



Open science - Fiche pratique

Trouver des jeux de
données de recherche

Pourquoi réutiliser des données ?

La réutilisation de données de recherche permet :

- de **gagner du temps** en évitant de collecter/produire/générer des données ;
- d'**élargir** l'objet de la recherche ;
- de faire une **méta-analyse** ;
- d'accéder à des **données uniques** qui ne peuvent être collectées/générées qu'une seule fois ;
- d'avoir une source d'**inspiration** pour d'autres mesures ou d'autres types de recherche ;
- de trouver des exemples de matériel pour l'**enseignement**.

Où trouver des données ?

De nombreuses options sont disponibles quant il s'agit de trouver des données de recherche. Voici une liste non-exhaustive – par ordre alphabétique - de sites existants, tous domaines de recherche confondus :

Dépôts de données de recherche		
BORIS Portal (Bern Open Repository and Information System)	EU Open Research Repository	Recherche Data Gouv (FR)
SWISSUbase (sciences sociales et linguistique, suite de FORSbase)	Yareta (Dépôt d'archives pour les hautes écoles genevoises)	Zenodo (CERN)
Plateformes d'accès aux publications et jeux de données		
BASE (Bielefeld Academic Search Engine)	OpenAIRE Explore (UE)	
Métamoteur (va chercher les informations dans plusieurs dépôts et plateformes existants)		
DataCite Commons (Consortium universitaire international)	Dataset Search (Google)	Dimensions (Digital science)
Octopus (UK)		Mendeley Data (Elsevier)
Open Government Data (OGD)		
data.europa.eu (EU)		data.gouv.fr (FR)
data.gov (USA)		opendata.swiss (CH)

Le répertoire de dépôts [re3data.org](#)  permet de trouver d'autres dépôts, notamment par discipline.

Comment évaluer un jeu de données ?

Avant d'utiliser un jeu de données, il faut prendre le temps d'évaluer son contenu. Voici quelques questions clés à se poser :

- **Fiabilité** de la source : qui a compilé les données ?
- **Conditions de production** : comment et pourquoi/dans quel but les données ont-elles été collectées/générées ?
- **Organisation** des données : les données sont-elles structurées et documentées ?
- **Accessibilité** : les données sont-elles réellement accessibles, c'est-à-dire téléchargeables ?
- **Réutilisation** : quelles sont les conditions de réutilisation (licence) ?
- **Compatibilité** : est-il possible de combiner ou enrichir ces données avec d'autres ensembles ?



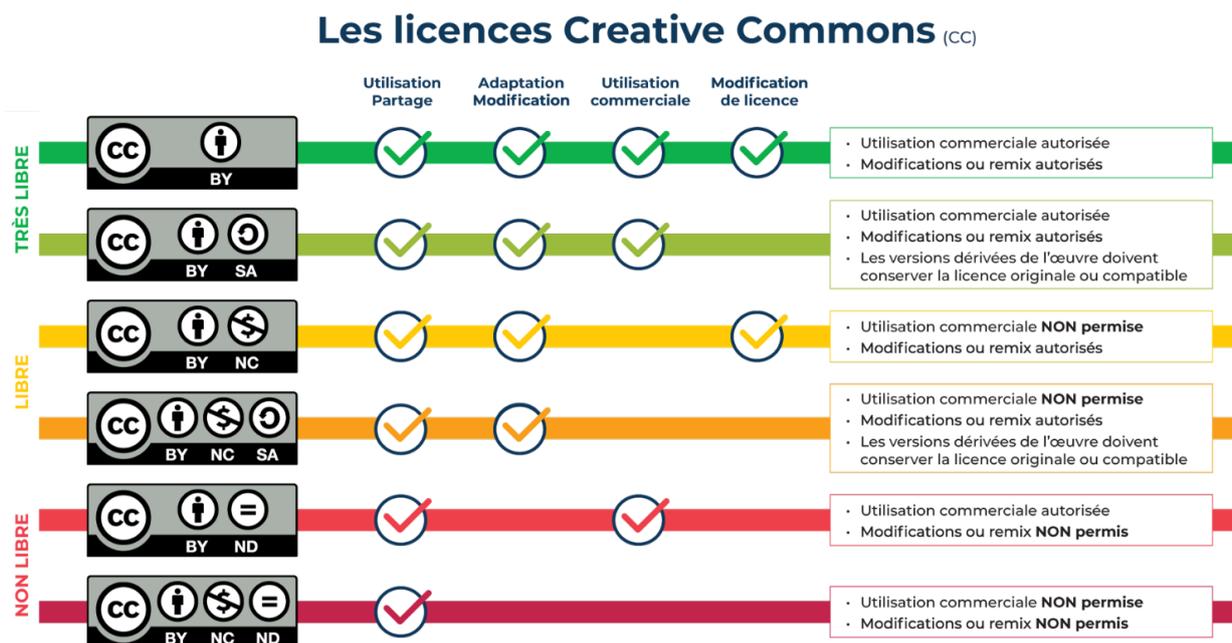
Le jeu de données doit être accompagné d'une documentation et de métadonnées complètes, qui décrivent notamment le projet, ses objectifs et méthodes. Le respect des principes FAIR (Faciles à trouver, Accessibles, Interopérables, Réutilisables) permet d'avoir une pré-évaluation plutôt positive du jeu de données.

Puis-je utiliser le jeu de données comme je le souhaite ?

Comme mentionné plus haut, il est indispensable de faire usage d'un jeu de données téléchargé en respectant ses conditions de réutilisation : celles-ci sont le plus souvent explicitées par une licence, c'est-à-dire un droit d'utiliser l'œuvre sous certaines conditions.

La licence est le plus souvent visible sur la page de téléchargement du jeu de données. Si ce n'est pas le cas, ce sont les conditions générales du dépôt qui s'appliquent.

Voici un exemple de licence souvent utilisée, les licences Creative Commons :



Comment citer un jeu de données ?

Il est essentiel de citer et d'attribuer correctement le travail des autres. La citation doit contenir au minimum les informations suivantes :

- Auteur.e(s)
- Titre du jeu de données
- Année de publication
- Dépôt de données de recherche
- Version du jeu (si indiquée)
- Identifiant persistant (DOI par ex.)



Si le jeu de données a un DOI, il est possible de créer une citation en utilisant l'outil [DOI Citation Formatter](#).

La méthode de citation (ordre, ponctuation) dépend du style de citation utilisé. Parfois, le dépôt indique comment les données doivent être citées.



En savoir plus

Pourquoi réutiliser les données de recherche ?

