

Bases de données bibliométriques ouvertes vs bases commerciales : vers une nouvelle ère ?

 openscience.pasteur.fr/2024/02/13/bases-de-donnees-bibliometriques-ouvertes-vs-bases-commerciales-vers-une-nouvelle-ere/

CeRIS - Institut Pasteur

13 février 2024

Une étude demandée par la cOAlition S concernant l'impact du plan S sur la communication scientifique a été l'occasion de tester la qualité de deux types d'échantillons de données bibliométriques : l'un provenant des **éditeurs commerciaux** (Elsevier et Clarivate) et l'autre provenant de **sources libres d'accès** (OpenAlex, Crossref, Unpaywall...). En effet, l'un des pré-requis de l'étude commandée par la cOAlition S à l'équipe scidecode science consulting était que les résultats soient diffusés sous licence ouverte.

L'équipe chargée de l'étude devait donc préalablement vérifier que, s'ils s'appuyaient sur des données bibliométriques ouvertes, elles soient de suffisamment bonne qualité pour effectuer leurs analyses. L'équipe scidecode a donc sélectionné un set de publications provenant de la Fiocruz, une institution publiant beaucoup en langue non-anglaise et avec peu de DOI. Les résultats ont montré que, sur ce set de publications, les **données issues des bases bibliométriques ouvertes** étaient **plus complètes** et de **meilleure qualité** que celles des bases commerciales.

Des problématiques liées à la couverture des bases commerciales ont déjà été rapportées, notamment la faible représentativité des monographies et des productions francophones. Récemment, il a été également signalé que Scopus et Web of Science n'indexent toujours pas certaines revues « diamant » comme le Journal of Open Source Software (JOSS), malgré les demandes répétées des équipes du journal.

Ces éléments expliquent en partie le fait que certaines institutions cherchent à **se désengager des bases de données commerciales** (le coût des abonnements étant un autre facteur). Ainsi, le **CNRS** vient de résilier son abonnement à Scopus (Elsevier) pendant que **Sorbonne Université** se désengage du Web of Science (Clarivate) et que le **CWTS Leiden** sort son nouveau classement basé uniquement sur des sources ouvertes (CWTS Leiden Ranking Open Edition 2023).

Si les bases commerciales ne réagissent pas et n'améliorent pas la couverture et la qualité de leurs données, ce phénomène de désengagement risque de s'accélérer. D'autant plus que de nombreuses institutions sont désormais engagées dans la CoARA (Coalition on Advancing Research Assessment), laquelle promeut, entre autres, la reconnaissance de la diversité des contributions à la recherche (par leur type, leur langue...). Un objectif qui nécessite des bases de données avec une couverture très étendue.