



Observer l'ouverture des publications scientifiques

Journée d'étude
Paris, Hcéres – 3 octobre 2022

SYNTHESE

Octobre 2022

SOMMAIRE

INTRODUCTION	5
SESSION 1 : OBSERVER L'OUVERTURE DES PUBLICATIONS SCIENTIFIQUES : DONNEES ET METHODES	5
Mesurer l'évolution de la science ouverte en France à partir de données fiables, ouvertes et maîtrisées sans recourir aux bases bibliométriques propriétaires	5
Le Baromètre lorrain de la science ouverte	7
Mesurer l'ouverture des publications scientifiques : apports et limites de différentes sources de données	8
BiblioLabs et BiblioHAL : rassembler les publications de son institution pour mieux les ouvrir	10
SESSION 2 : TABLEAU DE BORD DES DEPENSES D'APC : METHODES ET INDICATEURS	12
Mesurer les paiements d'APC sur le site universitaire lorrain	12
Pourquoi et comment suivre ses dépenses d'APC ?	13
Payer plus cher augmenterait-il l'impact sociétal et académique des publications ?	14
CLOTURE	15

INTRODUCTION

Frédérique Sachwald, *Directrice de l'OST - Hcéres*

Cette journée aborde la mesure de l'ouverture des publications en deux sessions. La première est consacrée à la mesure du degré d'ouverture des publications et la seconde à la mesure de l'évolution des coûts associés à travers les frais de publication d'article facturés par certaines revues (APC). Chacune de ces deux sessions présente des mesures à l'échelle nationale d'une part et au niveau d'institutions particulières d'autre part. Marin Dacos conclue cette journée en apportant une perspective d'ensemble sur l'évolution de l'ouverture des publications en France et sur les actions de politique publique.

L'ensemble des présentations power point des contributeurs est mise en ligne avec cette synthèse et permet de disposer des données très riches qui ont été présentées lors de la journée.

SESSION 1 : OBSERVER L'OUVERTURE DES PUBLICATIONS SCIENTIFIQUES : DONNEES ET METHODES

Présidente : Chérifa Boukacem-Zeghmouri, Université de Lyon 1

Mesurer l'évolution de la science ouverte en France à partir de données fiables, ouvertes et maîtrisées sans recourir aux bases bibliométriques propriétaires

Eric Jeangirard, *MESR-département des outils d'aide à la décision*

 [Présentation en pdf](#)

Le Baromètre de la science ouverte (BSO) est un outil de suivi de la politique nationale de science ouverte. Son originalité réside dans son approche ouverte, sans recours à des sources propriétaires, incompatibles avec le partage de données, la création de services ouverts et la valorisation de la diversité de la production scientifique française.

Les métadonnées nécessaires à la construction du BSO mais non disponibles directement dans les sources sont estimées par des techniques d'apprentissage automatique, en particulier pour les affiliations et les champs disciplinaires. En proposant du code, des données et des services ouverts, le BSO a contribué à la mise en place de dizaines de déclinaisons locales (établissements, laboratoires).

En 2021, une refonte complète du site (<https://barometredelascienceouverte.esr.gouv.fr/>) a été réalisée pour toucher un public plus large et fournir des indicateurs plus détaillés. D'abord centré sur la problématique de l'accès ouvert aux publications, le BSO étend son périmètre d'analyse en fonction de l'évolution de la politique publique, notamment en santé avec le suivi de l'ouverture des essais cliniques, et bientôt avec les codes et données de la recherche.

Le BSO a permis de vérifier que le taux de publications en accès ouvert a augmenté de plus de 20 points en 4 ans en France, depuis 2018 (publications 2017), avec une nette accélération la dernière année (2021).

La méthodologie utilisée pour construire le BSO est décrite dans une publication accessible sur HAL¹. Elle s'appuie sur un croisement de sources ouvertes :

- Des bases, telles que Pubmed, Unpaywall, Crossref, DOAJ, OpenAPC, liste officielle de Doi2 et métadonnées associées ;
- L'exploration systématique des informations sur les publications à partir d'Internet.

Pour calculer le taux d'accès ouvert des publications françaises, ne sont prises en compte que les publications répondant simultanément aux deux conditions suivantes :

- L'un des auteurs a une affiliation Française ;
- Elles présentent un DOI Crossref.

Quand l'affiliation n'est pas mentionnée dans Crossref (ce qui est fréquent) et ne peut pas être retrouvée dans une base comme PubMed, toutes les informations disponibles sur Internet sont mobilisées. Il s'agit en effet de constituer la base la plus complète possible de publications françaises avec, pour chacune, son statut ouvert ou non (Unpaywall). Pour définir ce dernier, l'un ou l'autre des critères suivants est pris en compte :

- La publication est disponible sur archive ouverte (type HAL) ;
- Elle est accessible en lecture sur le site de l'éditeur.

Une étude récente de L. Chaignon et D. Egret montre que les sources propriétaires comme WoS ou Scopus offrent une couverture moindre que le BSO mais obtiennent un taux d'accès ouvert des publications similaire à celui du BSO³.

Les sources utilisées par le BSO ne permettent pas d'identifier directement les disciplines/thématiques des publications. Une nomenclature en 10 postes est construite à partir des métadonnées des publications grâce à un algorithme d'apprentissage développé sur les bases Pascal et Francis.

Le taux d'accès ouvert est sensible à différents paramètres.

- Le taux d'AO dépend de la source (d'après l'étude de L. Chaignon et D. Egret, 57 % avec le périmètre BSO, 60 % avec les publications du WoS et 56 % avec Scopus) ;
- Le taux d'AO dépend de la date d'observation, certaines des publications pouvant voir leur accès ouvrir postérieurement à la date de publication, par exemple après une période d'embargo ou de barrière mobile ;
- Le taux d'AO dépend de la discipline, avec les plus forts taux pour la biologie et les plus faibles pour les SHS (dont les publications comportent rarement un DOI). Une discipline initialement faiblement ouverte comme la Chimie a fortement progressé au cours des dernières années.

Le BSO a étendu son analyse dans le domaine de la santé, avec un volet spécifique portant notamment sur les essais cliniques. Les prochaines étapes concernent :

- L'ouverture des données ayant contribué aux recherches, celle des logiciels utilisés ;
- L'élargissement des indicateurs en prenant en compte les thèses de doctorat ;
- La poursuite de la mise en place de BSO locaux, en établissement, avec la prise en compte des publications/notices déposées sur l'archive ouverte HAL ;
- De nouveaux indicateurs sur les financements.

Discussion

Lorsque l'affiliation institutionnelle des publications déposées sur HAL n'est pas précisée, une identification du pays est néanmoins menée en s'appuyant sur les éléments disponibles par ailleurs. Le BSO n'a cependant pas pour but d'analyser les collaborations internationales et les copublications.

¹ Laetitia Bracco, Anne L'Hôte, Eric Jeangirard, Didier Torny : A new framework for the French Open Science Monitor (BSO), 2022. <https://hal.archives-ouvertes.fr/hal-03651518>

² Digital object identifier

³ Lauranne Chaignon, Daniel Egret : "Identify scientific publications country-wide and measure their open access: the case of the French Open Science Barometer (BSO)", 2022. https://hal.archives-ouvertes.fr/search/index/q/*/authIdHal_s/lauranne-chaignon

La prolongation du travail du BSO au niveau local permet en particulier de contrôler et de préciser l'affiliation de façon beaucoup plus fine ; ce n'est qu'ainsi qu'il est possible de prendre en compte les évolutions du périmètre institutionnel français ces dernières années.

Le BSO ne conserve que les types principaux de publications : articles, communications à des congrès, chapitres de livre, monographies.

Le BSO exclut-il les éditeurs « prédateurs » ?

Il le fait pour des éditeurs ayant fait l'objet de poursuites aux Etats-Unis et retiré de Crossref. Le BSO n'a toutefois pas pour objectif de mesurer la qualité des publications scientifiques, ni de vérifier qu'elles ont bien fait l'objet d'une évaluation par les pairs.

Y a-t-il d'autres initiatives telles que le BSO à l'étranger ?

Aucune étude exhaustive n'a été menée à ce sujet mais oui, la Suisse, par exemple, fait apparemment un travail du même ordre, à partir de Unpaywall. Par ailleurs, le projet OpenAlex de OurResearch utilise aussi le scraping du web et des outils automatiques d'alignement pour alimenter une base de connaissance scientifique entièrement ouverte.

Le Baromètre lorrain de la science ouverte

Laetitia Bracco, *Mission Appui Recherche - Direction de la Documentation de l'Université de Lorraine*

 [Présentation en pdf](#)

En 2019, l'Université de Lorraine qui est résolument engagée dans une politique de science ouverte entreprend de tirer parti des ressources disponibles pour mettre en place un baromètre de la science ouverte (BSO) local. Les jeux de données du Baromètre de la science ouverte (BSO) comme les scripts en Python associés sont librement accessibles et permettent une déclinaison locale de cet outil de mesure.

La construction du BSO lorrain s'est faite en deux temps. Dans une première étape, la production de l'Université a été identifiée en croisant cinq sources au sein desquelles les affiliations des auteurs des publications sont relativement fiables : le Web of Science, PubMed, HAL-Université de Lorraine, Lens.org, et le suivi de paiements des APC. Les données sont ensuite nettoyées de manière automatisée afin d'obtenir une liste de DOI : 21 500 publications ont été ainsi identifiées pour la période 2016-2020. Dans une deuxième étape, des graphiques s'appuyant sur les données récoltées ont été générés en Python. Les DOI ont été enrichis en leur ajoutant le statut d'accès ouvert, grâce à un script issu d'Unpaywall et adapté par le ministère, ainsi que la discipline dont la publication relève. En vue d'une réutilisation simple par d'autres établissements, l'Université de Lorraine a utilisé des outils libres, Gitlab pour le versioning et Jupyter Notebook pour les commentaires du code. Par rapport au baromètre national, cette déclinaison locale, mise en ligne en juin 2020, présente en plus des indicateurs intéressant la politique de l'établissement : sur le paiement d'APC, sur l'accompagnement des chercheurs aux données de la recherche, sur la formation des doctorants à la science ouverte et sur la formation à HAL.

Un accompagnement et des formations par webinaires ont été proposées à une quinzaine d'universités, organismes de recherche et écoles souhaitant mettre en place leur propre baromètre. Une nouvelle version du baromètre, considérablement simplifiée, existe désormais. Au sein de l'Université de Lorraine par ailleurs, un accompagnement par communauté scientifique est réalisé.

Des résultats encourageants témoignent d'une prise de conscience des communautés scientifiques. Pour les publications toutes disciplines en 2020, le taux d'accès ouvert de l'Université de Lorraine est de 66 %, et particulièrement en chimie l'évolution est remarquable. On remarque, par ailleurs, une concentration des éditeurs, notamment d'Elsevier et de Springer ; le dépôt dans HAL n'est pas suffisant. Enfin, un travail est à réaliser sur les licences, car 41 % des publications n'en ont pas. Il reste à mesurer l'ouverture des codes logiciels et des données de la recherche. C'est l'objectif du projet BSO 3 (ministère, INRIA, Université de Lorraine). Une liste de diffusion existe, le club utilisateurs du baromètre.

Discussion

Quel est le niveau de fiabilité du BSO et de ses versions locales ?

Au niveau national, la détection automatique des affiliations fonctionne bien, car il s'agit de repérer un pays, mais au niveau local, la détection automatique d'une ville, voire d'un laboratoire, n'est pas encore prête. C'est pourquoi il est nécessaire de laisser à chaque établissement le soin de réaliser son propre corpus de publications.

Quel est le succès du baromètre au niveau des organismes de grande taille ?

Une page sur le site du BSO montre les établissements qui l'ont mis en place, par exemple, le CEA, l'INSERM. Il est possible de réaliser un baromètre à une échelle plus petite au niveau d'une unité de recherche.

Que faire quand l'établissement n'est pas mentionné ou que les publications sont mal signées ?

Il n'y a pas d'outil de repérage des publications en dehors des bases de données et du dépôt dans HAL. L'université de Lorraine rend obligatoire ce dépôt. Mais on n'a pas d'estimation des pertes. La meilleure solution pour que les chartes de signature soient respectées reste l'accompagnement.

Quel temps et quel coût de développement d'un BSO local ?

Le BSO lorrain a pris huit mois, temps d'apprentissage de Python compris. Aujourd'hui, il faut une journée, le temps de constituer sa liste de DOI et de l'envoyer au MESR.

Est-il possible d'utiliser le code de la première version du BSO lorrain ?

Le code est toujours à disposition et maintenu.

Mesurer l'ouverture des publications scientifiques : apports et limites de différentes sources de données

Marianne Lanoë et Abdelghani Maddi, OST – Hcéres

 [Présentation en pdf](#)

Différents indicateurs ont été proposés pour mesurer le taux d'accès ouvert des publications. Par ailleurs, les choix des bases de données et des périmètres de publications considérés influencent les résultats obtenus.

Une vue d'ensemble des publications en AO

Le taux de publication en AO a triplé au cours des 20 dernières années pour atteindre 47 % dans le monde. Autrement dit, presque une publication sur deux a au moins une version en AO. Cette augmentation est due à la forte augmentation du type « Gold », qui regroupe les deux modèles économiques : « auteur-payeur » et « Diamant ». En revanche, le type « Green » est resté stable entre 2000 et 2020.

L'analyse du taux de publication en AO par discipline (panels ERC) montre une très grande disparité en pour la publication en AO comme pour la voie choisie (verte, dorée, ...). Le taux d'AO varie de 23 % (SH5 : Cultures et production culturelle) à 74 % (LS3 : Biologie cellulaire, du développement et régénérative). En France, le taux de publication en AO est passé de 19 % à près de 70 % entre 2000 et 2020. Par ailleurs, il convient de noter le taux a enregistré un bond d'environ 10 % entre 2018 et 2019, à la suite du lancement du plan national de la science ouverte. Contrairement à la tendance

observée pour le monde, marqué par une forte augmentation de la part de la voie dorée, l'AO en France est tiré par la voie verte qui correspond aux publications ayant une version en libre accès dans une archive ouverte (institutionnelle, thématique ou autre). La voie dorée a également augmenté mais dans une moindre mesure.

Les disparités disciplinaires en termes de taux d'AO sont similaires en France et dans le monde en général. Les types d'AO varient aussi par discipline. Dans certaines disciplines, comme les mathématiques ou la physique, la voie verte est majoritairement utilisée. Ces disciplines ont développé au cours des 20 dernières années une forte culture de publication en AO. Leurs publications sont quasi-systématiquement déposées dans ArXiv (créé en 1991) et dans d'autres archives ouvertes.

La mesure du taux d'AO développée à l'OST

L'usage de taux bruts d'AO peut présenter des limites notamment quand il s'agit de comparer des acteurs (pays, institutions, etc.) dont les structures disciplinaires sont différentes, pour plusieurs raisons.

- La littérature en scientométrie a montré que chaque communauté disciplinaire a développé des pratiques de création et de diffusion des connaissances propres qui intègrent plus ou moins la problématique de la science ouverte.
- L'évolution des nécessite un temps certain et un environnement propice (création de revues Diamant, modes d'évaluation, financements, etc.).
- Pour un acteur (ex. institution) dont la communauté est active en libre accès (ou inversement), il est préférable de se comparer avec des acteurs qui partagent les mêmes caractéristiques. Cela permet de valoriser les efforts en matière de publication en AO et bien mesurer la progression.

Soit par exemple un établissement X qui a une forte orientation en biologie, dont plus de la moitié des publications sont en AO. Et soit un établissement Y avec une forte orientation SHS dont le taux d'AO est 35 %. X a un taux d'AO bien supérieur à Y. Mais si les deux établissements sont comparés à leurs homologues dans le monde, le diagnostic s'inverse : en effet, le taux d'AO en biologie est de 80 %, contre 20 % en SHS.

L'analyse au niveau institutionnel et au niveau national montre que les deux indicateurs (taux d'AO brut et indice normalisé) sont corrélés. L'indicateur normalisé permet d'apporter plus de précision dans la mesure où il permet des comparaisons plus adaptées. Ainsi, il est possible que le taux brut d'AO progresse pour un acteur sur une période donnée, mais si cette progression est moins rapide que celle du monde, son indice normalisé enregistrera une baisse. Ce type de tendance a été observés pour plusieurs établissements de l'ESR français, mais aussi pour un certain nombre de pays, à l'instar du Brésil.

Résultats selon les périmètres de publications considérés et les bases de données

Au cours des vingt dernières années, plusieurs dizaines de bases bibliographiques et bibliométriques ont été créées, dont certaines en libre accès. Cela offre une possibilité d'indépendance vis-à-vis des bases de données propriétaires telles que WoS ou Scopus. Cependant, avant d'envisager l'usage d'une telle ou telle source, il convient d'étudier la qualité des métadonnées concernant notamment les affiliations des auteurs, les types de documents, les disciplines, les supports de publication, etc.

Une analyse de la qualité du champ « type de document » dans HAL par exemple suggère qu'il mérite d'être précisé selon l'objectif de l'étude entreprise – en particulier s'il s'agit d'analyser des publications scientifiques : sur un échantillon de 100 notices, 28 % classées dans la catégorie « article dans une revue » correspondent à des articles de presse ou de magazines professionnels. De même, sur un échantillon des notices classées dans la catégorie « communication dans un colloque », 62 % correspondent à des conférences sans actes. Un résultat similaire est obtenu à partir de la base mondiale OpenAlex. D'autres explorations sur divers champs de ces bases de données (affiliations des auteurs par exemple) ont permis de conclure, qu'à ce stade, elles ne répondent pas aux standards de qualité permettant de conduire des études bibliométriques fiables. Mais certaines bases

comme OpenAlex sont en amélioration constante et elles sont susceptibles de concurrencer les bases de données propriétaires dans les années à venir.

Une autre question importante à considérer est celle des périmètres. Dans le cas de mesure des publications en AO, la première étape est de définir la nature de l'objet analysé. Autrement dit, doit-on considérer l'AO comme un objectif en soi ? Auquel cas, le périmètre à considérer sera constitué de toute la production scientifique quel que soit sa nature (création de nouvelles connaissances ou diffusion / évaluées par les pairs ou pas / publications dans des revues prédatrices /...). Ou bien considérer l'AO comme un moyen de tirer profit de la diffusion de la recherche scientifique pour répondre aux grands défis sociétaux et favoriser la dissémination de la science dans la société. Dans ce cas, il faudrait se limiter aux publications scientifiques soumises à une évaluation par les pairs. Enfin, il y a la question de savoir si l'analyse se limite aux publications disposant d'un DOI ou s'il faut intégrer les publications sans DOI dans les calculs.

Les résultats montrent que les publications qui ne rentrent pas dans le périmètre des publications scientifiques sont en général moins ouvertes que les publications scientifiques. Par ailleurs, l'intégration des publications sans DOI fait baisser le taux des publications en AO dans le WoS, alors qu'il n'a pas un grand impact dans les bases Unpaywall ou OpenAlex.

Discussion

Sur la représentation des modèles économiques dominants, le choix de l'ordre de priorité entre les différents types n'est pas neutre (pour les publications qui ont plus d'un type AO).

L'OST adopte une approche institutionnelle : les voies doré et hybride sont prioritaires car liées à des financements institutionnels, suivi du bronze qui provient des éditeurs de revues, et enfin la voie verte qui est majoritairement constituée des dépôts des auteurs dans des archives ouvertes (les revues peuvent également déverser les publications dans des archives ouvertes).

Sur la nomenclature disciplinaire, il serait utile de pouvoir classer les publications dans une nomenclature de type ERC par exemple sans devoir passer par les revues ou les classifications proposées par les bases propriétaires.

BiblioLabs et BiblioHAL : rassembler les publications de son institution pour mieux les ouvrir

Luc Bellier, *Directeur adjoint, resp. Pôle Numérique Recherche et Science Ouverte*
Henri Bretel, *Chargé de bibliométrie (DiBISO - Direction des Bibliothèques Université Paris-Saclay)*

 [Présentation en pdf.](#)

Observer l'ouverture des publications est une étape dans le processus de l'ouverture de la science. Observer l'ouverture suppose, en amont, de pouvoir observer les publications, la production scientifique. Elle permet, en aval, d'agir pour l'ouverture de la science.

Bibliolabs, l'outil de suivi des publications de l'université Paris-Saclay, a d'abord été conçu pour observer les publications scientifiques de l'université. Une attention particulière avait dès le départ été portée à leur ouverture, mais l'outil a pris une nouvelle dimension avec la création de BiblioHAL : en permettant aux chercheurs de comparer leur production telle qu'elle est vue par l'institution (dans Bibliolabs) et telle qu'elle est repérable dans HAL, cet outil apporte d'abord à chaque chercheur un état des lieux de sa visibilité en ligne. Au-delà de cette observation, il permet à l'utilisateur de déposer facilement sur HAL les publications qui n'y sont pas déjà, et à Bibliolabs de s'enrichir des publications qui n'étaient pas repérées par les bases de données traditionnelles. Bibliolabs et BiblioHAL permettent ainsi de couvrir les trois étapes : observer la science, observer l'ouverture de la science et agir pour l'ouverture de la science.

Un historique de BiblioLabs rappelle l'objectif du projet : développer un outil de collecte des publications scientifiques de l'Université Paris-Saclay pouvant servir de référence aux différentes composantes. L'unité d'observation des publications scientifiques étant le laboratoire, l'outil dispose d'un référentiel complet des laboratoires de Paris-Saclay. Ce périmètre validé des structures de recherche de Paris-Saclay permet d'identifier la production scientifique de l'université. Des travaux d'alignement d'identifiants internes aux laboratoires, nationaux, publics et privés viennent enrichir les données observées. Les référentiels de structure exploités pour l'enrichissement sont ceux de HAL, RNSR, IdRef, Scopus et WoS. Une connexion est aussi faite avec Cordis pour établir le lien entre les projets européens et les publications des laboratoires en passant par OpenAIRE.

En particulier BiblioHAL, un service fourni grâce aux données de BiblioLabs, permet de faire des dépôts dans Hal en utilisant un référentiel d'autorités fondé sur ORCID. Et grâce à l'ID-Hal il est possible d'identifier des publications dans HAL pour enrichir BiblioLabs.

L'interface d'administration des publications de BiblioLabs permet d'évaluer l'ouverture des publications de chaque chercheur qui s'y connecte, donc à terme de brosser le portrait de l'ouverture des structures.

Discussion

L'identifiant ORCID est-il imposé aux chercheurs ?

Le repérage prioritaire est celui du laboratoire (dans Scopus puis dans WoS) car malgré la charte de signature de l'université, cette information n'est pas suffisante et un identifiant chercheur est nécessaire. ORCID est géré par les chercheurs eux-mêmes avec parfois plusieurs identifiants pour un même chercheur et peu d'information rattachée (les anciennes publications par exemple). Pour l'imposer, il faut que les chercheurs aient confiance dans l'outil.

Est-ce que l'existence de plusieurs ORCID associés à un même chercheur donne lieu à un problème de cycle de vie ?

Le référentiel devrait s'améliorer mais il faudra de toute façon une curation. La fusion de différents identifiants ORCID d'un même chercheur est impossible dans le système : le passage par l'interface ORCID est nécessaire pour effectuer ces fusions.

Y-a-t-il un accompagnement des chercheurs dans l'utilisation de l'outil ?

Cette fonction d'aide est assurée à l'échelle de Paris-Saclay par un réseau de « référents recherche » issus des services de documentation. Les chercheurs acceptent volontiers un accompagnement, mais des difficultés techniques persistent qui freinent leurs usages. Un processus d'ouverture de l'outil à toute la communauté, pas uniquement à Paris-Saclay, est en cours. Il existe pour accompagner cette ouverture un groupe d'utilisateurs Bibliolabs coordonnant les évolutions futures de l'outil, mais il est complet et n'accepte actuellement plus de nouveaux utilisateurs. Le ministère peut fournir une aide pour le partage des outils développés par les universités. C'est en particulier une des missions d'Isabelle Blanc, l'administratrice ministérielle des données du MESR.

Existe-t-il des réflexions sur des alignements des identifiants établissements ? Quel identifiant utiliser pour les structures ?

Il y a un travail en cours sur le sujet au niveau du MESR. Des réflexions sont à développer sur les liens entre différents identifiants internationaux : ceux de Scopus, du WoS, Ringgold ID, ainsi que les identifiants GRID et ROR. L'université Paris-Saclay a retenu ORCID car Ringgold est moins ouvert, que ce soit à la lecture ou pour la modification et la correction des données.

La DOI-fication est fortement encouragée par le MESR, même si son utilisation réduit le nombre de documents repérés.

SESSION 2 : TABLEAU DE BORD DES DEPENSES D'APC : METHODES ET INDICATEURS

Présidente : Laetitia Bracco, Université de Lorraine

Mesurer les paiements d'APC sur le site universitaire lorrain

Jean-François Lutz, *Responsable de la Mission Appui Recherche (Direction de la Documentation de l'Université de Lorraine)*

 [Présentation en pdf](#)

Jean-François Lutz a diffusé un texte complémentaire de la présentation qui est joint sur le site, à l'instar des présentations ppt.

« Payer pour être publié » est une pratique ancienne mais le marché des APC a explosé et est devenu un enjeu pour la puissance publique. L'université de Lorraine s'est investie sur l'analyse de ces coûts.

Deux méthodes sont mises en œuvre pour estimer les coûts. Par des requêtes dans le logiciel comptable sur le nom du fournisseur : il faut ensuite vérifier les différentes factures pour identifier celles qui correspondent à des APC. Et cela permet d'obtenir le prix réellement payé (qui comporte des remises). Par recherche sur le WoS, des publications en OA identifiées à l'aide de Unpaywall. L'affiliation de l'auteur de correspondance doit être contrôlée. On ne relève que les articles et revues (les frais d'actes de conférence sont souvent groupés donc inutilisables).

La méthode 1 ne voit pas les publications OA diamant. La méthode 2 informe sur les publications de tout le site lorrain (Université de Lorraine et autres institutions). Les deux méthodes produisent donc des informations différentes, complémentaires.

L'augmentation des coûts est inquiétante : le coût annuel pour l'UL est actuellement supérieur à 700 K€ et le prix moyen par article à plus de 2 000 € : par exemple, en 10 ans, le coût d'un article dans PloS One a augmenté de plus de 30 %, de nouvelles revues comme celles de l'éditeur MDPI tirent les coûts vers le haut. Les accords transformants entre une organisation ou un collectif d'organisations et un éditeur (ex avec EDP Science) réduisent ces coûts (mais incitent parfois à déclarer deux auteurs de correspondance dont un dans l'institution ayant un tel accord).

Quelle stratégie d'établissement ?

Un investissement important est fait sur la sensibilisation des chercheurs et des unités avec la publication d'un « position paper » de la direction de l'Université en mars 2022, recommandant d'éviter les revues avec APC. Une position similaire du CNRS conforte la démarche.

L'université n'a pas réservé une ligne de dépense spécifique pour les APC et est en cours de définition d'une doctrine sur les accords transformants. Les effets de cette stratégie ne sont pas encore observables et seront suivis.

Discussion

Pourquoi ne pas regarder d'autres types de documents ?

Une politique d'un éditeur présent est de favoriser les autres types de documents (short reports, letters) qui contribuent au facteur d'impact. Mais l'UL centre ses analyses sur les articles et revues qui sont les plus nombreux. L'augmentation des coûts totaux d'APC est à combiner avec l'augmentation du nombre de publications. On peut le faire avec les données de la méthode 2.

Comment inciter les chercheurs à réduire le coût de publication ?

Si votre université impose de publier en OA (exemple de l'université de Genève) et que vous ne pouvez pas payer, vous pouvez déposer dans HAL un preprint ou la version publiée parfois sans embargo.

Une information utile aux chercheurs est de dire ce qu'est un montant raisonnable d'APC, c'est-à-dire compatible avec le coût réel. L'Inserm par exemple, diffuse ce type d'information. Attention aussi aux prix d'appel de certaines revues. Par contre, certaines agences de financement ne favorisent pas la réduction de ces coûts en les rendant éligibles à un financement.

Le soutien du Plan S (phase 2) aux *transformative journals* qui est dans sa 2^{ème} année contribuera à terme à maîtriser ces coûts mais ne constitue pas une solution rapide.

Que voit-on dans l'étude de l'UL sur les SHS sachant que certains chercheurs payent eux-mêmes les frais de publication ?

Des publications souvent en anglais sont dans le WoS et sont ainsi identifiées par la méthode 2 montrant que des chercheurs en SHS s'engagent dans la démarche OA en particulier dans le type diamant.

Pourquoi et comment suivre ses dépenses d'APC ?

Françoise Rousseau-Hans, *Chef du service valorisation de l'information (CEA) et Responsable du département Services et Prospectives de Couperin*

 [Présentation en pdf.](#)

L'enquête APC coordonnée par le consortium Couperin.org depuis 2016 a pour objectif de fournir une vision exhaustive des frais d'APC. Cette enquête permet d'évaluer la dépense des chercheurs français sur l'ensemble des frais de publication en OA : elle mesure les frais d'APC mais aussi les frais annexes (pages supplémentaires, couverture...). De plus, l'enquête permet de dédoubler et consolider les données remontées par les établissements.

Toutes les données sont intégrées dans la plateforme OpenAPC dans une perspective commune de transparence des coûts et d'évaluation de la dépense en matière d'"édition scientifique", c'est-à-dire des coûts d'abonnement et des frais de publication.

Les données de l'enquête APC combinées avec des données bibliométriques sont utilisées pour construire des indicateurs afin de suivre les dépenses d'APC rapportées au nombre d'articles publiés en accès ouvert, le coût unitaire moyen ou médian par article, les dépenses par éditeur, les principales revues utilisées, ou encore l'évolution des APC par revue. Ces données permettent d'une part d'observer l'ouverture des publications dans le cadre du suivi des politiques scientifiques et d'autre part de disposer de données objectives pour les négociations avec les éditeurs.

Toutefois la consolidation et la manipulation des données remontées par les établissements est coûteuse. Il serait plus efficace de disposer d'une base pivot pour l'ensemble de la France qui pourrait être enrichie collectivement et faciliterait la consolidation des données.

La démarche est illustrée par le suivi réalisé à partir de ces indicateurs au CEA. La bibliométrie institutionnelle réalisée au CEA pour suivre le contrat d'objectifs ou préparer les évaluations, les indicateurs concernant la science ouverte reposent pour l'instant sur un périmètre « Web of Science » pour maintenir une cohérence avec les indicateurs institutionnels. Deux nouveaux indicateurs s'appuyant sur HAL sont en cours de test.

Ces indicateurs ont permis de montrer notamment que les dépenses d'APC ont considérablement augmenté en 2021. Au CEA, cette augmentation est pour l'instant compensée par la baisse des dépenses d'abonnement. Au sein des différents éditeurs, ce sont les dépenses hors abonnement de

MDPI qui ont le plus fortement augmenté puisqu'elles ont doublé en 2021. Les dépenses hors abonnement d'éditeurs plus traditionnels comme Springer Nature (Nature Communications, Scientific reports,) se maintiennent.

Payer plus cher augmenterait-il l'impact sociétal et académique des publications ?

David Sapinho, OST – Hcéres

 [Présentation en pdf](#)

Cette étude se propose de mettre en évidence le lien potentiel entre le montant des APC et l'impact des publications, mesuré d'une part par les citations (impact académique), et d'autre part, par le retentissement qu'elles peuvent avoir dans la presse ou sur les réseaux sociaux (impact sociétal).

L'hypothèse sous-jacente est que les éditeurs pourraient appliquer une politique de prix qui tient compte de ce type d'indicateurs et que dans ce contexte, les chercheurs puissent être influencés par l'idée que le paiement de ces frais leur garantirait d'être publiés dans de meilleures conditions, augmentant ainsi la visibilité de leur recherche.

Pour réaliser cette étude, nous avons utilisé les données mises à disposition dans le cadre de l'initiative "OpenAPC", développée par l'Université de Bielefeld, qui recense les frais de traitement par article de plus de 250 institutions à travers le monde, dont une grande majorité en Europe. Les données recueillies sont directement fournies par les institutions concernées.

Pour la France, les données sont alimentées par l'intermédiaire du consortium Couperin, qui accueille au total 264 membres au moment de l'étude (113 universités et établissements assimilés, 29 organismes de recherche, 88 écoles, 4 bibliothèques dotées de la personnalité morale et 30 autres organismes ayant une mission d'enseignement supérieur ou de recherche. »), parmi lesquelles 43 ont fourni des informations sur le paiement des APC. L'étude a été réalisée sur la période 2006-2019.

La base de données comprenait, au moment de l'étude (janvier 2020), près de 110 000 publications venant de 6 941 journaux différents

L'impact académique des publications a été mesuré par leurs citations, en utilisant un indicateur normalisé qui tient compte de la diversité disciplinaire des articles (MNCS, Waldmann, 2011). La mesure des citations est issue de la base Web of Science de Clarivate Analytics.

Enfin, pour mesurer l'impact sociétal, nous avons utilisé les informations disponibles sur le site <https://www.altmetric.com/>; ce site recense des informations sur les retombées potentielles d'une publication hors du champ académique telles que les articles de presse, ou encore le nombre de reprise sur les réseaux sociaux, et propose un score sensé mesurer le degré d'attention d'un contenu.

Le lien entre les différentes bases de données a été fait en s'appuyant sur les DOI communs.

L'échantillon d'étude se composait finalement de 83 752 publications pour l'analyse de l'impact académique, et de 65 232 publications pour l'analyse de l'impact sociétal.

Les principaux résultats de cette étude sont les suivants :

- Le montant moyen des APC a fortement augmenté, passant de 1 000 € en moyenne en 2006 à près de 2 300 € en 2019 ; ce montant est d'autant plus élevé que les revues ont été récemment indexées ;
- Que ce soit en volume ou en prix, la pratique des APC varie de façon importante selon les disciplines : la recherche médicale et la biologie fondamentale sont les domaines où ils sont le plus pratiqués, avec des montants particulièrement élevés. Ces montants sont aussi élevés en chimie et en sciences pour l'ingénieur, bien que peu importants en volume. A l'inverse, les disciplines où ces montants sont les plus faibles sont les mathématiques et les sciences humaines ;
- Le montant des APC est 50 % plus élevé dans les revues hybrides ($\approx 2\,500$ €) en comparaison aux revues 100 % en accès ouvert ($\approx 1\,600$ €) ;
- Les éditeurs appliquant les tarifs les plus élevés sont plutôt des petits éditeurs spécialisés dans le domaine biomédical ;
- Il existe une corrélation entre le montant des APC et l'impact des publications, mais avec une forte hétérogénéité, suggérant que ce lien pourrait être expliqué par des facteurs endogènes non identifiés par cette étude.

Discussion

- Dans certains cas, pour des thématiques particulières, le fait de pouvoir être publié dans un délai très court avec des chances d'avoir une visibilité très forte est un argument fort, qui peut avoir des conséquences sur les retombées d'une recherche en termes de service rendu ; l'exemple donné est celui d'une étude réalisée dans le domaine des cancers pédiatriques ;
- Il est montré que la qualité des articles n'est pas lié aux citations, et comme il est montré ici aussi que le montant des APC n'est pas nécessairement liée au nombre de citations, ne serait-il pas envisageable de montrer qu'il n'y a aucun lien entre APC et qualité des articles ; Cette réflexion soulève la question de la mesure de la qualité de la production scientifique, et de l'évaluation par les pairs.
- Ne serait-il pas possible d'essayer de mesurer l'impact des APC en termes d'avantage de citations de l'OA ?
- L'OST a travaillé plus récemment sur une étude décrivant l'avantage de citations lié à l'accès ouvert (OACA) ; en revanche, la question particulière du paiement des APC dans ce contexte n'a pas été étudiée.

Marin Dacos suggère d'analyser plutôt l'ODCA (*open data citation advantage*) qu'il faudrait chercher à étudier. La recherche commence à s'intéresser à cette question.

CLÔTURE

Marin Dacos, *Coordinateur national de la Science ouverte, MESR*

Cette journée et nos débats nourris ont montré qu'une communauté s'est construite en France autour de la mesure de l'ouverture des publications et des coûts associés. Au-delà même, différents outils ont été co-construits et sont complémentaires, notamment entre le niveau national et les outils locaux adaptés aux besoins des acteurs, universités et organismes. Nous disposons donc d'une boîte à outils qui se diversifie progressivement. Les premiers outils ont été remplacés par une génération plus efficace ou plus simple, à mesure que des contributeurs nouveaux se sont joints à la communauté.

En parallèle de ces efforts couronnés de succès pour mesurer l'évolution de l'ouverture des publications, je souhaite souligner l'importance de la bibliodiversité qui doit permettre d'embrasser plus largement la variété des productions des chercheurs.

L'idée d'une base pivot des publications des chercheurs français pourrait aboutir à un beau jardin à la française. Nous n'y sommes pas et il faut admettre que l'indexation de toutes les publications scientifiques françaises ne peut être qu'une entreprise progressive. Elle reste à ce stade dépendante de multiples sources, mais nous progressons et les convergences émergent. La journée a d'ailleurs montré que nous disposons actuellement de multiples clés d'entrée dans le système qui correspondent à la diversité des usages. L'objectif de l'indexation la plus complète possible au niveau d'une institution est différent de l'objectif de l'OST de produire un indice d'ouverture des publications normalisé pour comparer les disciplines et les pays.

Isabelle Blanc, administratrice ministérielle des données, algorithmes et codes sources, notre « chief data officer » national, n'a malheureusement pas pu se joindre à nous aujourd'hui. Rappelons le cadre juridique dans lequel elle inscrit son action. La Déclaration des Droits de l'Homme et du Citoyen de 1789 prévoyait que « *la société a le droit de demander compte à tout agent public de son administration* ». La loi du 17 juillet 1978 mettait en place pour le citoyen « *la liberté d'accès aux documents administratifs* », mais cela signifie qu'il fallait une démarche active de demande d'accès aux données, trop souvent laborieuse et pas certaines d'aboutir. La loi d'octobre 2016 dite « *Loi pour une République numérique* » change de paradigme, en passant d'une logique de la demande d'un accès à une logique de l'offre de données publiques, la loi visant clairement à ouvrir très largement

les données publiques en amont des demandes des usagers. Enfin, en 2022, la loi dite 3DS du 21 février 2022, relative à la différenciation, la décentralisation, la déconcentration et portant diverses mesures de simplification de l'action publique, prévoit l'échange de données entre administrations. Est ainsi créé le principe « *dites-le nous une fois* », mettant en place une obligation pour les administrations d'échanger des données entre elles.

Par ailleurs, les Conclusions du Conseil sur l'évaluation de la recherche et sur la science ouverte, adoptées en juin 2022 dans le cadre de la présidence française de l'Union Européenne, estiment que "*les données et les bases de données bibliographiques utilisées pour évaluer la recherche devraient, en principe, être librement accessibles, et les outils et les systèmes techniques devraient permettre d'assurer la transparence*"⁴.

Afin de simplifier la vie des chercheurs, nous nous appuyerons notamment sur les identifiants ORCID. La France aura désormais un représentant dans la gouvernance d'ORCID. ORCID va progressivement faciliter la connexion à des infrastructures de recherche, à HAL, à recherche.data.gouv.fr, aux revues pour la soumission d'un article... ORCID facilitera également le dépôt des candidatures à un appel à projet du portail national des appels à projet. Dès cette fin d'année, les informations renseignées dans le profil ORCID par les chercheurs seront récupérables dans leur profil afin de leur éviter une nouvelle saisie.

En conclusion, cette journée a montré que se développent des convergences fortes en France. Cette expertise collective est un bien précieux, cultivons-le.

⁴ Évaluation de la recherche et mise en œuvre de la science ouverte - Conclusions du Conseil adoptées le 10 juin 2022, https://www.ouvrirelascience.fr/wp-content/uploads/2022/06/ST_10126_2022_INIT_fr.pdf, page 7.



2 rue Albert Einstein
75013 Paris, France
T. 33 (0)1 55 55 60 10
hceres.fr

Haut Conseil de l'évaluation de la recherche et de l'enseignement supérieur