

## Le Baromètre français de la Science Ouverte 2023

Selon l'édition 2023 du Baromètre de la Science Ouverte (BSO), 65 % des 160 000 publications scientifiques françaises parues en 2022 sont en accès ouvert en décembre 2023, soit un taux stable par rapport à l'année précédente. Le niveau d'ouverture des publications varie de manière significative d'une discipline à l'autre, mais chaque discipline reste à un taux d'ouverture stable. En revanche le mode d'ouverture via la plateforme éditeur augmente (+ 3pts) tandis que l'ouverture via archive ouverte recule (- 3 pts).

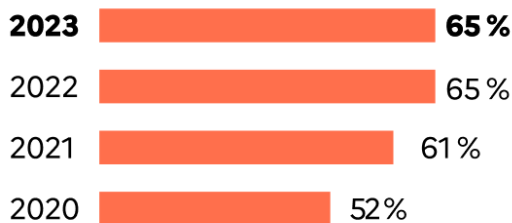
Dans le domaine de la santé, la part des essais cliniques qui partagent leurs résultats dans les 3 années après la fin de l'essai reste faible, mais cet indicateur évolue à la hausse passant de 46 % en 2014 à 52 % en 2020.

La part des publications partageant le code produit reste stable, légèrement inférieur à 20 %, tandis que la part des publications partageant un jeu de données augmente à 22 % (+2 pts).

### Stagnation des taux d'accès ouvert dans chaque discipline

Le taux d'ouverture des publications a beaucoup augmenté depuis 2018 mais il marque cette année un premier palier, avec une stagnation à 65 % entre 2022 et 2023.

Taux d'accès ouvert des publications, avec un DOI Crossref, parues durant l'année précédente par année d'observation



Les différences disciplinaires persistent, traduisant des pratiques et des usages différents d'une communauté scientifique à l'autre, avec par exemple 79 % des publications en accès ouvert en Mathématiques et 52 % en Sciences Sociales. La stagnation des taux d'accès ouvert se retrouve dans chaque discipline avec des taux d'ouverture très proches entre 2022 et 2023. Derrière cette stabilité se cache une dynamique différente entre les voies d'ouverture : l'ouverture via l'éditeur augmente à l'échelle globale (+3 pts) et dans chaque discipline tandis que la contribution des archives ouvertes diminue d'autant, et là aussi, dans chaque discipline (Tableau 1). C'est un signal faible car le taux d'accès ouvert aux publications reste quasiment stable et les contributions des deux voies d'ouverture sont à présent à parité. Néanmoins cette année marque un changement de régime dans les dynamiques internes entre les voies d'ouverture.

### Le taux de partage des résultats des essais cliniques reste stable

Déclarer l'existence des essais cliniques et poster leurs résultats dans des registres publics contribue à une plus grande transparence de la recherche en santé. Elle permet une circulation rapide des résultats, y compris lorsque ceux-ci ont été infructueux et ne font pas l'objet d'une publication scientifique. Le Baromètre analyse la proportion des essais cliniques qui partagent leurs résultats dans les 3 ans après la fin de l'essai. Cet indicateur passe de 46 % en 2014 à 52 % pour les essais terminés en 2020, avec toujours une forte disparité entre promoteurs industriels (74 %) et promoteurs académiques (28 %).

Tableau 1 : Taux d'accès ouvert des publications françaises parues dans l'année précédant la date d'observation

Discipline	Accès ouvert en 2023	Accès ouvert sur archive	Accès ouvert chez l'éditeur
Mathématiques	79 %	70 % (- 1 pt )	37 % (+ 3 pts)
Biologie fondamentale	75 %	62 % (- 2 pts)	58 % (+ 2 pts)
Sciences physiques, Astronomie	74 %	63 % (- 2 pts)	45 % (+ 4 pts)
Sciences de la terre, Ecologie	72 %	57 % (- 4 pts)	53 % (+ 4 pts)
Chimie	71 %	59 % (- 3 pts)	42 % (+ 5 pts)
Informatique	63 %	50 % (- 4 pts)	31 % (+ 4 pts)
Sciences de l'ingénieur	63 %	49 % (- 1 pt )	39 % (+ 5 pts)
Recherche médicale	58 %	39 % (- 1 pt)	50 % (+ 2 pts)
Sciences humaines	55 %	22 % (- 2 pts)	48 % (+ 4 pts)
Sciences sociales	52 %	28 % (- 2 pts)	39 % (+ 2 pts)
<b>Toutes disciplines</b>	<b>65 %</b>	<b>47 % (- 3 pts)</b>	<b>47 % (+ 3 pts)</b>

## Plus de 70 % des thèses de doctorat récentes en accès ouvert

Le baromètre assure le suivi de l'ouverture des thèses de doctorat, grâce notamment aux données de theses.fr (maintenu par l'Abes) et de HAL. La diffusion des thèses au format électronique est régie par arrêté ministériel.

Pour les thèses récentes, hormis celles confidentielles, le docteur ne peut s'opposer à la diffusion de son travail au sein de la communauté universitaire.

En revanche, le docteur peut s'opposer temporairement à la diffusion de sa thèse en accès libre sur internet en définissant une période d'embargo. Les thèses soutenues en 2021 sont en libre accès à 77 % (taux stable depuis 2017). La dernière année de soutenance, 2022, marque un taux de partage légèrement inférieur avec 71 %. C'est naturel car des embargos sont encore en cours. Comme pour les publications, les taux d'ouverture des thèses varient fortement d'une discipline à l'autre, avec par exemple plus de 94 % d'ouverture en Mathématiques et moins de 50 % en Droit et Littérature.

## Le partage des données, code et logiciels est faible mais la dynamique récente est à la hausse

L'obligation d'ouverture des données de la recherche publique est posée par la loi pour une République numérique de 2016. Le Baromètre de la Science Ouverte poursuit le suivi des données de la recherche ainsi que des code et logiciels déjà initié l'an passé : à partir du texte intégral des publications et grâce à des techniques d'intelligence artificielle en fouille de texte, les mentions de jeux de données et de code ou logiciels sont détectées

au sein même de la publication, puis caractérisées en établissant qu'il s'agit de mentions d'utilisation, de création, et/ou de partage. Ces indicateurs sont estimés à partir de modèles d'apprentissage profond, qui continuent à être développés et améliorés. Parmi les publications françaises parues en 2022 qui mentionnent la création d'un jeu de données, 22 % mentionnent son partage (contre 16 % en 2019 et 12 % en 2013). Les plus forts taux de partage s'observent en Sciences de la Terre, et Biologie (35 %), et les plus faibles en Chimie et Ingénierie (respectivement 9 % et 11 %). La dynamique à la hausse du partage des jeux de données se traduit aussi par l'augmentation de la présence d'un « Data Availability Statement » (paragraphe de la publication scientifique dédié à la disponibilité des données). Quasiment inexistant en 2013 (présent dans moins d'1 % des publications françaises), cette déclaration sur la disponibilité des données est présente dans plus de 26 % des publications françaises en 2022. Il est à noter que la présence d'un tel paragraphe ne garantit en rien le partage des données, mais la hausse de sa présence souligne l'évolution culturelle sur le rôle des données de la recherche et qu'au moins une partie des auteurs et des éditeurs sont attentifs à l'enjeu du partage des jeux de données liés aux publications scientifiques. Le taux de partage pour les code et logiciels est de 19 % en 2022 (stable). Le logiciel joue un rôle clé dans la recherche scientifique, dont il est à la fois un outil, un résultat et un objet d'étude. Il ressort d'ailleurs qu'en 2022, plus de 40 % des publications françaises utilisent du code ou un logiciel à des fins scientifiques.

Eric JEANGIRARD  
MESR-SIES

Depuis 2022, les établissements qui le souhaitent peuvent mettre en œuvre une déclinaison locale du Baromètre sur leur périmètre. Plus de 70 établissements disposent ainsi de leur déclinaison locale. Une communauté s'est construite autour de ces déclinaisons et s'est dotée d'une liste de discussion ouverte à tous : [bso-etablissements@groupe.renater.fr](mailto:bso-etablissements@groupe.renater.fr). Pour faciliter le suivi des jeux de données déposés dans les entrepôts, l'équipe du baromètre associant le MESR, Université de Lorraine et Inria met à disposition des établissements qui le souhaitent des outils de fouille pour les aider à lister et caractériser ces jeux de données. Ces outils développés s'améliorent dans le temps en bénéficiant des retours des établissements.

**Méthodologie** : La méthodologie du Baromètre de la Science Ouverte est détaillée dans les documents « A New Framework for the French Open Science Monitor », A. L'Hôte, E. Jeangirard, D. Torny and L. Bracco et « Large-scale Machine-Learning analysis of scientific PDF for monitoring the production and the openness of research data and software in France », A. Bassinet, L. Bracco, A. L'Hôte, E. Jeangirard, P. Lopez et al. Les outils développés sont en open source et disponibles sur GitHub.

**Sources** : Le Baromètre de la Science Ouverte repose uniquement sur des sources ouvertes, en particulier Unpaywall, HAL, theses.fr, DOAJ, OpenAPC, ORCID, clinicaltrials.org et EU Clinical Trials Register (EUCTR). Unpaywall est une base de données mondiale, ouverte, recensant plus de 100 millions de DOI avec leur méta-données disponibles (titre, auteur, éditeur ...) et leur type d'accès. ClinicalTrials.org et EUCTR sont les deux principaux registres publics en ligne d'études cliniques.

**Champ** : Publications avec un DOI (Digital Object Identifier) et dont au moins un des auteurs a une affiliation française. Essais cliniques et études observationnelles menés (au moins en partie) en France, et enregistré dans clinicaltrials.org ou EUCTR.

**Données et code source** : Les données du BSO sont disponibles sur le portail Open Data du MESR :

<https://barometredelascienceouverte.esr.gouv.fr/a-propos/opendata>

Pour plus d'informations, rendez-vous sur le site web du Baromètre de la Science Ouverte : <https://barometredelascienceouverte.esr.gouv.fr>