.

⁴ <https://sucupira.capes.gov.br/sucupira/public/consultas/coleta/veiculoPublicacaoQualis/listaConsultaGeralPeriodicos.jsf> (consultée le 27.04.2016).

⁵ Ce traitement spécifique a été largement critiqué et perçu comme une relégation de la recherche interdisciplinaire par rapport aux publications disciplinaires.

La sélection des revues scientifiques

Pour ces trois organisations, la production d'un classement spécifique aux SHS visait à inclure les nombreuses revues, en particulier non anglophones, dans lesquelles publient les chercheurs européens, australiens et norvégiens. Or la volonté de confectionner des outils qui englobent davantage de supports de publication se heurte à un problème inhérent à toute entreprise de ce type : formuler des critères pour définir ce qu'est une revue scientifique.

La référence à des procédures était constante dans les documents des trois organisations. Les standards procéduraux s'inscrivent dans une longue tradition de l'évaluation par les pairs : existence d'un comité de rédaction, anonymat des auteurs d'articles et/ou des évaluateurs, production systématique de rapports écrits, appui sur des experts extérieurs au comité de rédaction [Zuckerman et Merton, 1971; Baruch et al., 2008]. Mais l'appel à des références partagées sous l'appellation d'évaluation par les pairs ne suffisait pas toujours pour définir les modalités de sélection des revues. Les groupes d'experts étaient potentiellement tendus entre deux options : classer l'ensemble des supports existants qui déclarent s'aligner sur de telles procédures⁶ ou opérer un tri pertinent dans le cadre défini par chaque organisation.

Face à cette alternative, l'ARC a retenu comme standards minimaux l'existence d'un ISSN et d'une évaluation par les pairs. Elle a proposé à ses comités des premières listes extrêmement longues et ceux-ci ont classé la quasi totalité des revues proposées dans les limites de chaque cadre disciplinaire décrit plus haut et a finalement retenu 10 241 revues pour ses « draft rankings ».

De son côté, l'ERIH optait pour des critères d'inclusion détaillant davantage les standards attendus de publication :

All journals included must fulfill normal international academic standards, i.e. selection of articles is based on an objective review policy. [...] The journals must fulfill basic publishing standards (i.e. ISSN, timeliness of publication, complete bibliographic information for all cited references, full address information for every author [ERIH, 2007 : 1, italiques dans la version originale].

Si cela a conduit l'ERIH à la compilation de 12 212 revues, elle demandait à ses groupes d'experts d'adopter une politique de sélection en limitant partiellement les classements aux revues européennes et en indiquant de ne retenir que les « bonnes revues scientifiques ». En conséquence, moins de la moitié des revues présentes au départ ont été retenues dans les « initial lists » (soit 5 888 revues).

Enfin, l'UHR inscrivait le support de publication dans une nécessaire économie de diffusion de la connaissance vers des publics académiques :

An academic publication is defined according to four criteria; all four of these must be satisfied. The publication must: 1. present new insight; 2. be presented in a form that allows the research findings to be verified and/or used in new research activity; 3. be written in a language and have a distribution that make

⁶ Se présenter comme une revue à comité de lecture s'appuyant sur une évaluation par les pairs est aujourd'hui valorisé. Il existe néanmoins des positions dissidentes, où des revues refusent ces procédures au profit des seuls choix du comité éditorial comme *Medical Hypotheses*. Voir « Editor says no to peer review for controversial journal », *Nature*, 18.03.2010, <http://www.nature.com/news/2010/100318/full/news.2010.132.html> (consultée le 27.04.2016).

the publication accessible to most interested researchers; 4. appear in a publication channel (...) that has routines for external peer review [UHR, 2005 : section 3].

S'il s'agissait d'établir une vision exhaustive des lieux de production des chercheurs norvégiens, c'est bien une logique de diffusion qui dominait le processus de sélection. Aussi l'UHR recommandait d'éliminer les revues « locales » : définies par le fait que deux tiers des auteurs appartiennent à la même institution [UHR, 2005 : section 4.4], ces revues étaient exclues parce que l'UHR considérait qu'elles ne reposaient pas sur une évaluation externe et que leur prise en compte entraînerait un détournement de l'instrument en provoquant une inflation de publications à des fins purement évaluatives. C'est sur cette base que l'UHR a retenu plus de 6 500 revues.

Chaque organisation a donc d'abord adopté une définition propre de ce qu'est une revue scientifique comme support d'évaluation et de diffusion. Cette opération était conçue comme peu sélective, ne distinguant pas les technologies contrastées existant à l'intérieur même de l'espace des revues académiques [Pontille et Torny, 2015]. Dans cette logique, retenir une revue avant expertise constituait une certification de son processus de production, à la manière de l'audit pour les entreprises [Power, 1997]. En revanche, la sélection par les groupes d'experts était fortement ajustée aux objectifs prédéfinis de chaque organisation, façonnant donc des classements sensiblement différents. Chacune définissait les frontières de l'exhaustivité à partir de ses propres impératifs : un maximum de revues dans chaque domaine disciplinaire à l'ARC, une volonté d'exhaustivité sous contrainte de « non-localité » à l'UHR, des supports de publication exclusivement européens et de bonne qualité à l'ERIH.

Les définitions de la qualité

Loin de s'imposer naturellement, la qualité des revues académiques a fait l'objet de vastes discussions depuis au moins les années 1920 [Gross et Gross, 1927 ; Bradford, 1934]. Face à l'augmentation exponentielle du nombre de titres publiés, les sciences documentaires ont développé de nombreuses mesures pour les hiérarchiser, au sein des disciplines et entre disciplines. Le premier type de mesures pour ordonner les revues reposait sur des indices de production : nombre d'articles publiés, nombre de numéros par an, nombre d'années de parution... Complémentairement, des mesures d'audience ont été élaborées : nombre d'abonnements pour les revues imprimées, nombre de consultations en ligne pour les revues électroniques, nombre de téléchargements des articles. Au croisement de la production et de l'audience, les mesures de citation, communément appelés bibliométrie à partir des années 1970 [Pritchard, 1969], se sont multipliées ces dernières années [Bollen et al., 2009 ; Cronin et Sugimoto, 2014, 2015]. La recherche de la « bonne » mesure de la qualité conduit à des raffinements méthodologiques progressifs, tout comme à l'invention de nouvelles mesures complémentaires ou alternatives [Wouters et Costas, 2012 ; Wilsdon et al. 2015]. Dans cette tradition, classer une revue revient à sélectionner les indices qui rendent compte de manière proximale de sa qualité, à l'instar du travail des statisticiens sociaux [Desrosières, 1995].

Si les organisations reconnaissaient que les revues sont un composite d'éléments variés, elles ont rejeté l'usage d'outils quantitatifs pour élaborer leurs classements. En effet, non seulement elles ont fait appel à des groupes d'experts pour établir les listes, mais elles ne leur ont fourni aucune mesure sur les revues. En revanche, elles leur ont donné des consignes écrites permettant d'identifier la qualité relative des revues à partir d'une double

définition. D'une part, elles ont proposé une définition intrinsèque de leur catégorisation, et notamment de la catégorie dédiée aux meilleures revues. Elles ont ainsi élaboré une série de descripteurs pour qualifier les revues ou les articles qu'elles contenaient. Aussi, l'ARC définissait de cette manière les revues les plus prestigieuses :

Typically an A* journal would be one of the best in its field or subfield in which to publish and would typically cover the entire field/subfield. Virtually all papers they publish will be of a very high quality. These are journals where most of the work is important (it will really shape the field) and where researchers boast about getting accepted. Acceptance rates would typically be low and the editorial board would be dominated by field leaders, including many from top institutions [ARC, 2008].

Pour sa part, l'ERIH considérait que la catégorie des meilleures revues regroupait celles qui honoraient d'autres critères :

Journals category A: i.e. high-ranking international publications with a very strong reputation among researchers of the field in different countries, regularly cited all over the world [ERIH, 2007 : 1].

Enfin, l'URH adoptait une définition du meilleur niveau de publication, spécifiquement pour les disciplines des groupes B et C ⁷, fondée uniquement sur une exigence d'internationalisation des supports :

Level 2 channels must be perceived as the leading publication channels in a wide variety of academic contexts, must publish the most outstanding works by researchers from different countries [UHR, 2005 : section 5.2].

Ces différents descripteurs dessinaient le prototype d'une « revue de très grande qualité », même si leur évaluation concrète pose problème, en particulier pour juger de la « qualité », de « l'importance » ou de « l'originalité » qui sont des critères nécessairement dépendant des univers de références associés [Lamont, 2009]. De tels jugements sont le plus souvent relatifs à un ensemble de revues et non absolus, comme l'illustrent bien les qualifications « one of the best » ou « leading ». Néanmoins, ce jugement relatif peut être substantifié, en particulier sous la forme d'une échelle de réputation. Ainsi, en France, l'AERES a indiqué en 2008 dans le court argumentaire qui accompagnait la liste générale des revues que « la classification A, B, C correspond à l'étendue du rayonnement des revues » [Pontille et Torny, 2012]. Une définition intrinsèque des lieux de publication, fondée sur des critères descriptifs aussi fins soient-ils, ne garantit donc absolument pas d'aboutir à un classement de revues identique.

D'autre part, les organisations ont introduit une définition relationnelle de leur classement qui s'est principalement exprimé par la mise en place de quotas de répartition. L'ERIH recommandait « that in category A, only 10 to 25% of the total list should appear; this percentage target will differ from one discipline to another » [ERIH, 2007 : 1]. Le critère de l'UHR était plus strict puisqu'en aucun cas, les revues classées en « level 2 » ne devaient représenter plus de 20% du total de la production – et non des titres – pour chaque discipline. Pour sa part, l'ARC définissait un cadre encore plus contraignant en donnant une clé de répartition complète où la catégorie A* contenait 5% du classement total, la catégorie A 15%, la catégorie B 30%, et la catégorie C 50%. L'importance relative des catégories de revues était ainsi conçue de plusieurs façons d'une organisation à l'autre.

⁷ Pour le groupe A, il était nécessaire que les revues soient également référencées par le Web of Science.

Or la coexistence de ces deux définitions n'allait pas de soi. Elle générait une tension qui a très nettement orienté le statut de la dernière catégorie. L'objectif relativement ouvert de l'ARC, qui consistait à lister un maximum de revues, a eu pour conséquence de définir la catégorie C négativement : « Tier C includes quality, peer reviewed, journals that do not meet the criteria of the higher tiers » [ARC, 2008]. Les revues regroupées dans cette catégorie, qui constitue la moitié de la liste, sont celles qui ne sont ni A*, ni A, ni B. En adoptant une démarche explicitement sélective, l'ERIH considérait au contraire qu'au terme du processus la catégorie C, réservée aux revues « with an important local / regional significance in Europe », contenait de bonnes revues :

The Expert Panels emphasise that high quality research appears in journals throughout all three categories. [...] The "C" category represents the real European added value of ERIH. Unlike the journals in the two other categories, "C" journals listed in ERIH fully reflect the linguistic and cultural diversity of Humanities research production in Europe. For many of them, ERIH offers the first opportunity to gain wide visibility also within less specialised research communities [ERIH FAQ, 2008: "Does the categorisation A, B and C reflect differences in quality?" and "Do the C journals make up some kind of residual category?"].

Même si l'ERIH défendait sa catégorie la plus basse, l'enjeu du classement était bien de promouvoir des publications dans des revues de « haute qualité ». Il en est de même au Brésil pour le classement des publications destiné aux doctorants, dont la structure pyramidale visait à valoriser les deux catégories les plus hautes sur les huit créées [Rocha-e-Silva, 2010]. En Norvège, l'objectif explicite de la création de deux niveaux de publication par l'UHR était d'inciter la publication dans la plus haute, et c'est sur la base du taux de publication national en « level 2 » que l'instrument est jugé comme une réussite [Gilhus et Sivertsen, 2010].

Le processus de sélection des revues conformes à certaines exigences académiques, la définition des frontières de l'exhaustivité et les critères érigés pour les classer sur une échelle de valeur différaient donc très nettement d'un cas à l'autre. Dans ces conditions, les trois organisations ont produit des instruments qui se distinguaient aussi bien par la taille (plus de 12 000 revues à l'ARC, contre environ 6 000 à l'ERIH et l'UHR) que par les catégories de classement (deux niveaux à l'UHR, trois classes à l'ERIH, quatre à l'ARC). Autre aspect plus directement lié à leur usage, ces classements de revues étaient également distincts dans leur présentation même : alors que l'UHR donnait accès à une base de données en ligne où les utilisateurs peuvent opérer des tris multicritères, l'ERIH publiait des classements séparés, consultables sous forme de fichiers PDF, pour chacune des quinze disciplines retenues, tandis que l'ARC proposait un tableau unique organisé autour des codes de multiples « champs de recherche » (FoR). Sur le site web des trois organisations, le classement de revues s'appréhendait alors soit comme un objet unique, soit il se déclinait en plusieurs sous-listes. Parallèlement, l'AERES expérimentait plusieurs formules : en 2008, l'outil était disponible sur un site au format html, et se présentait à la fois par ensembles disciplinaires et une liste générale, résultat de leur agrégation ; lors du processus de révision amorcé en 2009, les listes ont ensuite pris la forme de fichiers PDF séparés les uns des autres⁸.

⁸ <http://www.aeres-evaluation.fr/Publications/Methodologie-de-l-evaluation/Listes-de-revues-SHS-de-l-AERES>, consultée le 27.04.2016.

Excellence internationale et pertinence nationale

Chaque pays possède donc son propre classement de revues, qui présente des critères issus d'une histoire institutionnelle propre, mais aussi des facteurs beaucoup plus circonstanciels. Ces initiatives portent néanmoins des valeurs communes : l'international, la qualité, l'excellence. En affichant de telles valeurs, orientées vers la promotion des productions en SHS, elles ne font pas qu'élaborer des outils jugés alternatifs aux indices de mesures bibliométriques. L'évolution propre à chaque classement est marquée par une tension entre l'horizon d'une excellence scientifique internationale et un espace de pertinence immédiatement national.

Le mouvement vers une transnationalisation grandissante est d'abord présent du côté des promoteurs de classements : les Norvégiens ont observé ce qu'avait mis en place les Australiens [Sivertsen, 2009] pour éviter de produire un outil favorisant la quantité plutôt que la qualité ; des chercheurs danois [Lillkalf, 2009] et suédois [Sandström et Sandström, 2009] ont examiné la reprise, pour leur pays, de ce qui est désormais dénommé le « modèle norvégien » [Sivertsen, 2016] et la construction d'une « Nordic list » mutualisée est en cours ; l'AERES a reproduit le modèle de l'ERIH [Merllié, 2009] tout en modifiant le sens des jugements portés sur les revues⁹. Loin d'une opération isolée et totalement indépendante, chaque production d'un nouveau classement de revues s'inscrit dans un monde déjà peuplé par d'autres, sur lesquels les promoteurs peuvent s'appuyer avant d'élaborer le leur. Ensuite, ce mouvement transnational existe également du côté des instances classées, en particulier des revues [Pontille et Torny, 2010]. Celles-ci n'hésitent pas à adopter des positions stratégiques¹⁰, à faire de leur place dans un classement un label de qualité, voire à multiplier les formes de label sur leur site web (e.g. facteur d'impact, rangs dans différents classements). Toutefois, l'appui sur plusieurs classements n'est pas uniquement orienté vers l'attractivité. Les revues peuvent aussi s'en saisir dans une perspective critique pour faire valoir leur présence ou leur bonne position dans tel classement afin de la voir améliorer dans un autre. C'est notamment le cas de plusieurs revues qui étaient bien positionnées dans celui de l'ERIH et mal notées ou absentes de celui de l'AERES. Enfin, les bibliomètres eux-mêmes participent à ce mouvement en comparant le contenu de ces différents classements de revues [Hicks et Wang, 2011], l'enjeu étant d'aboutir *in fine* à la couverture pertinente des productions des SHS et au classement le mieux ajusté à la « qualité » des revues. Ces classements sont donc devenus des objets publics transnationaux, mobilisables dans diverses situations et comparables pour servir différents objectifs : produire un nouveau classement ici, améliorer une position dans un autre là, évaluer sa pertinence ailleurs.

L'espace de prescription des classements de revues demeure toutefois national (UHR, ARC) ou potentiellement continental (ERIH). Leur déploiement dans la mise en œuvre de politiques publiques s'est d'ailleurs accompagné de réactions critiques et de contestations dans les territoires concernés, par anticipation d'effets induits ou par constat des conséquences après les usages de ces classements de revues dans des évaluations. Ainsi,

⁹ Alors que pour l'ERIH, la catégorie « C » contenait des revues de qualité régionale, pour l'AERES seules les catégories A et B étaient qualifiantes. La valeur des revues notées « C » était donc équivalente à celles qui ne figuraient pas dans le classement. Voir [Pontille et Torny 2012].

¹⁰ La révision du classement australien au printemps 2011 a été de ce point de vue exemplaire, avec des appels lancés par différentes revues à la participation de leurs lecteurs, à l'échelle mondiale, afin de maintenir ou d'améliorer leur cotation.

la défense de la production scientifique en « langue vernaculaire » a émergé comme problème mobilisateur. En combinant un critère intrinsèque d'internationalité et un quota de 20%, l'UHR n'a laissé quasiment aucune place dans sa meilleure catégorie pour des revues publiant en norvégien. Cela a provoqué des débats parlementaires et une pétition de 223 chercheurs en SHS pour la défense de leur langue de travail et d'expression, posant directement la question de l'avenir du norvégien comme langue académique [Brock-Utne, 2007]. Un mouvement similaire de défense des publications en langue française a vu le jour face au classement de revues de l'AERES. Et si les revues australiennes publient en anglais, elles n'en sont pas moins menacées par l'hégémonie des revues britanniques et américaines [Gläser, 2004], très présentes dans les catégories les plus valorisées du classement de l'ARC. Au-delà de la langue d'expression, la question de l'ajustement par rapport à un standard scientifique purgé de considérations locales ou nationales est clairement en jeu.

Parmi les disciplines de SHS dont les productions ont des lectorats extra-académiques, les tensions entre excellence internationale et pertinence nationale sont particulièrement visibles pour le droit. Les chercheurs de cette discipline publient dans des lieux qui n'ont pas pour référence l'international – à l'exception du droit comparé, du droit international et de la théorie du droit. Aussi, lorsque les mêmes critères que pour les autres disciplines sont appliqués à leurs revues, les différents objectifs ne sont pas simultanément tenables. Dans le cas norvégien, valoriser des revues « leaders » revient à renoncer au critère de diversité de l'origine des auteurs puisque 99 % d'entre eux sont américains [UHR, 2005 : section 5.5]. Dans le cas de l'ARC, les problèmes concrets auxquels peuvent répondre les juristes australiens et qui fondent leur utilité sociale ne peuvent pas figurer dans les articles publiés dans des revues anglaises ou américaines [Svantesson, 2009]. En favorisant ces dernières, les classements de l'ARC perdraient de vue leur mission première.

Les tensions entre un horizon international et une pertinence nationale ne travaillent pas uniquement la dynamique interne de chaque classement de revues particulier. Elles sont simultanément au cœur de l'évolution du monde engendré par cette technologie d'évaluation.

Quel(s) monde(s) pour les SHS ?

Examiner en détail les conditions de production de la technologie d'évaluation « classement de revues » n'éclaire pas uniquement leurs principales caractéristiques. C'est aussi le moyen de mettre l'accent sur les points d'achoppements (frontières disciplinaires, critères d'inclusions, modalités de hiérarchies...) inhérents à l'outil de catégorisation. Mais les trois cas étudiés ici, de même que leurs équivalents dans d'autres pays, sont également conçus comme des outils en constante révision, ce qui permet de rendre visibles les processus d'exclusion ou de minoration de certains supports, voire de communautés entières [Pontille et Torny 2010]. Ces révisions ont en effet impliqué des participations variables des instances du monde scientifique (chercheurs, universités, sociétés savantes, revues), et ont entraîné des modifications substantielles du contenu des listes et, parfois même de l'instrument. Outre l'abandon des quotas entre les différentes catégories par l'ARC dans le cas australien, la hiérarchie entre les revues a progressivement disparue des autres classements de revues. Ainsi, dans le cas français, après un mouvement critique d'envergure en 2009, l'AERES a

abandonné sa catégorisation (A, B, C) au profit d'un périmètre de scientificité prenant généralement la forme d'une simple liste non hiérarchisée¹¹. Suite aux importantes critiques lors de la révision du classement en 2011, l'ARC a également annoncé l'abandon de la catégorisation (A*, A, B, C) au profit d'une liste ordonnée par fréquence de publication des divers supports¹². Enfin, les listes de l'ERIH, après une première révision en 2011, se présentent elles aussi sans hiérarchie depuis 2014¹³.

La participation d'un grand nombre d'acteurs aux processus de révision découle de leurs anticipations de la multitude d'usages de cette technologie d'évaluation : recrutement et promotion des individus, choix des supports de publication, politique de promotion des revues, financement des laboratoires et des universités. Si certains usages sont parfois consubstantiels à la production de l'outil, comme dans le cas norvégien, les acteurs mobilisés considèrent toujours que ces instruments, avec leurs formes d'incitation et de reconnaissance propres, modifient sensiblement leur monde social. Si l'élaboration d'un classement présuppose l'existence d'un lien entre la qualité des revues et celle des articles, l'analyse détaillée de la production, proposée ici, permet de dessiner trois mondes profondément différents en fonction du design de la technologie d'évaluation.

Le premier est un monde hiérarchisé autour d'un nombre limité de revues qualifiées de « centrales ». Ici, bien qu'il s'agisse de dépasser le Web of Science, celui-ci sert quasi-systématiquement de point de départ et de socle commun aux classements, les revues présentes dans cet index étant le plus souvent les mieux classées [Hicks et Wang, 2011]. Le deuxième monde est, quant à lui, beaucoup plus vaste et diversifié, incluant potentiellement toute revue à comité de lecture, à condition que celle-ci soit validée par sa présence effective dans une liste pour des disciplines données. Mais ces deux mondes partagent une caractéristique que soulignent les critiques des classements de revues : ils pointent souvent l'absence des ouvrages et chapitres d'ouvrages, alors même qu'ils constituent une part importante de la production, notamment dans les humanités. Aussi le troisième monde les inclut explicitement, à l'image du modèle Norvégien. Non seulement ce sont des supports qui comptent, mais la pondération prévue dans le modèle permet de valoriser davantage les ouvrages que les articles.

L'espace de variations des classements de revues en SHS est donc toujours bordé par deux options : favoriser une « bonne recherche » qui, sous l'effet d'avantages cumulatifs [Merton, 1968], risque de conduire à une science (hyper)normale soutenant des dispositions de conformité sociale chez les chercheurs [Gendron, 2008] ; encourager l'émergence des communautés minoritaires (linguistiques, disciplinaires, interdisciplinaires) et promouvoir la diversité des méthodes, théories et objets, au risque de mener à des formes de relativisme ou d'archipelisation de la recherche.

¹¹ À l'exception des listes Économie, Gestion, et sciences du sport STAPS qui ont maintenu jusqu'en 2016 des catégorisations propres pour distinguer la qualité des revues.

¹² Cela n'empêche pas leur usage cinq ans après, malgré leur caractère « dépassé » aux yeux des producteurs de classement.

¹³ Elles sont gérées par l'organisme norvégien qui entretient les listes de l'UHR. C'est sans doute la raison pour laquelle le critère « deux tiers des auteurs appartiennent à la même institution » pour exclure les revues « locales » est désormais en vigueur pour ces listes ERIH (<https://dbh.nsd.uib.no/publiseringskanaler/erihplus/>, consulté le 27.04.2016).

Références

- Aksnes D.W., Taxt R.E., 2004, « Peer reviews and bibliometric indicators: a comparative study at a Norwegian university », *Research Evaluation*, 13, 1, 33-41.
- Alexander J. K., Lecoutre M., Scherer R. F., 2007, « A global comparison of business journal ranking systems », *Journal of Education for Business*, 82, 6, 321-328.
- Australian Research Council, 2008, « Tiers for the Australian Ranking of Journals ».
- Auranen O., Nieminen M., 2010, « University research funding and publication performance – An international comparison », *Research Policy*, 39, 6, 822-834.
- Baruch Y., Konrad A. M., Aguinis H., Starbuck W. H. (eds.), 2008, *Opening the Black Box of Editorship*, New York, Palgrave Macmillan.
- Bollen J., Van De Sompel H., Hagberg A., Chute R., 2009, « A Principal Component Analysis of 39 Scientific Impact Measures », *PLoS One*, 4, 6, e6022-e6022.
- Bradford S. C., 1934, « Sources of Information on Specific Subjects », *Engineering: An Illustrated Weekly Journal*, 137, 3550, 85-86.
- Brock-Utne B., 2007, « Language of instruction and research in higher education in Europe: highlights from the current debate in Norway and Sweden », *International Review of Education*, 53, 4, 367-388.
- Bruneau W., Savage D., 2002, *Counting Out the Scholars: How Performance Indicators Undermine Universities and Colleges*, Toronto, Lorimer.
- Bruno I., 2008, « La recherche scientifique au crible du benchmarking. Petite histoire d'une technologie de gouvernement », *Revue d'Histoire Moderne et Contemporaine*, 55, 4, 28-45.
- Butler L., 2004, « What happens when funding is linked to publication counts? », in *Handbook of Quantitative Science and Technology Research*, Dordrecht, Kluwer Academic Publishers, 389-405.
- Cronin B., Sugimoto C.R. (eds.), 2014, *Beyond Bibliometrics. Harnessing Multidimensional Indicators of Scholarly Impact*, Cambridge, MIT Press.
- Cronin B., Sugimoto C.R. (eds.), 2015, *Scholarly Metrics Under the Microscope: From Citation Analysis to Academic Auditing*, Medford, New Jersey, Information Today.
- Desrosières A., 1995, « Classer et mesurer : les deux faces de l'argument statistique », *Réseaux*, 71, 11-29.
- Donovan C., 2005, « The governance of social science and everyday epistemology », *Public Administration*, 83, 3, 597-615.
- Encinas de Munagorri R., 2002, « Quel statut pour l'expert ? », *Revue Française d'Administration Publique*, 103, 379-389.
- European Research Index for the Humanities, 2007, « Summary guidelines », Septembre 2007, 2p.
- Gendron Y., 2008, « Constituting the academic performer: the spectre of superficiality and stagnation in academia », *European Accounting Review*, 17, 1, 97-127.
- Giles M.W., Wright G.C., 1975, « Political scientists' evaluations of sixty-three journals », *Political Science*, 8, 3, 254-256.
- Gilhus N.E., Sivertsen G., 2010, « Publishing affects funding in neurology », *European Journal of Neurology*, 17, 1, 147-151.
- Gläser J., 2004, « Why are the most influential books in Australian sociology not necessarily the most highly cited ones? », *Journal of Sociology*, 40, 3, 261-282.
- Glenn N.D., 1971, « American sociologists' evaluation of sixty-three journals », *American Sociologist*, 6, 4, 298-303.
- Gross P.L.K., Gross E.M., 1927, « College Libraries and Chemical Education », *Science*, 66, 1713, 385-389.
- Hicks D., 1999, « The difficulty of achieving full coverage of international social science literature and the bibliometric consequences », *Scientometrics*, 44, 2, 193-215.
- Hicks D., 2012, « Performance-based university research funding systems », *Research Policy*, 41, 2, 251-261.
- Hicks D., Wang J., 2011, « Coverage and overlap of the new social sciences and humanities journal lists », *Journal of the American Society for Information Science and Technology*, 62, 2, 284-294.
- Katz J.S., 1999, « Bibliometric indicators and the social sciences », Brighton, E. Sussex.
- Johnson J.L., Podsakoff P.M., 1994, « Journal influence in the field of management: an analysis using Salancik's index in a dependency network », *The Academy of Management Journal*, 37, 5, 1392-1407.

- Lamont M., 2009, *How Professors Think : Inside the Curious World of Academic Judgment*, Cambridge, Mass., Harvard University Press.
- Lemerrier C., 2010, « Classer les revues de sciences humaines et sociales : comment ? pour quoi ? », *MethIS*.
- Lillkål M., 2009, *Evaluation of Journal Ranking for Research Evaluation Purposes in the Social Sciences*. Master thesis, Royal School of Library and Information Science.
- Merllié D., 2009, « La machine de l'AERES à classer les revues de philosophie. Un démontage statistique », *Revue Philosophique de la France et de l'Étranger*, 134, 1, 91-99.
- Merton R. K., 1968, « The Matthew effect in science », *Science*, 159, 3810, 56-63.
- Moore W.J., 1972, « The relative quality of economics journals: a suggested rating system », *Western Economic Journal*, 10, 2, 156-169.
- Polonsky M.J., Jones G., Kearsley M.J., 1999, « Accessibility: an alternative method of ranking marketing journals? », *Journal of Marketing Education*, 21, 3, 181-193.
- Pons-Novell J., Tirado-Fabregat D., 2010, « Is there life beyond the ISI Journal lists? The international impact of Spanish, Italian, French and German economics journals », *Applied Economics*, 42, 6, 689-699.
- Pontille D., Torny D., 2010, « The Controversial policies of journal ratings : evaluating social sciences and humanities », *Research Evaluation*, 19, 5, 347-360.
- Pontille D., Torny D., 2012, « Rendre publique l'évaluation des SHS : les controverses sur les listes de revues de l'AERES », *Quaderni*, 77, 11-24.
- Pontille D., Torny D., 2015, « From Manuscript Evaluation to Article Valuation: The Changing Technologies of Journal Peer Review », *Human Studies*, 38, 1, 57-79.
- Pritchard A., 1969, « Statistical Bibliography or Bibliometrics? », *Journal of Documentation*, 25, 4, 348-349.
- Power M.P., 1997, *The Audit Society : Rituals of Verification*, Oxford, Oxford University Press.
- Rocha-e-Silva M., 2010, « Impact factor, SCImago indexes and the Brazilian journal rating system: where do we go from here? », *Clinics*, 65, 4, 351-355.
- Sandström U., Sandström E., 2009, « The field factor: towards a metric for academic institutions », *Research Evaluation*, 18, 3, 243-250.
- Sivertsen G., 2008, « Experiences with a bibliometric model for performance based funding of research institutions », *Excellence and Emergence. A New Challenge for the Combination of Quantitative and Qualitative Approaches*, Book of Abstracts, 10th International conference on science & technology indicators, Vienna, 126-128.
- Sivertsen G., 2009, « A performance indicator based on complete data for the scientific publication output at research institutions », *International Society for Scientometrics and Informetrics Newsletter*, 6, 1, 22-28.
- Sivertsen G., 2016, « Publication-based funding: the Norwegian model », In Ochsner M., Sven E. H., and Hans-Dieter D. (eds.), *Research Assessment in the Humanities - Towards Criteria and Procedures*, Springer, pp. 79-90.
- Steward M.D., Lewis B.R., 2010, « A comprehensive analysis of marketing journal rankings », *Journal of Marketing Education*, 32, 1, 75-92.
- Svantesson D.J.B., 2009, « International ranking of law journals: can it be done and at what cost? », *Legal Studies*, 29, 4, 678-691.
- UHR 2005, « A Bibliometric Model for Performance-based Budgeting of Research Institutions », Recommendation from the committee appointed by the Norwegian Association of Higher Education Institutions on assignment from the Ministry of Education and Research, 51 p. (http://www.uhr.no/documents/Rapport_fra_UHR_prosjektet_4_11_engCJS_endelig_versjon_av_hele_oversettesen.pdf).
- Wilsdon, J., et al. (2015). *The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management*, Higher Education Funding Council for England, (<http://www.hefce.ac.uk/pubs/rereports/Year/2015/metrictide/Title,104463,en.html>).
- Wouters P., Costas R., 2012, *Users, Narcissism and Control – Tracking the Impact of Scholarly Publications in the 21st Century*, Utrecht, SURFfoundation.
- Zuckerman H.A., Merton R.K., 1971, « Patterns of evaluation in science: institutionalisation, structure and functions of the referee system », *Minerva*, 9, 1, 66-100.