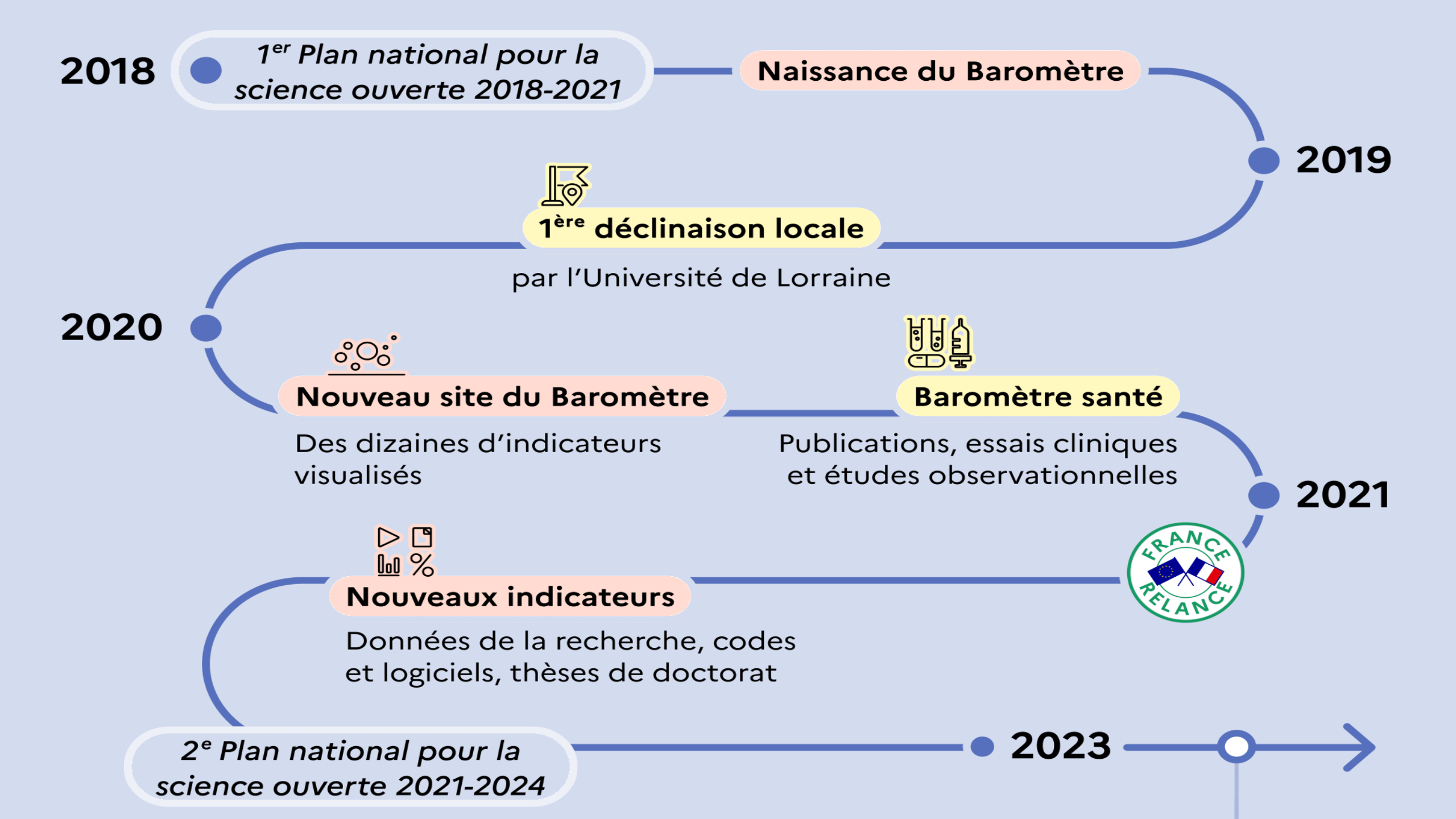
Baromètre de la science ouverte et utilisation d'OpenAlex pour la bibliométrie

Théophile Franzen, DRED/SCD

Baromètre de la science ouverte et utilisation d'OpenAlex pour la bibliométrie

Théophile Franzen, DRED/SCD

I. Baromètre de la science ouverte


2018
1er Plan national pour la science ouverte 2018-2021
Naissance du Baromètre
2019
1ère déclinaison locale
par l'Université de Lorraine
2020
Nouveau site du Baromètre
Des dizaines d'indicateurs visualisés
Baromètre santé
Publications, essais cliniques et études observationnelles
2021
FRANCE RELANCE
Nouveaux indicateurs
Données de la recherche, codes et logiciels, thèses de doctorat
2e Plan national pour la science ouverte 2021-2024
2023

Méthodologie

- Récupération d'une liste de DOI sur différentes bases (Web of Science, The Lens, Pubmed)
- Dédoublonnage des DOI via un script en code Python
- Récupération du code de la collection HAL de l'Université
- Récupération du code NNT pour les thèses de l'établissement

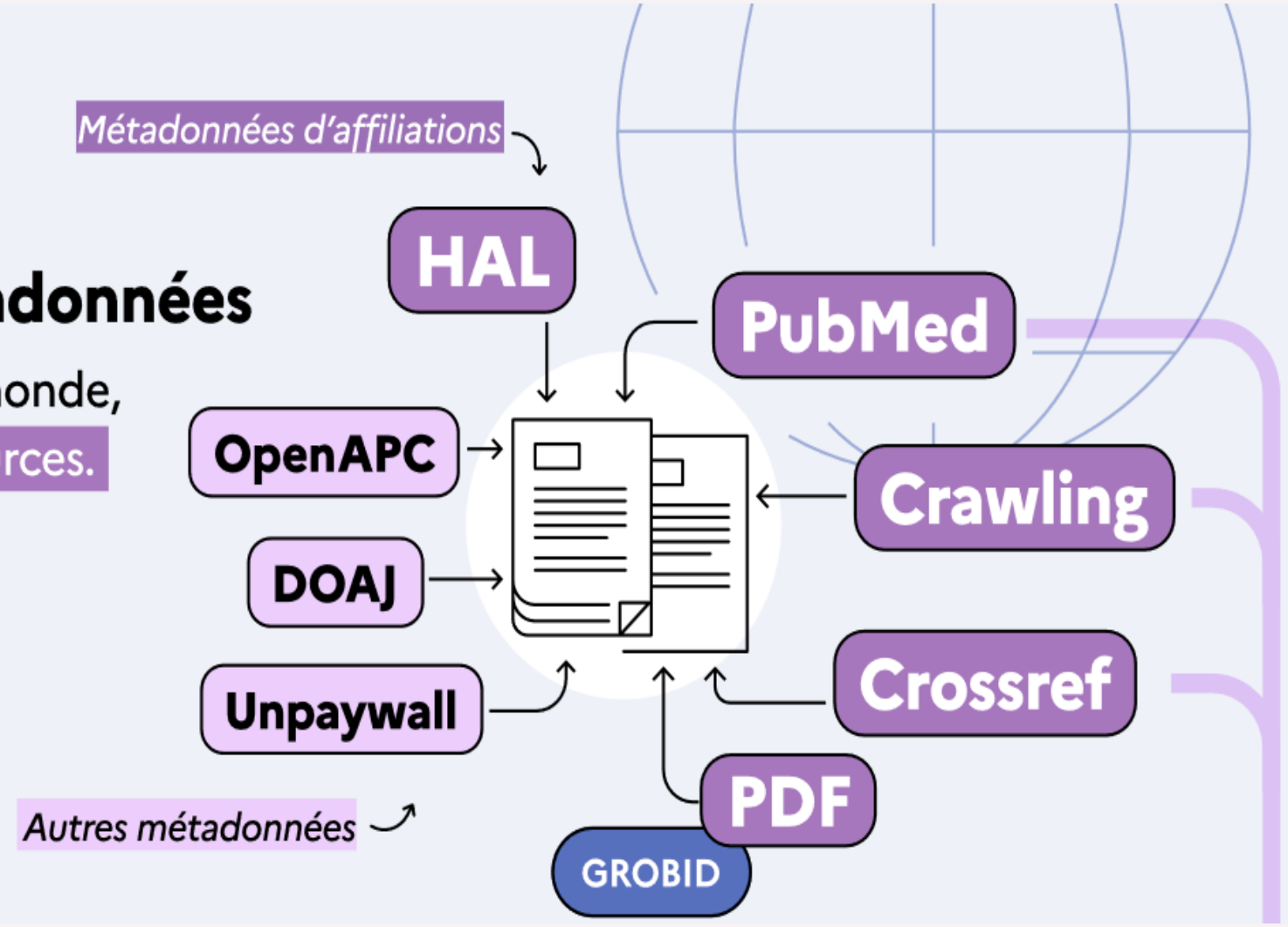
Toutes ces informations sont ensuite transmises à l'équipe du BSO dans un fichier Excel

UNIVERSITÉ LUMIÈRE LYON 2

#1 Collecter...

un maximum de métadonnées

Pour chaque publication du monde,
consolidation de plusieurs sources.



#2 Détecter

Université Lumière Lyon 2

#3 Enrichir...

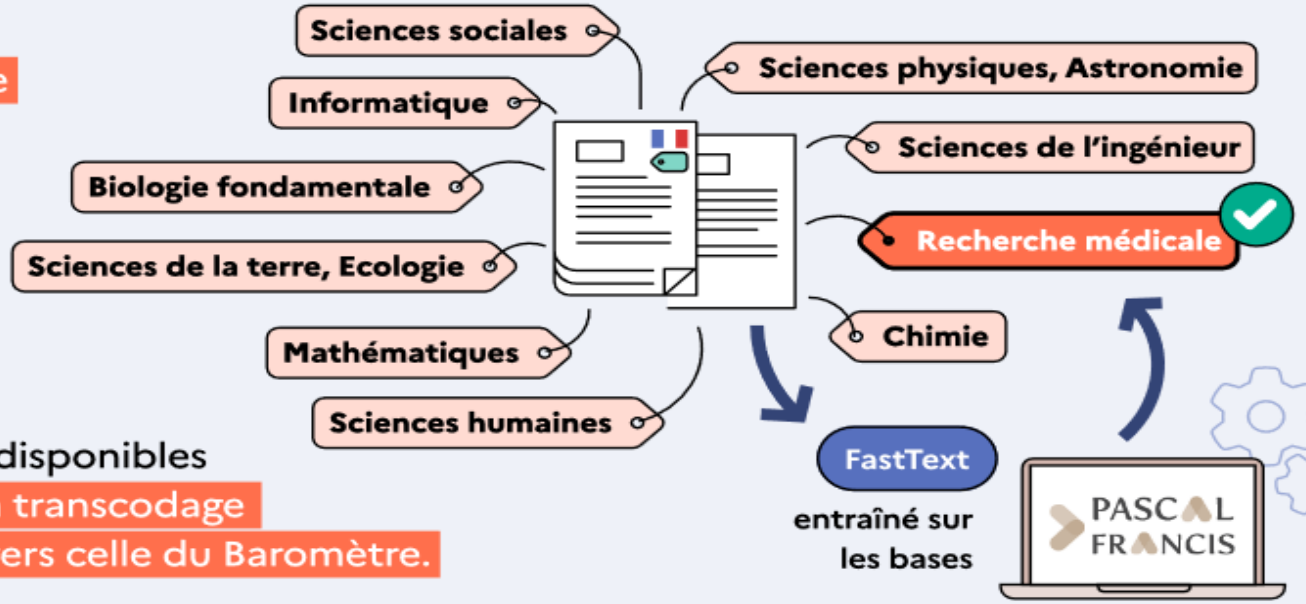
... le statut d'ouverture

Pour les DOI crossref :
les informations viennent d'**Unpaywall**

Pour les publications dans HAL (sans DOI) :
les informations viennent de **HAL**

... la classification disciplinaire

Grâce à un algorithme de machine learning de classification automatique (fastText) basé sur le titre, le résumé et le nom de la revue.



Si des métadonnées sont disponibles dans HAL, on effectue un transcodage de la nomenclature HAL vers celle du Baromètre.

UNIVERSITÉ LUMIÈRE LYON 2

Statut d'ouverture des publications

« Le baromètre prend en compte les publications ayant un DOI Crossref et/ou un identifiant HAL. Un identifiant unique est nécessaire pour éviter les doublons. De plus, pour savoir si une publication est en accès ouvert, un outil de "découverte" d'accès ouvert est nécessaire. Unpaywall fournit cet outil pour les DOI Crossref, et HAL fournit des métadonnées concernant l'accès ouvert pour les publications avec un identifiant HAL. A notre connaissance, il n'existe pas d'autre outil de découverte d'accès ouvert applicable à d'autres identifiants. Pour cette raison, les DOI non Crossref ne sont pas pris en compte, comme par exemple les DOI Datacite »

Source : FAQ du site du Baromètre Science Ouverte

Résultats essentiels Lyon 2

- 68,2% des publications de 2023 sont en accès ouvert, une progression conséquente depuis l'année dernière où l'ouverture n'était que de 61,5% (66,8% au niveau national).

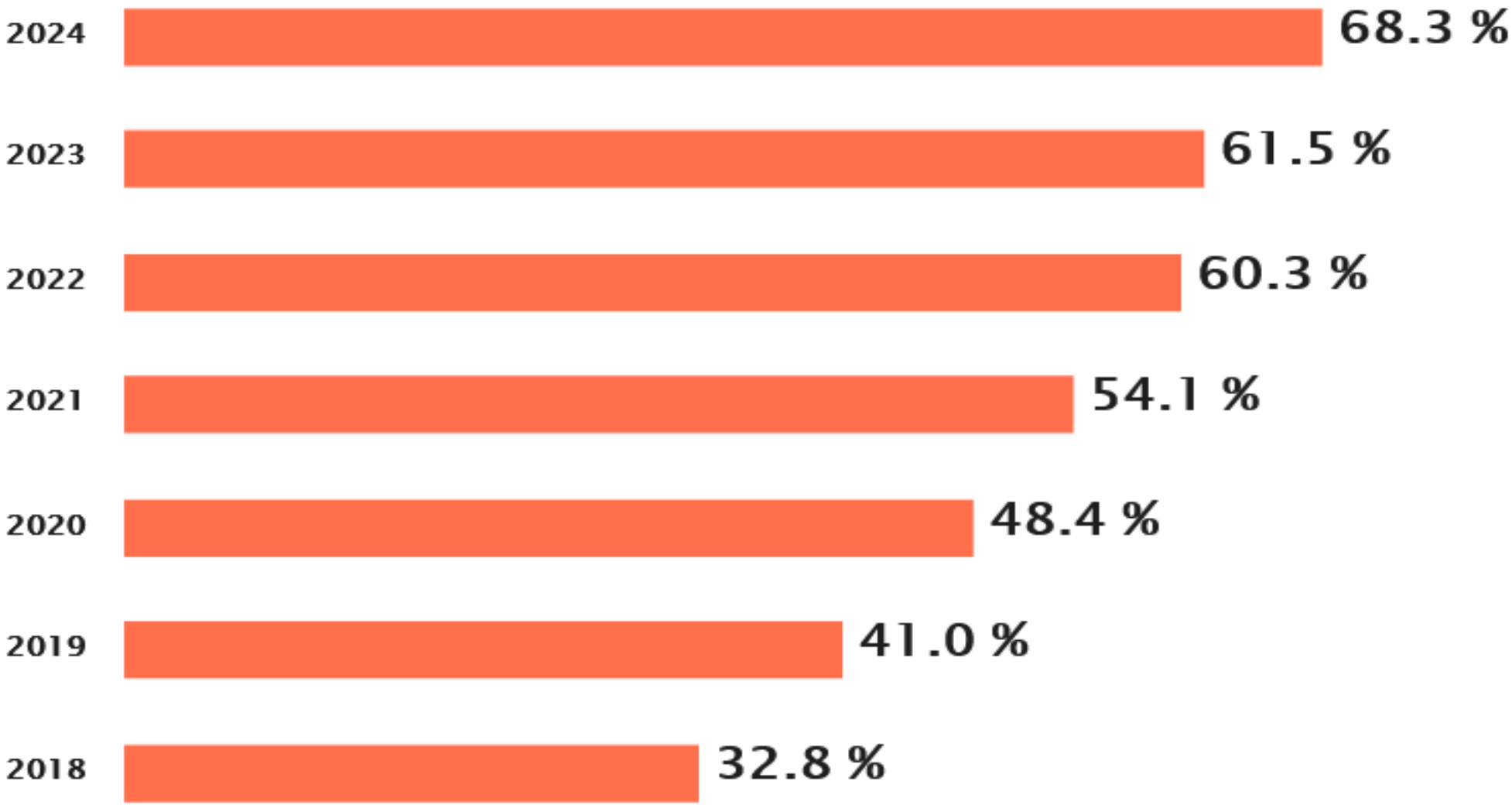
- S'il existe de fortes disparités entre les disciplines, on observe une augmentation de l'accès ouvert de manière générale

- On observe une hausse significative des dépôts dans des archives ouvertes et notamment dans HAL

- Les éditeurs choisis par les chercheurs mettent de plus en plus les publications à disposition en accès ouvert avec un passage de 44% à 53% de publication en OA

UNIVERSITÉ LUMIÈRE LYON 2

Université Lumière Lyon 2 (ULL2) : Taux d'accès ouvert des publications scientifiques de l'ULL2, avec un DOI Crossref, parues durant l'année précédente par année d'observation

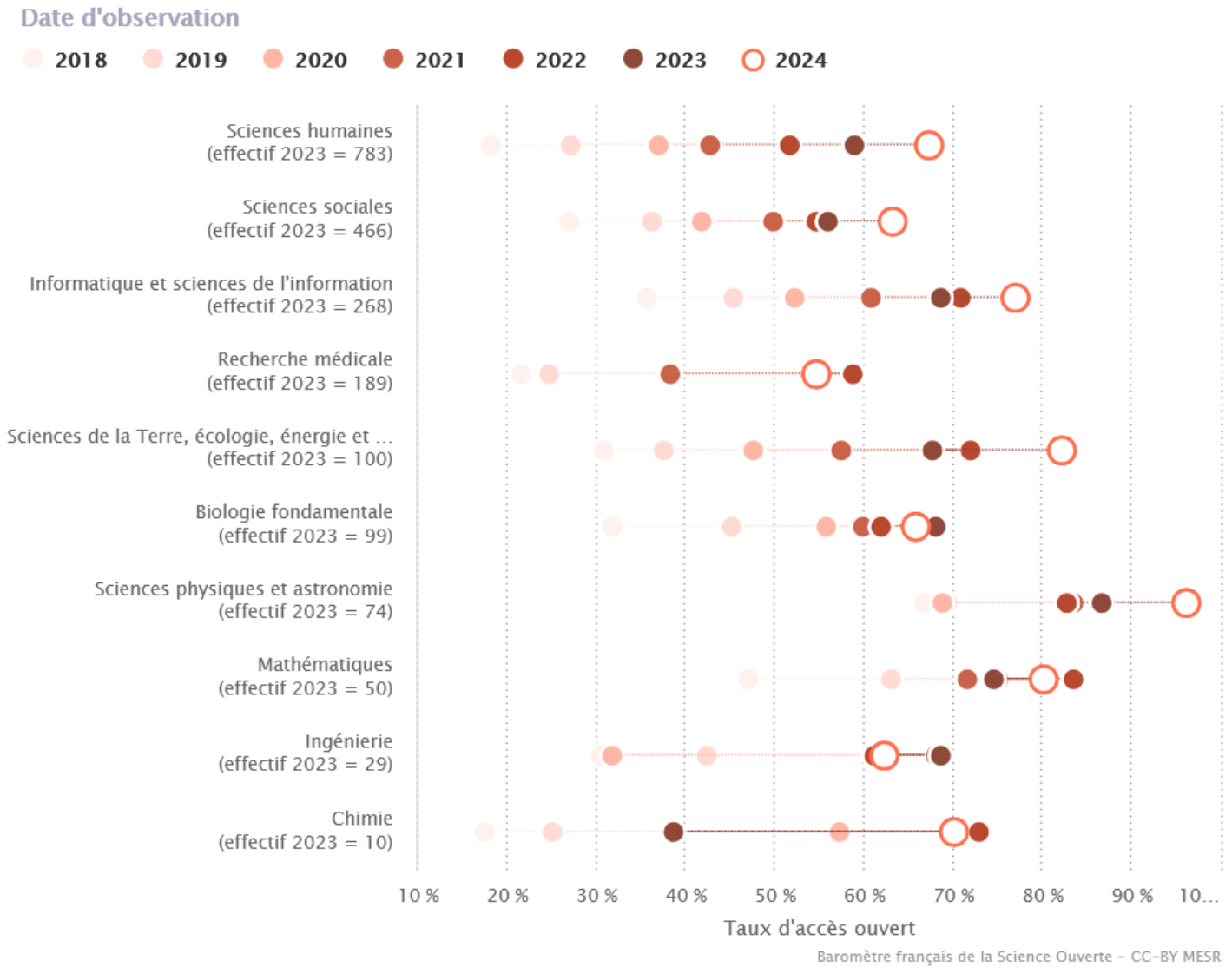


Baromètre français de la Science Ouverte – CC-BY MESR

l'ULL2, avec un DOI Crossref, en accès ouvert pour chaque discipline par année d'observation

Trier par :

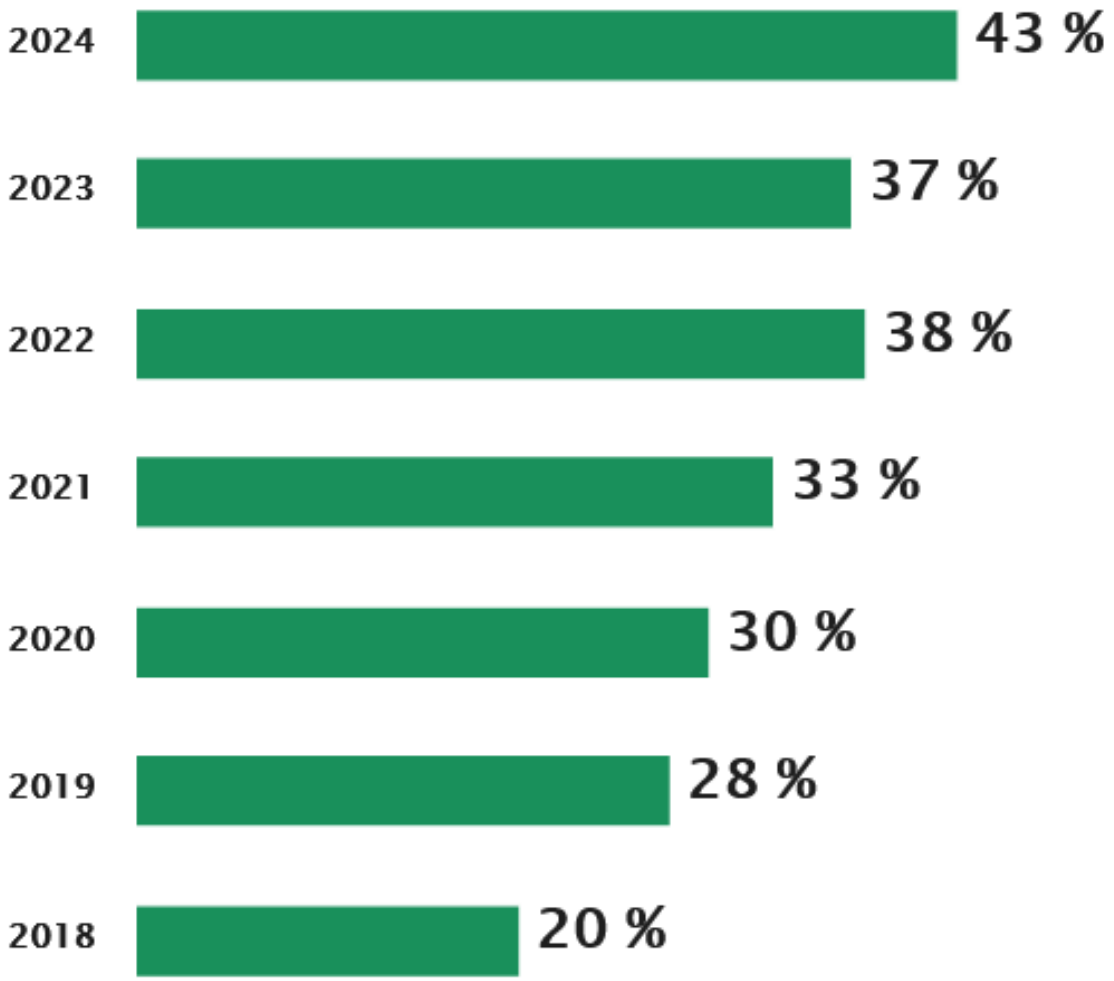
◯ Plus fort taux d'accès ouvert ◯ Plus forte progression ◉ Plus grand effectif en 2023



Baromètre français de la Science Ouverte – CC-BY MESR

Université Lumière Lyon 2 (ULL2) : Taux de publications scientifiques de l'ULL2, avec un DOI Crossref, ouvertes et hébergées sur une archive ouverte par année d'observation

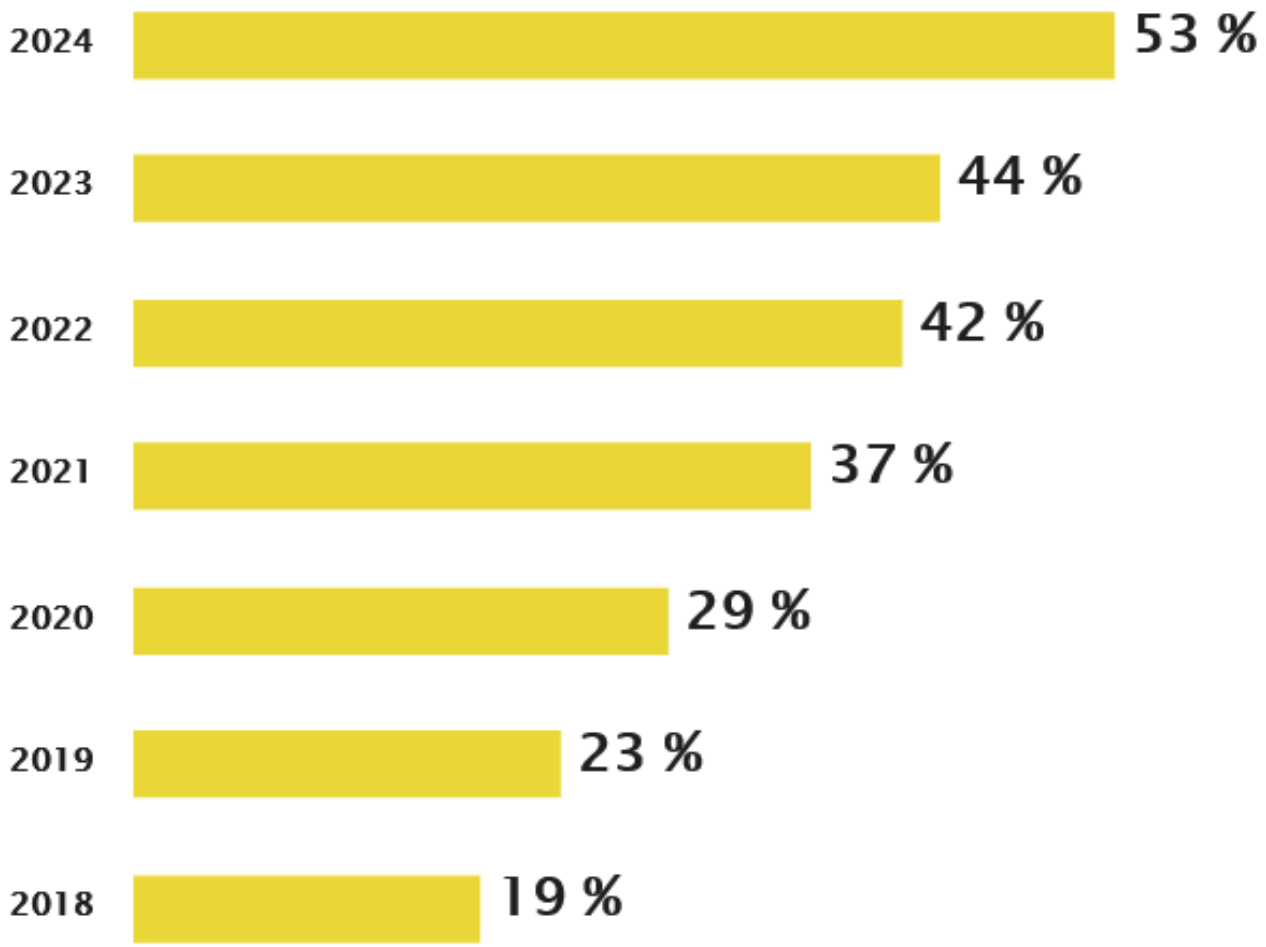
Sélectionner une archive ouverte Toutes les archives ouvertes



Baromètre français de la Science Ouverte – CC-BY MESR

Université Lumière Lyon 2 (ULL2) : Part des publications scientifiques de l'ULL2, avec un DOI Crossref, mises à disposition en accès ouvert par leur éditeur, par année d'observation, pour les publications parues durant l'année précédente

Sélectionner un éditeur Tous les éditeurs



Baromètre français de la Science Ouverte – CC-BY MESR

Conclusion et remarque

- La politique de Science Ouverte de l'Université porte ses fruits.
- Les deux écueils principaux du BSO sont l'utilisation seulement des DOI Crossref et le regroupement des SHS en seulement deux disciplines.
- Le BSO est maintenant déclinable à la demande pour chaque laboratoire.
- L'équipe du BSO travaille également pour inclure le statut d'ouverture des données.
- L'équipe du BSO réfléchit à utiliser également OpenAlex l'an prochain, ce qui pourrait améliorer la détection de DOI de l'Université.

UNIVERSITÉ LUMIÈRE LYON 2

II. Utilisation d’OpenAlex pour la bibliométrie à l’Université

OpenAlex c'est quoi ?

- Une base de données bibliographiques et bibliométriques fondée en 2021.
- Créée et maintenue par OurResearch une organisation à but non lucratif qui fournit aux chercheurs et aux personnels des bibliothèques des outils (Unpaywall ou Unsub).
- Totalement ouvert et gratuit,
- Dispose d'une interface utilisateur et d'une API.

UNIVERSITÉ LUMIÈRE LYON 2

Objectifs de cette étude

- Argumenter le choix d'OpenAlex comme base utilisable pour la bibliométrie à l'Université.
- Estimer la quantité de publications affiliées à l'Université sur OpenAlex, notamment par rapport à HAL.
- Quantifier le travail d'amélioration des affiliations dans OpenAlex.
- Présenter les usages d'OpenAlex pour la bibliométrie, en comparaison avec HAL.

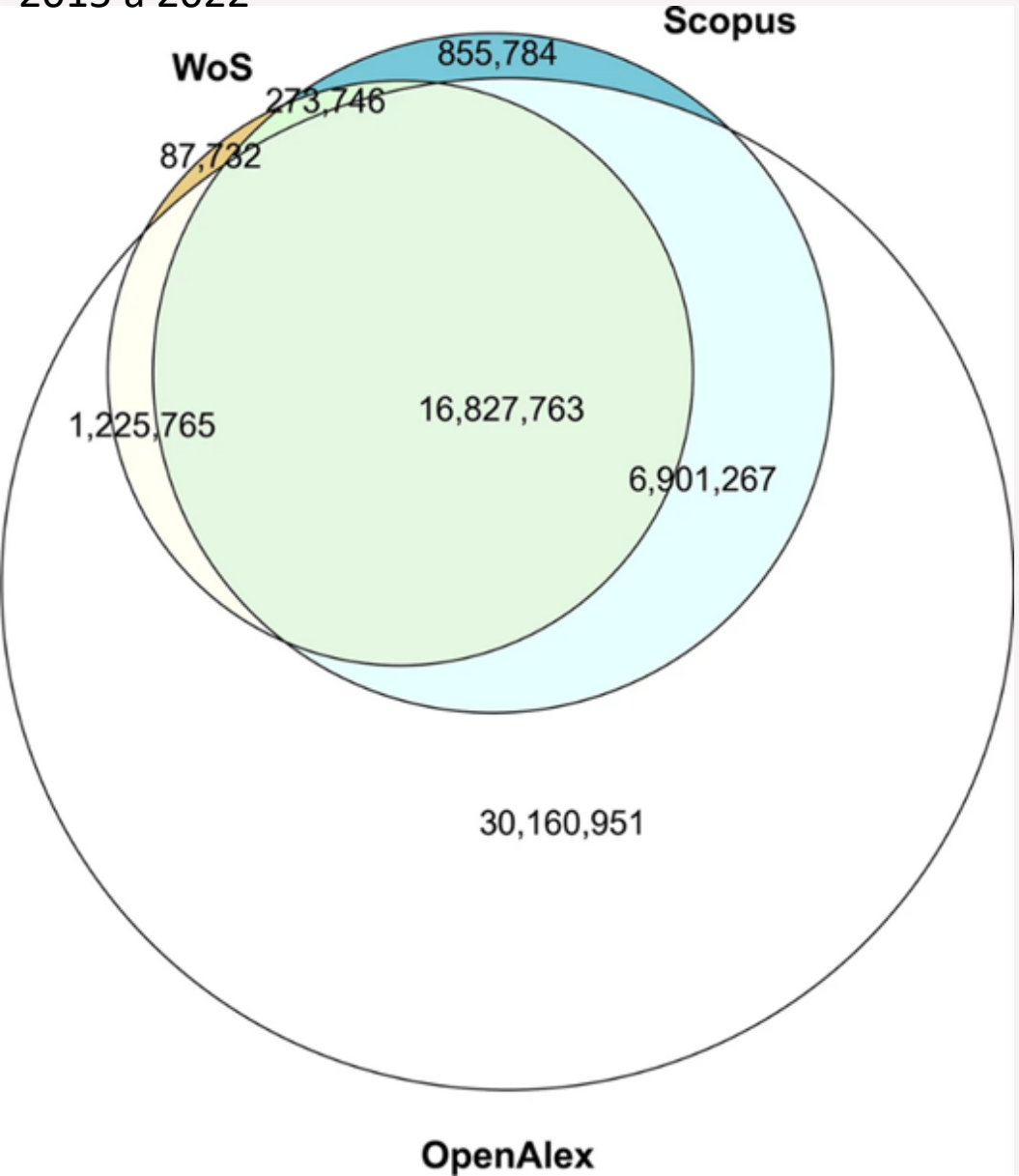
UNIVERSITÉ LUMIÈRE LYON 2

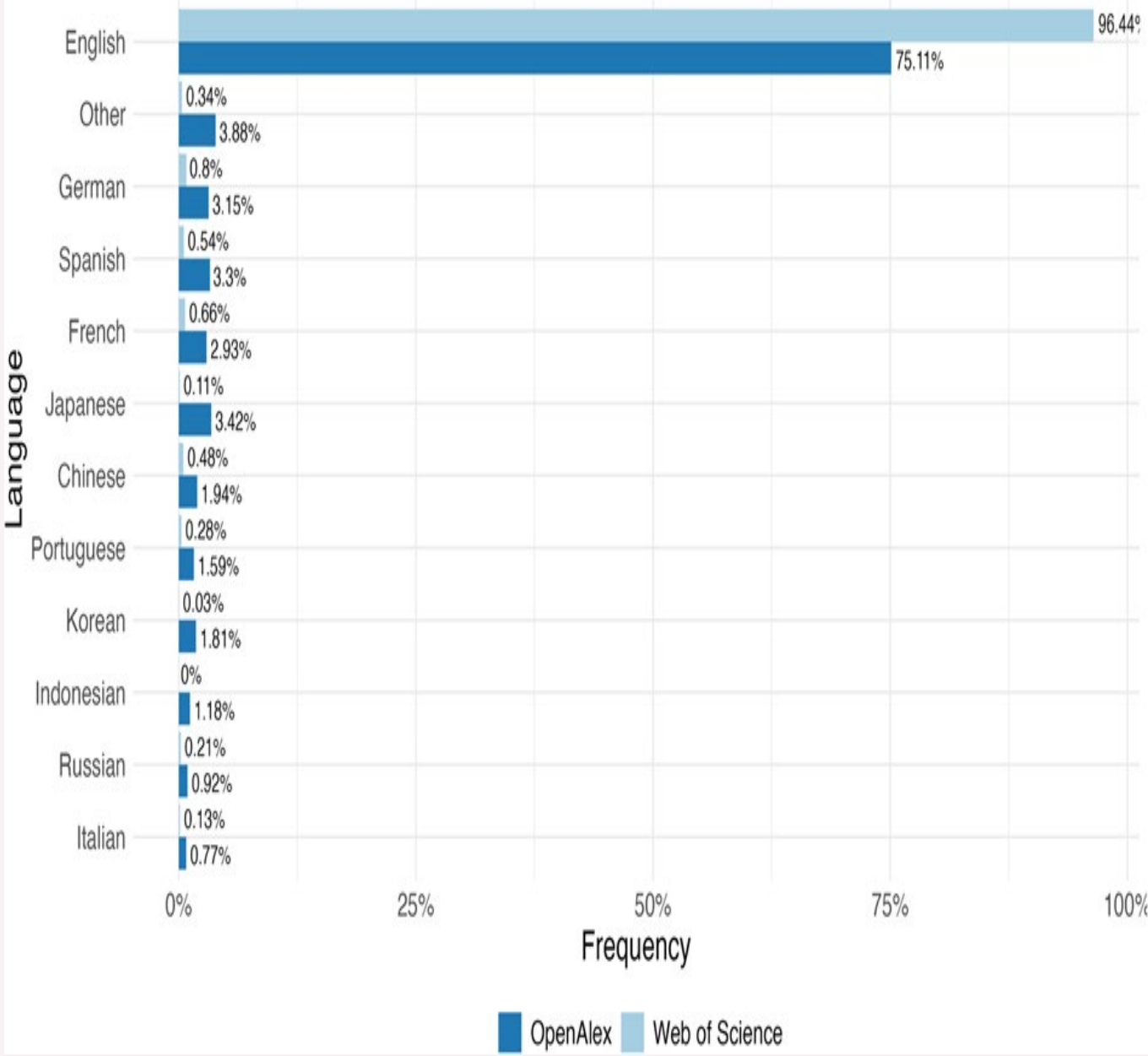
Pourquoi choisir OpenAlex ?

- Gratuité d'utilisation complète à la fois pour la consultation et l'extraction des données
- Grande représentativité des publications en SHS et des publications issues de laboratoires français, car OpenAlex référence plus de documents dans des langues non-anglophones
- Climat en France propice à OpenAlex : partenariat avec le ministère en février 2024, création du *works magnet* par l'équipe du BSO permettant la correction des affiliations.
- Correspond au critère d'utilisation des bases de données bibliographiques et bibliométriques de CoARA.

UNIVERSITÉ LUMIÈRE LYON 2

Diagramme de Venn du nombre de DOI unique (sans déduplication) dans les bases de données Scopus, WoS et OpenAlex pour des publications de 2015 à 2022





Pourcentage de publications par langue déclarées par le WoS et Open Alex pour des parutions entre 2000 et 2020

Méthode et outils

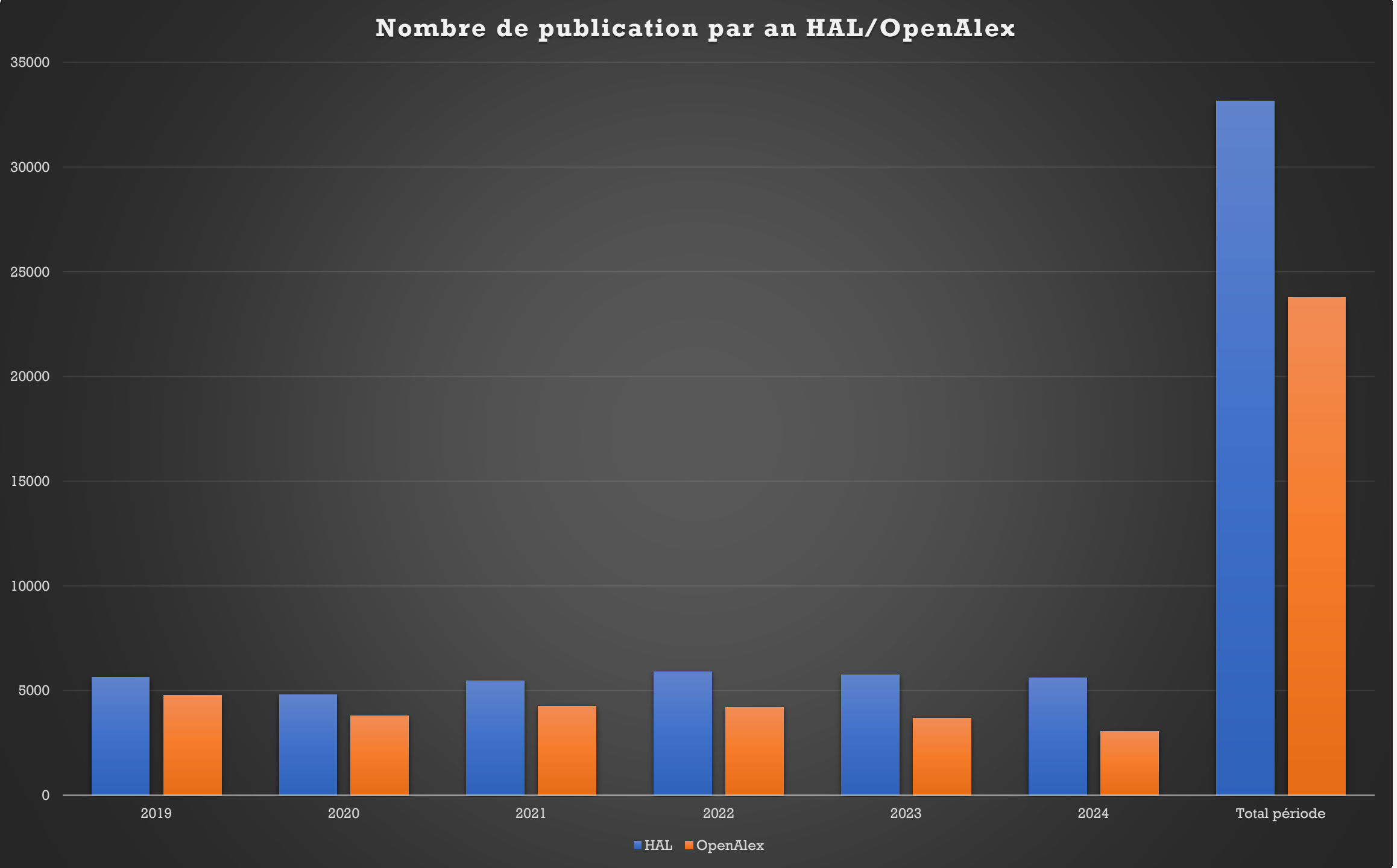
- Interrogation des API de HAL et OpenAlex via des scripts en code Python pour l'extraction des données bibliographiques et bibliométriques.
- Extractions des données en .csv manipulé ensuite avec Excel.
- Comparaison des deux bases via un outil développé par l'Université de Nantes : C2LabHAL.

UNIVERSITÉ LUMIÈRE LYON 2

Alternate names Université Lyon II, Lyon II, Lumière University Lyon 2, Université Lumière Lyon II

Parent institution Université de Lyon

Child institutions Archéologie et Archéométrie, Archéorient, Center for Research in Psychopathology and Clinical Psychology, Centre Max Weber, Centre de Recherche en Linguistique Appliquée, Centre de Recherches Critiques sur le Droit, Conception de l'Action en Situation, Decision & Information Sciences for Production Systems, Développement Individu Processus Handicap Education, Education, Cultures, Politiques, Entrepôts, Représentation et Ingénierie des Connaissances, Environnement, ville, société, Equipe de recherche de Lyon en Sciences de l'Information et de la Communication, Fédération Informatique de Lyon, Groupe d'Analyse et de Théorie Economique Lyon St Etienne, Groupe de Recherche en Psychologie Sociale, Histoire et Sources des Mondes Antiques, Histoire, Archéologie, Littératures des Mondes Chrétiens et Musulmans Médiévaux, Institut Rhônalpin des Systèmes Complexes, Institut d'Asie Orientale, Institut d'Histoire des Représentations et des Idées dans les Modernités, Institut de Recherche sur l'Architecture Antique, Institut supérieur d'étude des religions et de la laïcité, Interactions, Corpus, Apprentissages, Représentations, Laboratoire Aménagement Économie Transports, Laboratoire Dynamique du Langage, Laboratoire Passages XX_XXI, Laboratoire d'Informatique en Images et Systèmes d'Information, Laboratoire d'anthropologie des enjeux contemporains, Laboratoire d'études rurales, Laboratoire de Recherche Historique Rhône-Alpes, Laboratoire d'Étude des Mécanismes Cognitifs, Lettres et Civilisations Etrangères, Maison de l'Orient et de la Méditerranée Jean Pouilloux, Maison des Sciences de l'Homme Lyon St-Étienne, Persée, Sciences et Technologies des Cultures et Sociétés Numériques, Transversales, Triangle : Action, Discours, Pensée politique et économique, Typologie et Universaux Linguistiques (**less**)


Nombre de publication par an HAL/OpenAlex
35000
30000
25000
20000
15000
10000
5000
0
2019
2020
2021
2022
2023
2024
Total période
HAL
OpenAlex

Affiliation dans OpenAlex

- Nombreux problèmes d'affiliations avec des laboratoires qui n'ont pratiquement aucune publication, ou qui en perdent énormément entre HAL et OpenAlex.

- A l'échelle de l'Université, via C2LabHAL, 67% des publications affiliées à Lyon 2 dans OpenAlex sont également dans HAL.

- Les 33% restants ne sont pas forcément de mauvaises affiliations, ce sont juste des publications qui ne sont pas dans HAL.

Université Lumière Lyon 2

Travail sur les affiliations

- Avec 67% des affiliations correctes à l'échelle de l'Université le travail de correction des affiliations n'est pas titanesque.
- Les affiliations sont facilement corrigeables avec *works magnet,* qui identifie les publications grâce au RoR et qui est directement en contact avec OpenAlex.
- OpenAlex est actuellement très réactif sur la correction des affiliations.

UNIVERSITÉ LUMIÈRE LYON 2

HAL, OpenAlex et la bibliométrie

- HAL : peu adapté aux études bibliométriques poussées.
- OpenAlex met à disposition de nombreux outils et métadonnées pour la bibliométrie sur son interface utilisateur.
- L'API de OpenAlex permet d'avoir accès à plus d'outils de recherche bibliographique et bibliométrique.
- Des outils externes ont été développés pour permettre l'exploitation bibliométrique extraite d'OpenAlex.

UNIVERSITÉ LUMIÈRE LYON 2

Works

Pyannote.Audio: Neural Building Blocks for Speaker Diarization
2020 · Hervé Bredin, Ruiqing Yin, et al. · *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*
Cited by 257 PDF

Human sound systems are shaped by post-Neolithic changes in bite configuration
2019 · Damián E. Blasi, Steven Moran, et al. · *Science*
Cited by 194

Different languages, similar encoding efficiency: Comparable information rates across the human communicative niche
2019 · Christophe Coupé, Yoon Mi Oh, et al. · *Science Advances*
Cited by 181 PDF

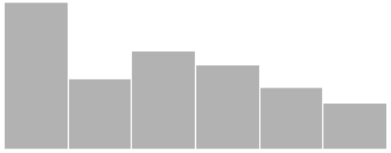
Proceedings Language Resources and Evaluation Conference (LREC) 2020
2020 · Nicoletta Calzolari, Frédéric Bechet, et al.
Cited by 123 PDF

Difference and repetition in language shift to a creole. The expression of emotions
2020 · Maïa Ponsonnet
Cited by 94 PDF

Stats

1 179 results

year



open access

44,4%
523

topic

☐	French Language Learning Methods	97
☐	Discourse Analysis and Argumentation Studies	89
☐	Linguistic and Sociocultural Studies	42
☐	Phonetics and Phonology Research	40
☐	Categorization, perception, and language	38

More...

institution

☐	Centre National de la Recherche Scientifique	1 179
☐	Université Lumière Lyon 2	1 179
☐	Université de Lyon	1 179
☑	Laboratoire Dynamique du Langage	1 179
☐	Normandie Université	266

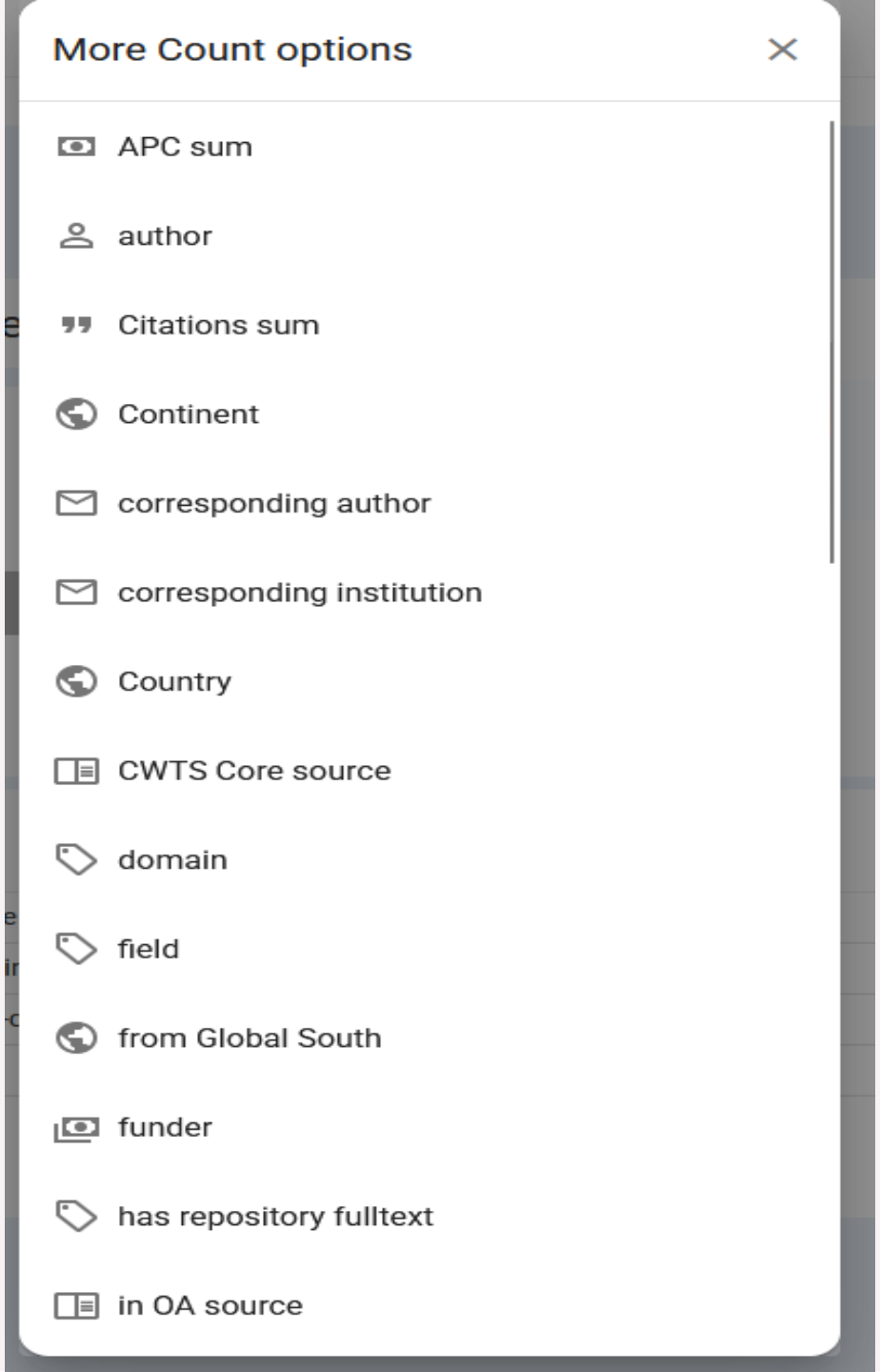
More...

type

☐	article	568
☐	preprint	435
☐	book-chapter	94
☐	book	23
☐	other	17

More...

UNIVERSITÉ LUMIÈRE LYON 2



UNIVERSITÉ LUMIÈRE LYON 2

Les outils externes

- Le *works magnet* (développé par le ministère)
- Deux librairies : Python et R, ayant des fonctions spécifiques pour l'utilisation des données d'OpenAlex.
- VOSViewer : un outil de création de graphes directement interrogeable via une requête API d'OpenAlex
- Lodex : un outil développé par l'INIST, avec un plugin spécialement destiné à l'étude des données extraites d'OpenAlex.

UNIVERSITÉ LUMIÈRE LYON 2

Conclusion

- Malgré des soucis d'affiliation, OpenAlex permet déjà de faire des études bibliométriques simples.
- Avant tout travail pour des études bibliométriques plus poussées, il faudra cependant corriger les affiliations
- L'utilisation d'OpenAlex nous permettrait de développer des indicateurs plus en lien avec la Science Ouverte et plus vertueux conformément à l'engagement pris par l'Université dans CoARA.

UNIVERSITÉ LUMIÈRE LYON 2