



HAL
open science

Différences disciplinaires en contexte de Science ouverte. Étude avec les publications de l'archive ouverte HAL

Maxence Larrieu, Joachim Schöpfel

► To cite this version:

Maxence Larrieu, Joachim Schöpfel. Différences disciplinaires en contexte de Science ouverte. Étude avec les publications de l'archive ouverte HAL. 8ème Conférence Document Numérique et Société: Communication scientifique et science ouverte: opportunités, tensions et paradoxes, Université de Liège, Jun 2022, Liège, France. hal-03760316

HAL Id: hal-03760316

<https://hal.science/hal-03760316>

Submitted on 29 Aug 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Différences disciplinaires en contexte de Science ouverte. Étude avec les publications de l'archive ouverte HAL

Maxence Larrieu*, Joachim Schöpfel**

*Université Paris Cité, Direction Générale Déléguée des Bibliothèques et Musées
<https://orcid.org/0000-0002-1834-3007>

**GERiCO, Groupe d'Études et de Recherche Interdisciplinaire en Information et COmmunication,
Université de Lille
<https://orcid.org/0000-0002-4000-807X>

Résumé

L'article présente les résultats d'une étude sur les variations disciplinaires des publications dans HAL. Notre recherche fait suite à d'autres études empiriques qui confirment l'inégalité de développement de la Science ouverte selon les disciplines, par rapport au degré d'adoption de l'accès ouvert, et à l'importance relative des vecteurs de publication utilisés par les chercheurs pour publier. L'étude porte sur les publications récentes (2018-2021) dans HAL, regroupant les articles de revue, communications, ouvrages, chapitres, directions d'ouvrage et preprints. L'analyse est effectuée sur les métadonnées de 563 741 publications selon deux axes, l'un concernant les moyens de communications des disciplines (langue et type de document), la porosité avec l'oligopole de l'édition scientifique, et l'autre sur l'acculturation à l'archive ouverte HAL. Les résultats montrent deux groupes de trois disciplines s'inscrivant en contraste. Les sciences humaines, sociales et les mathématiques possèdent une forte diversité typologique, une forte autonomie oligopolistique et une faible visibilité dans WoS/Scopus. À l'inverse, la biologie, la chimie et la médecine possèdent une forte concentration typologique, une forte dépendance à l'oligopole et une très bonne visibilité dans WoS/Scopus.

Introduction

Les recommandations pour la Science ouverte de l'UNESCO (2021) viennent durablement marquer le mouvement de la Science ouverte. Il y a vingt ans, la Déclaration de Budapest (BOAI 2002) venait qualifier des modalités d'accès à la littérature scientifique. De nos jours, le passage de l'accès ouvert à la science ouverte vient marquer une assise politique. La Science ouverte se structure ainsi au niveau politique, incluant les organisations internationales, gouvernements, financeurs de la recherche et établissements.

Un des points d'attention de la Science ouverte se trouve dans les spécificités des disciplines. Comment réaliser une politique nationale, ou institutionnelle, en prenant en compte la variété des pratiques des disciplines scientifiques ? Dans ce contexte plusieurs études ont ainsi vu le jour afin d'éclairer l'adoption de la Science ouverte par les disciplines, aussi bien du côté des archives que de celui des plateformes de publication. Pour donner un aperçu de ces disparités, prenons l'exemple d'arXiv, l'archive ouverte pionnière conçue pour les sciences de la physique. Deux décennies après la réalisation de l'archive, le dépôt des publications, avant relecture par les pairs (preprint), est devenu "la norme" (Ginsparg, 2011), le 'workflow quotidien' (Creaser 2010) pour les chercheurs de la discipline. Les mathématiques et la physique sont ainsi les disciplines où l'acculturation aux archives est la plus forte.

Notre contribution s'inscrit entre ces deux strates, entre des politiques dont la granularité est nécessairement limitée et des pratiques, des cultures (Séverin *et al.* 2020), de publication variant selon les disciplines. Son objectif est d'éclairer les pratiques disciplinaires dans le contexte de la Science ouverte. Au regard de l'ascension politique de cette dernière, notre objectif est moins d'éclairer les modalités d'accès des publications, que les moyens de communication scientifique (type de document et langue) et le positionnement des disciplines vis-à-vis de l'oligopole mondial de l'édition scientifique – quelle est l'importance de ce dernier dans la diffusion des connaissances des disciplines ? À cette fin nous avons recours à une méthode quantitative en utilisant l'archive ouverte nationale HAL¹. Elle nous permet, comme nous le verrons, d'apporter un autre regard sur les publications, un regard non focalisé sur les articles de revue et les publications rédigées en anglais, un regard capable de rendre compte d'une diversité, typologique, linguistique et éditoriale.

État de l'art

Les disciplines scientifiques font partie des variables indépendantes du développement de la Science ouverte. Le récent Baromètre français de la science ouverte² souligne par exemple “le niveau d'ouverture des publications varie de manière significative d'une discipline à l'autre, en fonction de la sensibilité des communautés scientifiques et de la diversité de leurs pratiques”, des variations qui “s'observent également dans la trajectoire d'évolution du niveau d'ouverture dans le temps” (cf. Jeangirard 2019 et Bracco *et al.* 2022). Les chiffres du Baromètre européen montrent de plus que ces différences disciplinaires dépendent de la politique et du développement particulier de la science ouverte dans chaque pays³.

Plusieurs études empiriques confirment l'inégalité de développement de la science ouverte au sein des disciplines, à deux niveaux (cf. la synthèse de Severin *et al.*, 2020) : (i) L'adoption à l'accès ouvert, ainsi le pourcentage de publications en accès ouvert diffère selon les disciplines. (ii) Les vecteurs de publications pour publier en accès ouvert, dont l'importance varie selon les disciplines. En résumé (cf. le détail chez Severin *et al.*, 2020) :

- Les chercheurs en sciences médicales, en chimie et en génie civil (ingénierie) privilégient la publication d'articles dans des revues 100% en libre accès, avec ou sans frais de publication (voie dorée ou *gold road*) ;
- La diffusion des résultats via le dépôt dans une archive ouverte (voie verte ou *green road*) joue un rôle plus important pour les chercheurs en physique, en mathématiques, en informatique et en astronomie ;
- Les sciences sociales et humaines semblent partagées entre d'une part, les sciences de la société avec une préférence pour la voie verte (dépôt dans les archives ouvertes) et d'autre part, les humanités avec une approche mixte (voie verte et voie dorée).

Concernant HAL, plusieurs études ont révélé des différences disciplinaires. Baruch (2007) a fait remarquer qu'en physique et en mathématiques, « l'initiative d'arXiv a fait de l'auto-archivage la norme, en particulier en physique des hautes énergies ; cependant, d'autres domaines ne suivent pas le même schéma », un constat confirmé par Wojciechowska (2017) et Mahé (2017). Aussi, les directeurs d'unités du CNRS en mathématiques, physique et informatique soutiennent davantage HAL que dans d'autres domaines (Schöpfel *et al.* 2016).

D'autres recherches montrent des variations entre disciplines au niveau de la volumétrie, des types de publication, du taux d'ouverture, de la présence du texte intégral, d'une collection HAL et les modalités d'alimentation (cf. Prime-Claverie & Mahé, 2013; Mahé & Prime-Claverie, 2017; Schöpfel *et al.*, 2019; Schöpfel, 2020).

Une raison à ces différences est la présence d'archives disciplinaires et de leurs interconnexions avec HAL. Relevons arXiv pour la Physique, RePEc pour l'économie (Magron, 2017), Pubmed Central pour la médecine

1 HAL <https://hal.archives-ouvertes.fr/>

2 Baromètre française de la science ouverte <https://barometredelascienceouverte.esr.gouv.fr/publications/disciplines?id=disciplines.dynamique-ouverture>

3 EU Open Science Monitor https://ec.europa.eu/info/research-and-innovation/strategy/goals-research-and-innovation-policy/open-science/open-science-monitor/trends-open-access-publications_en

(Roberts, 2001). Un autre facteur est la mobilisation des chercheurs eux-mêmes et la présence (ou l'absence) d'équipements et de personnels de l'information. Ces variables ne sont pas indépendantes ; pratiques individuelles, choix collectifs, politiques d'établissements, pratiques disciplinaires font partie du même écosystème. À partir de leur synthèse de la littérature scientifique, Severin et al. (2020) résumant que "les forces motrices de la transition de l'édition savante vers l'accès ouvert sont multiples et entremêlées", telles que les initiatives de sensibilisation ascendantes au sein de la communauté scientifique, la politique (mandats) des agences de financement, des gouvernements, des universités et des organismes de recherche qui obligent les chercheurs de toutes les disciplines à mettre leurs résultats de recherche en accès ouvert, ou encore, au moins pour les sciences de la vie et la médecine, les fonds suffisants pour couvrir les frais de publication (APC) dans les revues en accès ouvert.

Méthode

Au regard des études quantitatives réalisées sur le sujet, deux aspects apparaissent prépondérants dans la réalisation de la méthode : les sources et la correspondance disciplinaire.

Sources

Les sources utilisées pour établir des statistiques sur les publications sont d'une importance capitale, car elles influencent directement les résultats. Des variations de résultats se retrouvent ainsi dans les outils de monitoring. Par exemple le tableau de bord "Curtin Open Knowledge Initiative" mentionne que 64 % des publications françaises de 2019 sont en accès ouvert, quand celui de OpenAire indique 60 % et le Baromètre français de la Science ouverte 59 %. Une des principales raisons de ces variations se trouve dans les sources utilisées⁴. Par exemple dans (Maddi, 2020), les disciplines physique et mathématiques – connues pour avoir une forte acculturation à l'accès ouvert (Jeangirard, 2019) (Creaser, 2010) – apparaissent avec un taux d'accès ouvert en deçà de la moyenne des disciplines. La raison se trouve dans la source utilisée, le Web of Science (WoS), lequel indexe des revues scientifiques et non les preprints non publiés dans une revue (Maddi, 2021).

Ces deux exemples nous montrent l'importance des sources, ce qui revient à dire que les sources de données des publications ne sont pas neutres. Tout comme le Manifesto de Leiden (Hicks, 2015) qualifie le *h-index* de *database dependent*, il faut souligner que les résultats de statistique sur les publications dépendent des sources utilisées ; ils prennent sens en considération des limites de celles-ci.

HAL, l'archive ouverte nationale

L'archive ouverte pluridisciplinaire HAL, qui a fêté ses 20 ans l'an dernier, est maintenant reconnue et stabilisée dans le paysage scientifique français. Relevons : 2016, HAL devient une infrastructure nationale de recherche ; 2019, elle se substitue à l'archive institutionnelle de l'INRAE ; 2020, le CNRS fait de HAL la source pour l'évaluation de ses chercheurs (CRAC/RIBAC) ; 2021, l'ANR rend obligatoire le dépôt dans HAL pour les publications issues des projets financés et un HALathon à échelle nationale est organisé par l'association CasuHAL. Soulignons également que depuis 2021, l'État français soutient HAL dans le cadre du Programme d'investissements d'avenir (PIA), comme un équipement scientifique nécessaire au maintien de l'excellence et du leadership de la recherche nationale (Équipements structurants pour la recherche, EquipEx+), pour le projet HALiance (HAL interoperability, accessibility and network centric environment). Ces événements témoignent de la stabilité politique de l'archive et présument de son utilisation par la communauté scientifique. Concernant l'utilisation, il est important de relever que l'alimentation de HAL ne fonctionne pas uniquement sur le modèle d'auto-archivage. L'alimentation directe des archives par les auteurs (auto-archivage) est un des principes originels de l'accès ouvert : c'est la première stratégie mentionnée dans la Budapest Initiative for Open Access (BOAI, 2002), le mode de fonctionnement de arXiv, lui-même modèle de l'archive ouverte HAL (Baruch 2007). A l'issue de ses 20 ans, HAL a la particularité d'être aussi alimentée par des procédés automatisés et par des

⁴ Une autre raison de ces variations se trouve dans la date d'observation par rapport aux années de publications. Le Baromètre français de la Science Ouverte montre ainsi que les publications de 2019 étaient en accès ouvert à hauteur de 52 % en 2020, puis de 59 % en 2021

personnels supports. Pour l'alimentation automatisée citons le projet Collex-Persée Droit2HAL⁵, lequel effectue un versement et enrichissement des métadonnées Dalloz, et le contrat national avec Elsevier⁶ qui verse des publications de l'éditeur dans HAL⁷. L'alimentation par les personnels support est effectuée au niveau institutionnel (Mahé *et al.*, 2017), avec parfois des méthodes également automatisées⁸.

Dans notre méthode le choix de HAL s'est fait en considération de cette visibilité accrue de l'archive. Il paraît prévisible en effet que la couverture de HAL augmente au fil des années. Néanmoins, actuellement l'archive couvre moins de publications françaises qu'on en trouve dans le WoS ou Scopus. Nos résultats sont donc à comprendre à la lumière de ces variations, lesquelles sont précisées en liminaire de la section dédiée.

Enfin, HAL a également été choisie pour des raisons techniques et politiques. Rendre compte des publications françaises avec HAL relève d'une dimension cohérente, car l'archive est alimentée directement par les acteurs de l'Enseignement Supérieur et de la Recherche. Par rapport aux outils traditionnels commerciaux, qui sont eux alimentés 'par le haut', avec un comité d'experts sélectionnant les sources à indexer, l'archive ouverte peut être qualifiée d'endogène, car alimentée directement par les acteurs de la communauté académique⁹. Comme le soulignent Pölonen *et al.* (2020), construire des statistiques à échelle nationale en dépendant du WoS ou de Scopus pose des problèmes de perspective, d'intérêt économique (Tennant 2018) et, ajoutons, d'équité¹⁰.

Techniquement enfin, avec ses métadonnées ouvertes et son API (Application Programming Interface), HAL permet une exploitation efficace des données. L'extraction, le nettoyage, l'enrichissement et l'exploitation des données ont été réalisés avec un code python, partagé sous licence ouverte sur la plateforme Github¹¹. Soulignons enfin que l'usage de métadonnées ouvertes et le partage d'un accès à ces données est une des recommandations de la déclaration DORA¹².

Correspondance disciplinaire

La correspondance disciplinaire des publications est un sujet à part entière qui, comme pour le choix des sources, ne possède pas de réponse parfaite. Fixer une nomenclature et rattacher des publications à celle-ci, reste, à l'heure de l'inter, pluri et transdisciplinarité, un procédé artificiel. Dans la littérature – voir (Shu, 2019) pour un état de l'art – il est d'usage de différencier deux méthodes, l'une effectuée au niveau des revues, l'autre au niveau des publications. Nous retrouvons ces deux méthodes dans le paysage politique français : l'OST¹³ se base sur le Web of Science (Maddi *et al.*, 2021), qui intègre une correspondance effectuée au niveau des revues¹⁴ ; le Baromètre français de la Science ouverte effectue lui une correspondance au niveau des publications – à l'aide d'une algorithmie utilisant le *machine learning* et les bases bibliographiques du CNRS, PASCAL et FRANCIS (Jeangirard, 2019).

5 Voir <https://www.collexpersee.eu/projet/droit2hal/>

6 Pour le périmètre des publications, types de dépôt et modalités techniques voir le billet de blog de l'unité qui maintient HAL, le Centre pour la Communication Scientifique Direct (CCSD) <https://www.ccsd.cnrs.fr/2022/02/accord-national-avec-l-editeur-elsevier-mise-en-oeuvre-des-imports-dans-hal-la-suite/>

7 Pour prolonger relevons le compte EDP Science dans HAL https://hal.archives-ouvertes.fr/search/index/q/*/contributorId_i/301560

8 A l'université Versailles et celles de Rennes 1 des versements automatiques sont réalisés au fil de l'eau voir (Larrieu *et al.* 2021) et <https://scienceouverte.univ-rennes1.fr/centralisation-par-le-scd>

9 Le contrat avec Elsevier n'atténue pas ce constat, car c'est le CCSD qui effectue les versements et non la multinationale.

10 Voir (Tennant *et al.* 2019) (Mongeon *et al.* 2016) et (Singh 2021) pour un aperçu sur les problèmes de couverture et d'équité de WoS et Scopus, où l'anglais est la langue hégémonique des publications indexées, où la répartition géographique des comités éditoriaux revues est concentrée.

11 Voir <https://github.com/ml4rrieu/disciplines-differencies-hal>

12 La déclaration de San Francisco sur l'évaluation de la recherche (DORA) a établie en 2012 une vingtaine de recommandations sur les méthodes d'évaluation des résultats de la recherche scientifique, la 12e étant "Provide the data under a licence that allows unrestricted reuse, and provide computational access to data, where possible" <https://sfdora.org/read/>

13 L'Observatoire des Sciences et Techniques est un département du HCERES (Haut Conseil de l'évaluation de la recherche et de l'enseignement supérieur)

14 "Based on the citation pattern and selection by experts, journals are assigned to one or multiple WoS categories (Pudovkin & Garfield, 2002) but the detailed methodology has never been published". (Shu 2019)

HAL se structure notamment avec un référentiel des disciplines¹⁵, lesquelles sont renseignées au moment du dépôt. Toutes les publications sont ainsi rattachées à un domaine primaire (métadonnée *primaryDomain_s*) et éventuellement à d'autres domaines. Notre étude utilise ainsi une correspondance au niveau des publications. Le référentiel HAL des disciplines s'est avéré trop fin pour notre étude et nous avons effectué une correspondance avec la nomenclature du Baromètre français (Bracco *et al.*, 2021), qui utilise dix disciplines (cf. annexe). Utiliser cette nomenclature nous permet de comparer les volumétries avec le Baromètre de la Science ouverte, ce qui est essentiel pour contextualiser les résultats¹⁶. La correspondance est effectuée à l'aide d'une table qui relie les deux cartes disciplinaires¹⁷. Enfin, pour ne pas surcharger l'écrit nous nous référons aux disciplines en utilisant le premier terme de leur nom complet.

Périmètre

L'étude porte sur les publications présentes dans HAL publiées entre 2018 et 2021 compris. Une particularité de l'archive ouverte est d'inclure une forte variété des productions scientifiques, par exemple logiciels, rapports et cartes géographiques. Notre étude visant à éclairer les cultures de publications, elle inclut les articles de revue, communications dans des congrès, ouvrages, chapitres, directions d'ouvrage et preprints.

Les résultats présentés sont un instantané des publications présentes dans HAL. Les données ont été recueillies durant le mois de mai 2022 et concernent 563 741 publications publiées entre 2018 et 2021.

L'importance des sources pour les résultats obtenus n'échappe évidemment pas à notre démarche : les limites de nos résultats viennent directement de la représentativité des disciplines dans HAL. Afin d'éclairer ces limites le tableau 1. compare le volume de publications de l'année 2018 par disciplines entre deux sources, HAL et la source utilisée par le Baromètre français de la Science ouverte, CrossRef (Bracco *et al.*, 2022), c'est-à-dire les publications ayant un DOI de cette agence internationale.

Disciplines	HAL	BSO	Rapport HAL / BSO
Biology (fond.)	20154	24147	0.8
Chemistry	9135	7223	1.3
Computer and information sciences	13924	11479	1.2
Earth, Ecology, Energy and applied biology	11089	13675	0.8
Engineering	13879	5772	2.4
Humanities	41517	15092	2.7
Mathematics	5696	4407	1.3
Medical research	5048	53333	0.1
Physical sciences, Astronomy	13693	18803	0.7
Social sciences	18357	10846	1.7
Total	152492	164787	0.93

Tableau 1. Comparaison de la distribution disciplinaire des publications de 2018 entre HAL et le Baromètre français de la Science Ouverte (BSO).

Nous constatons d'abord avec intérêt que le nombre de publications pour les deux sources est du même ordre de grandeur (~ 150k par an) ; utiliser HAL pour rendre compte des cultures de publication n'est donc pas moins pertinent que d'autre source comme CrossRef, lequel possède lui-même une meilleure volumétrie que les outils commerciaux comme WoS/Scopus – cf. (Visser M. *et al.*, 2021). Le tableau montre ensuite des variations

15 Voir <https://aurehal.archives-ouvertes.fr/domain/index>

16 Si aucune grille disciplinaire ne peut s'avérer parfaite, une limite de celle utilisée peut se trouver dans les regroupements entre Sciences humaines et Sciences sociales, qui laissent par exemple de côté la spécificité du droit – voir (Severin *et al.* 2020) pour une autre grille disciplinaire et (Mahé *et al.* 2017) pour un focus sur les Sciences Humaines et Sociales.

17 Voir <https://github.com/ml4rrieu/disciplines-differencies-hal>

importantes pour *Medical*, 10 fois plus de publication dans le BSO, *Engineering* et *Humanities*, où l'on trouve plus du double de publication dans HAL

Résultats

Types de publication

Le graphique *Diversité des types de publications par disciplines* (figure 1) est directement obtenu à partir de la typologie documentaire de HAL.

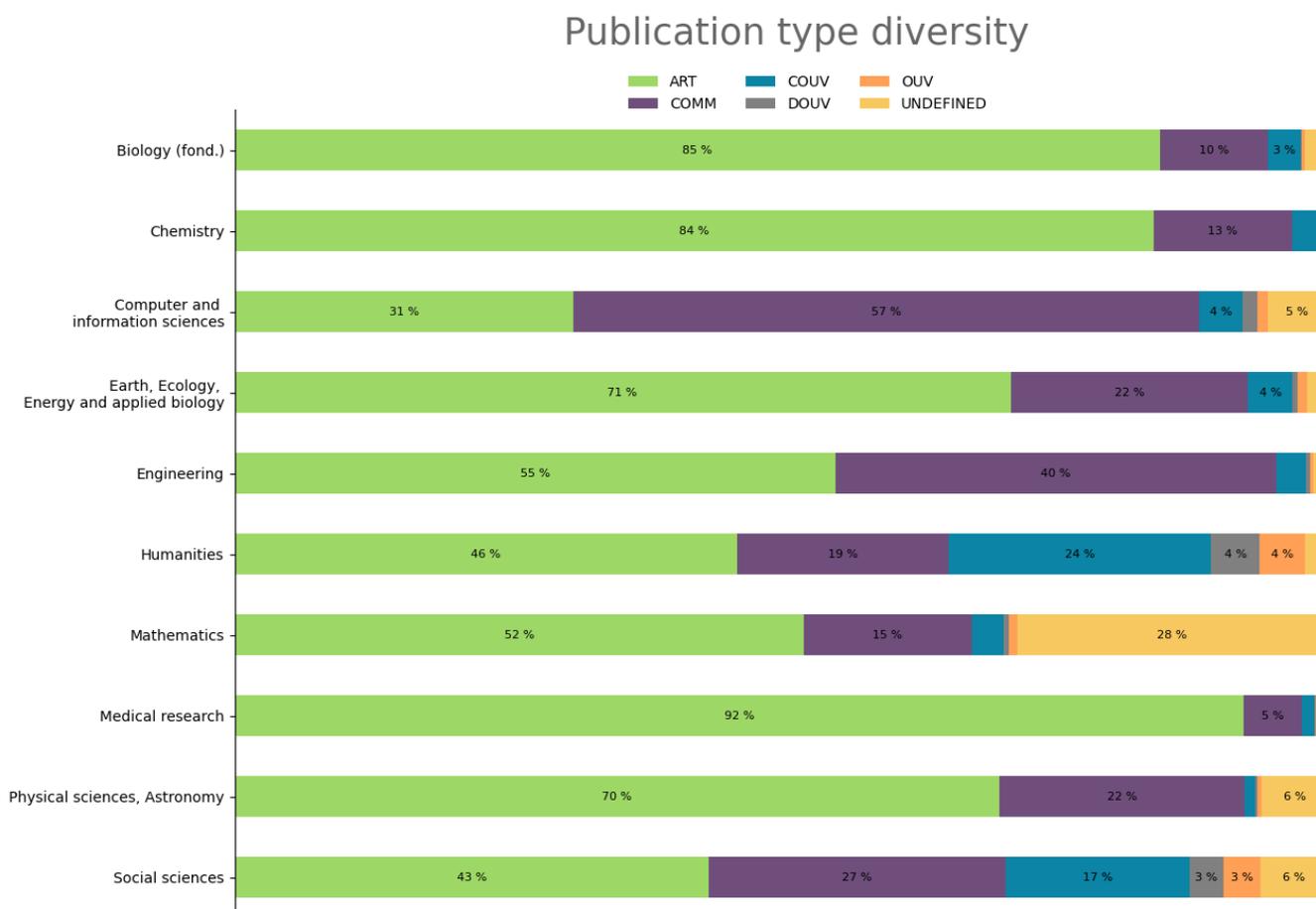


Figure 1. Diversité des types de publications par disciplines . Article de revue (ART), chapitre d'ouvrage (COUV), ouvrage (OUV), communication dans un congrès (COMM) , direction d'ouvrage (DOUV) et preprint (UNDEFINED)

Les résultats laissent apparaître trois groupes :

- Un premier à forte diversité, *Humanities*, *Social* et *Mathematics*, où les articles ne représentent pas plus de la moitié des publications ;
- Un deuxième à diversité moyenne, *Computer*, *Engineering*, *Physical* et *Earth* ;
- Un dernier à faible diversité, *Medical*, *Biology* et *Chemistry*, où les articles de revue sont présents à plus de 90 %.

Le graphique relève certaines spécificités disciplinaires : les preprints sont très présents (28 %) en *Mathematics* ; les proceedings (communications) sont fortement utilisés en *Computer* (57 %) et *Engineering* (40%) - pour *Computer* c'est le premier type ; les *Humanities* et *Social* où les ouvrages (ouvrage, chapitre et direction d'ouvrage) sont les plus présents.

Langue

Quelles sont les langues de publication des disciplines ? Le graphique *Diversité des langues de publication par disciplines* (figure 2) utilise la métadonnée dédiée et obligatoire lors du dépôt dans HAL. Les résultats révèlent la forte particularité des *Humanities* et *Social*. Là où les disciplines ont recours à plus de 85% à l'anglais, la première langue des publications des *Humanities* et *Social* est le français.

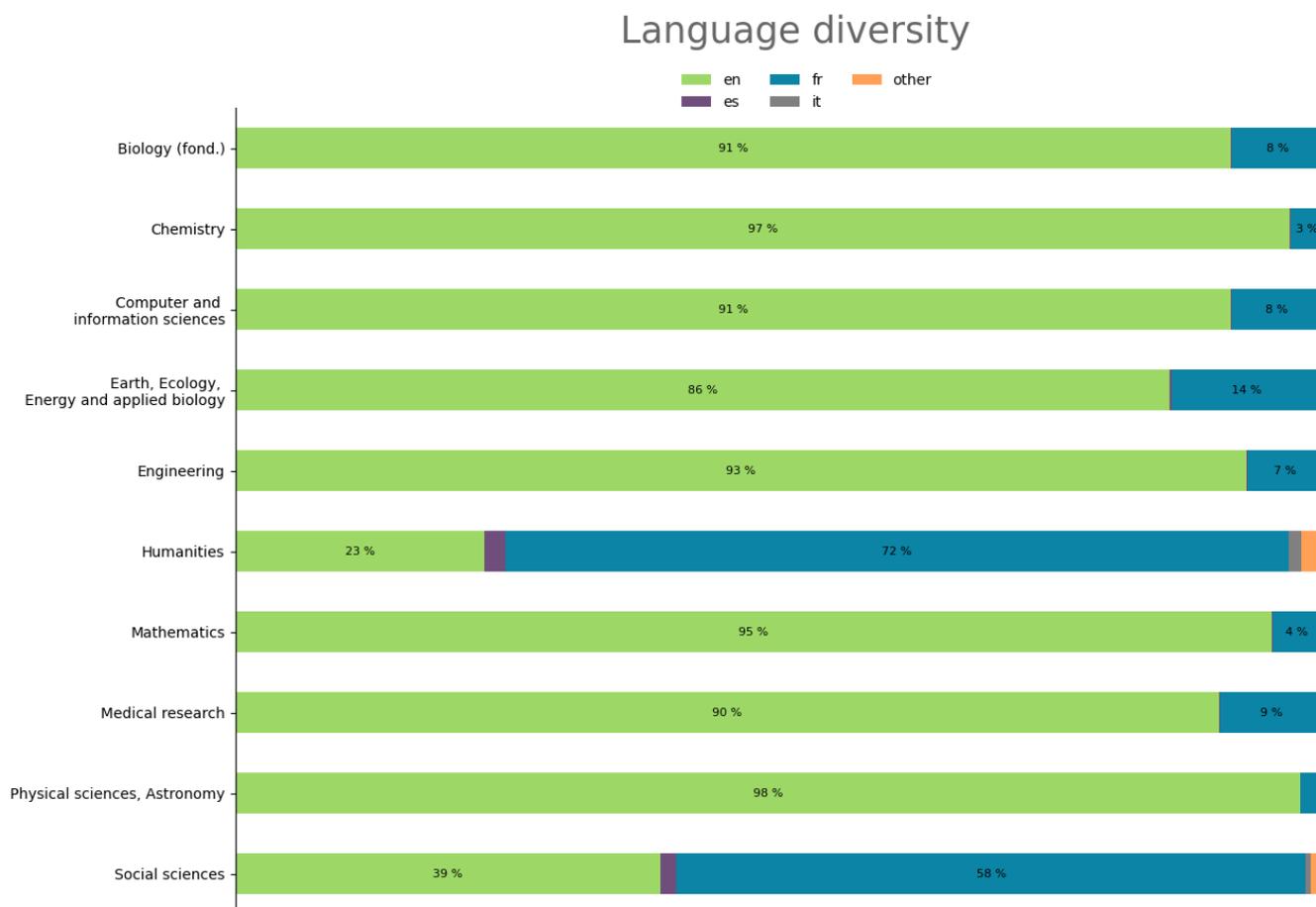


Fig. 2. Diversité des langues de publication par disciplines.

Marché de l'édition et oligopole

Le graphique *dépendance des disciplines à l'oligopole* (figure 3) éclaire la dépendance des disciplines à l'égard de l'oligopole de l'édition scientifique mondiale d'une part, et des éditeurs en accès ouvert commerciaux de l'autre. L'oligopole regroupe les six plus grands éditeurs scientifiques commerciaux : Elsevier, Springer Nature, Wiley, American Physical Society, Taylor & Francis et Sage Publications (Larivière et al., 2015). Les éditeurs en accès ouvert commerciaux sont ceux qui publient intégralement en accès ouvert et qui ne sont pas des sociétés sans but lucratif. A la suite de notre expérience et de l'étude (Khoo, 2019) sont inclus Frontiers, MDPI, Hindawi, BMC et IOP Publishing¹⁸. Le graphique a nécessité un travail de nettoyage et de regroupement en amont – fusionner les variantes de noms et effectuer des regroupements comme Hindawi pour Wiley.

¹⁸ Actuellement, Hindawi est la propriété de Wiley et BMC de Springer Nature. Pour ce graphique seulement ces regroupements n'ont pas été effectués.

Oligopolistic dependency

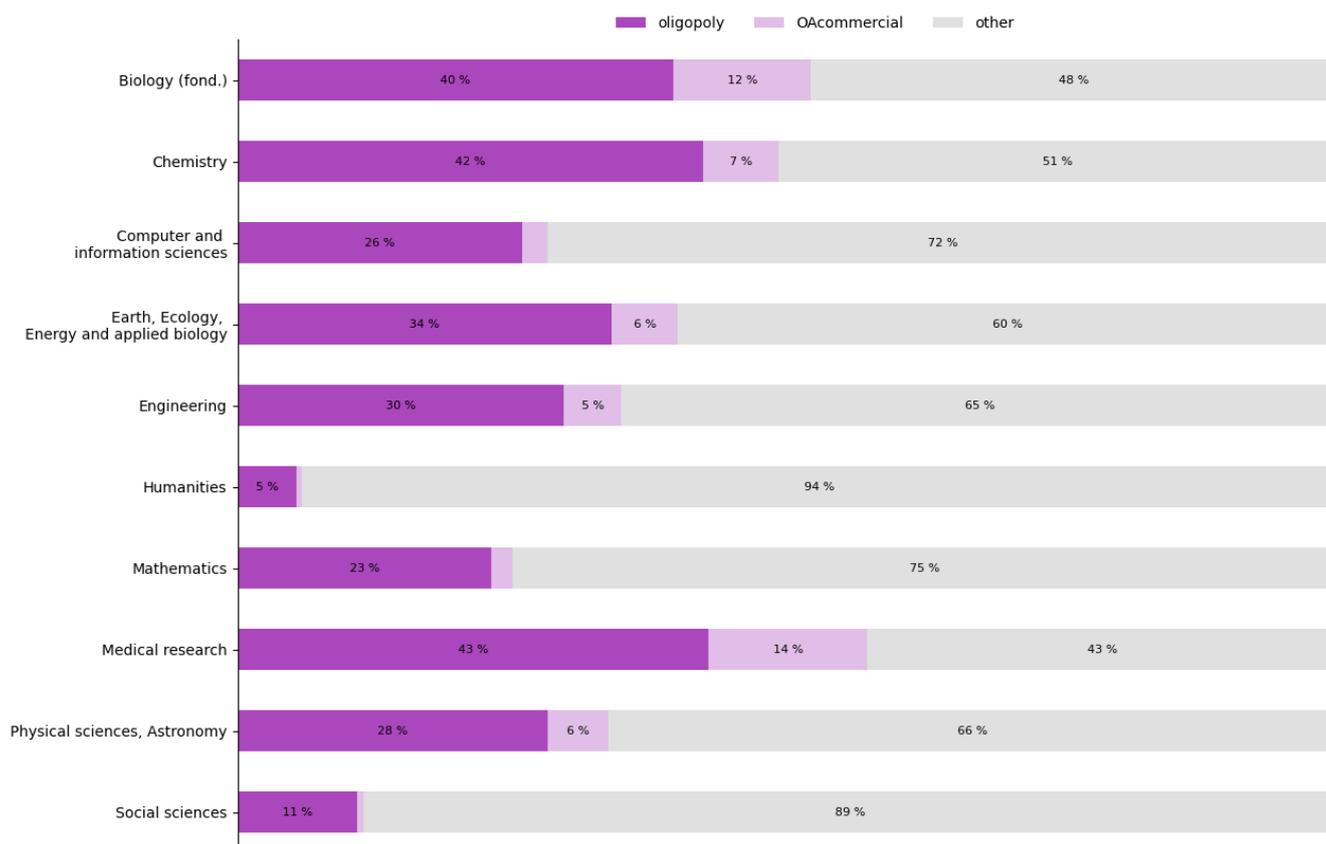


Figure 3. Dépendance des disciplines à l'oligopole. Editeurs de l'oligopole : Elsevier, Springer Nature, Wiley, American Physical Society, Taylor & Francis et Sage Publications. Editeurs commerciaux en accès ouvert Frontiers, MDPI, Hindawi, BMC et IOP Publishing

Le graphique laisse apparaître trois groupes :

- Un premier où l'oligopole et les éditeurs commerciaux en accès ouvert publient au minimum la moitié des publications : *Medical, Biology et Chemistry* ;
- Un deuxième où la dépendance se situe entre 20 et 40 %, *Earth, Engineering, Physical, Computer et Mathematics* ;
- Un dernier où l'oligopole est peu présent, *Humanities et Social* avec seulement 5 % pour les premières.

Visibilité

La figure 4 *Visibilité des disciplines* concerne la présence de DOI sur les publications et l'indexation des sources des publications dans le WoS et Scopus. L'indexation dans WoS/Scopus a été obtenue avec les listes des ISSNs indexés dans les bases commerciales.

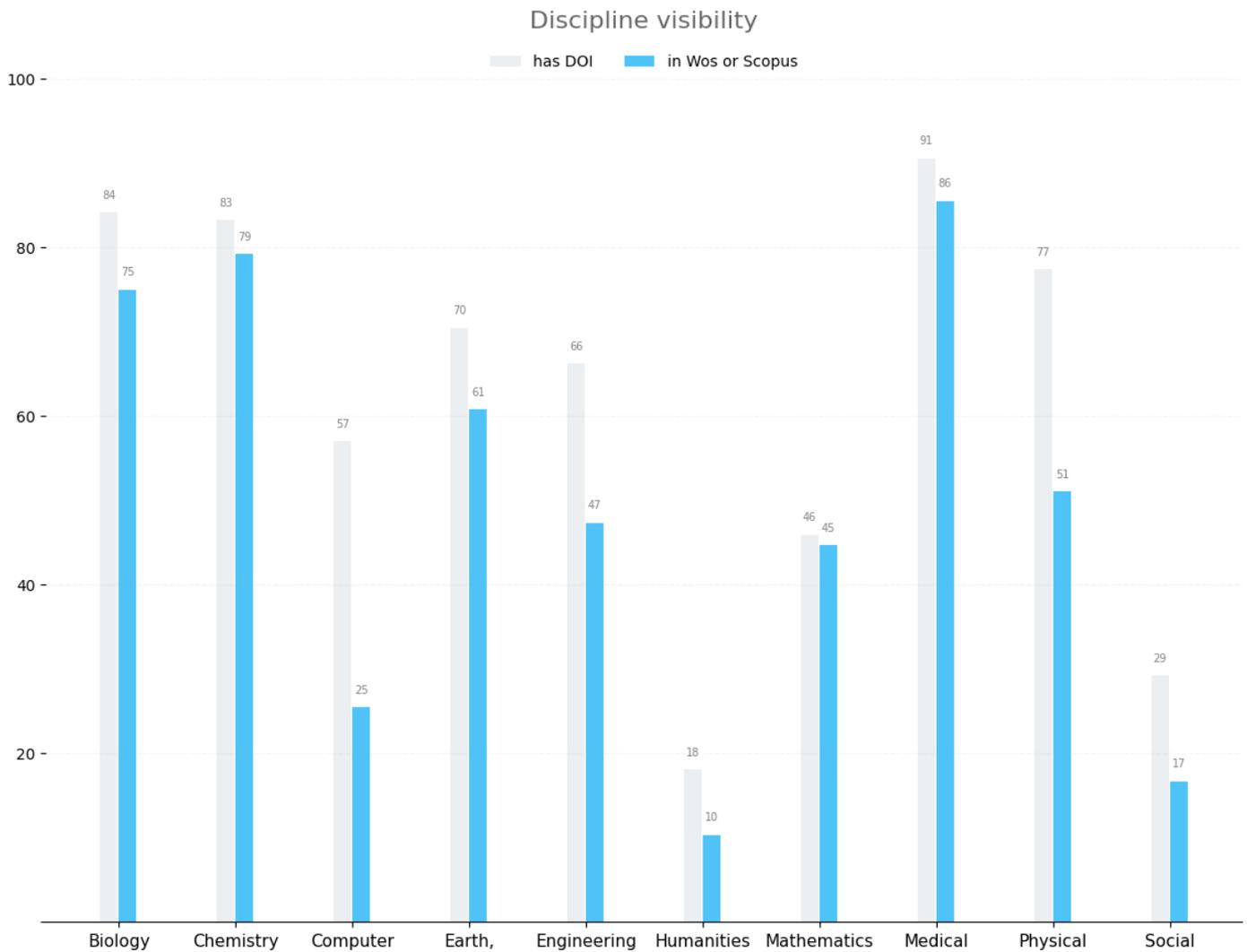


Figure 4. Visibilité des disciplines. En gris le pourcentage de publication avec DOI ; en bleu le pourcentage de publications indexées dans WoS ou Scopus

Pour les DOI, on relève trois disciplines dont le taux de couverture est inférieur à 60 %, par ordre décroissant, *Computer*, *Mathematics*¹⁹ et *Humanities*. La visibilité dans le WoS/Scopus laisse apparaître trois groupes :

- Un premier où l'indexation est supérieure à 70 % : *Medical*, *Chemistry*, *Biology* ;
- Un second compris entre 45 et 61 % : *Earth*, *Physical* et *Mathematics* ;
- Un dernier inférieur à 30 % : *Computer*, *Social* et *Humanities*.

Comme nous le verrons, ces résultats sont à mettre en regard des types de publication.

Acculturation à l'archive ouverte HAL

Quel est le taux des publications déposées en texte intégral dans HAL et le taux d'auto-archivage des disciplines ? Le graphique *Taux d'accès ouvert par disciplines et Auto-archivage* (figure 5) est réalisé directement avec les données de HAL. Le taux d'accès ouvert est obtenu avec la métadonnée *openAccess_bool*, une métadonnée booléenne qui est valide pour les publications déposées dans HAL sans embargo et pour celles disposant d'un accès ouvert sur le web identifié via Unpaywall²⁰, arXiv ou Pubmed Central. La donnée 'Hal file'

¹⁹ La faible présence de DOI pour les Mathematics n'est qu'une question de mise à jour dans HAL, car arXiv a ajouté des DOIs à tous ses dépôts en début d'année 2022, cf. <https://blog.arxiv.org/2022/02/17/new-arxiv-articles-are-now-automatically-assigned-DOIs/>

²⁰ Unpaywall est un service de l'organisation à but non lucratif Our Research ; c'est l'outil industriel mondial qui analyse et met à disposition des informations d'accès ouvert sur les publications à partir de leurs DOI.

décrit la présence du texte intégral et ‘hal file self archiving’ indique si le dépôt a été réalisé par un des auteurs (métadonnée *selfArchiving_bool*).

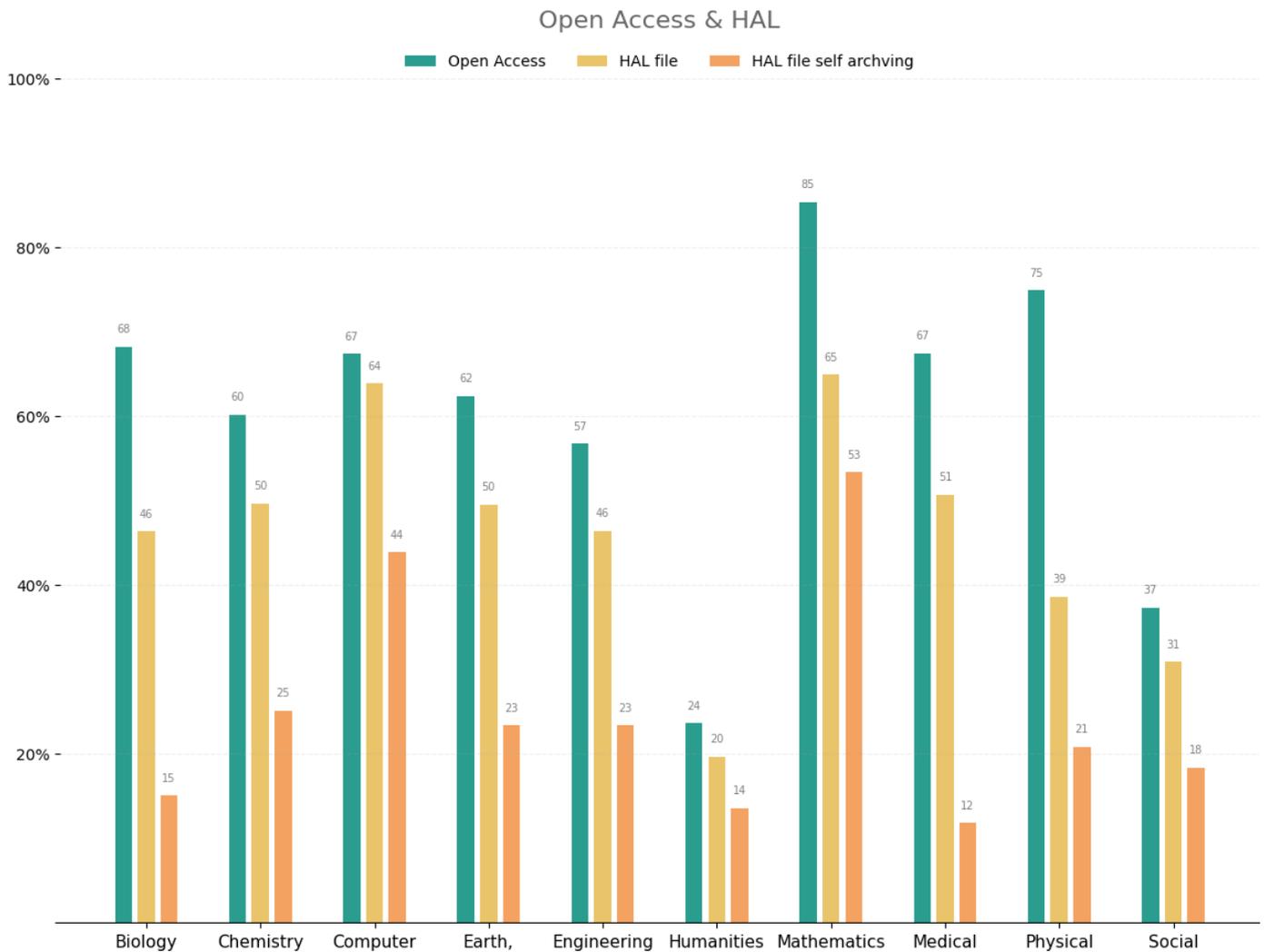


Figure 5. Taux d'accès ouvert par disciplines et auto-archivage. En vert le pourcentage de publication en accès ouvert ; en jaune le pourcentage de publications déposées en intégral dans HAL ; en orange le pourcentage de publications déposées en texte intégral par les auteurs eux-mêmes.

Nous observons un taux d'accès ouvert élevé de plus de 60 % sauf pour les *Social* et *Humanities* (37 et 24 %). Les taux obtenus pour chacune des disciplines sont plus élevés que ceux du Baromètre français de la Science ouverte : nombre des publications absentes de HAL sont donc en accès fermé. La présence du texte intégral, en regard du taux d'accès ouvert, est également élevée. Deux groupes se distinguent :

- Les disciplines où la différence entre les deux est inférieure ou égale à 20 % – c'est-à-dire que l'accès ouvert se fait à plus de 80 % via le texte intégral dans HAL – *Chemistry, Computer, Earth, Engineering, Humanities, Mathematics, Medical* et *Social* ;
- Les disciplines où l'écart est supérieur à 20 %, *Physics, Biology*.

La lecture de l'auto-archivage peut se faire par rapport aux taux de texte intégral. Se distinguent les disciplines où les dépôts sont essentiellement faits par les auteurs (> 65 %), *Mathematics, Humanities* et *Computer* ; celles où environ la moitié des dépôts est faite par les auteurs, *Chemistry, Earth, Engineering, Physical, Social*, et enfin celles où l'auto-archivage est faible (< 30 %) *Biology, Medical*.

Analyses

Deux axes permettent d'analyser ces résultats. Le premier regroupe les graphiques de diversité typologique, dépendance oligopolistique et visibilité dans les bases bibliométriques commerciales, le deuxième concerne l'acculturation à l'archive ouverte HAL.

Une nette corrélation apparaît entre la diversité typologique et la dépendance oligopolistique, autour de laquelle deux groupes disciplinaires se forment : *Humanities*, *Social* et *Mathematics* possèdent une forte diversité typologique et une très faible dépendance à l'oligopole. À l'opposé *Biology*, *Chemistry* et *Medical* possèdent une faible diversité typologique et une forte dépendance oligopolistique. La lecture se poursuit dans les résultats sur la visibilité. *Humanities* et *Social* possèdent les plus faibles visibilités dans WoS/Scopus et inversement pour *Biology*, *Chemistry* et *Medical*. Dans ce premier axe, les autres disciplines, *Earth*, *Engineering*, *Physical* et *Computer* forment un schéma moins marqué où la corrélation entre les analyses est présente mais moins évidente. *Humanities* et *Social* se démarquent amplement par la langue avec plus de la moitié des publications en langue française.

La lecture de l'accès ouvert et de HAL est moins évidente, car une publication peut être simplement signalée dans HAL (métadonnées) et en accès ouvert ailleurs sur le Web. *Humanities* et *Social* se démarquent encore nettement avec les taux d'accès ouvert les plus faibles (inférieur à 40 %). Ces taux sont à interpréter avec la forte présence des livres et la faible présence de DOI. D'une part, le mouvement de Science ouverte s'est d'abord façonné avec les articles de revue – pensons à Sherpa Roméo, le DOAJ et la Loi pour une République numérique – d'autre part, l'absence de DOI freine inévitablement l'identification de version en accès ouvert sur le Web.

Comparées aux pourcentages d'accès ouvert, deux disciplines apparaissent avec des faibles taux de dépôt en texte intégral dans HAL : *Physics* et *Biology* ont plus de 20 % de leurs publications en accès ouvert en dehors de HAL. Pour *Physics* ce taux est à reconsidérer avec arXiv, c'est une des disciplines les plus acculturées aux archives (Wojciechowska, 2017), nous faisons ainsi l'hypothèse que nombre de ces publications sont déposées dans arXiv. *Biology* et *Medecine* possèdent les taux les plus bas d'auto-archivage, avec seulement un tiers des dépôts en texte intégral fait par les auteurs. Ces disciplines sont donc peu acculturées aux archives ouvertes (Schöpfel 2016) (Creaser et al., 2010). Cette faible utilisation est à mettre en regard de Pubmed Central, une archive en sciences de la vie dont l'alimentation, à la différence de arXiv, n'est pas effectuée par les auteurs mais directement par les maisons d'édition (Roberts, 2001).

Conclusion

Avec une méthode quantitative, l'étude éclaire des différences entre disciplines au niveau des pratiques de communications, de porosité avec l'oligopole de l'édition scientifique et d'acculturation avec l'archive ouverte nationale. Son originalité se trouve dans la source et la grille d'analyse. Utiliser HAL pour faire des statistiques sur les publications reste peu habituel. L'archive a été choisie car elle apporte un autre regard sur les publications par rapport aux outils traditionnels : elle n'est pas alimentée 'par le haut', la littérature n'est pas issue de la sélection d'un ensemble de revues par un comité d'experts. En considérant la sur représentation de la langue anglaise et des articles de revues, les problèmes de concentration géographiques des outils commerciaux, HAL nous apparaît finalement plus équitable et plus cohérente à la science ouverte – l'équité étant une de ses valeurs constitutives (Unesco, 2021).

Les résultats montrent d'abord la spécificité des *Humanities* et *Social* : seules disciplines à utiliser le français comme première langue, la plus forte utilisation des livres (chapitre, ouvrage, direction d'ouvrage), les moins visibles dans les outils bibliométriques commerciaux, avec les couvertures en DOI les plus faibles et, par conséquence, avec taux d'accès ouvert les plus faibles.

L'analyse montre ensuite deux groupes de trois disciplines s'inscrivant en contraste. *Humanities*, *Social* et *Mathematics* possèdent une forte diversité typologique, une forte autonomie oligopolistique et une faible visibilité dans WoS/Scopus. À l'inverse, *Biology*, *Chemistry* et *Medical* possèdent une forte concentration typologique (> 84 % d'article), une forte dépendance à l'oligopole et conséquemment les plus hautes visibilités dans WoS/Scopus. L'analyse met ainsi en lumière les corrélations entre type de publication, dépendance

oligopolistique et visibilité dans les outils bibliométriques. Ces corrélations sont d'autant plus présentes que les disciplines se démarquent (*Humanities* et *Medical*). Les résultats concernant l'acculturation aux archives ouvertes concordent avec la littérature : les Mathematics sont fortement acculturés, contrairement à *Biology* et *Medical* où des taux d'auto-archivage sont les plus bas.

Notre étude apporte une contribution dans l'éclairage des cultures de publication des disciplines. D'autres analyses pourraient approfondir le travail réalisé. Nous pensons par exemple à une étude plus fine sur le type d'accès ouvert (avec la typologie habituelle Diamond Open Access, Gold APC, Hybrid etc.), pour éclairer la facilité des disciplines à payer pour publier en accès ouvert. Enfin, la grille d'analyse utilisée est actuellement liée au marché de l'édition scientifique. De façon complémentaire il serait intéressant de la porter vers l'infrastructure ouverte, avec par exemple la place et l'utilisation des revues diamants ou l'indexation des livres dans les disciplines.

Bibliographie

Baruch, P. (2007). Open access developments in France: the HAL Open Archives System. *Learned Publishing*, 20(4), 267–282. <https://doi.org/10.1087/095315107x239636>

BOAI (2002) *Budapest Initiative of Open Access* <https://www.budapestopenaccessinitiative.org/read/>

Bracco, L., l'Hôte, A., Jeangirard, E., & Torny, D. (2022). *Extending the open monitoring of open science*. <https://hal.archives-ouvertes.fr/hal-03651518/>

Creaser, C., Fry, J., Greenwood, H., Oppenheim, C., Probets, S., Spezi, V., & White, S. (2010). Authors' awareness and attitudes toward open access repositories. *New Review of Academic Librarianship*, 16(S1), 145-161. <https://doi.org/10.1080/13614533.2010.518851>

Ginsparg, Paul. "arXiv at 20." *Nature* 476.7359 (2011): 145-147. <https://doi.org/10.1038/476145a>

Hicks, D., Wouters, P., Waltman, L., De Rijcke, S., & Rafols, I. (2015). Bibliometrics: the Leiden Manifesto for research metrics. *Nature*, 520(7548), 429-431. <https://doi.org/10.1038/520429a>

Jeangirard, E. (2019,). Monitoring Open Access at a national level: French case study. In *ELPUB 2019 23rd edition of the International Conference on Electronic Publishing* <https://hal.archives-ouvertes.fr/hal-02141819>

Khoo, S. (2019). Article processing charge hyperinflation and price insensitivity: An open access sequel to the serials crisis. *Liber Quarterly*, 29(1). <https://doi.org/10.18352/lq.10280>

Larivière, V., Haustein, S., & Mongeon, P. (2015). The oligopoly of academic publishers in the digital era. *PLoS one*, 10(6), e0127502. <https://doi.org/10.1371/journal.pone.0127502>

Larrieu, M., & Pain, D. (2021). Ouvrir l'accès aux dernières publications de son établissement avec HAL. *Bulletin des bibliothèques de France*. <https://hal.archives-ouvertes.fr/hal-03339754/>

Maddi, A. Measuring open access publications: a novel normalized open access indicator. *Scientometrics* 124, 379–398 (2020). <https://doi.org/10.1007/s11192-020-03470-0>

Maddi, A., Lardreau, E., & Sapinho, D. (2021). Open access in Europe: a national and regional comparison. *Scientometrics*, 126(4), 3131-3152. <https://doi.org/10.1007/s11192-021-03887-1>

- Magron, A. (2017). Utiliser les archives ouvertes pour valoriser ses travaux - l'exemple de HAL-SHS. In C. Schnedecker & A. Aleksandrova (Eds.), *Le doctorat en France : mode(s) d'emploi* (pp. 227–239). Frankfurt a.M., Peter Lang https://hal.archives-ouvertes.fr/sic_01697164
- Mahé, A. (2017). Les pratiques informationnelles des chercheurs dans l'enseignement supérieur et la recherche : regards sur la décennie 2000-2010. In G. Chartron, B. Epron, & A. Mahé (Eds.), *Pratiques documentaires numériques à l'université* (pp. 12–41). Villeurbanne, Presses de l'Enssib <https://doi.org/10.4000/books.pressesensib.1171>
- Mahé, A., & Prime-Claverie, C. (2017). Qui dépose quoi sur Hal-SHS? Pratiques de dépôts en libre accès en sciences humaines et sociales. *Revue française des sciences de l'information et de la communication*, (11). <https://doi.org/10.4000/rfsic.3315>
- Mongeon, P., & Paul-Hus, A. (2016). The journal coverage of Web of Science and Scopus: a comparative analysis. *Scientometrics*, 106(1), 213–228. <https://doi.org/10.1007/s11192-015-1765-5>
- OST. (2021). *La position scientifique de la France dans le monde et en Europe, 2005-2018*. Paris: Hcéres. <https://www.hceres.fr/fr/actualites/parmi-les-pays-de-recherche-intensive-la-france-conserve-un-profil-scientifique>
- Pölonen, J., Laakso, M., Guns, R., Kulczycki, E., & Sivertsen, G. (2020). Open access at the national level: A comprehensive analysis of publications by Finnish researchers. *Quantitative Science Studies*, 1(4), 1396–1428. https://doi.org/10.1162/qss_a_00084
- Prime-Claverie, C., & Mahé, A. (2013). Sites de dépôt en libre accès et formes de médiations : quelles évolutions ? In J. Boustany, É. Broudoux, & G. Chartron (Eds.), *La médiation numérique : renouvellement et diversification des pratiques*. Actes du colloque Document numérique et société, Zagreb 2013 (pp. 125–139). <https://doi.org/10.3917/dbu.chron.2013.01.0125>
- Roberts, R. J. (2001). PubMed Central: The GenBank of the published literature. *Proceedings of the National Academy of Sciences*, 98(2), 381–382. <https://doi.org/10.1073/pnas.98.2.381>
- Schöpfel, J. (2020). L'usage de la plateforme HAL par des unités de recherche. *I2D - Information, Données & Documents*, 3(3), 167–198. <https://doi.org/10.3917/i2d.203.0167>
- Schöpfel, J., Ferrant, C., André, F., & Fabre, R. (2016). Ready for the future? A survey on open access with scientists from the French National Research Center (CNRS). *Interlending & Document Supply*, 44(4), 141–149. <https://doi.org/10.1108/ILDS-06-2016-0023>
- Schöpfel, J., Prost, H., & Ndiaye, I. E. H. (2019). Going Green. Publishing Academic Grey Literature in Laboratory Collections on HAL. *GL21 International Conference on Grey Literature*, 22-23 October 2019, Hannover (Germany). <https://hal.archives-ouvertes.fr/hal-02300017>
- Severin A, Egger M, Eve MP and Hürlimann D. Discipline-specific open access publishing practices and barriers to change: an evidence-based review [version 2; peer review: 2 approved, 1 approved with reservations]. *F1000Research* 2020, 7:1925 <https://doi.org/10.12688/f1000research.17328.2>
- Shu, F., Julien, C. A., Zhang, L., Qiu, J., Zhang, J., & Larivière, V. (2019). Comparing journal and paper level classifications of science. *Journal of Informetrics*, 13(1), 202–225. <https://doi.org/10.1016/j.joi.2018.12.005>
- Singh, V. K., Singh, P., Karmakar, M., Leta, J., & Mayr, P. (2021). The journal coverage of Web of Science, Scopus and Dimensions: A comparative analysis. *Scientometrics*, 126(6), 5113–5142. <https://doi.org/10.1007/s11192-021-03948-5>

Tennant, J. (2018). *Complaint to the European Ombudsman about Elsevier and the Open Science Monitor*. Zenodo. <https://doi.org/10.5281/zenodo.2554199>

Tennant, J. P., Crane, H., Crick, T., Davila, J., Enkhbayar, A., Havemann, J., Kramer, B., Martin, R., Masuzzo, P., Nobes, A., Rice, C., Rivera-López, B., Ross-Hellauer, T., Sattler, S., Thacker, P. D., & Vanholsbeeck, M. (2019). Ten Hot Topics around Scholarly Publishing. *Publications*, 7(2), 34. <https://doi.org/10.3390/publications7020034>

Unesco (2021), *UNESCO Recommendation on Open Science*, reports. <https://unesdoc.unesco.org/ark:/48223/pf0000379949.locale=en>

Visser, M., van Eck, N. J., & Waltman, L. (2021). Large-scale comparison of bibliographic data sources: Scopus, Web of Science, Dimensions, Crossref, and Microsoft Academic. *Quantitative Science Studies*, 2(1), 20-41. https://doi.org/10.1162/qss_a_00112

Wojciechowska, A. (2017). Pratiques documentaires et pratiques d'auto-archivage des mathématiciens et informaticiens en France. In G. Chartron, B. Epron, & A. Mahé (Eds.), *Pratiques documentaires numériques à l'université* (pp. 90–120). Villeurbanne, Presses de l'Enssib <https://books.openedition.org/pressesenssib/1176>

Tous les liens ont été vérifiés le 2022-05-04

Annexe

Discipline (en)	Discipline (fr)
Biology	Biologie (fond.)
Chemistry	Chimie
Computer and information sciences	Informatique
Earth, ecology, energy and applied biology	Sciences de la terre, environnement, énergie, biologie appliquée
Engineering	Sciences de l'ingénieur
Humanities	Sciences humaines
Mathematics	Mathématiques
Medical research	Sciences médicales
Physical sciences, astronomy	Sciences physiques, astronomie
Social sciences	Sciences sociales

Tableau 2. Traduction des dix disciplines de la nomenclature disciplinaire utilisée.