

# Comment ouvrir et mesurer la science vont de pair ?

Ouvrir et mesurer la science vont de pair. S'il faut combattre les études décontextualisées et les indicateurs réductionnistes, il convient également d'encourager le libre accès à toutes les données de la science pour alimenter des études scientométriques pertinentes.

**La scientométrie (l'étude quantitative des sciences et des technologies) s'appuie sur des outils bibliométriques permettant la collecte, l'analyse et la mesure des productions scientifiques** (articles, livres ou brevets). Cette quantification de l'activité de recherche est parfois décriée lorsqu'elle conduit à la production de classements internationaux d'universités ou qu'elle est maladroitement utilisée pour l'évaluation des chercheurs. Certains fustigent l'évaluation technocratique de leurs activités et s'opposent aux obligations de dépôt de leurs travaux en archive ouverte, dans HAL notamment. En réalité, la bibliométrie n'a pas eu besoin de l'accès ouvert pour exister. En revanche, elle en a besoin pour progresser et produire des études fiables et pertinentes. Inversement, le mouvement du libre accès a besoin de mesures bibliométriques pour évaluer l'impact des incitations mises en œuvre et trouver des voies d'amélioration.

## OUVRIRE POUR MESURER MIEUX

Les études scientométriques, lorsqu'elles sont correctement menées et expliquées, constituent un moyen d'observer l'activité scientifique. Les publications et les métadonnées qui leur sont associées constituent le matériau de base de ces études, dont dépendent leur richesse et leur pertinence.

Pour constituer ses corpus, le bibliomètre a le choix d'interroger des dizaines de bases bibliographiques généralistes ou disciplinaires. Néanmoins, les bases commerciales de *Clarivate (Web of Science)* et d'*Elsevier (Scopus)* sont les plus couramment utilisées, concurrencées par *Dimensions<sup>1</sup> (Digital Science)* dont la plupart des fonctions sont utilisables gratuitement. Mais ces bases ont des biais de couverture : elles favorisent notamment les publications en anglais, privilégient les sciences et techniques au détriment des sciences sociales et des humanités ou encore préfèrent les articles aux monographies. C'est en croisant les bases mais surtout en puisant dans des archives ouvertes que ces biais peuvent être atténués. Que vaut en effet l'analyse de la recherche d'un établissement ou d'un pays si un pan complet de sa production n'est pas pris en considération ? Comment peut-on parler d'étude interdisciplinaire si les sciences sociales sont a priori

oubliées ? Pour jouer pleinement leur rôle, il faut néanmoins que les archives ouvertes disposent de métadonnées de qualité. Les opérations de dédoublement et d'alignement restent souvent une gageure pour le bibliomètre qui dépend de ces bases, de leur ouverture et aussi de leur pérennité. En effet, Microsoft a récemment annoncé la fermeture prochaine du *Microsoft Academic Graph*, une base bibliographique librement accessible et sur laquelle reposaient de nombreux outils, comme par exemple *The Lens<sup>2</sup>*. Heureusement, *OurResearch* (société à but non lucratif et acteur déjà bien connu de l'*Open Access* pour son outil *Unpaywall<sup>3</sup>*) est déjà prêt à développer un successeur, *OpenAlex*, et à ouvrir l'accès aux données et au code, ce qui permettrait de garantir la pérennité de l'outil.

## VERS UN ÉCOSYSTÈME OUVERT

La bibliométrie s'intéresse à la circulation des connaissances et étudie les réseaux de citations ; il s'agit techniquement de parcourir les bibliographies et d'attribuer les citations aux documents qui y sont cités. Mais ces données de citation sont parfois soumises à des licences d'utilisation selon les éditeurs propriétaires des documents. Pour libérer les liens de citations et permettre au plus grand nombre de les exploiter, l'initiative *I4OC<sup>4</sup> (Initiative for Open Citations)* a été lancée en 2017 et ce sont désormais plus de 2700 éditeurs qui ont rejoint le mouvement et versé leurs citations, laissant entrevoir une meilleure évaluation de la recherche avec cette base ouverte, structurée et réutilisable conformément aux principes du libre accès.

Pour des études bibliométriques à l'échelle d'un pays ou d'une ville, il est nécessaire d'exploiter la ligne de signature que les chercheurs associent à leur publication pour mentionner leur affiliation. *Scopus* et le *Web of Science* ont des systèmes avancés qui reconnaissent les organismes mentionnés, au prix d'efforts concédés en amont par les professionnels de l'IST pour fournir des informations facilitant ce repérage. Mais des initiatives pour mettre en œuvre des identifiants uniques et ouverts se développent (comme *ROR<sup>5</sup>* basé sur *GRID (Digital Science)*, faisant écho aux *RNSR* et *AuréHAL* français). Ces identifiants pérennes, à l'instar des *DOIs* pour les

[1] <https://app.dimensions.ai>

[2] <https://www.lens.org>

[3] <https://unpaywall.org>

[4] <https://i4oc.org>

[5] <https://ror.org>

publications ou des identifiants Orcid pour désambiguïser les noms de chercheurs, sont essentiels à la constitution d'un écosystème ouvert et répondent au premier critère des principes FAIR (*Findable, Accessible, Interoperable, Reusable*) qui exige l'attribution d'un identifiant aux métadonnées.

Enfin, au-delà du libre accès aux métadonnées, le libre accès au résumé et au texte intégral des publications ou brevets permet aussi d'alimenter des indicateurs bibliométriques innovants qui exploitent les données textuelles. L'École des Ponts a par exemple mis au point des requêtes<sup>6</sup> pour faire correspondre ses publications aux Objectifs de développement durable de l'Onu. Cela démontre tout l'intérêt de l'initiative I4OA<sup>7</sup> (*Initiative for Open Abstracts*) qui promeut l'ouverture et la mise à disposition des résumés d'articles et de chapitres d'ouvrage.

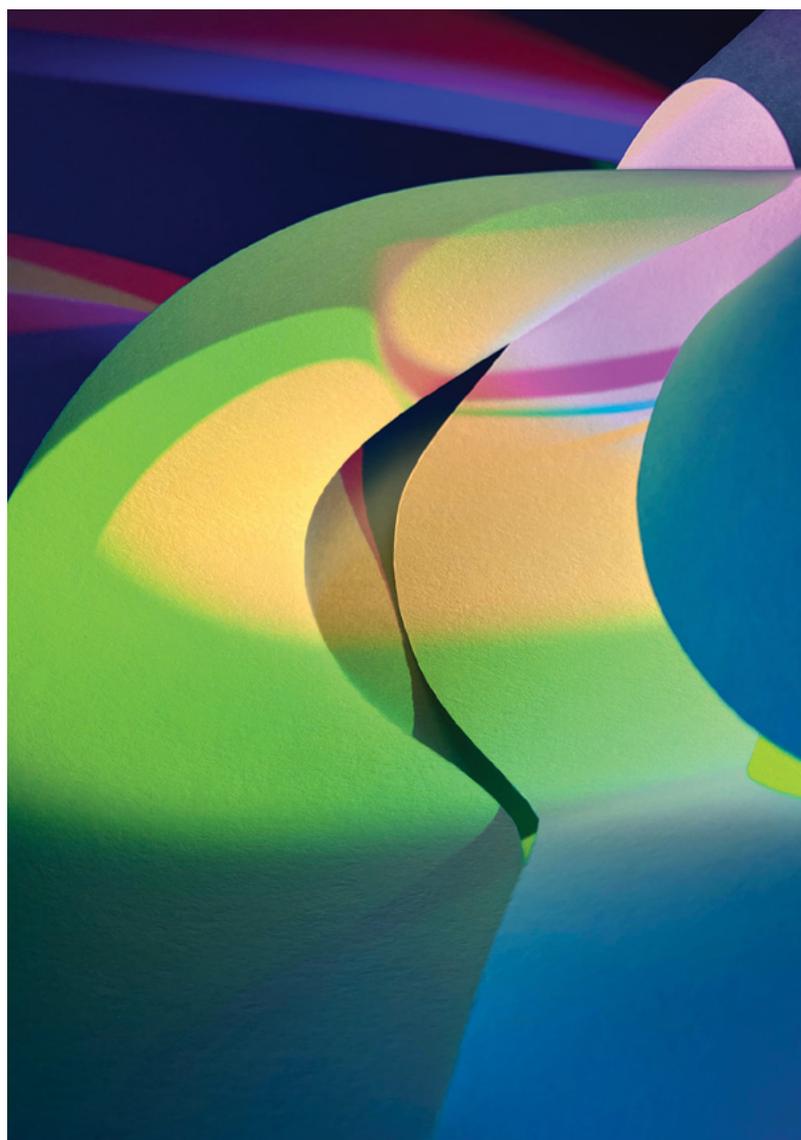
L'ouverture du texte intégral constitue aussi la formidable opportunité d'analyser automatiquement le contexte des citations (la phrase notamment) en offrant une alternative à un comptage brut avec l'attribution d'une polarité (positive, critique ou neutre envers le travail cité) à l'instar de ce que l'outil *Scite.ai* a développé à partir d'accords avec les éditeurs pour constituer sa base de textes. La mise à disposition de ces enrichissements pourrait même s'inscrire dans le Web sémantique comme l'offre l'ontologie CiTO<sup>8</sup>.

Ces projets vertueux sont rendus possibles par le libre accès aux productions scientifiques, aux métadonnées qui les décrivent et aux infrastructures qui les exposent. Vue sous un autre angle, la bibliométrie peut à son tour être un élément facilitateur du libre accès.

## MESURER POUR OUVRIR DAVANTAGE

Tout d'abord, il faut être lucide, de nombreux bibliomètres sont aussi des acteurs du libre accès et ils ne renient pas leurs engagements en faveur de l'ouverture de la science lorsqu'ils s'investissent dans la réalisation d'un rapport bibliométrique. Les bibliomètres sont donc attachés à la qualité des données des archives ouvertes et sont souvent à la manœuvre pour faire les vérifications et tampons dans HAL.

Par ailleurs, la bibliométrie permet de mesurer le taux de disponibilité d'un corpus et de cibler ainsi les disciplines ou laboratoires qui ont une marge de progression en vue des objectifs fixés, par les établissements ou par le récent deuxième Plan national pour la science ouverte. Ces baromètres de la science ouverte (qui sont d'ailleurs plutôt des thermomètres) sont plus justes lorsqu'ils sont déclinés à l'échelle d'une institution avec des professionnels locaux qui connaissent mieux le périmètre de leurs unités de recherche. À l'École des Ponts par exemple, c'est le corpus servant de base aux analyses bibliométriques qui est utilisé. Cela



© Unsplash - Susan Gold

permet de s'affranchir de la dépendance aux DOIs dont souffre le baromètre national et de prendre la température sur un corpus qui frôle l'exhaustivité. Ces mesures régulières informent de l'effet des incitations, permettent d'organiser des campagnes ciblées de collecte de *postprints*, et renseignent sur le type de libre accès (green ou gold) et la mise à disposition de jeux de données associés. Enfin elles ont démontré que les publications en libre accès ont plus d'impact en termes de citations, mais aussi de reprises dans les media ou Wikipedia.

### FRÉDÉRIQUE BORDIGNON

Responsable du Pôle IST à l'École des Ponts,  
chercheuse au LISIS (Laboratoire Interdisciplinaire  
Sciences, Innovations, Sociétés)  
frederique.bordignon@enpc.fr



École des Ponts  
ParisTech

[6] Bordignon, F. (2021). Dataset of search queries to map scientific publications to the UN Sustainable Development Goals. Data in Brief, 106731.

<https://doi.org/10.1016/j.dib.2021.106731>

[7] <https://i4oa.org>

[8] <https://sparontologies.github.io/cito/current/cito.html>