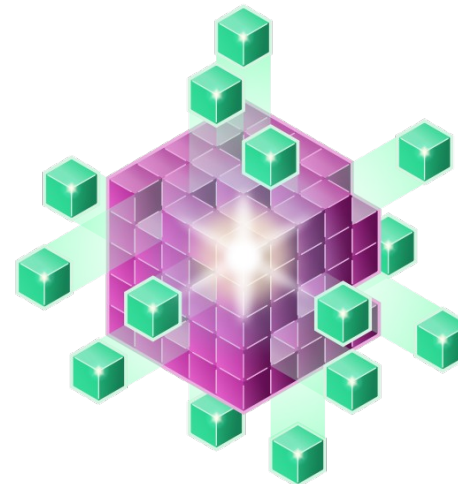




Trouver une info dans une botte de données !





Trouver des données qu'on puisse utiliser !

- Pour sa thèse
- Pour vérifier des résultats de recherche, pour comparer
- Pour compléter une approche
- Pour une nouvelle recherche
- Dans le cadre d'un projet collaboratif ...



Dans votre recherche, utilisez-vous des données déjà existantes ?



Comment trouver des données utiles à sa recherche ?

Votre avis ?



Différents lieux de diffusion des données

- des « **entrepôts** »
- des **données directement liées aux publications**

2 enjeux :

- Trouver et réutiliser des données pour sa propre recherche
- Permettre à vos collègues, votre jury, la communauté, de trouver, vérifier et réutiliser vos données



Focus entrepôts

Définition

Service en ligne permettant le dépôt, la description, la conservation, la recherche et la diffusion des jeux de données.

Plus de 2500 entrepôts !!



Différents **types d'entrepôts** :

Généralistes

Zenodo (CERN)

Nationaux

Recherche Data Gouv (France)

Thématiques

PANGAEA (Sciences de l'environnement)

Nakala (Sc Humaines)

Progedo (Sciences sociales)

PubChem (Chimie)



Données publiques

Data gouv : plateforme ouverte des données publiques françaises

Meteo France

Legifrance (droit français : lois et jurisprudence)

Data europe : le portail officiel des données européennes

US data gouv : U.S. Government's open data

UNESCO Institute for Statistics (UIS)



Des **fonctionnalités** utiles pour faciliter l'identification et la récupération des données utiles ..

Comment trouver un jeu pertinent ?

- Moteur de recherche

- cartes

- Filtres (sujets, type, format, organismes, classifications, etc)

- outils de recherche avancée (plugin, requêtes API, webservices, paquets R et python, etc..)



- Que vais-je vraiment trouver dans ces données ?

Description détaillée : titre, auteur, objectifs, résumés, mots-clés/taxonomie, couverture, format, type, volume, paramètres, read-me, financeurs, etc.)

voir par ex :

<https://entrepot.recherche.data.gouv.fr/dataset.xhtml?persistentId=doi:10.57745/SCOY6D>

<https://doi.pangaea.de/10.1594/PANGAEA.842378>

<https://www.icpsr.umich.edu/web/ICPSR/studies/37193/summary>



Y-a-t-il d'autres ressources intéressantes en lien ?

- **Données en rapport** : données utilisées par d'autres (« Users interested in this dataset were also interested in ») ou collection de données
- **Publications associées** (voir ici)






















Qu'ai-je le **droit** d'en faire ?

Licences

License (for files):

[Creative Commons Attribution 4.0 International](#)

Les licences Creative Commons

		Utilisation Partage	Adaptation Modification	Utilisation commerciale	Modification de licence	
TRÈS LIBRE						<ul style="list-style-type: none"> Utilisation commerciale autorisée Modifications ou remix autorisés
						<ul style="list-style-type: none"> Utilisation commerciale autorisée Modifications ou remix autorisés Les versions dérivées de l'œuvre doivent conserver la licence originale ou compatible
LIBRE						<ul style="list-style-type: none"> Utilisation commerciale NON permise Modifications ou remix autorisés
						<ul style="list-style-type: none"> Utilisation commerciale NON permise Modifications ou remix autorisés Les versions dérivées de l'œuvre doivent conserver la licence originale ou compatible
NON LIBRE						<ul style="list-style-type: none"> Utilisation commerciale autorisée Modifications ou remix NON permis
						<ul style="list-style-type: none"> Utilisation commerciale NON permise Modifications ou remix NON permis



BY

ATTRIBUTION

Vous pouvez retenir, réutiliser, réviser, remixer et redistribuer.

L'auteur doit être cité



SA

PARTAGE DANS LES MÊMES CONDITIONS

Vous pouvez retenir, réutiliser, réviser, remixer et redistribuer.

Partage sous licence compatible



NC

POUR USAGE NON COMMERCIAL

Vous pouvez retenir, réutiliser, réviser, remixer et redistribuer.

Pour usage non commercial



ND

PAS DE MODIFICATION

Vous pouvez retenir, réutiliser et redistribuer.

Aucune modification permise



Comment peut-on **les récupérer** ?

Modalités d'export métadonnées (RIS, bibteX) et données (zip, etc.)

Comment vais-je **les citer** dans ma thèse ? Comment mes données seront citées dans d'autres publications ?

Ohio State University. Center for Human Resource Research. National Longitudinal Surveys of Labor Market Experience, 1966-1992. Inter-university Consortium for Political and Social Research [distributeur], 2008-05-21. <https://doi.org/10.3886/ICPSR07610.v1>



Des **moteurs de recherche** « données » pour trouver des données :

- DataCite Commons
- Dataset Search (Google)



Des **moteurs de recherche** « généralistes » pour lier les informations

Dimensions et Lens (lien avec les publications et brevets)

Bielefeld Academic Search Engine (BASE) : filtre dataset/software

OpenAire Explore (Union européenne : liens publications données projets de recherche)



Répertoire d'entrepôts

Re3data (Registry of Research Data Repositories)

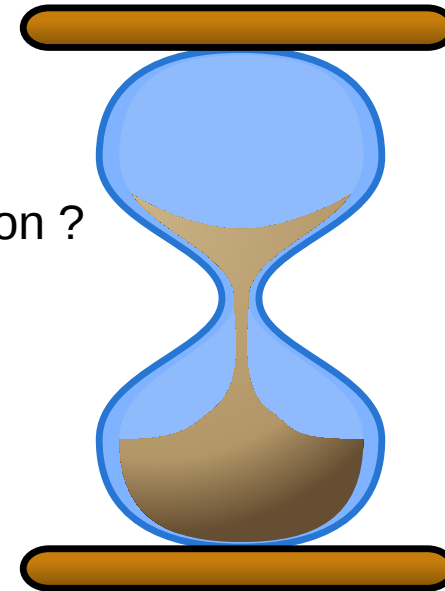
Repository Finder (DataCite)

Dataacc.org (physique, chimie, science de l'environnement)



A vous de jouer !

- Trouvez les données climatologique de base mensuelles de Météo France pour un département
et ...
- Vérifiez leur utilisation possible
 - Est-ce que j'ai le droit ?
 - Est-ce que ça va être simple de les utiliser ?
 - Est-ce que je sais à quoi elles correspondent ?
 - Est-ce que je pourrai les citer dans ma future publication ?
 - **Vous avez 15 mn !!!**





- Mon directeur de thèse me demande d'utiliser les données de certains collègues
 - 10.5281/zenodo.4572733
- Super, même pas besoin de les chercher ...
- Mais que puis-je en faire ?

Petite astuce :

- Utiliser un DOI resolver :
 - <https://dx.doi.org/>
Exemple :
dx.doi.org/10.5281/zenodo.4572733
- Ou intégrer un plugin à Chrome comme "DOI Resolver" et à FireFox comme "CNRI Handle Extension for Firefox"
 - Exemple :
[doi:10.5281/zenodo.4572733](https://doi.org/10.5281/zenodo.4572733)





- **Votre avis ?**





- **Si je vous dis que les données doivent être « FAIR » ?**

- Équitables
- En rapport avec des fées
- Findable, Accessible, Interoperable, Reusable
- Faciles, Attractives, Intéressantes, Rapides

- **Et concrètement ?**



F
indable



A
ccessible



I
nteroperable



R
eusable





Des idées pour rendre ses données :

Trouvables ?

Accessibles ?

Interopérables ?

Réutilisables ?



Findable / Trouvable



Données **faciles à trouver**.

- possédant un identifiant unique et pérenne
- décrites par des métadonnées riches
- enregistrées ou indexées dans une source interrogeable

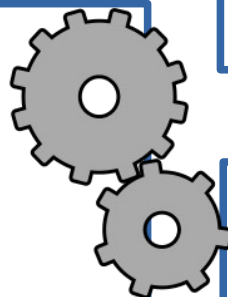
Accessible / Accessible



Données **ou au moins méta-données** facilement accessibles.

- entrepôt de confiance, pérenne, certifié
- définir les conditions d'accès et la licence de diffusion
- si embargo ou accès restreint : méta-données accessibles

Interoperable / Interopérable

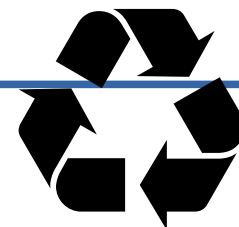


Facile à combiner avec d'autres jeux de données, par les humains **et** les systèmes informatiques

- formats libres et ouverts
- mise à disposition du code source si le logiciel de traitement existe
- standards de métadonnées et vocabulaire standardisés

Reusable / Réutilisable

Prêtes à être **réutilisables** pour une future recherche y compris via des méthodes informatiques





Rendre ses données FAIR est un moyen :

- Pour bien gérer ses données tout au long de la thèse
- Pour que ses propres données soient réutilisables ...
 - par soi-même
 - par d'autres (votre directeur de thèse, votre labo, votre jury et la communauté)
- Donc pour faciliter le travail de recherche



- **Le minimum à FAIR :**
 - Organiser ses données
 - Les nommer
 - Les décrire : donner le plus d'informations contextuelles possible (objectifs, méthodes)
 - Privilégier les formats ouverts
 - Définir les droits d'accès et attribuer une licence
 - Lier avec ses autres productions (publications, thèse, codes ...)
- **Comment FAIR ? C'est le menu de ces deux jours !**