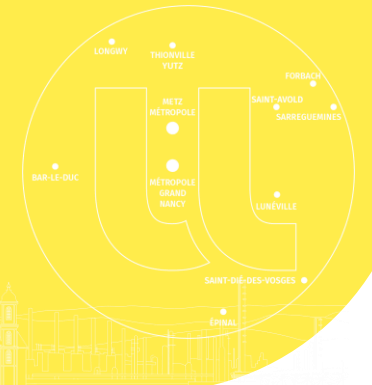


Trouver des données de recherche

Formation proposée par l'atelier de la donnée ADOC Lorraine



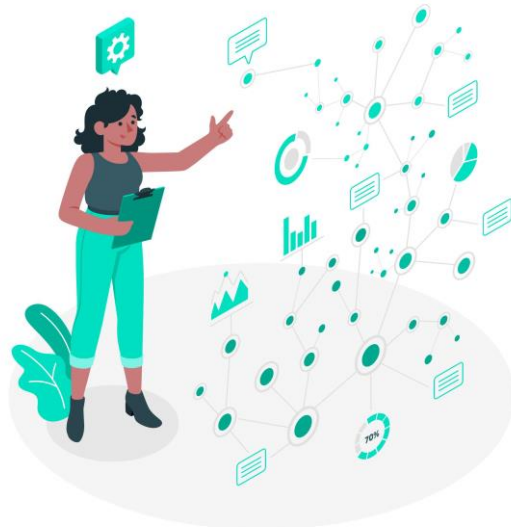
Que sont les données de la recherche ?

« Les données de la recherche sont définies comme des **enregistrements factuels** (chiffres, textes, images et sons), qui sont utilisés comme **sources principales pour la recherche scientifique** et sont généralement reconnus par la communauté scientifique comme nécessaires pour **valider les résultats de la recherche** »

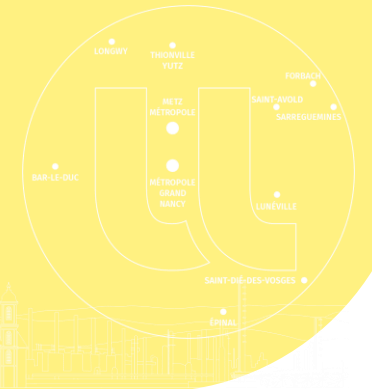
OCDE (Organisation de coopération et de développement économiques)

Pourquoi cet atelier ?

- Une ouverture des données de la recherche de plus en plus grande
 - La possibilité d'accéder à des milliers de jeux de données réutilisables
 - Une nouvelle donne pour la recherche : les jeux de données entrent dans l'état de l'art
- > Comment trouver des données dans un océan de bases, sites, portails ?



Contexte et enjeux des données



Quels outils à votre disposition à l'Université de Lorraine ?

Pour permettre la gestion de ces données, plusieurs infrastructures sont disponibles :

- La [B'UL](#) pour le **travail collaboratif** et le stockage de **faible volumétrie**
- [PETA](#) pour le **stockage volumineux** pendant la durée du projet (PETA peut aussi accueillir autre chose que des données)
- [DOREL](#) pour **préserver** vos données à la fin du projet

Pour vous aider, consultez la boîte à outils de la Science Ouverte de l'UL, rubrique « Données » : <https://scienceouverte.univ-lorraine.fr/boite-a-outils/>

Quel contexte réglementaire ?

- Horizon Europe : accès libre aux publications, encouragement à l'ouverture des données, plan de gestion de données (PGD)
- France (ANR), Suisse (FNS), États-Unis (NIH) : le PGD devient obligatoire
- Principe « aussi ouvert que possible, aussi fermé que nécessaire »
- Des [ateliers « plan de gestion de données »](#) vous sont proposés régulièrement



Dans le modèle ANR du PGD, la 1^{re} partie porte sur les **données produites ou réutilisées**

- Entrée en jeu des entrepôts de données pour trouver des données qui peuvent vous servir
- Dépôt des données de plus en plus demandé par les éditeurs : **ne pas déposer dans leurs propres entrepôts !**

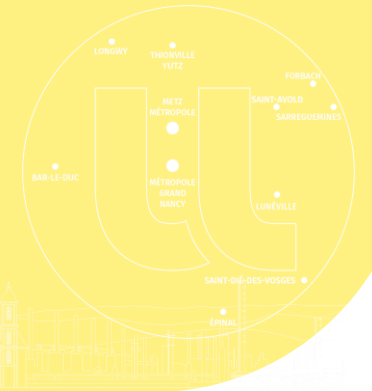


Pourquoi déposer dans un entrepôt ?

- **Conservation** des données
- **Visibilité** des données
- **Interopérabilité** des données
- **Réutilisation** et **citation** facilitées
- Gestion des modalités de **partage** des données
- **Respect des recommandations** des financeurs et institutions sur l'ouverture des données
- **Reproductibilité** de la recherche, **intégrité**
- **Valorisation** des données

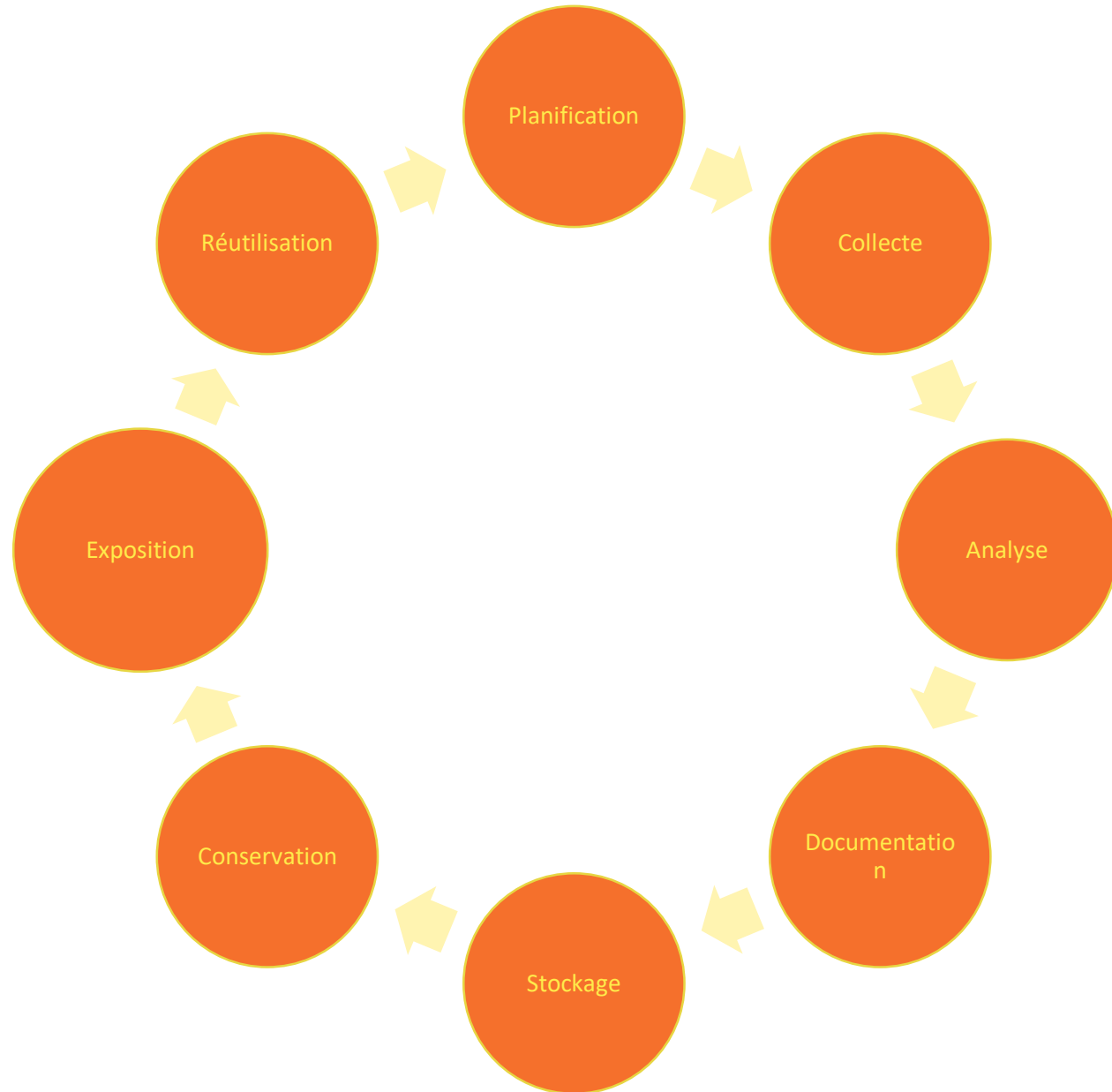


Les catalogues de données



CAT OPIDOR

- Outil de l'Inist
- Services français uniquement
- Permet de trouver des entrepôts mais pas que

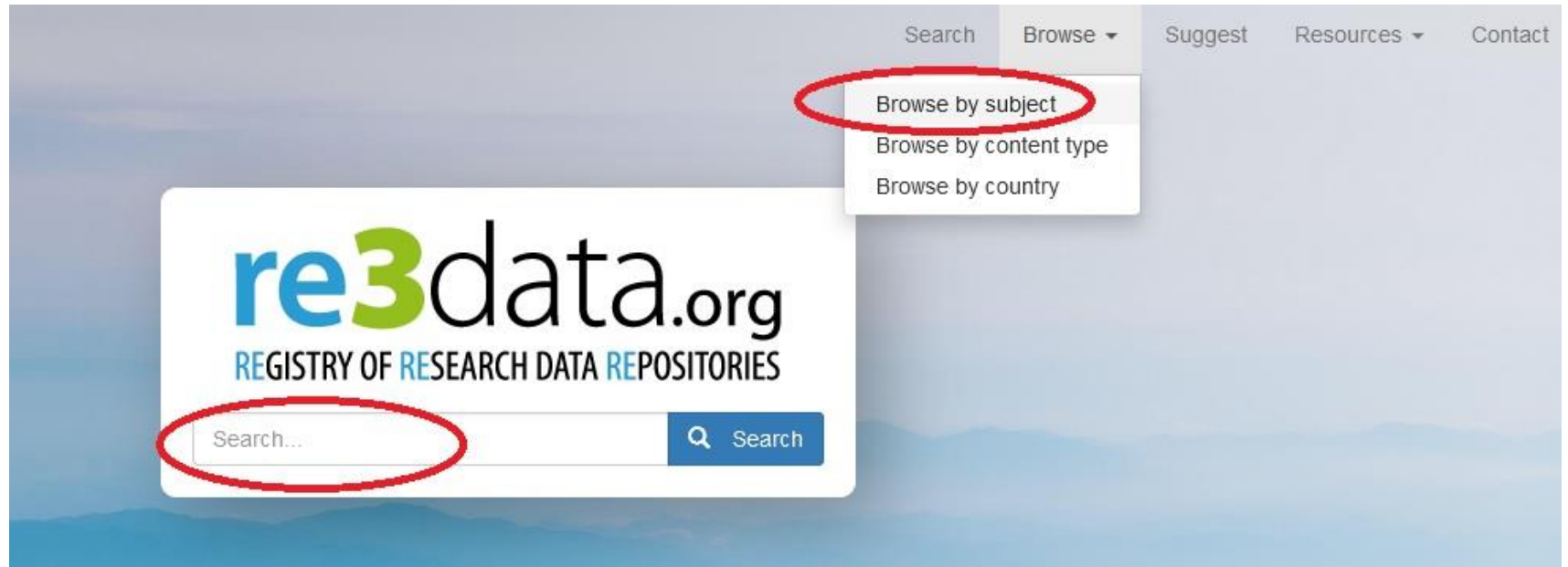


[Site de Cat OPIDoR](#)

Typologie des outils : les annuaires d'entrepôts

Ils permettent de trouver des entrepôts de données (et pas directement des données) :

- [Re3data](https://re3data.org)



- [Open Access Directory](https://openaccess.directory)

Typologie des outils : les annuaires de jeux de données

Ils permettent de trouver directement des jeux de données déposés dans de nombreux entrepôts différents :

- [Google Dataset Search](#)

Google



Dataset Search

Rechercher des ensembles de données



Essayez [coronavirus covid-19](#) ou [water quality site:canada.ca](#).

[Découvrez le projet Recherche d'ensembles de données.](#)

Typologie des outils : les annuaires de jeux de données

OpenAIRE EXPLORE

The screenshot displays the OpenAIRE EXPLORE search interface. At the top, the logo 'OpenAIRE EXPLORE' is visible on the left, and navigation links for 'Search', 'Deposit', 'Link', and 'Data sources' are on the right. A search bar is highlighted with an orange border, containing the text 'Advanced search in' and a dropdown menu set to 'Research products'. Below the search bar, the 'SEARCHING FIELDS' section shows 'Any field' selected, and the 'TERMS' section shows 'includes' selected. A search button with a magnifying glass icon and the text 'SEARCH →' is located at the bottom right of the search area.

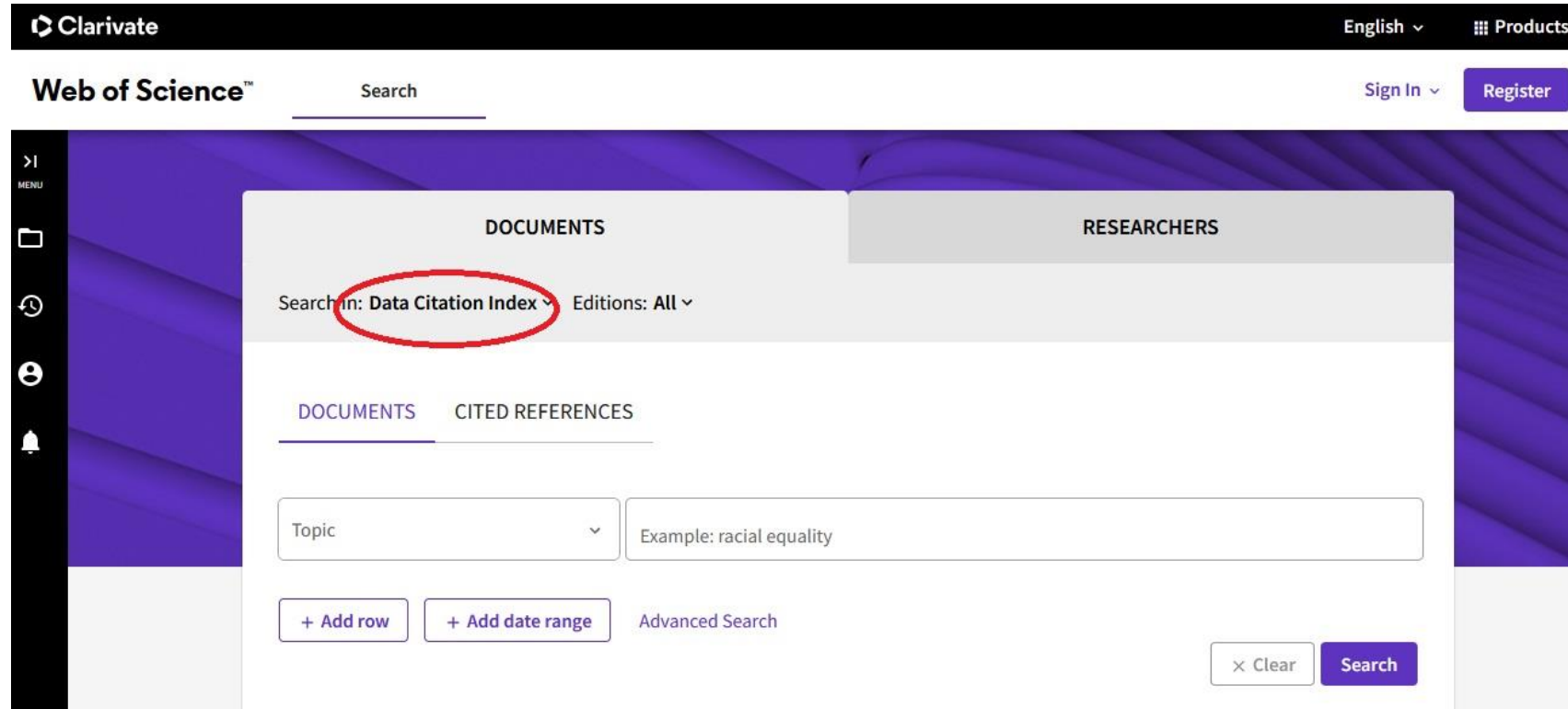
Below the search area, the 'Filters' section is visible, with a 'Clear All' link. The 'Access (40)' filter is expanded, showing radio button options for 'Open Access (4,017,108)', 'Restricted (30,950)', 'Closed Access (14,743)', 'Embargo (5,710)', 'Creative Commons Attribut... (831)', and 'Open Source (764)'. A 'View all >' link is provided. The 'Type (4)' filter is also expanded, with radio button options for 'Publications', 'Research data' (which is checked and highlighted with an orange border), 'Research software', and 'Other research products'.

The main content area displays '58,119,105 Research Products'. A 'RESEARCH DATA' filter is active, indicated by a pill-shaped button with an 'x' icon. Below this, the 'Results per page' is set to '10' and 'Sort by' is set to 'Relevance'. Two search results are shown:

- Peptidase M28**
Research Data » Bioentity • 2017
http://www.uniprotlo... | Link to | Share | Cite
- Opioid receptor kappa 1**
Research Data » Bioentity • 2019
http://www.uniprotlo... | Link to | Share | Cite

Typologie des outils : les annuaires de jeux de données

- Web of Science : depuis le [catalogue des bibliothèques](#), aller sur le WoS et entrer en recherche avancée, puis sélectionner le *Data Citation Index*



The screenshot displays the Web of Science search interface. At the top, the Clarivate logo is on the left, and 'English' and 'Products' are on the right. Below the logo, 'Web of Science' and 'Search' are visible. On the right side of the header, there are 'Sign In' and 'Register' buttons. The main search area is divided into 'DOCUMENTS' and 'RESEARCHERS' tabs. The 'DOCUMENTS' tab is active, and the search dropdown menu is open, showing 'Data Citation Index' selected and circled in red. Below the search bar, there are tabs for 'DOCUMENTS' and 'CITED REFERENCES'. The search input field contains the text 'Example: racial equality'. At the bottom of the search area, there are buttons for '+ Add row', '+ Add date range', and 'Advanced Search'. On the far right, there are 'Clear' and 'Search' buttons.

Typologie des outils : les entrepôts

Il en existe des milliers, disciplinaires ou généralistes. Quelques exemples :

Pluridisciplinaires :

- [Recherche Data Gouv](#)
- [Zenodo](#)
- [Dryad](#)
- [Mendeley Data](#)
- [Figshare](#)

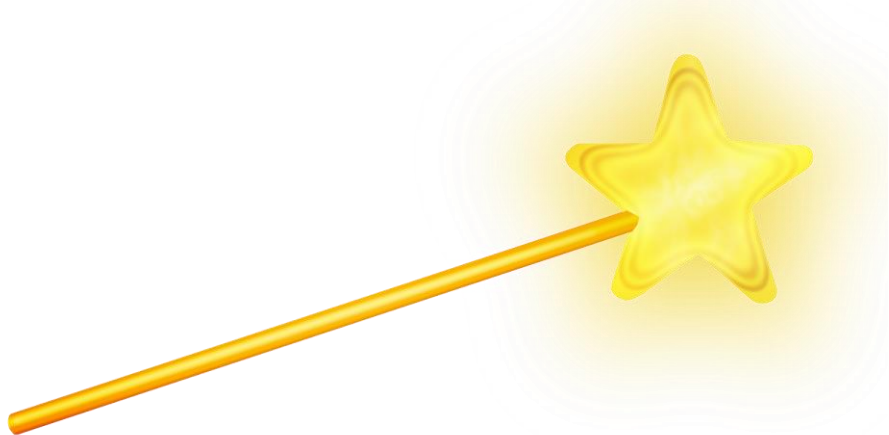
Spécialisés :

- [Crystallography Open Database](#)
- [Nakala](#)
- [Pangaea](#)
- [Ortolang](#)
- [Quetelet Progedo](#)



Citer des jeux de données

- ☛ **Zenodo** : fonction « cite as »
- ☛ **Zotero** : le type de document « dataset » est à présent accepté
- ☛ Avec un simple DOI, l'outil magique : **DOI Citation Formatter**
<https://citation.crosscite.org/>



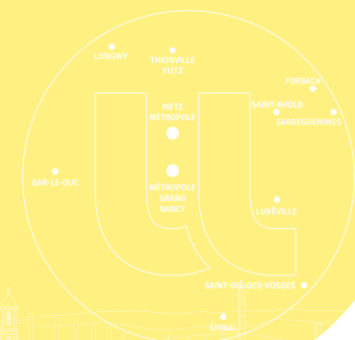
Cite as

Banda, Juan M., Tekumalla, Ramya, Wang, Guanyu, Yu, Jingyuan, Liu, Tuo, Ding, Yuning, Artemova, Katya, Tutubalina, Elena, & Chowell, Gerardo. (2023). A large-scale COVID-19 Twitter chatter dataset for open scientific research - an international collaboration [Data set]. In *Epidemiologia* (Version 162, Vol. 2, Number 3, pp. 315–324). Zenodo. <https://doi.org/10.5281/zenodo.7834392>

Start typing a citation style...



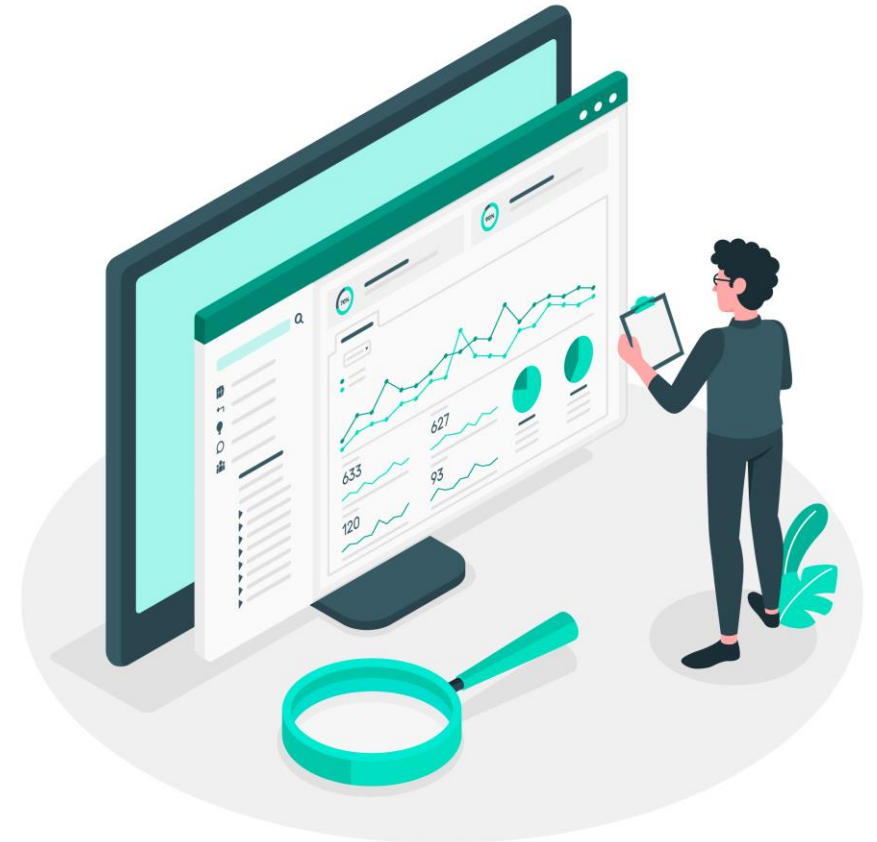
Exercices pratiques



Trouver des données

Cherchez dans votre domaine ; en cas de panne d'inspiration, quelques idées :

- Sur le **Data Citation Index**, trouver des données sur les défauts d'arcs électriques
- Sur **Dataset Search**, trouver des données d'accélérateurs de particules
- Sur **Dryad**, trouver des données sur la migration des baleines
- Sur **Zenodo**, trouver des données sur les polymères
- Sur **Mendeley Data**, trouver des données sur les pesticides nocifs utilisés en Guadeloupe



Trouver des données : propositions

- Sur le Data Citation Index, trouver des données sur les défauts d'arcs électriques : « [arcing faults](#) »
- Sur Dataset Search, trouver des données d'accélérateurs de particules : « [synchrotrons](#) »
- Sur Dryad, trouver des données sur la migration des baleines : « [whales migration](#) »
- Sur Zenodo, trouver des données sur les polymères : « [polymers](#) » puis restriction à « [dataset](#) »
- Sur Mendeley Data, trouver des données sur les pesticides nocifs utilisés en Guadeloupe : « [chlordecone](#) »

> Trouver des vocabulaires scientifiques : <https://www.loterre.fr/>

Où déposer ses données ?

Vous avez apprécié de trouver des données pour faire avancer votre projet, maintenant déposez les vôtres !

- L'Université de Lorraine dispose son espace dans Recherche Data Gouv → [DOREL](#)
- Pour les laboratoires du pôle Otelo (CRPG, GeoRessources, LIEC, LSE) → [Ordar](#)
- Comment choisir un entrepôt ?
 - Ne pas déposer dans un entrepôt d'un éditeur
 - Privilégier un entrepôt disciplinaire
 - S'il n'y en a pas, choisir DOREL. Des [ateliers DOREL](#) vous sont proposés régulièrement
 - Ne pas hésiter à solliciter l'atelier de la donnée ADOC Lorraine : donnees-recherche@univ-lorraine.fr



Conclusion

- Trouver des données ou identifier des entrepôts nécessite une certaine gymnastique
- Mais plus on en fait, plus c'est facile !
- **Et pour toute question : donnees-recherche@univ-lorraine.fr**
- Merci !



Pour aller plus loin

- Rechercher des données : fiche pratique proposée par DoRANum <https://doranum.fr/acces-visualisation/rechercher-donnees-10-13143-9k8k-s875/>
- Trouver des jeux de données via des bases : fiche pratique CoopIST (CIRAD) <https://coop-ist.cirad.fr/gerer-des-donnees/trouver-des-jeux-de-donnees-via-des-bases/1-l-interet-des-jeux-de-donnees>
- Kit de survie dans le monde des données de la recherche : <https://arche.univ-lorraine.fr/course/view.php?id=48249>
- Produire, gérer, stocker, diffuser les données de la recherche : <https://arche.univ-lorraine.fr/course/view.php?id=48259>