

Publier un *Data paper*,

en 5 points

1. Qu'est-ce qu'un <i>Data paper</i>	1
2. Objectif et enjeu du <i>Data paper</i>	2
3. Structure et contenu du <i>Data paper</i>	2
4. Choisir la revue	2
5. Evaluer un <i>Data paper</i>	4
Liens utiles.....	4

1. Qu'est-ce qu'un *Data paper*

Un *Data paper* (autres dénominations : *Data descriptor*, *Data article*, *Data Briefs*, *Resource Announcements*, *Data Resource Profile*) est un article scientifique évalué par les pairs et citable. Il décrit un jeu de données (*data*, *data set*), la méthode ayant permis de l'obtenir et le potentiel de réutilisation de ce jeu. Le *Data paper* ne décrit pas de résultats de recherche et ne contient ni discussion, ni conclusion.

Publier un *Data paper* implique de rendre accessible le jeu de données :

- le mieux est de le déposer dans un entrepôt de données (*data repository*) ; le lien (URL, DOI) vers les données et le nom de l'entrepôt sont mentionnés dans le *Data paper*. Il existe tout un panel d'entrepôts, institutionnels, pluridisciplinaires, disciplinaires ou thématiques, dans lesquels les données peuvent être déposées (voir la fiche CoopIST [Déposer des données de recherche dans un entrepôt](#)) ;
- certaines revues acceptent d'associer directement les données à l'article (à l'intérieur de l'article ou en fichiers supplémentaires). Cette pratique est déconseillée car peu adaptée à la réutilisation des données.

Le *Data paper* est ouvert à tous types de données, brutes ou dérivées (observations, mesures, enquêtes, séquences, cartes, images, simulations informatiques, etc.), à condition que le jeu de données constitué ait une valeur scientifique, ou un impact environnemental, sociétal, économique, culturel.

Ce type d'article informe la communauté scientifique de l'existence, de l'originalité, de la qualité et de la disponibilité d'un jeu de données. Il valorise le travail de ses auteurs en expliquant l'importance des données produites et leur potentiel de réutilisation pour de futures recherches.

A noter : les *Database papers* sont plutôt considérés comme des articles de recherche car ils doivent présenter au moins une application de la base de données. Les *Software papers* sont l'équivalent des *Data papers* pour publier des codes, scripts ou logiciels.

2. Objectif et enjeu du *Data paper*

L'objectif du *Data paper* est de mettre à disposition des chercheurs des données de qualité, compréhensibles, interprétables et réutilisables.

Publier un *Data paper* s'inscrit dans la dynamique mondiale d'ouverture de la science (*Open science*) et répond aux [principes FAIR](#) qui préconisent que les données soient faciles à trouver, accessibles, interopérables et réutilisables. Par exemple, en France, le [décret n° 2021-1572 du 3 décembre 2021](#) relatif au respect de l'intégrité scientifique demande aux établissements publics de recherche de promouvoir « la mise à disposition des méthodes et protocoles, des données et des codes sources associées aux résultats de recherche afin d'en garantir la traçabilité et la reproductibilité ».

L'enjeu du *Data paper* est la réutilisation des données rendues publiques. Cela implique de mettre en œuvre de bonnes pratiques pour que les données soient décrites et documentées correctement, c'est-à-dire selon les standards disciplinaires reconnus et pratiqués par les différentes communautés scientifiques.

3. Structure et contenu du *Data paper*

Le *Data paper* est généralement structuré en 3 points :

- description du jeu de données et de son contexte,
- description des méthodes d'obtention,
- explication du potentiel de réutilisation des données.

Les instructions aux auteurs de la revue proposent souvent un modèle (*template*) de *Data paper*.

Le *Data paper* doit contenir tous les éléments décrivant le jeu de données pour qu'il puisse être compris et réutilisé par une personne extérieure à l'étude : type de données, lieu et contexte de collecte, protocoles expérimentaux, équipements, variables étudiées, format de fichiers, standards de métadonnées ainsi que toutes informations explicitant les tableaux de données (contenus des lignes et des colonnes, signification des abréviations, unités de mesure).

Si certaines de ces informations sont déjà publiées, par exemple les protocoles, il suffit de mentionner le nom du protocole avec la référence qui lui donne accès. Le *Data paper* peut aussi renvoyer à un protocole accessible dans l'entrepôt où est déposé le jeu de données : dans l'entrepôt, il est en effet possible d'associer un (ou des) fichier(s) de données avec un (ou des) fichiers texte, par exemple un fichier *Read me* contenant de multiples informations, le fichier Questionnaire d'enquête, le fichier Dictionnaire des variables, le protocole d'obtention, etc.

Le *Data paper* explique pourquoi le jeu de données est important, en quoi il peut être utile à d'autres chercheurs ou à d'autres acteurs (société civile, décideurs, acteurs privés, organisations non gouvernementales...) et comment il pourrait être réutilisé.

Pour vérifier si votre *Data paper* est complet, relisez-le en contrôlant qu'il met à la disposition des lecteurs tous les éléments pour comprendre et interpréter sans ambiguïté les fichiers de données ; contrôlez également que ces fichiers soient accessibles dans un format ouvert.

4. Choisir la revue

Le choix de la revue doit être réfléchi avant de se lancer dans la rédaction du *Data paper*. La revue détermine le modèle de *Data paper*, oriente vers certains entrepôts de données et précise, le cas échéant, le format et les standards disciplinaires à appliquer.

Dans un premier temps, ce choix se fait selon le domaine scientifique et le lectorat (vos futurs utilisateurs) et, dans un second temps, après avoir lu les instructions aux auteurs des revues et pris en compte les critères décrits ci-après (voir aussi la fiche CoopIST [Choisir la revue où publier un article](#)).

De nombreuses revues publient des *Data papers* :

- elles sont multidisciplinaires, disciplinaires, ou thématiques ;
- elles peuvent être dédiées (*Data journal* ne publiant que des *Data papers*) ou classiques (publiant différents types d'articles dont des *Data papers*) ;
- comme toute revue, elles sont en libre accès ou non, avec ou sans frais de publication (*Article processing charges*, APC), et indexées par diverses bases de données bibliographiques.

Voici quelques critères spécifiques à la publication d'un *Data paper*.

Echelle du jeu de données. Certaines revues ne publient que des jeux de données à grande échelle (à une large échelle géographique, ou couvrant une espèce entière, ou un génome complet) ayant un intérêt pour une large communauté, ou ayant de nombreuses possibilités de réutilisation et d'application dans divers domaines de recherche (exemples de revues : *Global Ecology and Biogeography*, *GigaScience*, *Nature Biotechnology*).

Structure du *Data paper*. Certaines revues imposent un modèle structuré par un standard de métadonnées disciplinaires (exemples de revues : *Ecology*, *Open Health Data*, *Annals of Forest Science*). D'autres ont un modèle simple (exemples : *Data in Brief*, *Geoscience Data Journal*, *Scientific Data*). Certaines ne proposent pas de modèle, la rédaction est libre mais le *Data paper* est limité à 2 à 3 pages ou 1000 à 2000 mots (exemples : *Phytopathology*, *Hydrological Processes*, *European Journal of Lipid Science and Technology*). Quelques revues publient des articles très complets décrivant des données (exemples : *Water Resources Research*, *International Journal of Epidemiology*), incluant parfois les parties *Results*, *Discussion*, *Conclusion* (exemples : *Ethnobiology Letters*, *The Plant Journal*).

Localisation des données. Le dépôt des données dans un entrepôt (*data repository*) digne de confiance est la pratique recommandée ; de plus en plus de revues ayant adopté les [Transparency and Openness Promotion \(TOP\) Guidelines](#) l'encouragent. Généralement, la revue recommande aux auteurs un choix d'entrepôts. Certaines revues disposent d'un entrepôt mais rares sont celles qui imposent un entrepôt précis. Certaines revues proposent encore d'associer les données directement à l'article, mais cette pratique est déconseillée.

Modalités de diffusion des données. En plus de la licence de diffusion appliquée aux articles par la revue (voir fiche CoopIST [Connaître et utiliser les licences Creative Commons](#)), certaines revues imposent une licence de diffusion pour les données (exemples : *Biodiversity Data Journal*, *Scientific Data*), ce qui peut limiter le choix de l'entrepôt de données. Ce sont souvent les licences de type **CC0** (pas d'obligation de citer les auteurs) ou **CC-BY** (obligation de citer les auteurs). D'autres revues laissent le choix de la licence, avec la possibilité, par exemple, d'exclure une utilisation commerciale des données (cas de la licence **CC-BY-NC** et dérivées). Certaines revues acceptent une période d'embargo sur l'accès aux données pour laisser le temps aux auteurs de publier un article de recherche.

Critères d'évaluation par les pairs. Souvent précisés dans les instructions aux auteurs, ils peuvent aussi constituer un critère de choix de la revue (voir [point 5](#)).

Pour vous aider, le site [Où publier](#) du Cirad regroupe plus de 2300 revues interrogeables par thèmes scientifiques et par divers critères dont les « types d'articles ».

Sur ce site CoopIST, vous pouvez consulter une [sélection de revues publiant des *Data papers*](#) régulièrement actualisée. Trois autres listes sont disponibles : site du [Système mondial d'information sur la biodiversité](#), Wiki de l'[University of Edinburgh](#) et Wiki de [Forschungsdaten.org](#).

Attention : si vous souhaitez aussi publier un article de recherche lié à votre jeu de données, vérifiez qu'un *Data paper* préalable ne soit pas un frein. Pour cela, consultez les instructions aux auteurs des

revues ciblées pour le *Data paper* et pour l'article de recherche. Par exemple, [Scientific Data](#) et [F1000research](#) précisent les revues qui acceptent la publication préalable d'un *Data paper*.

5. Evaluer un *Data paper*

Tous les *Data papers* sont examinés par les pairs avant publication mais les critères d'évaluation varient selon la revue. Ils sont plus ou moins approfondis selon qu'ils concernent uniquement le texte du *Data paper*, ou qu'ils s'intéressent aussi aux fichiers de données (organisation, format et accès), au dépôt dans l'entrepôt, à l'organisation des métadonnées et parfois des données, et, dans certains cas, aux données elles-mêmes (exemples : [Geoscience Data Journal](#), [Earth System Science Data](#)).

Voici les principaux critères d'évaluation :

- la portée scientifique ou sociétale des données dans le champ scientifique de la revue ;
- la rigueur et l'adéquation de la méthode par rapport aux pratiques de la discipline ;
- l'exhaustivité de la description permettant à d'autres chercheurs de comprendre, d'interpréter et de réutiliser les données ;
- la qualité, l'intégrité et la fiabilité des données ;
- la conformité avec les standards disciplinaires ;
- le potentiel de réutilisation des données.

En exemples, voir les *Instructions for reviewers* de la revue [Ecology](#) ou les *Guidelines for Reviewers* de [Journal of Environmental Quality](#).

Liens utiles

CINES, 2017. Les formats de fichier. <https://www.cines.fr/archivage/des-expertises/les-formats-de-fichier/>

CNRS, 2021. Guide de bonnes pratiques sur la gestion des données de recherche. Publier un Datapaper pour valoriser et expliciter les données. <https://mi-gt-donnees.pages.math.unistra.fr/guide/07-publier.html>

CNRS, 2020. Groupe de travail *Atelier Données* : Foire aux questions. <https://mi-gt-donnees.pages.math.unistra.fr/site/FAQ.html>

DoRANum, 2018. La minute *Publier un Data paper*. <https://doi.org/10.13143/4mhn-mq42>

DoRANum, 2020. Format ouvert ou fermé ? <https://doi.org/10.13143/mcwq-qs64>

DoRANum, 2020. Webinaire *Data paper* - Une incitation à la qualification et à la réutilisation des jeux de données. <https://doranum.fr/2020/11/20/supports-du-webinaire-data-paper-une-incitation-a-la-qualification-et-a-la-reutilisation-des-jeux-de-donnees/>

Software Sustainability Institute, 2021. In which journals should I publish my software? <https://www.software.ac.uk/which-journals-should-i-publish-my-software>

Laurence Dedieu

Délégation à l'information scientifique et technique, Cirad

Octobre 2014, dernière mise à jour 6 janvier 2022

Comment citer ce document :

Dedieu L., 2022. Publier un *Data paper*, en 5 points. Montpellier (FRA) : CIRAD, 5 p. <https://doi.org/10.18167/coopist/0057>

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 remixer, 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.