

Comment partager des données associées à une publication ?

 opendscience.pasteur.fr/2025/02/03/comment-partager-des-donnees-associees-a-une-publication/

CeRIS - Institut Pasteur

3 février 2025

[Mise à jour d'un article initialement publié en septembre 2022]

Vous êtes en train de soumettre un article à une revue et on vous demande de partager les données associées ? Quelles solutions s'offrent à vous et comment faire ?

La première étape consiste à **identifier les données qu'il est nécessaire d'associer à votre publication**. Pour cela, passez en revue votre manuscrit et notamment les figures/tableaux/graphiques, en vous posant ces questions : quelles sont les données qui ont été utilisées pour générer ce résultat ? Quelles sont les données qui permettent de comprendre et/ou reproduire ce résultat ? Quelles données supplémentaires pourraient être utiles à la compréhension de l'étude et de son contexte ?

Une fois que vous avez identifié les données à partager, demandez-vous s'il existe des **contraintes juridiques ou éthiques** qui pourraient empêcher ou limiter le partage de ces données. Pour en savoir plus, retrouvez notre article sur le sujet.

Ensuite, pour partager vos données, les revues vous proposent généralement deux solutions :

1. Envoyer les données directement à la revue de façon à ce qu'elles apparaissent sous forme de **supplementary data** ou **supplementary materials** ;
2. Déposer les données dans un ou plusieurs **entrepôt(s) de données** (nous vous expliquons ce que c'est ici et comment en trouver un ici).

Nous vous conseillons de **privilégier la deuxième solution**. Pourquoi ?

- Pour que votre jeu de données constitue un **produit de recherche** citable (grâce à l'identifiant attribué par l'entrepôt) et trouvable via des moteurs de recherche ;
- Pour que votre jeu de données soit **compréhensible et réutilisable** : un entrepôt de données vous permet de décrire précisément le jeu de données, par des métadonnées structurées et par de la documentation. De plus, vous pouvez déposer dans l'entrepôt toutes les données qui vous semblent utiles à la compréhension de l'étude et de son contexte, même si elles ne sous-tendent pas directement l'article.
- Pour que votre jeu de données soit **accessible sur le long terme**, dans un environnement non-commercial (indépendant des acteurs commerciaux de l'édition scientifique) ;

- Pour pouvoir **contrôler l'accès aux données** si nécessaire (données accessibles sur demande ou via un *Data Access Committee* par exemple) ou leur attribuer une **licence de diffusion** différente de celle de l'article. De plus, la plupart des entrepôts vous permettent de conserver votre jeu de données en accès privé avant publication de l'article associé et de fournir un lien privé aux *reviewers* pour qu'eux seuls puissent accéder au jeu de données.

Attention, certains éditeurs peuvent vous proposer de déposer vos données dans **leur propre entrepôt**. Retrouvez dans cette note de décryptage produite par le GTSO Données de Couperin les **risques encourus** lors du dépôt de données dans un **entrepôt d'éditeur commercial**.

Et ensuite ?

Il vous reste désormais à indiquer dans l'article la nature des données partagées, la façon d'y accéder (en particulier l'identifiant ou l'url permettant d'accéder aux données déposées dans un entrepôt), les éventuelles restrictions d'accès, etc. Ceci prend généralement la forme d'un **Data Availability Statement** et vous trouverez des conseils pour le rédiger dans cet article.

Pour aller plus loin : Partager les données liées aux publications scientifiques – Guide pour les chercheurs, Ouvrir la Science, licence CC-BY