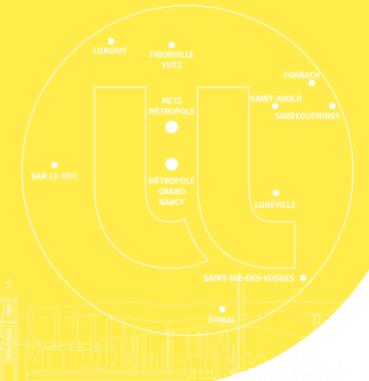


Trouver des données de recherche dans son domaine disciplinaire

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 ?

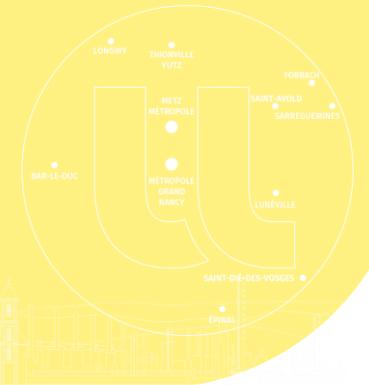
- Une ouverture des données de la recherche de plus en plus grande
 - La possibilité d'accéder à des milliers de jeux de données réutilisables
 - Une nouvelle donne pour la recherche : les jeux de données entrent dans l'état de l'art
- > Comment trouver des données dans un océan de bases, sites, portails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





UNIVERSITÉ
DE LORRAINE

Contexte et enjeux des données



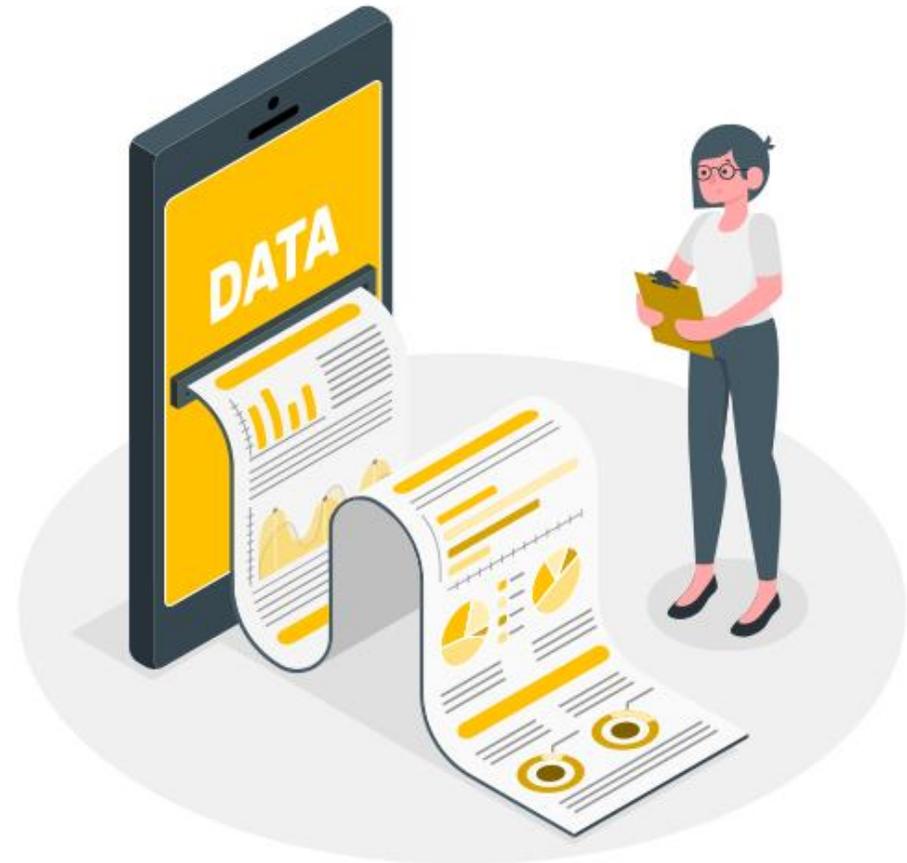
Quel contexte réglementaire ?

- 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 !**

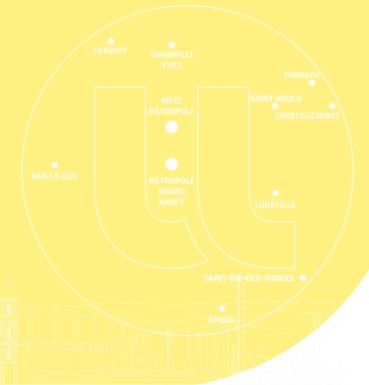


Pourquoi déposer dans un entrepôt ?

- **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

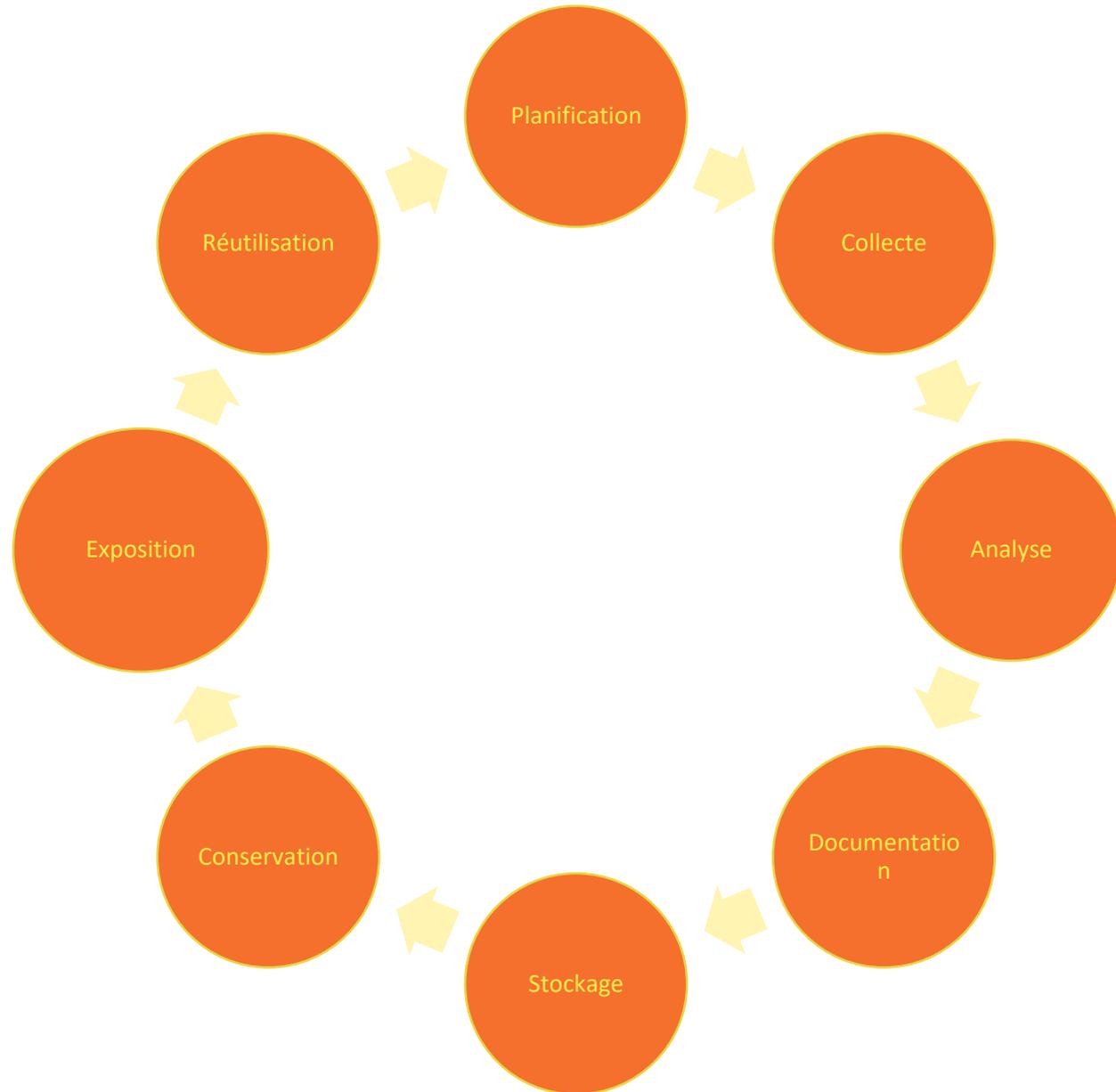


Les catalogues de données



CAT OPIDOR

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

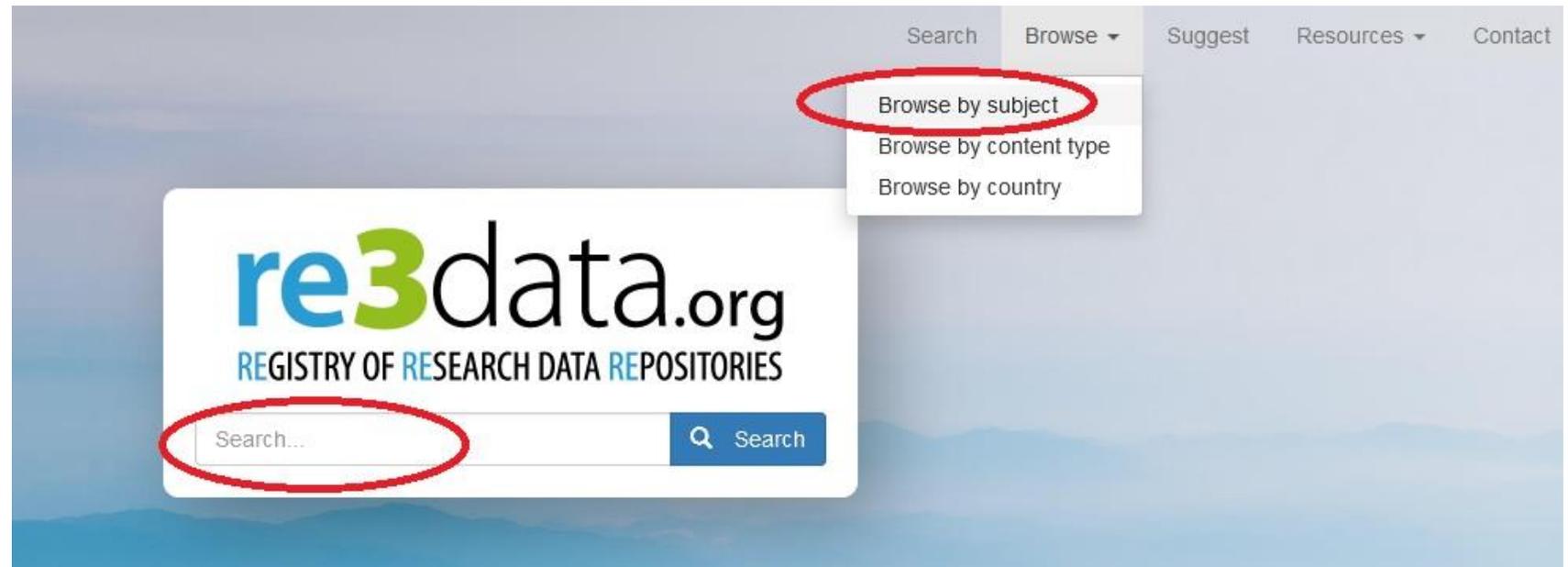


[Site de Cat OPIDoR](#)

Typologie des outils : les annuaires d'entrepôts

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

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



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. The main search area features a blue header with 'Advanced search in' followed by a dropdown menu set to 'Research products'. Below this, there are sections for 'SEARCHING FIELDS' (set to 'Any field') and 'TERMS' (set to 'includes'), with a text input field for 'Type keywords...'. A dark blue search button with a magnifying glass icon and the text 'SEARCH →' is positioned at the bottom of the search area.

Below the search area, the results page is shown. On the left, there are two filter sections: 'Filters' with a 'Clear All' link, and 'Type (4)' with a 'Clear' link. The 'Access (40)' filter includes radio buttons for 'Open Access (4,017,108)', 'Restricted (30,950)', 'Closed Access (14,743)', 'Embargo (5,710)', 'Creative Commons Attribut... (831)', and 'Open Source (764)'. The 'Type (4)' filter includes radio buttons for 'Publications', 'Research data' (which is checked and highlighted with an orange box), 'Research software', and 'Other research products'. On the right, the total number of results is '58,119,105 Research Products'. There is a 'RESEARCH DATA' button with an 'x' icon. Below this, there are controls for 'Results per page' (set to 10) and 'Sort by' (set to Relevance). Two search results are displayed: 'Peptidase M28' (Research Data, Bioentity, 2017) and 'Opioid receptor kappa 1' (Research Data, Bioentity, 2019). Each result includes a URL, a 'Link to' icon, and 'Share' and 'Cite' icons.

Typologie des outils : les annuaires de jeux de données

- [OpenAlex](#)

The screenshot displays the OpenAlex website interface. At the top, there is a search bar with the text "Search OpenAlex" and a "Works" dropdown menu. Below the search bar, there is a section for "Unsaved search" with a dropdown arrow and notification, settings, and share icons. The main content area shows "No filters applied" and a blue plus icon. The "Works" section lists three items:

- Protein Measurement With The Folin Phenol Reagent**
1951 · OliverH. Lowry, NiraJ. Rosebrough, et al. · *Journal of Biological Chemistry*
Cited by 319 900
- R: A language and environment for statistical computing.**
2014 · R Core Team · *MSOR Connections*
Cited by 300 300
- Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4**
1970 · Ulrich K. Laemmli · *Nature*
Cited by 254 300

The "Stats" section shows "264 900 000 results" and a table of filters:

type	count
<input type="checkbox"/> article	199 200 000
<input type="checkbox"/> book-chapter	22 710 000
<input type="checkbox"/> dataset	7 675 000
<input type="checkbox"/> preprint	6 619 000
<input type="checkbox"/> dissertation	6 110 000

The "institution" filter is also visible, showing a list of institutions with their respective counts:

institution	count
<input type="checkbox"/> Centre National de la Recherche Scientifique	2 720 000
<input type="checkbox"/> Chinese Academy of Sciences	1 343 000
<input type="checkbox"/> United States Department of Energy	1 286 000
<input type="checkbox"/> University of London	1 147 000
<input type="checkbox"/> Russian Academy of Sciences	1 018 000

Typologie des outils : les entrepôts

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

Pluridisciplinaires :

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

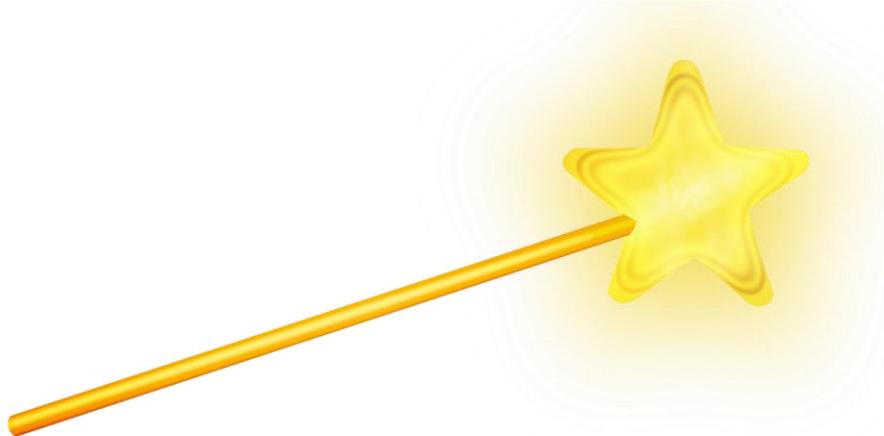
Spécialisés :

- [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/>



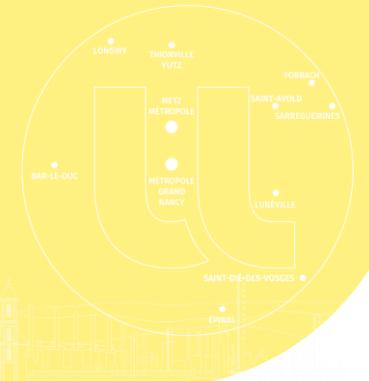
Cite as

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...



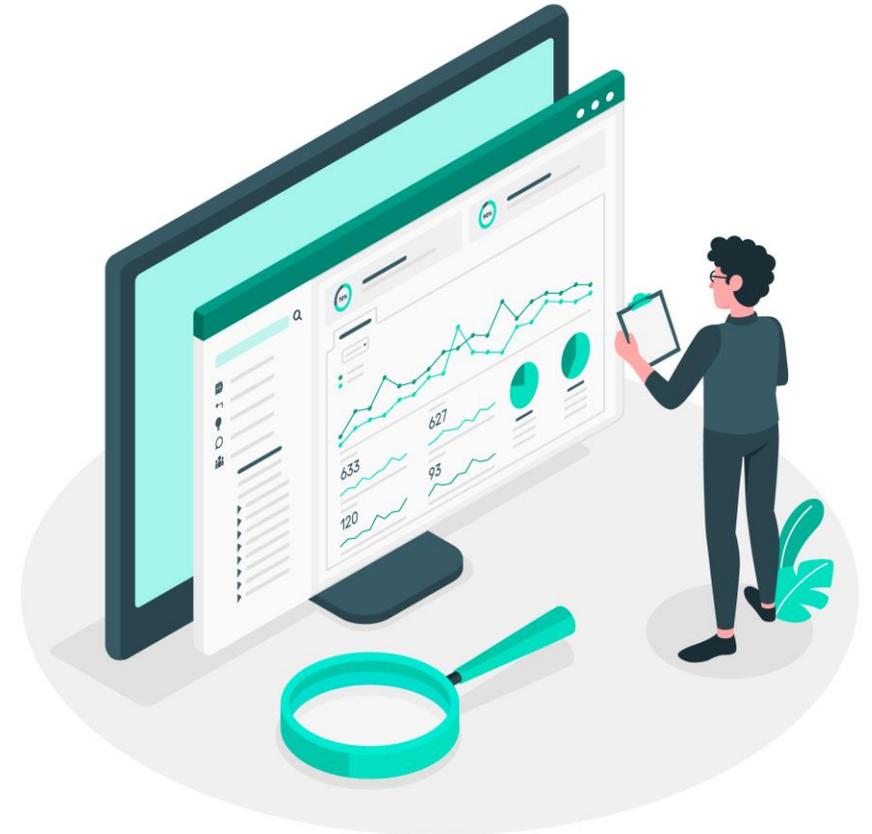
Exercices pratiques



Trouver des données

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

- Sur **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 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/>

Et maintenant, à vous de déposer !



Quels outils à votre disposition à l'Université de Lorraine pour produire et gérer vos données ?

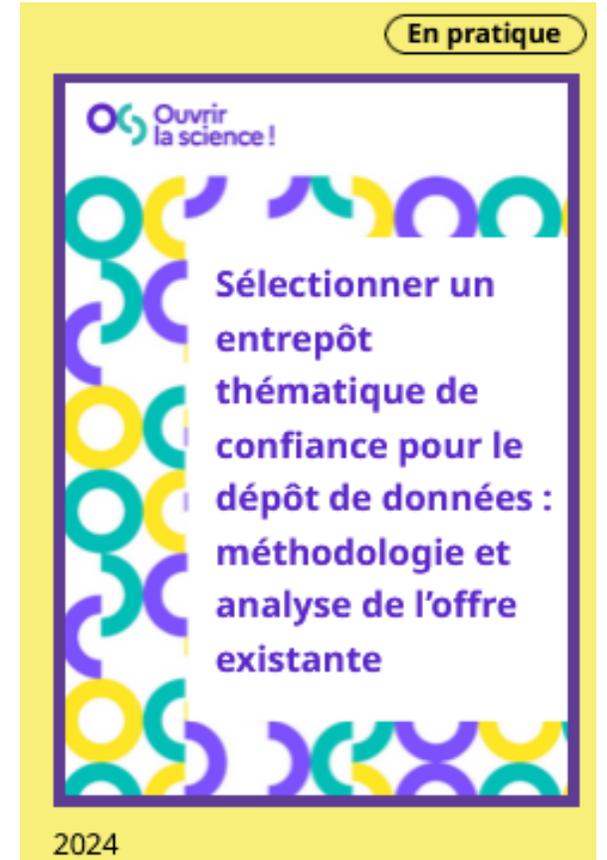
Voir la [cartographie des outils sur le site universitaire lorrain](#). En résumé :

- Enquêtes : LimeSurvey
- Calcul intensif : EXPLOR
- Envoi de données volumineuses : FileSender
- Forge logicielle : Gitlab
- Cahier de laboratoire électronique : eLabFTW
- Stockage pour données <20 Go : B'UL
- Stockage pour données de forte volumétrie : PETA

Où déposer ses données ?

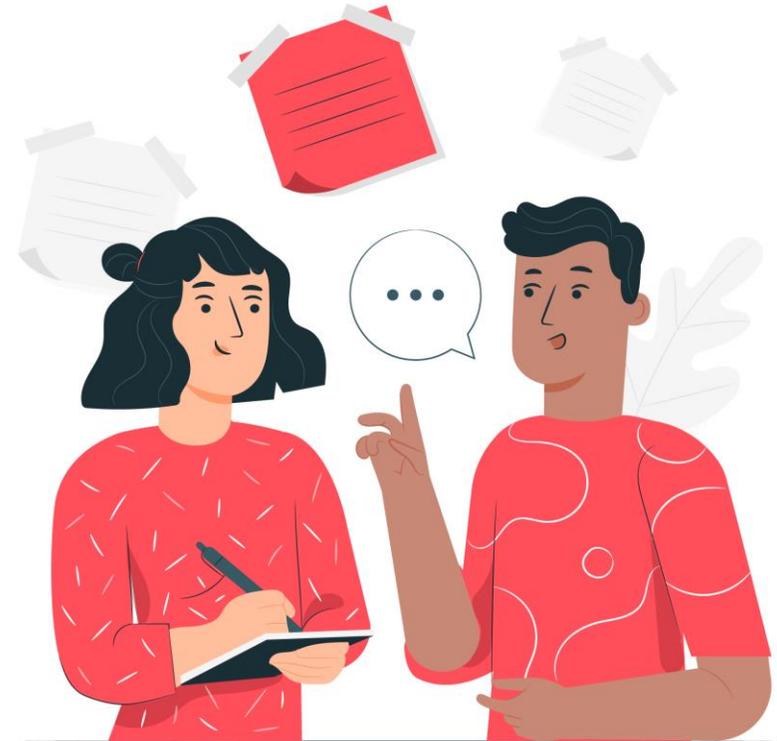
Vous avez apprécié de trouver des données pour faire avancer votre projet, maintenant déposez les vôtres !

- Trouver un entrepôt de données disciplinaire de confiance → [voir la liste](#)
- Si pas d'entrepôt dans votre domaine, l'Université de Lorraine dispose son espace dans Recherche Data Gouv → [DOREL](#)
- Des [ateliers « Publier des données dans un entrepôt »](#) 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>