

Comment citer un jeu de données ou un logiciel dans ses publications ?

 opendscience.pasteur.fr/2025/04/24/comment-citer-un-jeu-de-donnees-ou-un-logiciel-dans-ses-publications/

CeRIS - Institut Pasteur

24 avril 2025

[Mise à jour d'un article initialement publié en avril 2021]

Vous avez trouvé un jeu de données intéressant dans un entrepôt et vous l'avez réutilisé pour produire un résultat scientifique ? Vous avez généré des données à l'aide d'un logiciel développé par d'autres scientifiques ? Dans ce cas, il est important de **citer de manière univoque ces données et logiciels dans vos publications**. Ceci permettra d'en identifier le (ou les) auteur(s), d'en faciliter l'accès et de permettre la reproductibilité de vos résultats et analyses.

Mais comment faire concrètement ?

Les jeux de données et les logiciels étant considérés comme des **produits de recherche légitimes et citables**, ils doivent être cités dans le texte et inclus dans la **liste complète des références**, au même titre que les autres résultats de recherche (articles, livres, thèses...).

La forme de la citation peut varier (par exemple en fonction des recommandations du journal dans lequel vous souhaitez publier), mais elle doit inclure les éléments suivants :

- **Créateur(s) / auteur(s)**,
- **Nom du jeu de données ou du logiciel**,
- **Date** de publication ou de diffusion,
- **Localisation** : entrepôt ou archive dans lequel le jeu de données ou le logiciel est accessible,
- **Versión** utilisée ou sous-ensemble utilisé,
- **Identifiant** fourni par l'entrepôt ou URL s'il n'y a pas d'identifiant.

Nous recommandons également d'inclure la **licence** associée au jeu de données ou au logiciel.

Exemples :

Citation d'un jeu de données dans une publication :

- Citation au sein du texte : « Les graphiques de la figure X montrent la distribution des mesures sélectionnées à partir des données principales [Auteur(s), Année, sous-ensemble réutilisé] »

- Citation dans la liste des références : Auteur(s), Année, Titre du jeu de données, Entrepôt de données, Version, Identifiant/URL.

Retrouvez dans ce guide des exemples concrets de citation de jeux de données et dans cet article des exemples de citation de logiciels.

Points d'attention :

Si vous avez réutilisé un **sous-ensemble** d'un jeu de données ou d'un logiciel, il est important de préciser les informations permettant au lecteur de retrouver l'élément concerné.

- Certains entrepôts attribuent un **identifiant spécifique à chaque fichier** d'un jeu de données (c'est le cas de l'entrepôt Recherche Data Gouv qui attribue un DOI à chaque fichier), vous pouvez donc utiliser cet identifiant, en complément de la référence complète du jeu de données.
- Si vous souhaitez citer un élément d'un logiciel présent dans **Software Heritage** (répertoire, code, révision...), vous pouvez facilement générer un *SoftWare Heritage persistent IDentifiers (SWHIDs)* pour le désigner précisément (nous vous en parlons ici).

Si un jeu de données ou un logiciel existe en **plusieurs versions**, assurez-vous de citer la bonne.

- Pour certains entrepôts, le **DOI change à chaque version** (c'est le cas de Zenodo par exemple, voir ce dataset). Il est donc possible de citer précisément la version utilisée.
- Pour d'autres (comme Recherche Data Gouv par exemple), le DOI ne change pas. Il est donc indispensable d'indiquer le **numéro de version** dans la citation.

Pour aller plus loin :

- Data Citation Synthesis Group: **Joint Declaration of Data Citation Principles**. Martone M. (ed.) San Diego CA: FORCE11; 2014 <https://doi.org/10.25490/a97f-egy>
- Smith AM, Katz DS, Niemeyer KE, FORCE11 Software Citation Working Group. 2016. **Software citation principles**. PeerJ Computer Science 2:e86 <https://doi.org/10.7717/peerj-cs.86>