

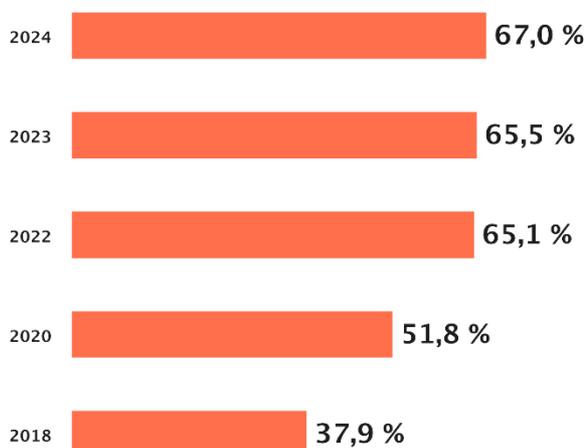
Le Baromètre français de la Science Ouverte 2024

Selon l'édition 2024 du Baromètre de la Science Ouverte (BSO), 67 % des 160 000 publications scientifiques françaises parues en 2023 sont en accès ouvert en décembre 2024, soit un taux en légère hausse (+ 1,5 point) 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 la tendance observée l'an passé se poursuit, le mode d'ouverture via la plateforme éditeur augmente (+ 4 points) tandis que l'ouverture via archive ouverte stagne. 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 et enregistre une baisse de 8 points (de 52 % à 44 % des essais partageant leurs résultats dans les 3 ans). La part des publications partageant le code produit reste stable, légèrement sous 20 %, tandis que la part des publications partageant un jeu de données augmente à 25 % (+ 3 points). Le baromètre introduit cette année le suivi des politiques de science ouverte des établissements : en 2024, 57,5 % des établissements considérés avaient une politique de science ouverte.

Faible augmentation du taux d'accès ouvert

Le taux d'ouverture des publications a beaucoup augmenté depuis 2018 mais l'atteinte d'un plateau se confirme, avec une faible augmentation à 67 % soit + 1,5 point entre 2023 et 2024.

Taux d'accès ouvert des publications 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 80 % 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 2023 et 2024. Derrière cette stabilité se cache une dynamique différente entre les voies d'ouverture : l'ouverture via l'éditeur augmente à l'échelle globale (+ 4 points en un an et + 7 points en 2 ans) tandis que la contribution des archives ouvertes diminue ou stagne, et là aussi, dans chaque discipline (Tableau 1). Cette dynamique interne est un prolongement du signal déjà présent l'an dernier.

Le taux de partage des résultats des essais cliniques reste faible et baisse

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 et fait l'objet d'une réglementation européenne. 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 52 % à 44 % pour les essais terminés en 2021, avec toujours une forte disparité entre promoteurs industriels (69 %) et promoteurs académiques (21 %).

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

Discipline	Accès ouvert en 2024	Accès ouvert sur archive	Accès ouvert chez l'éditeur
Mathématiques	80 %	67 % (- 4 points)	42 % (+ 4 points)
Chimie	78 %	64 % (+4 points)	53 % (+ 10 points)
Sciences physiques, Astronomie	77 %	61 % (- 3 points)	53 % (+ 7 points)
Biologie fondamentale	77 %	61 % (- 1 point)	63 % (+ 3 points)
Sciences de la terre, Ecologie	71 %	53 % (- 4 points)	58 % (+ 4 points)
Informatique	65 %	52 % (+1 point)	31 % (- 1 point)
Sciences de l'ingénieur	65 %	49 % (+ 0 point)	41 % (+ 2 points)
Recherche médicale	60 %	37 % (- 1 point)	54 % (+ 5 points)
Sciences humaines	60 %	26 % (+ 4 points)	51 % (+ 3 points)
Sciences sociales	52 %	26 % (- 1 point)	41 % (+ 2 points)
Toutes disciplines	67 %	47 % (0 point)	52 % (+ 5 points)

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'Agence bibliographique de l'enseignement supérieur - Abes) et de HAL. La diffusion des thèses au format électronique est régie par arrêté ministériel.

Pour les thèses récentes, hormis les thèses 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 2022 sont en libre accès à 77 % (taux stable depuis 2017). La dernière année de soutenance, 2023, marque un taux de partage légèrement inférieur avec 72 %. 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 90 % d'ouverture en Mathématiques et moins de 50 % en Droit et en Littérature.

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

Le Baromètre de la Science Ouverte poursuit le suivi des données de la recherche ainsi que des codes et logiciels déjà initié l'an passé : à partir du texte intégral des publications et grâce à des techniques d'intelligence artificielle mobilisant la 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 2023 qui mentionnent la création d'un jeu de données, 25 % 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 (38 %), et les plus faibles en Chimie et en Ingénierie (respectivement 14% et 12 %).

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 inexistante en 2013 (présent dans moins d'1 % des publications françaises), cette déclaration sur la disponibilité des données est présente dans plus de 37 % des publications françaises en 2023. 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 codes et logiciels est de 19 % en 2023 (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 : bs0-etablissements@groupe.renater.fr Pour faciliter le suivi des jeux de données déposés dans les entrepôts, l'équipe du baromètre 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.gov, EU Clinical Trials Register (EUCTR) et Clinical Trials Information System (CTIS). Unpaywall est une base de données mondiale, ouverte, recensant plus de 100 millions de DOI avec leur métadonnées disponibles (titre, auteur, éditeur ...) et leur type d'accès. ClinicalTrials.gov, EUCTR et CTIS sont les 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.gov, EUCTR ou CTIS.

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> et <https://frenchopensciencemonitor.esr.gouv.fr/> pour la version anglaise.