RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

Normandie Université

Atelier créer un baromètre science ouverte pour son unité de recherche

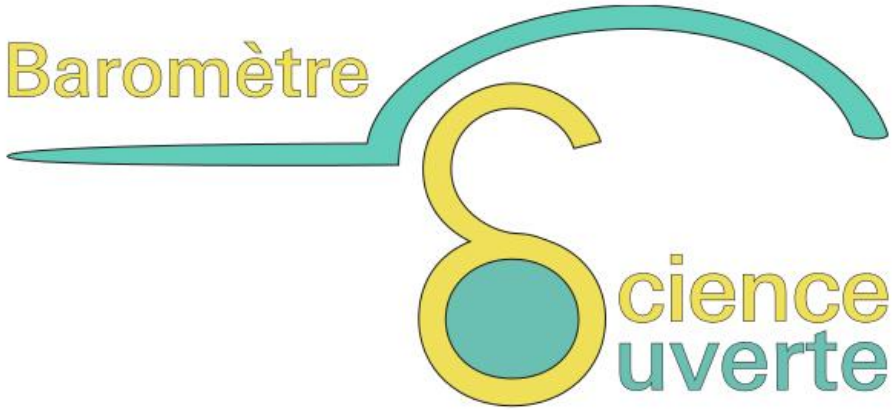
OAW 2023 – journée des référents HAL Normands

16 novembre 2023

Programme

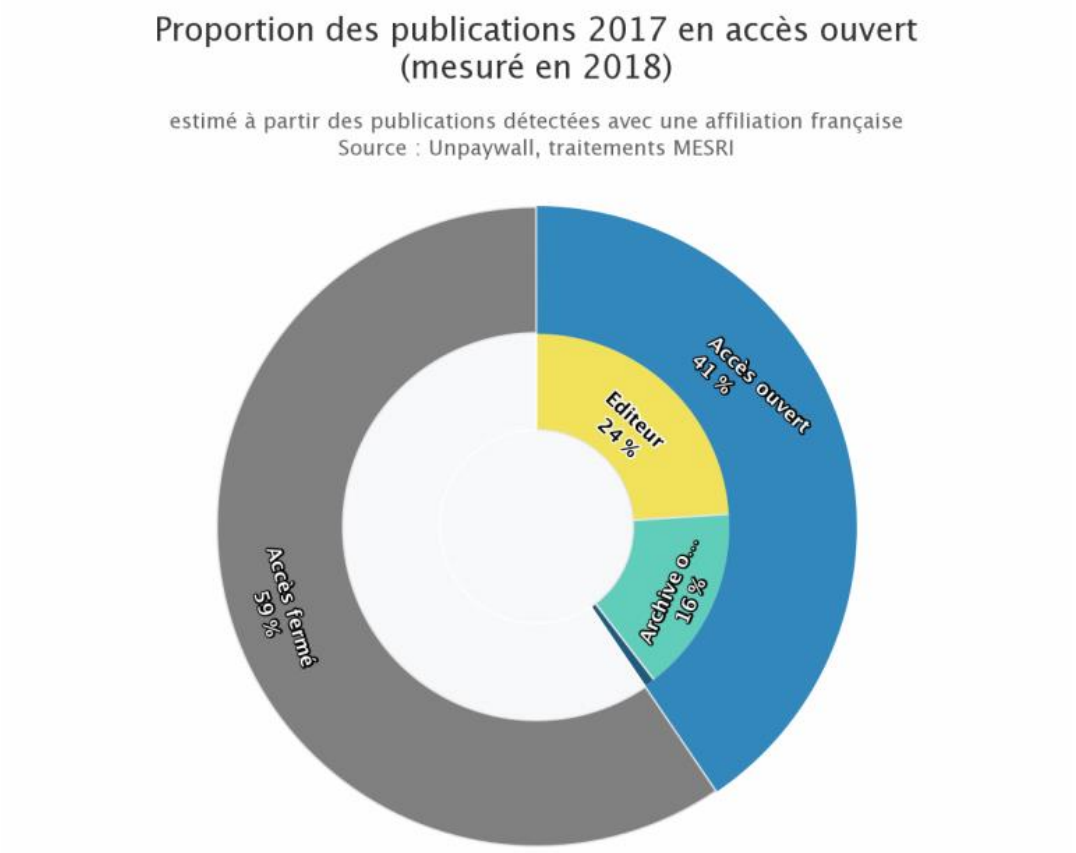
- Présentation du baromètre pour la science ouverte : historique et fonctionnement
- Les sources de données
- Finalisation de la liste de DOI avec le notebook jupiter

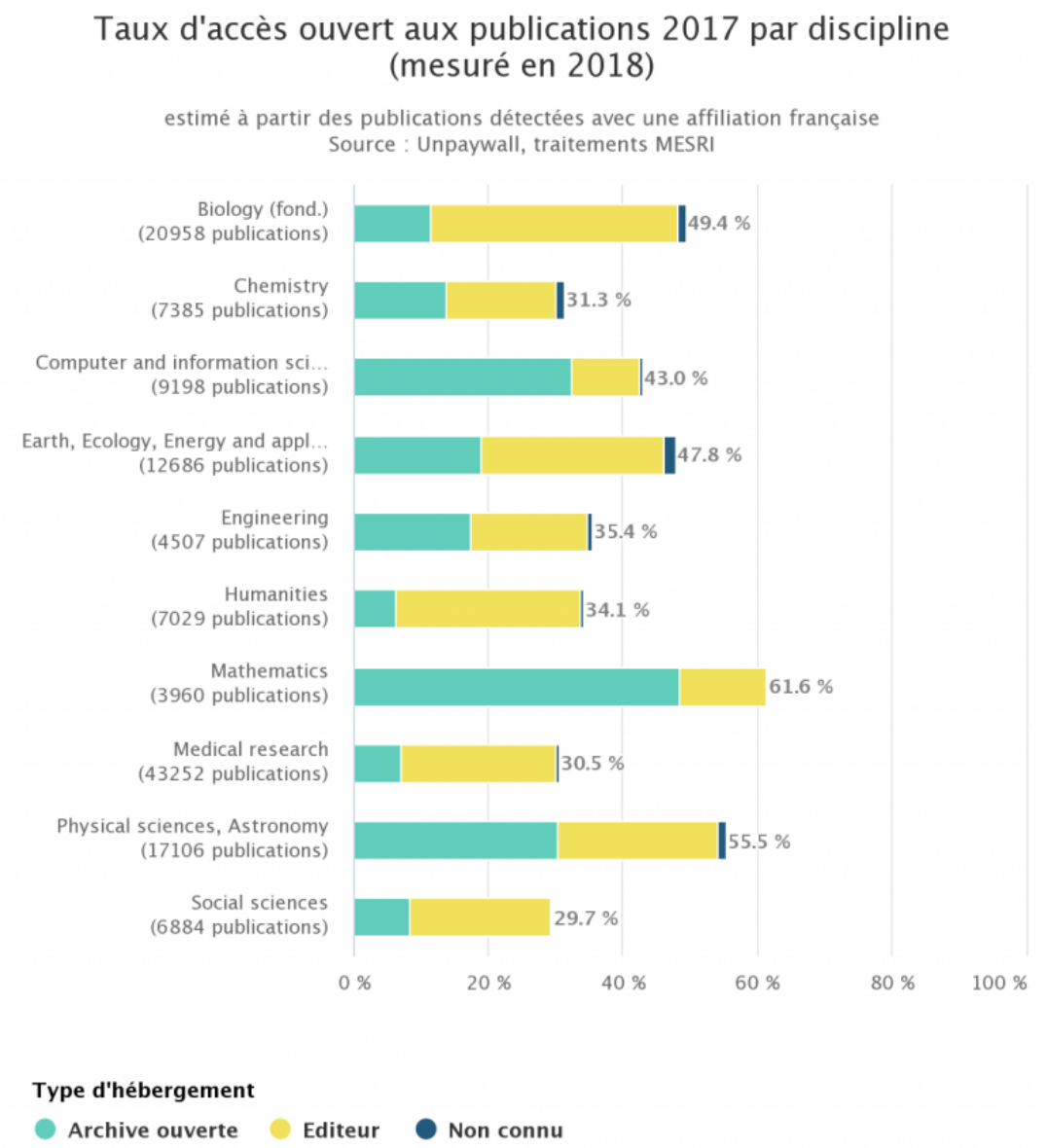
Présentation du BSO



- **Première édition en 2019**
- **Objectif :** mesurer l'accès ouvert des publications françaises et sa progression par type d'accès (voie dorée et voie verte) et par discipline.
- Réalisé par le MESRI. Indicateur important dans le cadre du **Plan national pour la science ouverte**, dont le pilier n° 1 est l'ouverture des publications.
- **Fonctionnement :** liste de publications possédant un DOI dont l'un des auteurs à une affiliation française. L'information sur l'ouverture provient d'Unpaywall.
- Publication du nouveau baromètre une fois par an

Présentation du BSO





https://www.ouvrirlascience.fr/41-de-publications-francaises-en-acces-ouvert/

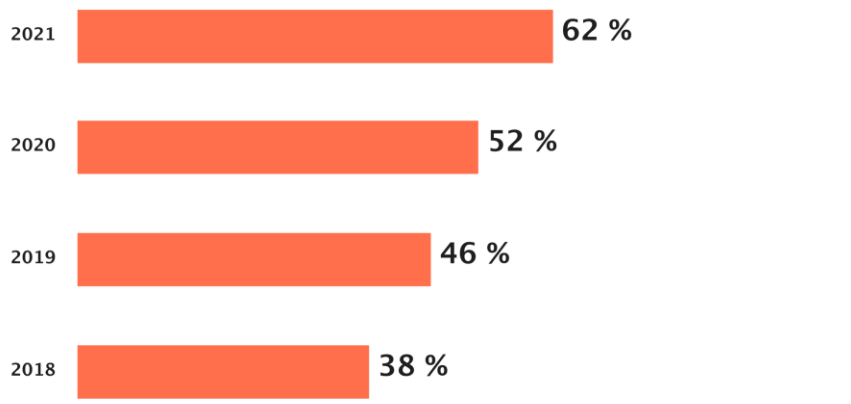
Normandie Université

Nouvelle version du BSO en 2021

- Deux baromètres : général et spécifique Santé (suite au Covid)
- Données mises à jour tous les trimestres
- De nouveaux indicateurs (modèle économique des revues) et nouveaux graphiques (voie privilégiée par discipline)
- Faciliter la mise en place de « baromètres locaux » > expérimentation

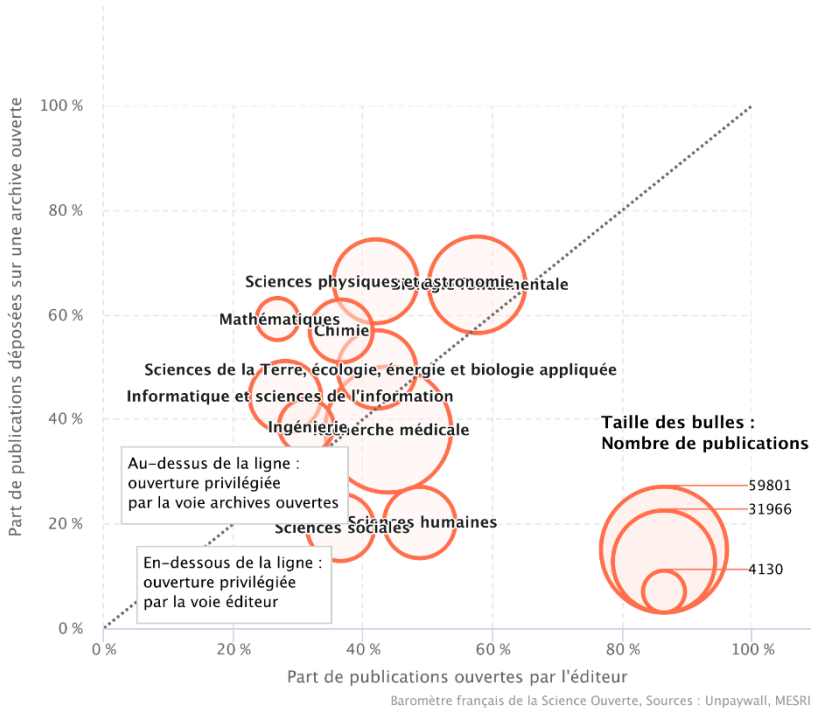
Nouvelle version du BSO en 2021

Taux d'accès ouvert des publications scientifiques françaises parues durant l'année précédente par date d'observation



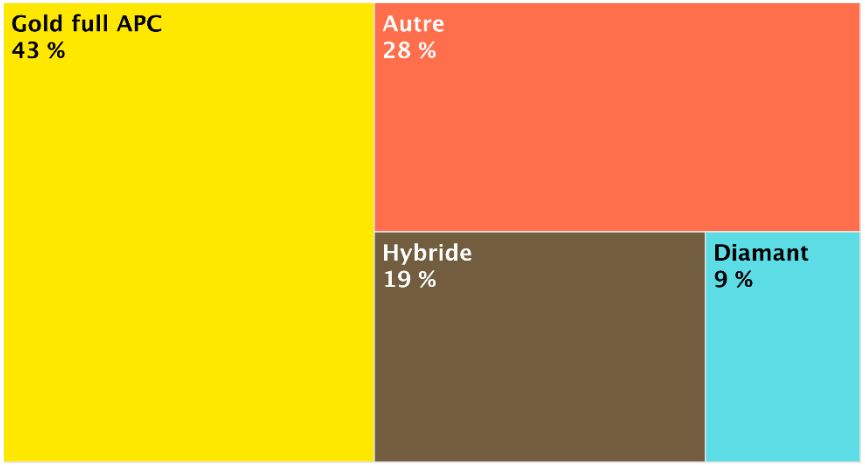
Baromètre français de la Science Ouverte, Sources : Unpaywall, MESRI

Positionnement des disciplines en fonction des voies privilégiées pour l'ouverture de leurs publications (publications de 2020)



Baromètre français de la Science Ouverte, Sources : Unpaywall, MESRI

Répartition des modèles économiques pour les articles parus en 2020 et diffusés en accès ouvert par leur éditeur



Baromètre français de la Science Ouverte, Sources : Unpaywall, HAL, MESR,

https://barometredelascienceouverte.esr.gouv.fr/publications/editeurs?id=publishers.type-ouverture

Normandie Université

Nouvelle version du BSO en 2023

› De nouveaux indicateurs sont proposés en 2023 :

- thèses de doctorat
- données de la recherche
- codes et logiciels

- pour les publications, option possible sur l'intégration des dépôts dans HAL

Normandie Université

Baromètre : déclinaisons locales

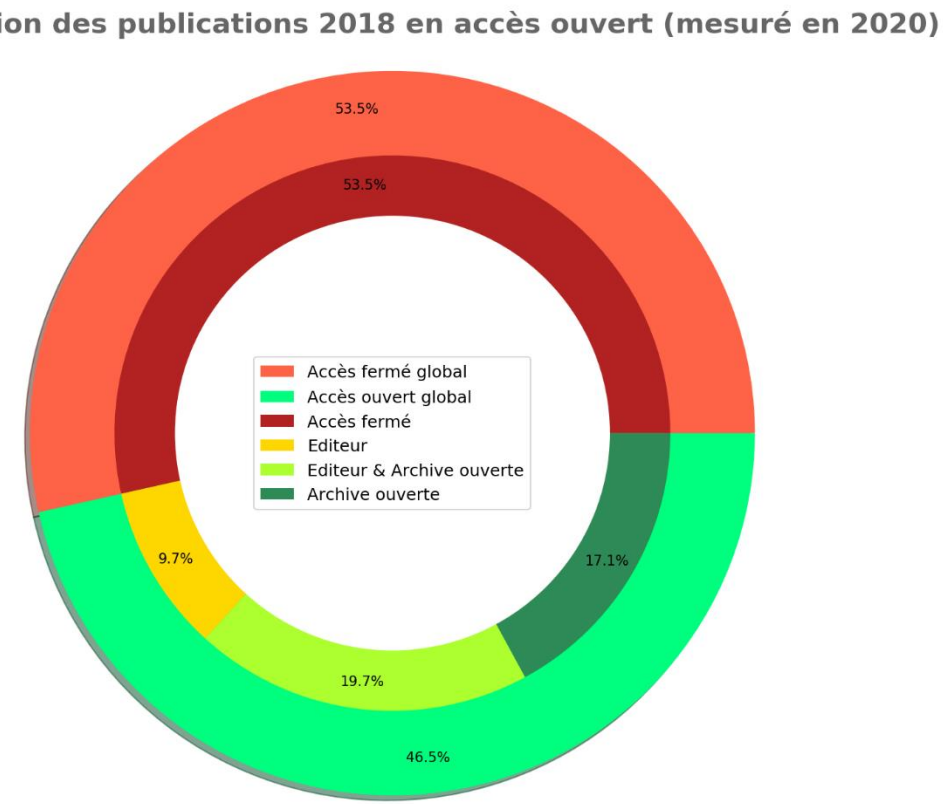
- Le code du baromètre national étant librement accessible, il a été repris en 2020 par l'Université de Lorraine pour proposer une déclinaison des indicateurs à l'échelle de cet établissement.

https://scienceouverte.univ-lorraine.fr/barometre-lorrain-de-la-science-ouverte/

- Le code lorrain a été conçu pour être réutilisable par d'autres établissements ou structure de recherche en France (notebook Jupiter)

Baromètre : déclinaisons locales

› Une première expérimentation du baromètre normand en 2020 :

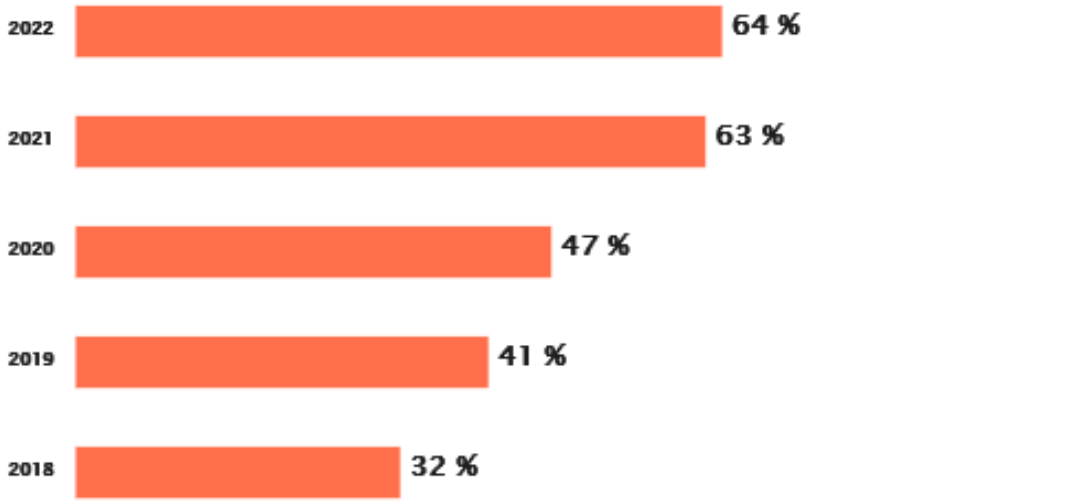


Normandie Université

Baromètre : déclinaisons locales

› Version validée et valorisée pour l'édition 2022 :

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



Baromètre français de la Science Ouverte – CC-BY MESR

Commentaire

Ce graphique présente, pour chaque date d'observation depuis 2018, le taux d'accès ouvert des publications scientifiques de NU, avec un DOI Crossref, parues durant l'année précédente.
Ainsi, 63 % des publications scientifiques de NU, avec un DOI Crossref, publiées en 2020 étaient en accès ouvert en 2021 (date d'observation). Et 64 % des publications scientifiques de NU, avec un DOI Crossref, publiées en 2021 étaient ouvertes en 2022. Le taux d'accès a donc augmenté de 1 point(s) en une année seulement.

Normandie Université

Baromètre local : fonctionnement

- La première étape : **constituer une liste de DOI.**

  Pour cela :

  - Définir un périmètre : ComUE, établissement, laboratoire
  - Définir un périmètre de temps : à partir de quand ? Exploiter des sources de données pour obtenir un corpus le plus exhaustif possible
  - Passer les données par le Notebook de l'université de Lorraine pour avoir une liste globale de DOI dédoublonnée

- La seconde étape : transmettre la liste au BSO national (format CSV ou Excel)

  - L'équipe du BSO national pourra exploiter la liste de DOI et transmettre les graphiques qui en découlent
  - Intégration possible ensuite dans des rapports ou sites web

Baromètre local : le périmètre

- Indicateur qui peut être créé à l’échelle d’une unité de recherche ou d’une fédération de recherche, même si certains graphiques seront sans doute moins exploitables (disciplines notamment).

Normandie Université

Baromètre local : le périmètre de temps

- Le BSO national recommande de remonter au moins à 2017 (possible jusque 2013).

- Pour le BSO normand, nous sommes remontés à la création du portail HAL : en 2017

Normandie Université

Baromètre local : les thèses de doctorat

- Il faut fournir : un code établissement et/ou une liste des NNT.

- **Fonctionnement en Normandie :**

Code établissement global pour le doctorat en Normandie : NORM

Chacun des 4 établissements inscrivant des doctorants à une déclinaison du code :

- Univ. Rouen = YYYYNORMRXXX
- Univ. Caen = YYYYNORMCXXX
- Univ. Le Havre = YYYYNORMLHXX
- INSA = YYYYNORMIRXX

Une liste des NNT est constituable à partir de theses.fr.

En utilisant la recherche avancée et en filtrant sur Etablissement de soutenance : « Normandie » et labratoire ou équipe de recherche « Nom du laboratoire » par exemple + les thèses soutenues

Un export CSV est ensuite possible. Il suffit de nettoyer le fichier en ne conservant que la colonne contenant les NNT.

Fichier global

› Exemple pour l'INSA de Rouen :

doi	hal_struct_id	hal_coll_code	hal_id	nnt_etab	nnt_id
10.1016/j.chemgeo.2016.10.031	301288	INSA-ROUEN			2021NORMIR10
10.1371/journal.pone.0168349					2021NORMIR07
10.1016/j.jpowsour.2016.10.037					2021NORMIR08
10.1016/j.jpowsour.2016.10.035					2021NORMIR09
10.1021/acs.jpcc.6b09974					2021NORMIR06
10.1021/acs.jpcb.6b09664					2020NORMIR01

Baromètre local : les sources de données

- Le notebook de l'Université de Lorraine permet d'exploiter plusieurs sources de données :
  - Le Web of Science : Abonnement Unicaen, Univ. De Rouen et Le Havre
  - Scopus (plus d'abonnement depuis 2022) > mais les UMR y ont accès via le portail BibCNRS
  - HAL
  - Pubmed (moteur de recherche dans le domaine de la santé)
  - Lens (moteur de recherche international sur les brevets et travaux académiques)
  - Suivi des APC : listes de DOI en fonction des APC payés par les établissements

Baromètre local : les sources de données

› Les fichiers doivent être créés par année avec une dénomination particulière. Exemple : **« wos_normandie_2017 »**

https://gitlab.com/Cthulhus_Queen/barometre_scienceouverte_universitedelorraine/-/blob/master/01_nettoyage_donnees.ipynb

Baromètre local : les requêtes à préparer

HAL :

- Faire les exports avec ExtrHAL > URL : https://halur1.univ-rennes1.fr/ExtrHAL.php
- Indiquer l’identifiants Auréhal de la structure
- Dans "choix des listes et dates de publications", choisir la période année par année, prendre tous les articles de revue (sauf vulgarisation), toutes les communications (sauf grand public), ouvrages ou chapitres ou direction d'ouvrages et toutes les autres productions puis valider.
- En haut de la liste de résultats, cliquez sur le bouton VOSviewerDOI. La liste des DOI s'affiche, il ne vous reste plus qu'à copier cette liste dans un fichier csv avec "doi" comme nom de colonne
- Nommage des fichiers : hal_accronymelabo_année

Baromètre local : les requêtes à préparer

WoS :

- Pour une unité de recherche, faire une recherche par affiliation : ex.: AD=(LPCCAEN) OR AD=(LPC CAEN) OR AD=(LPCC) OR AD=(corpusculaire CAEN) OR AD=(physique CAEN)
- Ajouter le filtre : « Year Published » > choisir année par année
- Exporter les résultats en cliquant sur « export » puis « fast 5k »
- Nommage des fichiers : wos_accronymelabo_année

PubMed :

- Recherche avancée
- Requête de type (à coller dans la *Query box* et en cherchant toutes les formes possibles) :

(("ISTCT"[Affiliation] OR "umr6030"[Affiliation] OR "umr 6030"[Affiliation] OR "Imagerie et Stratégies Thérapeutiques pour les Cancers et Tissus cérébraux"[Affiliation]) AND "2017"[Date - Publication] : "2017"[Date - Publication])

- Cliquez sur "save" puis "all results" et choisir le format CSV et cliquer sur « create file »
- Nommage des fichiers : pubmed_accronymelabo_année

Normandie Université

Baromètre local : les requêtes à préparer

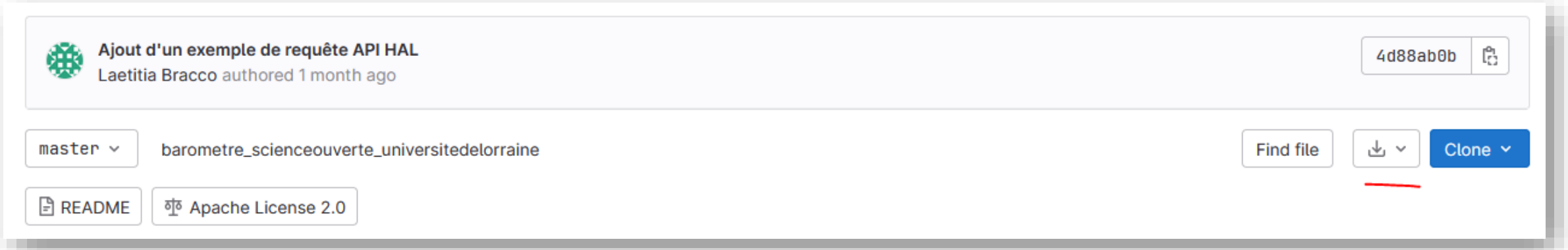
Lens.org :

- https://www.lens.org/
- onglet « scholarly works » et lancer la recherche
- En haut, faire Edit query > choisir le critère "institution" et chercher le nom du laboratoire, son numéro, son ROR puis valider
- Filtrer par année
- Dans le menu "export", choisir toutes les publications (attention c'est 1000 par défaut), exporter le champ DOI uniquement, au format CSV.
- On reçoit un mail quand l'export est prêt. L'export contient deux informations : le Lens ID et le DOI. Il faut séparer les deux informations au sein de la colonne, en sélectionnant la colonne puis en cliquant sur "Données", "Convertir", "Délimité", "Virgule" et terminer. Puis supprimer la colonne "Lens ID" et renommer la colonne "DOI" en "doi". Enfin, supprimer les lignes vides en sélectionnant la colonne, cliquer sur "Rechercher et sélectionner", "Sélectionner les cellules", "Cellules vides" et les supprimer.
- Nommage des fichiers : lens_accronymelabo_année

Etape 1

- Téléchargez l'ensemble du Baromètre lorrain sur le bureau en utilisant le bouton "Download". Dé-zippez l'archive. Sur les PC Unicaen, le dézipper dans le dossier de Téléchargement

- Lien vers le dossier : https://gitlab.com/Cthulhus_Queen/barometre_scienceouverte_universitedelorraine

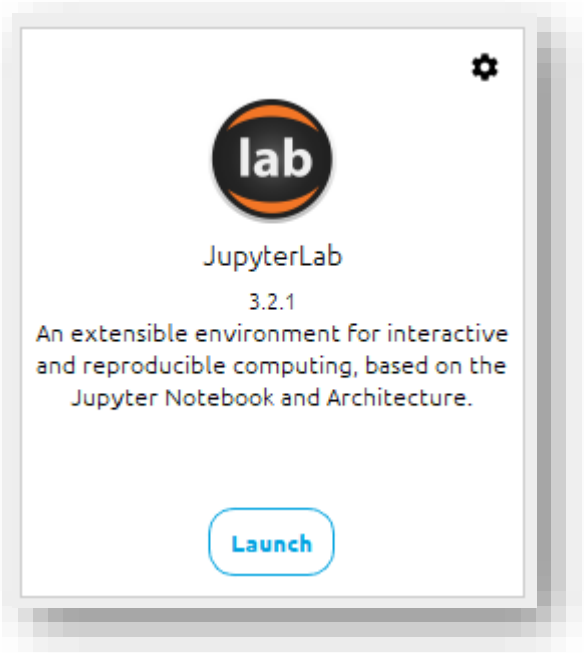


Etape 2

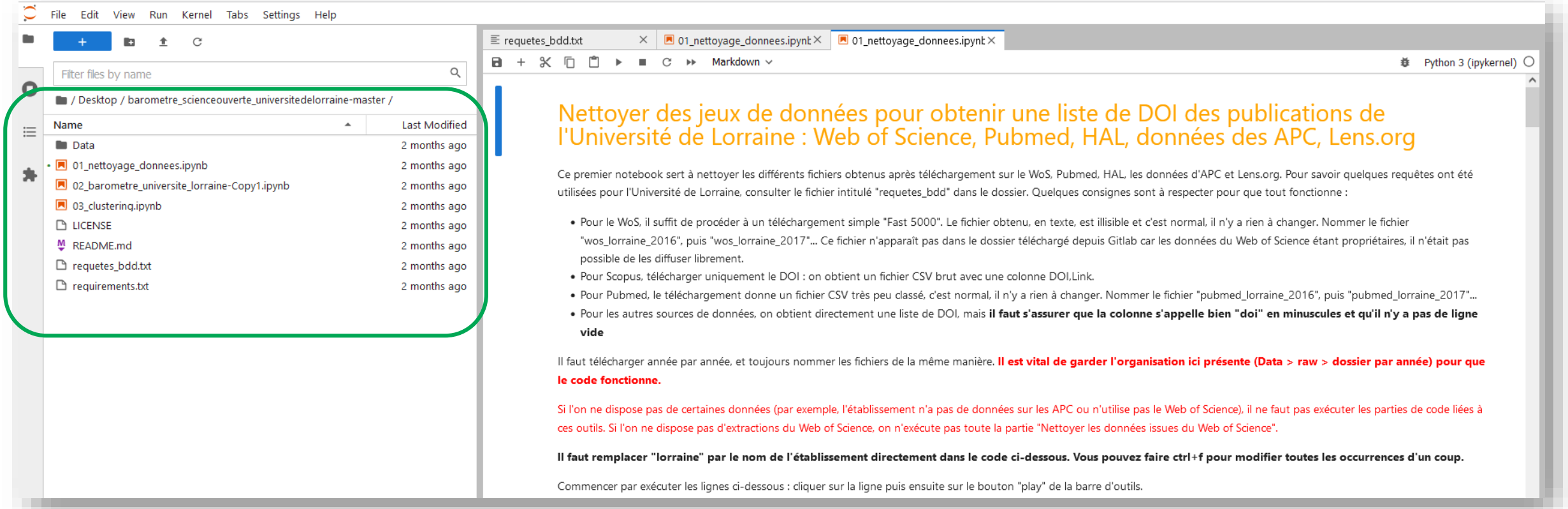
- Installez la suite Anaconda Navigator (https://www.anaconda.com/products/individual).

Etape 3

- Lancez Anaconda puis Jupyter Lab. Le dossier du Baromètre lorrain téléchargé sur le bureau apparaît sur la partie gauche de l'écran.



Etape 3



Normandie Université

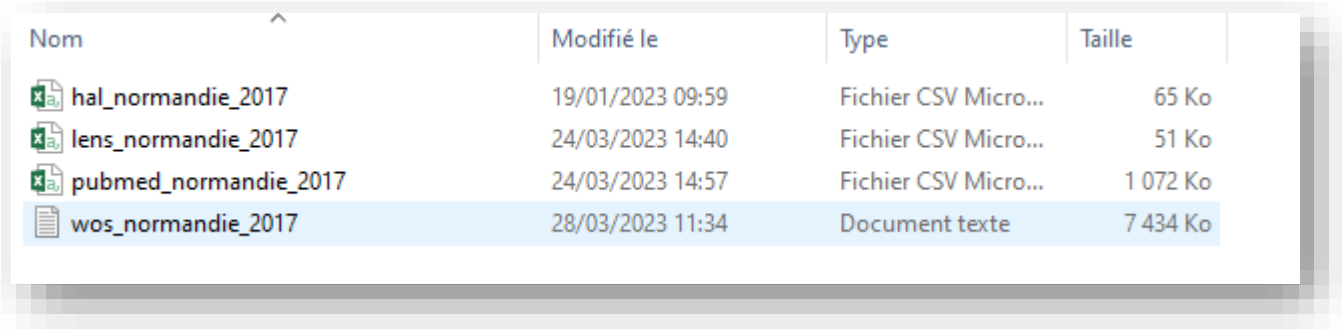
Etape 4

- Faire les extractions dans les différentes bases de données souhaitées pour faire le corpus de votre établissement. La liste des requêtes utilisées est dans le fichier "requetes_bdd". Les extractions dans les bases de données doivent se faire exactement comme c'est expliqué dans le notebook qui s'appelle "01_nettoyage_donnees".

- Cf. diapo 3

Etape 5

- Dans le dossier qui est sur le bureau, enlever les fichiers lorrains dans le dossier Data/raw et les remplacer par les vôtres. Attention, il faut reproduire très exactement le nommage des fichiers. Dans le code lorrain par exemple, le fichier 2016 pour PubMed s'appelle "pubmed_lorraine_2016". Il faut aussi effacer le contenu du dossier 'outputs' pour enlever les graphiques et résultats lorrains (mais conserver le dossier vide).



Normandie Université

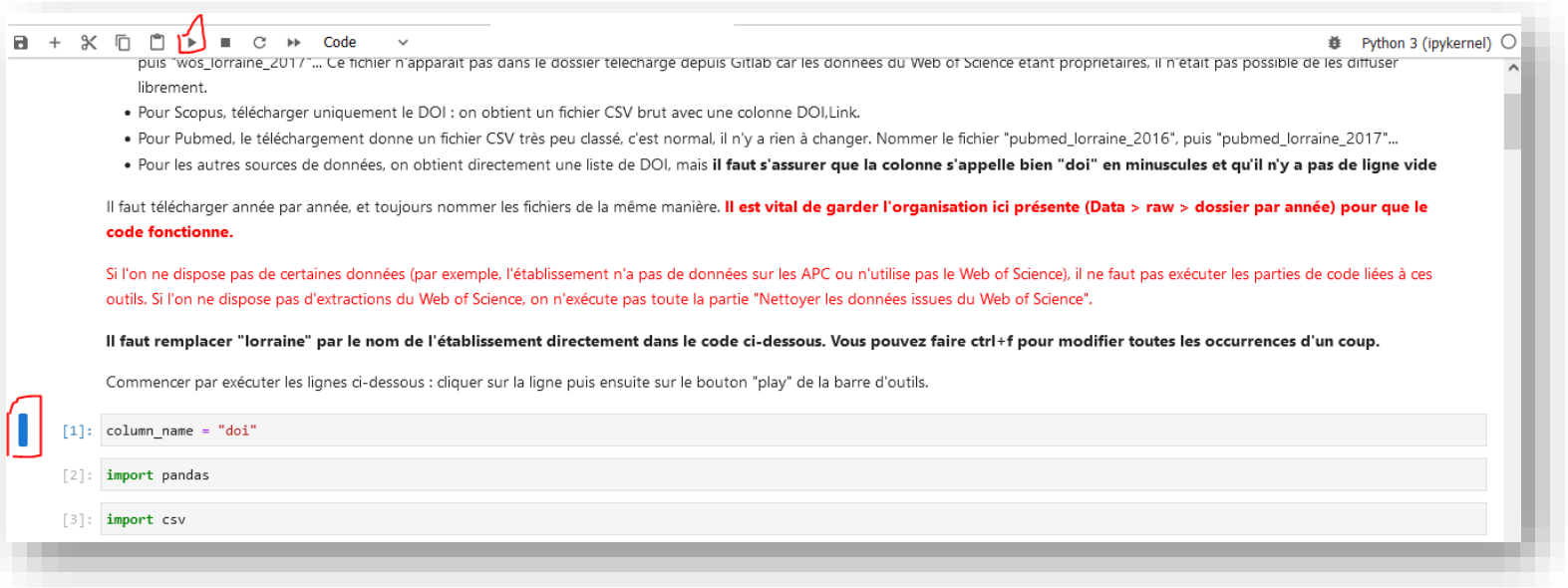
Etape 6

- Ouvrir le notebook "01_nettoyage_donnees". Remplacer toutes les occurrences de "lorraine" par le nom du laboratoire (attention : utiliser le même nom que pour les extractions de bases de données).

- Vous pouvez faire une recherche dans le notebook pour repérer les occurrences « Lorraine » (Ctrl F)
- Modification des dates : remplacement des dates de la Lorraine pour correspondre à notre période : 2016 > 2017

Etape 7

- Exécutez le notebook "01_nettoyage_donnees" en lisant bien les instructions à chaque étape. Pour exécuter les lignes de code les unes après les autres, il suffit de cliquer sur la première ligne tout en haut du code puis cliquer successivement sur le bouton "play" de la barre d'outils. Il faut bien lire les instructions.



Normandie Université

Etape 7

- Pour les étapes Scopus et APC, il ne faut pas lire les lignes.

- Dans le fichier Data > Outpouts : vous allez pouvoir récupérer le fichier unique de DOI dédoublonnés sur 2017-2022.

Merci pour votre attention !

Normandie Université