

Distinguer dans un article les jeux de données produits des jeux de données cités, en 5 points

1. Intérêt de mentionner les jeux de données dans un article
2. Mention des jeux de données produits pour l'article
3. Mention des jeux de données cités par l'article
4. Mention des jeux de données dans les bases bibliographiques
5. Prendre en compte les données dès la rédaction de l'article

Liens utiles

1. Intérêt de mentionner les jeux de données dans un article

Un article de recherche rend compte de résultats originaux issus de travaux scientifiques. Les données produites et sous-tendant (on dit aussi « étayant ») les résultats présentés dans l'article peuvent exister sous forme de jeux de données (*dataset*), bases de données (*database*), logiciels, codes source, modèles, etc. (voir la fiche CoopIST [S'initier en ligne aux données de la recherche et à leur gestion](#)).

Les éditeurs de revues, les agences de financement, les institutions exigent de plus en plus que ces données produites dans le cadre d'un travail de recherche aboutissant à une publication soient rendues publiques et accessibles aussi librement que possible. Cette ouverture répond à plusieurs objectifs : comprendre et évaluer la recherche menée, permettre sa reproductibilité, réutiliser les données, appliquer les principes de transparence et de respect d'intégrité scientifique.

A côté des données produites au cours de ce travail, d'autres données ont pu être une source d'information ou d'inspiration : par exemple pour définir l'objectif ou le point de départ de la recherche, choisir la méthode ou analyser les résultats. Ces autres données doivent être citées comme toutes les publications ayant inspiré ce travail.

Pour l'auteur comme pour le lecteur, les mentions des données produites et des données source d'inspiration doivent être lisibles sans être confondues.

2. Mention des jeux de données produits pour l'article

Les conseils aux auteurs des revues indiquent de plus en plus comment mentionner les jeux de données étayant les résultats de l'article. Dans l'article, cette mention peut faire l'objet d'une section spécifique, ou s'insérer dans une section préexistante comme *Supplementary Materials*, ou combiner les deux, ou encore ne relever d'aucune section spécifique.

Certaines revues ou sites d'éditeurs proposent des modèles pour différents cas de jeux de données. Exemple : le tableau [Sample Data Availability Statements](#) de l'éditeur Oxford Academic.

Lorsqu'une section spécifique de l'article est dédiée à la déclaration de disponibilité des données produites [*Data Availability Statements*, *Data Access Statement (DAS)*, *Data Availability*, *Data Access*, *Open Research...*], elle indique si ces données sont accessibles ou non, et si oui, où et à quelles conditions. Lorsque les données sont accessibles, les auteurs doivent mentionner les liens adéquats vers les entrepôts ou les fichiers concernés. Généralement, la mention de l'entrepôt où sont déposées les données est également demandée, avec ou sans lien.

Exemples d'éditeurs sans section spécifique pour les jeux de données dans un article :

Oxford Academic

Conseils aux auteurs [Data availability statement](#)

Exemple d'article : <https://doi.org/10.1093/sysbio/syy070> ; jeux de données mentionnés dans la section *Results* via un lien de type *Supplementary Material* permettant d'accéder aux jeux de données dans l'entrepôt Dryad.

Taylor & Francis

Conseils aux auteurs [Writing a data availability statement](#) (Home > Sharing and citing data > How to share your data > Writing a data availability statement)

Exemple d'article : <https://doi.org/10.1080/10256016.2021.1946532> ; jeu de données accessible dans la section *Supplemental* avec un lien via l'icône et le nom de l'entrepôt (Figshare) hébergeant le jeu.

Exemples d'éditeurs avec une section spécifique pour les jeux de données dans un article :

BioMedCentral, Part of Springer Nature

Conseils aux auteurs [Availability of data and materials](#) (Editorial Policy > availability of data and materials)

Exemple d'article : <https://doi.org/10.1186/s13071-020-04522-3> ; jeux de données mentionnés dans la section *Availability of data and materials* via leurs identifiants de la database GenBank ; l'accès aux notices des jeux de données nécessite d'ouvrir la base et d'y rechercher les identifiants.

Elsevier

Conseils aux auteurs [Data Availability Statement](#)

Exemple d'article : <https://doi.org/10.1016/j.compag.2020.105624> ; jeux de données mentionnés dans la section *Research data for this article* via leurs liens vers l'entrepôt Mendeley Data.

Nature

Conseils aux auteurs [Availability of data](#)

Exemple d'article : <https://doi.org/10.1038/s41467-020-19818-2> ; la section *Data availability* renvoie vers les données génomiques dans la base GISAID et vers un fichier de *Supplemental Data*, la section *Code availability* vers les codes via le lien dans le GitHub.

PLoS

Conseils aux auteurs [Data Availability](#)

Exemple d'article : <https://doi.org/10.1371/journal.pntd.0008821> ; son encart *Data Availability* renvoie aux jeux de données dans Zenodo via leurs liens DOI.

Springer Open

Conseils aux auteurs [Data Availability Statements: Guidance for authors and editors](#)

Exemple d'article : <https://doi.org/10.1186/s12284-022-00559-3> ; sa section *Availability of Data and Materials* renvoie vers les données et scripts via un lien dans le GitHub.

Wiley

Conseils aux auteurs [Sharing and Citing your Research Data](#) (Author Resources > Journal Authors > Open Access > Data Sharing & Citation > [Data Sharing Policy](#))

Exemple 1 d'article : <https://doi.org/10.1002/ece3.5046> ; sa section *Data Accessibility* renvoie aux données via leur lien DOI vers l'entrepôt Movebank Data Repository.

Exemple 2 d'article : <https://doi.org/10.1111/gcb.16302> ; sa section *Open Research/Data Availability Statement* renvoie vers le lien DOI de la database dans l'entrepôt Zenodo et vers le lien de ses mises à jour dans le GitHub.

3. Mention des jeux de données cités par l'article

Comme les publications citées par un article, les jeux de données ayant inspiré le travail de recherche doivent être cités, d'une part par leur citation dans le texte de l'article, et d'autre part par la référence bibliographique complète dans la liste des références de la publication (*Bibliography, References*).

Pour les formats de références, reportez-vous à la fiche CoopIST [Citer un jeu de données scientifiques](#).

A noter que l'éditeur Wiley ([Wiley's Data Citation Policy](#)) demande aux auteurs d'ajouter dans le manuscrit l'indication [dataset] avant la référence du jeu de données cité afin qu'il soit identifié lors de l'évaluation de l'article. Cette indication est ensuite enlevée dans l'article final publié.

4. Mention des jeux de données dans les bases bibliographiques

L'agence et base de données [Crossref](#) d'enregistrement d'identifiant numérique pérenne DOI (Digital Object Identifier) d'une publication demande aux éditeurs, lors de l'enregistrement du DOI et des métadonnées (auteur, titre, éditeur...) de leurs publications, de préciser la relation entre une publication et chacun des jeux de données, logiciels, etc., liés :

- si la publication est associée à son jeu de données, la mention de celui-ci doit être introduite dans les métadonnées saisies par *isSupplementedBy* suivi de l'identifiant DOI ou du lien URL du jeu de données ;
- si la publication cite un jeu de données comme source d'information, le lien de la publication au jeu de données doit être introduite par la mention *References*. Si la publication liste dans sa bibliographie les jeux de données cités, ceux-ci peuvent être déclarés dans le champ *Citation_list*.

La base de données bibliographique commerciale [WoS Core Collection](#) (Clarivate) signale parfois le lien vers les jeux de données associés dans les notices des publications qu'elle indexe :

- lors d'une recherche bibliographique, le filtre *Associated Data* permet de limiter les résultats aux publications associées à des données ;
- sur la page d'affichage de la notice de publication, le lien *Associated data* en haut de la notice ouvre une nouvelle page avec la description et le lien d'accès au jeu de données.

La base de données bibliographique commerciale [Scopus](#) (Elsevier), accessible gratuitement via le portail [Research4life](#) aux institutions de recherche des pays les moins avancés, signale le lien vers les jeux de données associés dans les notices des publications indexées :

- pas de possibilité de recherche bibliographique par un filtre de présence de jeux de données associés ;
- mais sur la page d'affichage d'une notice de publication indexée comme étant associée à des jeux de données, le lien *Related research data* à droite de la référence ouvre dans une nouvelle fenêtre le lien vers la référence du jeu de données.

Certaines archives ouvertes de publications enregistrent les liens (DOI, URL) des jeux de données associés aux publications déposées. Exemple de l'archive institutionnelle du Cirad [Agritrop](#) :

- possibilité de recherche par le DOI et/ou l'URL du ou des jeux de données associés aux publications ;

- la page de la notice d'une publication affiche, quand ils existent, le ou les liens des jeux de données associés (voir l'aide en ligne : https://agritrop.cirad.fr/aide_jeux_donnees/).

5. Prendre en compte les données dès la rédaction de l'article

Voici une proposition de démarche en 6 points pour prendre en compte les données avant et pendant la rédaction de l'article scientifique :

- rechercher dans les conseils aux auteurs de la revue les recommandations pour mentionner et citer des jeux de données dans un article ; si absentes, se reporter à la fiche CoopIST [Citer un jeu de données scientifiques](#) ;
- prendre connaissance de la politique générale de la revue ou de l'éditeur, si elle existe, en matière de données de recherche. Cette politique peut être indiquée dans la rubrique de politique de partage d'information ou de libre accès (*Open Policy, Open Access, Sharing, etc.*) ;
- considérer et mentionner chaque jeu de données (codes sources, logiciels, bases de données, modèles) produit au cours du travail de recherche et sous-tendant les résultats ;
- citer les jeux de données externes – qui ont été source d'inspiration pour définir l'objectif de la recherche, choisir les méthodes, ou comparer avec les résultats de l'article ;
- mentionner la citation dans le texte et la référence dans la section dédiée (exemples : section *Data availability* s'il s'agit d'une référence d'un jeu de données étayant l'article ; liste des *References* s'il s'agit d'une référence d'un jeu de données cité) ;
- une fois l'article publié, si celui-ci doit faire l'objet d'un dépôt par l'auteur dans une archive ouverte (institutionnelle, thématique, nationale ou internationale), saisir dans les champs prévus (s'ils existent) les métadonnées demandées (comme l'URL ou le DOI) du ou des jeux de données associés étayant les résultats de l'article.

Liens utiles

Crossref. 2022. Data and software citation deposit guide.

<https://www.crossref.org/documentation/reference-linking/data-and-software-citation-deposit-guide/>

Ministère de l'enseignement supérieur, de la recherche et de l'innovation. 2022. Partager les données liées aux publications scientifiques. Guide pour les chercheurs. Paris (FRA) : MESRI, 9 p.

<https://www.ouvrirelascience.fr/partager-les-donnees-liees-aux-publications-scientifiques-guide-pour-les-chercheurs/>

Marie-Claude Deboin, Frédérique Rambour

Délégation à l'information scientifique et technique, Cirad

13 octobre 2022

Comment citer ce document : Deboin Marie-Claude, Rambour Frédérique. 2022. Distinguer dans un article les jeux de données produits des jeux de données cités, en 5 points. Montpellier (FRA) : CIRAD, 4 p. <https://doi.org/10.18167/coopist/0079>

Cette œuvre est mise à disposition selon les termes de la Licence Creative Commons : Attribution - Pas d'Utilisation Commerciale - Partage dans les Mêmes Conditions 4.0 International, disponible en ligne : <http://creativecommons.org/licenses/by-nc-sa/4.0/deed.fr> ou par courrier postal à : Creative Commons, 171 Second Street, Suite 300, San Francisco, California 94105, USA. Cette licence vous permet de remix, arranger, et adapter cette œuvre à des fins non commerciales tant que vous créditez l'auteur en citant son nom et que les nouvelles œuvres sont diffusées selon les mêmes conditions.