

La conversation scientifique ouverte

Analyse I/IST n°29 - Mars 2020

L'évaluation ouverte ou la révision ouverte de travaux pourrait être la prochaine révolution du libre accès.

Alors que le Plan S prévoit que la plupart des publications scientifiques soient librement accessibles en ligne d'ici 2021, un nouveau mouvement d'ouverture prend de l'ampleur sous le nom d'*open peer review* ou littéralement « évaluation par les pairs ouverte ». Il ne s'agit plus seulement d'ouvrir les productions scientifiques mais leurs coulisses : l'échange entre auteurs et évaluateurs conduisant à la publication. Même les revues les plus prestigieuses suivent de près ces évolutions : *Nature* vient tout juste de rendre possible la diffusion des rapports d'évaluation¹.

En ce sens, l'évaluation ouverte ne se limite pas à la mise à disposition de documents internes au processus de publication. D'importantes recherches menées depuis quelques années dans le cadre de grands programmes comme *Peere*² mettent en évidence les implications éditoriales, épistémologiques et économiques profondes d'un changement de régime d'évaluation et de la remise en question des procédures classiques de *peer review*. Cette note de synthèse revient sur un débat fondamental autour du futur de l'édition scientifique.

La prochaine révolution du libre accès ?

Les politiques d'évaluation ouverte se développent significativement depuis deux ans. C'est la conclusion principale des données inédites³ collectées par Dietmar Wolfram, Peiling Wang et Hyoungjoo Park à partir d'une analyse semi-automatique des métadonnées éditoriales de dizaines de milliers de revues :

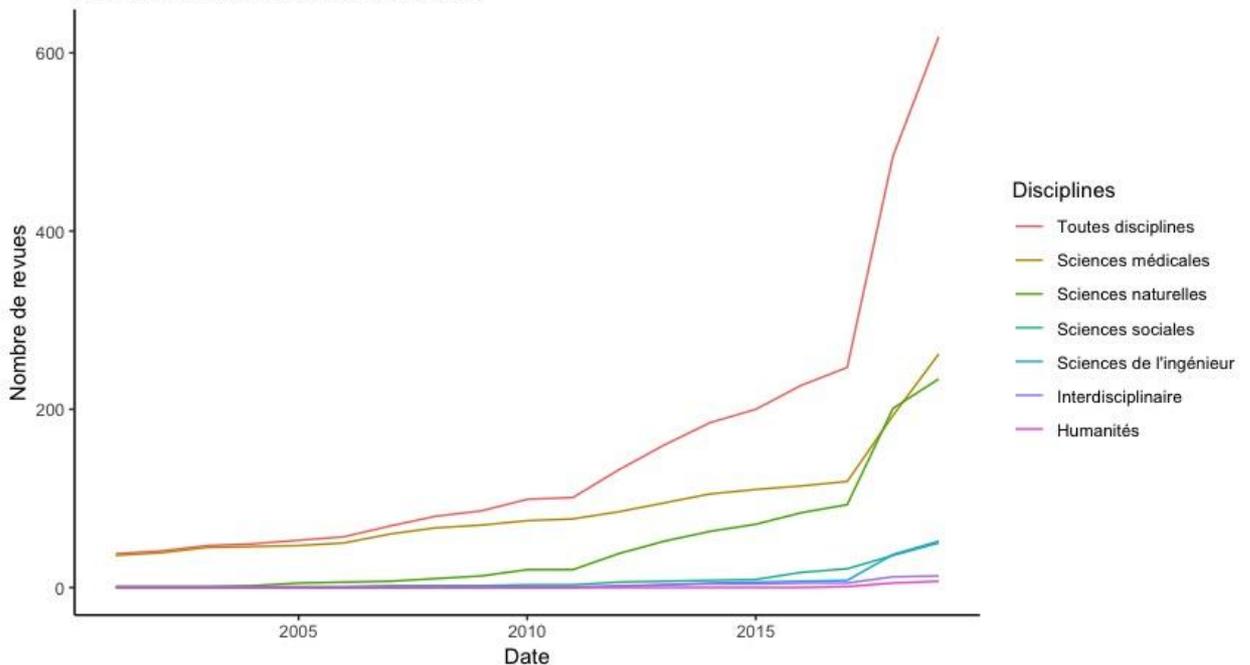
¹ <https://www.nature.com/articles/d41586-020-00309-9> Pour l'instant la procédure est expérimentale.

² <https://www.peere.org/>

³ Ces données inédites nous ont été communiquées avant publication.



Développement des politiques d'évaluation ouvertes Un essor sans précédent depuis deux ans



Source : Dietmar Wolfram, Peiling Wang & Hyoungjoo Park (données inédites)

La pratique de l'évaluation ouverte est déjà attestée en 2001, avec 38 revues dont 36 en sciences médicales. Elle connaît une croissance continue pendant les quinze années suivantes, puis une soudaine accélération à partir de 2017, passant de 247 revues à 618 en 2019. Parallèlement, l'ancrage disciplinaire de l'évaluation ouverte se diversifie. Il s'agit d'abord d'une innovation spécifique des revues de médecine sous l'impulsion du *British Medical Journal* (95% des revues en 2001) qui se répand en sciences naturelles (234 revues en 2019), et, dans une moindre mesure, en sciences de l'ingénieur (52 revues en 2019) et en sciences sociales (50 revues en 2019). Des acteurs importants du libre accès s'y sont récemment convertis. Depuis mai 2019, PLOS permet de divulguer « l'historique du peer review » à titre optionnel⁴.

Dans l'absolu, l'évaluation ouverte reste encore marginale. Dans une première version de leur recherche réalisée uniquement à partir des données du *Directory of Open Access Journal*, Dietmar Wolfram, Peiling Wang et Hyoungjoo Park notent que « le nombre de revues avec une évaluation ouverte représente une part toujours infime de l'ensemble des revues (Wolfram et al., 2019). ».

Néanmoins, la dynamique de croissance de ces deux dernières années met clairement en évidence le potentiel de cette nouvelle pratique éditoriale. Ce développement est d'ailleurs suivi de près par les grands éditeurs scientifiques traditionnels. Ces derniers multiplient les enquêtes préliminaires depuis quelques années. Elsevier a mené une expérience pilote à grande échelle de 2014 à 2018 : cinq revues de médecine et de sciences naturelles ont

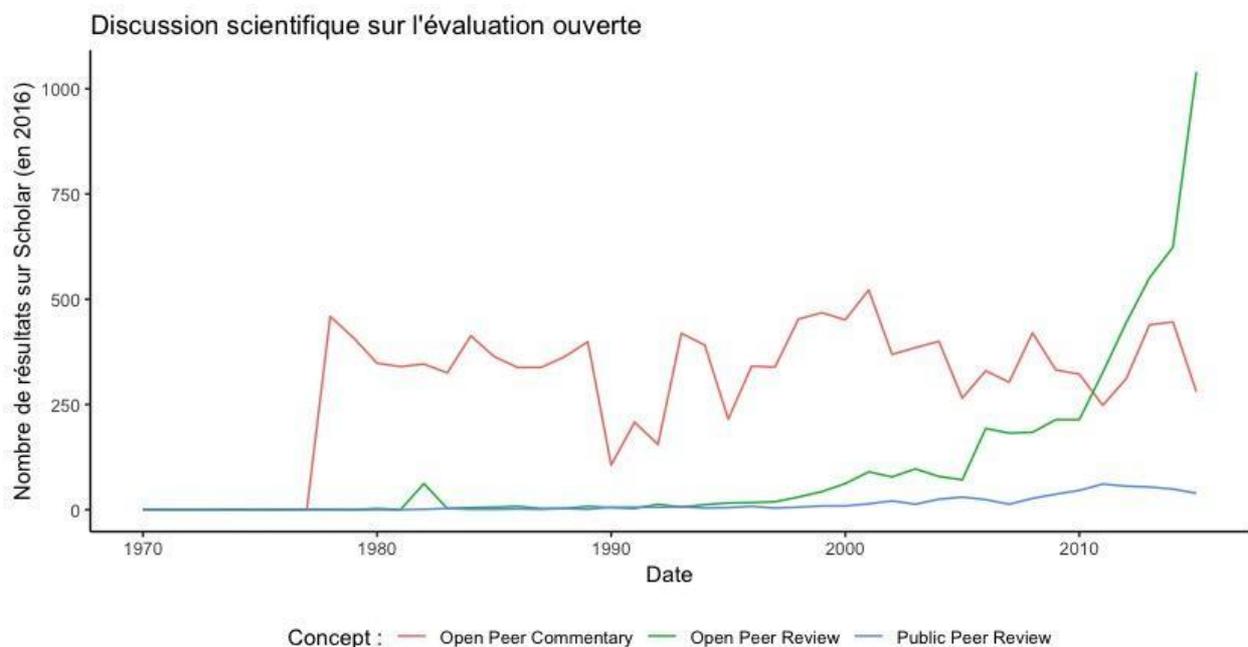
⁴ Madison Crystal, « PLOS Journals Now OPEN for Published Peer Review », *blog.plos.org*, 22 mai 2019

EPRIST

généralisé l'évaluation ouverte (*Agricultural and Forest Meteorology, Annals of Medicine and Surgery, Engineering Fracture Mechanics, Journal of Hydrology: Regional Studies, International Journal of Surgery*)⁵. Springer a tiré parti de l'acquisition du pionnier de l'évaluation ouverte, le *British Medical Journal* pour mener une analyse comparative⁶. Wiley a sondé ses collaborateurs : si le modèle d'évaluation en double aveugle reste largement plébiscité, l'évaluation ouverte est significativement soutenue⁷. En France, Open Edition a mis en œuvre des dispositifs plus innovants d'évaluation ouverte sur la revue *Vertigo* ou sur certaines monographies d'*Open Edition Books*, à partir de systèmes d'annotations en continu. Au vu des tendances actuelles, l'évaluation ouverte pourrait être l'une des formes d'évaluation proposées dans la plupart des grandes revues scientifiques d'ici quelques années. Mais, de quelle forme d'évaluation ouverte parle-t-on ?

L'émergence progressive de modèles d'évaluation ouverte

L'évaluation ouverte a émergée bien antérieurement à la numérisation de l'édition scientifique et au développement du libre accès. Le graphique ci-dessous montre le nombre annuel de résultats dans Google Scholar des trois principales expressions utilisées historiquement : *open peer commentary*, *open peer review* et *public review*.



⁵ <https://www.elsevier.com/reviewers-update/story/innovation-in-publishing/is-open-peer-review-the-way-forward>

⁶ <https://www.springernature.com/gp/advancing-discovery/blog/blogposts/lessons-learned-from-open-peer-review--a-publisher-s-perspective/16123780>

⁷ <https://www.wiley.com/network/latest-content/how-do-researchers-feel-about-open-peer-review>

EPRIST

Le concept apparaît soudainement en 1978, lors du lancement de la revue *Brain & Behavioral Science* sous la direction de Stevan Harnad qui deviendra plus tard l'une des grandes figures du libre accès. Harnad reprend un dispositif expérimenté alors depuis près de vingt ans par une revue en anthropologie, *Current Anthropology* : la publication de commentaires extérieurs associés à l'article. Il théorise cette pratique sous le nom d'« open peer commentary » en 1978 puis, quatre ans plus tard, à l'occasion d'une édition spéciale consacrée au peer review, d'« open peer review ».

Du fait de son ancienneté, le débat sur l'évaluation ouverte est indissociable du débat plus global sur l'évaluation par les pairs. En 1978, les procédures aujourd'hui classiques de *peer review* sont encore des innovations récentes (Pontille and Torny 2015). Ainsi l'évaluation en double aveugle n'a été introduite à titre expérimental qu'en 1955 dans l'*American Sociological Review*. Jusqu'aux années 1970, l'évaluation restait surtout une prérogative de l'éditeur dans d'importantes revues comme *Nature*. Les expériences de Stevan Harnad et de ses collègues constituent surtout une alternative à des normes d'évaluation encore en voie de standardisation.

Le format de l'*open peer commentary* est très différent des formes actuelles d'évaluation ouverte. Il matérialise plutôt une conversation scientifique autour d'une publication qu'une *évaluation*, et annonce plutôt les commentaires de blogs scientifiques. Vingt ans plus tard, le *British Medical Journal* se convertit à l'évaluation ouverte selon un autre modèle (Smith 1999). L'enjeu est alors d'ouvrir un processus d'évaluation déjà préalablement normalisé, en rendant accessible la correspondance entre les évaluateurs et les auteurs, ainsi que leurs identités.

Initialement très influent, le modèle du *British Medical Journal* va être progressivement concurrencé par d'autres approches sous l'effet de trois facteurs conjoints :

- 1) La dissémination progressive de l'évaluation ouverte hors des sciences médicales requiert des acculturations à d'autres environnements disciplinaires ;
- 2) La divulgation des échanges entre évaluateurs et auteurs change profondément leur nature - de correspondances internes, ces textes se muent graduellement en publication scientifique secondaire, parfois dotés d'un identifiant DOI spécifique - ;
- 3) La conversion numérique des revues élargit le champ des possibles en matière de modèles éditoriaux ; il devient notamment envisageable de rendre public le processus d'écriture des articles scientifiques, sur le modèle de plateformes collaboratives comme Wikipédia.

La recherche sur le *peer review* a progressivement intégré cette diversification des pratiques. Les comparaisons systématiques entre des formes canoniques d'évaluation ouverte et d'évaluation en aveugle ont laissé place à des typologies élaborées. En 2017, une étude majeure de Tony Ross-Hellauer identifie 22 modèles différents d'évaluation ouverte dans la littérature scientifique, correspondant aux combinaisons attestées de 7 traits fondamentaux:

EPRIST

identités publiques, rapports ouverts, participation ouverte⁸, interaction ouverte, version préalable de l'article (*pre-print*) ouverte, versions finales commentées ouverte et plateforme ouverte (Ross-Hellauer, 2017, p. 9).



Données : Ross-Hellauer, 2017

Les 22 modèles identifiés par Ross-Hellauer, par combinaison de 7 « traits ».

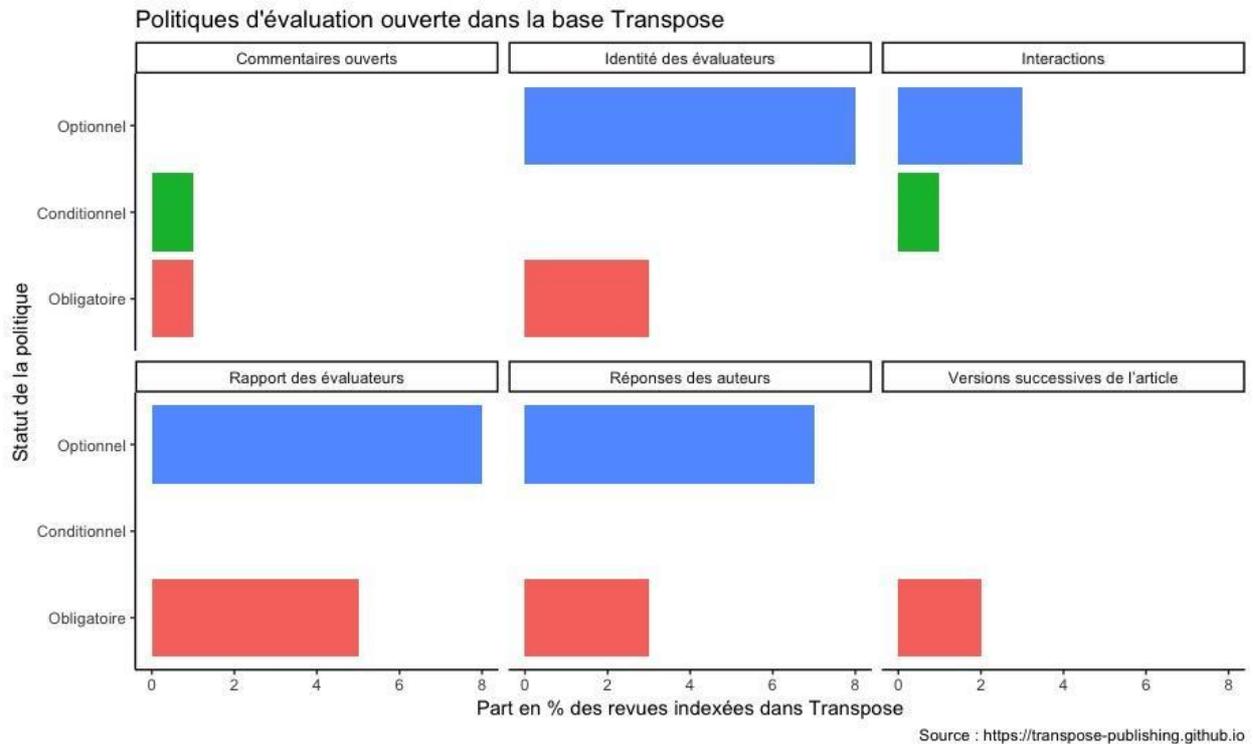
L'adoption de ces modèles est en partie conditionnée par les cultures disciplinaires préexistantes. Les communautés en sciences humaines et sociales semblent ainsi valoriser davantage les commentaires externes et les processus de publication ouverts, quand leurs collègues en sciences techniques et médicales semblent davantage enclins à divulguer leur identité.

L'approche initiale de Ross-Hellauer reposait sur une méta-analyse de la littérature scientifique sur l'évaluation ouverte. Depuis l'année dernière, cette recherche a été étendue aux pratiques effectivement constatées dans les revues. La base TRANSPOSE⁹ recense ainsi par crowdsourcing les politiques éditoriales de 3202 revues. Pour qualifier l'évaluation, l'architecture de la base reprend la description par « traits » introduite par Ross-Hellauer. Dans ce corpus, 468 revues utilisent au moins une forme d'évaluation ouverte, soit 14,5% des revues actuellement indexées. Nous avons extrait et nettoyé ces métadonnées afin d'évaluer le degré d'adoption de chaque forme, en proportion de l'ensemble des revues¹⁰.

⁸ Généralement sous la forme de commentaires externes ouverts après la publication de la recherche scientifique.

⁹ <https://transpose-publishing.github.io>

¹⁰ Les données générées par crowdsourcing ne sont pas complètement normalisées et admettent parfois des propriétés complémentaires que nous avons retirées.



Par rapport à la proposition initiale de Ross-Hellauer, les données sont encore affinées puisqu'il est possible de différencier les politiques obligatoires, optionnelles ou « conditionnelles » (généralement réservées à une certaine catégorie d'articles). Ces distinctions rendent compte de la diffusion récente de l'évaluation ouverte comme procédure complémentaire et alternative plutôt que comme norme impérative.

Sans beaucoup de surprise, les pratiques les plus anciennes sont aussi les mieux installées. La republication des rapports des évaluateurs, de la réponse des auteurs ou de l'identité des évaluateurs sont relativement fréquentes (près de 10% des revues, en agrégeant également les politiques non contraignantes). Les initiatives plus expérimentales relevant d'une pratique scientifique « wikipédienne » sont également attestées, même si généralement à titre optionnel ou conditionnel.

L'évaluation ouverte nuit-elle à la participation ?

C'est probablement l'aspect le plus critiqué de l'évaluation ouverte : la procédure démotiverait les évaluateurs en retirant la protection de l'anonymat. L'effet négatif serait notamment significatif pour des scientifiques déjà confrontés à des discriminations professionnelles ou sexuelles. Si cet enjeu est discuté dès les années 1990, il est particulièrement sensible dans le contexte actuel. Les revues peinent en effet de plus en plus à recruter des évaluateurs, en partie en raison de l'amplification de la production scientifique. Il est aujourd'hui beaucoup plus facile de mesurer l'impact social et scientifique de l'évaluation ouverte : la multiplication des expérimentations a généré de nombreuses

données empiriques. Bravo et al. (2019) ont ainsi pu récemment accéder à 138 117 textes d'évaluation de 21 647 manuscrits scientifiques dans les cinq revues participant au programme pilotes d'évaluation ouvert d'Elsevier¹¹ et dans cinq revues en évaluation fermée à titre de comparaison. L'analyse automatisée du corpus par plusieurs techniques de *text & data mining* a permis de créer des indices de « polarité » des évaluations et de reconstituer le genre et le statut professionnel de l'évaluateur.

Après avoir contrôlé les résultats avec d'autres variables communes aux deux corpus, les auteurs n'identifient pas d'impact significatif global de l'évaluation ouverte sur la participation. De manière plus inattendue, la pratique semble être davantage favorisée par les jeunes chercheurs : « Les chercheurs seniors avec un statut de professeur ont moins consenti à évaluer durant l'expérience pilote, alors que les jeunes chercheurs, avec ou sans statut de docteur, sont devenus plus volontaires (Bravo et al. 2019, 3). » L'attribution d'un DOI et la possibilité de valoriser l'évaluation dans son curriculum académique pourrait ainsi avoir eu un impact positif sur la participation.

En accord avec l'approche par « traits » de Tony Ross-Hellauer, l'étude n'a pas seulement testé un modèle uniforme d'évaluation ouverte mais différentes modalités. Dans l'expérience pilote d'Elsevier, les évaluateurs avaient la possibilité de divulguer ou non leur identité. Cette option n'a été que très peu utilisée (dans 8% des cas) : « nos résultats suggèrent que le voile de l'anonymat est également fondamental pour l'évaluation ouverte ». Ces données *en situation* amplifient une tendance déjà perceptible dans des sondages plus théoriques¹².

Parmi l'éventail des modèles possibles, l'évaluation ouverte anonyme pourrait ainsi s'imposer comme la procédure privilégiée par les éditeurs traditionnels comme Elsevier au cours des années à venir : elle maintient une continuité avec les formes de *peer review* classique.

Cette formule de compromis se heurte toutefois à des évolutions techniques discordantes. Les plateformes d'édition scientifique ont récemment développé des outils d'identification automatisée d'évaluateurs, afin de repérer plus facilement des experts potentiels spécialisés dans tel ou tel sujet. Il pourrait être ainsi techniquement envisageable de rapprocher automatiquement les textes d'évaluation avec des articles de recherche signés, ce qui remettrait en cause sur la durée la préservation de l'anonymat.

L'évaluation ouverte anonyme pénalise par ailleurs les initiatives plus expérimentales. C'était l'une des conclusions de l'étude qualitative réalisée par Julien Bordier à partir d'une expérimentation d'annotation ouverte menée sur la revue *Vertigo*. Dans ce cadre « l'anonymat du commentaire ouvert ne présente pas d'intérêt particulier. Le processus de commentaire autour d'une pré-publication est pertinent dans la mesure où les uns et les

¹¹ Ces cinq revues sont : *Agricultural and Forest Meteorology*, *Annals of Medicine and Surgery*, *Engineering Fracture Mechanics*, *Journal of Hydrology: Regional Studies* et *International Journal of Surgery*

¹² Une enquête menée en 2013 sur 4000 chercheurs montrait déjà un soutien limité à la divulgation de l'identité des évaluateurs : la proposition n'était approuvée que par 20% des sondés (Mulligan, Hall and Raphael 2013)

autres peuvent identifier leurs domaines d'expertises respectifs (Bordier 2016, 50). » Un rapport récent réalisé par *Publons*¹³ témoigne ainsi d'un « saut » générationnel : les nouvelles formes d'évaluations, même sans anonymat, sont nettement plus soutenues par de jeunes chercheurs et peuvent constituer une motivation en soi.

Dans un contexte aussi changeant, les études pilotes ne peuvent manifestement apporter des résultats pertinents que dans une perspective de court terme, de l'ordre de quelques années. L'évaluation ouverte anonyme pourrait être prochainement intégrée dans les dispositifs de publication de la plupart des grands éditeurs traditionnels.

Vers une évaluation sans revue ?

La conversation scientifique théorisée par Stevan Harnad dès 1978 s'est déjà généralisée, mais en s'exportant hors des revues. Longtemps après la fin du processus de publication, les articles de recherche continuent d'être discutés, commentés et critiqués sur les réseaux sociaux généralistes ou spécialisés, sur des plateformes de blogging comme Hypotheses ou sur des forums participatifs d'évaluation collaborative comme *PubPeer*.

Une grande enquête menée en 2015 et 2016 sur les usages numériques de 20 000 chercheurs de 142 pays, met en évidence une transformation rapide des interactions académiques (Kramer and Bosman 2016). Près d'un tiers des sondés (6 395) publient ou ont publié un blog scientifique ou généraliste, 5601 sondés ont un compte sur Twitter et 3112 sont actifs sur un site de peer-reviewing indépendant (*Publons*, *PubPeer*, *PeerCommunity*, *Review Commons*, *Rubriq...*)¹⁴. Ces tendances se sont probablement intensifiées au cours des cinq dernières années.

L'avènement des plateformes de peer-reviewing est probablement le phénomène qui a suscité le plus d'attention. Ces acteurs court-circuitent frontalement le processus de publication des revues et en révèlent les déficiences. Plusieurs cas de fraude scientifique majeurs ont ainsi d'abord émergé sur *PubPeer* et ses homologues, à l'image de l'affaire Olivier Voinnet¹⁵.

Au-delà de la critique des procédures existantes, de nouveaux projets s'inscrivent explicitement dans une nouvelle approche de la publication scientifique. Créé en 2017, *PeerCommunity* propose un service bénévole d'évaluation des *preprint* déposés sur les archives ouvertes. Au terme du processus, les articles ne sont pas publiés dans une revue mais *recommandés* par au moins deux évaluateurs et diffusés, ainsi que les rapports d'évaluation,

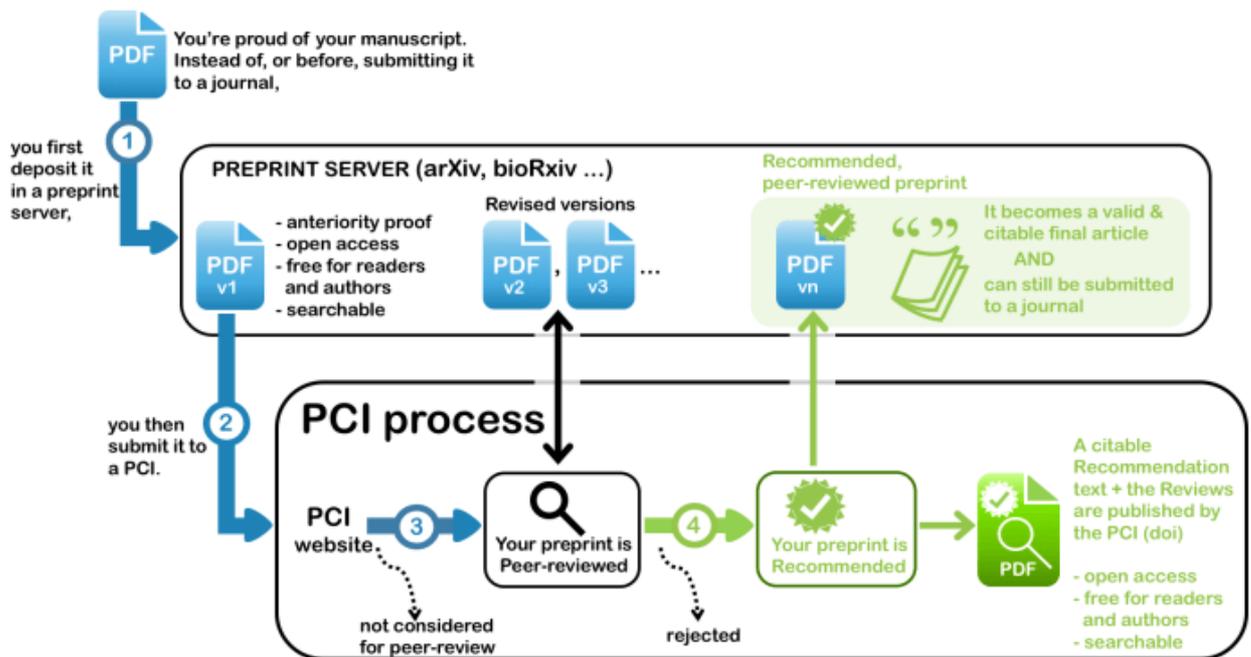
¹³ <https://publons.com/static/Publons-Global-State-Of-Peer-Review-2018.pdf>

¹⁴ Ces statistiques ne sont pas mentionnées dans l'article mais ont été recalculées à partir des données mises à disposition.

¹⁵ Sur le déroulé de l'affaire, voir l'excellent enquête de David Larousserie en six volets publiées dans le Monde en 2019 : https://www.lemonde.fr/sciences/article/2019/01/15/l-affaire-olivier-voinnet-une-loupe-posee-sur-l-integrite-scientifique_5409373_1650684.html

EPRIST

sur l'une des plateformes de *PeerCommunity*¹⁶. Le projet est en effet pluriel : il se revendique de l'appel de Jussieu pour la bibliodiversité et s'est décliné en plusieurs communautés indépendantes et autorégulées (*Peer Community in Ecology*, *Peer Community in Genomics*, *Peer Community in Circuit Neuroscience...*).



Le processus de publication de PeerCommunity.

Les archives ouvertes pourraient également s'imposer d'ici quelques années comme des acteurs-clés du processus d'évaluation. S'émancipant de leur fonction classique de dépôt, ces infrastructures commencent à republier les rapports d'évaluation ouverts des revues classiques ou des plateformes autonomes. Depuis octobre 2019, *bioRxiv* a officialisé un partenariat inédit de ce type avec deux éditeurs et deux plateformes, *Transparent Review in Preprint*. L'un des partenaires est une nouvelle structure créée deux mois plus tard, *Review Commons* : le site agrège les évaluations ouvertes de 17 revues en biologie.

Dans des configurations de ce type, les revues ne sont qu'un intermédiaire parmi d'autres, plutôt que l'épicentre de l'édition scientifique. L'intérêt récent des revues scientifiques pour l'évaluation ouverte peut être considéré comme une réponse à cette concurrence émergente, en rapatriant dans le processus éditorial interne des nouvelles pratiques qui ont fait leurs preuves.

La mise en œuvre de l'évaluation ouverte amène à réinterroger de nombreux usages et attendus préexistants, pas forcément explicites¹⁷. En ce sens elle incarne peut-être une révolution éditoriale plus profonde que le passage au libre accès en remettant en question la notion même de revue scientifique et sa pertinence dans le nouvel écosystème de publication en ligne.

¹⁶ <https://peercommunityin.org/how-does-it-work-2/>

¹⁷ Ce retour critique est très perceptible dans les recommandations émergentes pour la création de systèmes d'évaluation ouvertes (Ross-Hellauer and Görögh 2019).

Bibliographie

Bordier, Julien. 2016. "Open Peer Review: From an Experiment to a Model." <https://hal.archives-ouvertes.fr/hal-01283582>.

Bravo, Giangiacomo, Francisco Grimaldo, Emilia López-Iñesta, Bahar Mehmani, and Flaminio Squazzoni. 2019. "The Effect of Publishing Peer Review Reports on Referee Behavior in Five Scholarly Journals." *Nature Communications* 10 (1): 1–8. <https://doi.org/10.1038/s41467-018-08250-2>.

Kramer, Bianca, and Jeroen Bosman. 2016. "Innovations in Scholarly Communication - Global Survey on Research Tool Usage." *F1000Research* 5 (April): 692. <https://doi.org/10.12688/f1000research.8414.1>.

Mulligan, Adrian, Louise Hall, and Ellen Raphael. 2013. "Peer Review in a Changing World: An International Study Measuring the Attitudes of Researchers." *Journal of the American Society for Information Science and Technology* 64 (1): 132–61. <https://doi.org/10.1002/asi.22798>.

Pontille, David, and Didier Torny. 2015. "From Manuscript Evaluation to Article Valuation: The Changing Technologies of Journal Peer Review." *Human Studies* 38 (1): 57–79. <https://doi.org/10.1007/s10746-014-9335-z>.

Ross-Hellauer, Tony, and Edit Görögh. 2019. "Guidelines for Open Peer Review Implementation." *Research Integrity and Peer Review* 4 (1): 4. <https://doi.org/10.1186/s41073-019-0063-9>.

Smith, Richard. 1999. "Opening up BMJ Peer Review." *BMJ : British Medical Journal* 318 (7175): 4–5.