

Publier un data paper

Les « data journals » sont des revues dont les sujets de publication sont exclusivement consacrés aux problématiques de création et de gestion des données. Publier dans un data journal, est une voie de valorisation de ses jeux de données et une manière d'augmenter sa citabilité (Thomson Reuters Data Citation Index).

La structuration d'un data paper

Une partie descriptive du ou des jeux de données brutes qui compose le « data paper »

Cette partie se présente comme un article scientifique avec une introduction présentant le contexte et les questions de recherche posées, une partie matériel et les méthodes qui ont permis l'obtention des données, une description des données et des conseils pour la réutilisation.

Une partie composée des données mises à disposition sous forme de fichiers associés (Excel par exemple) aux articles (le data paper proprement dit)

- Les jeux de données sont déposés dans des entrepôts de données dans des formats répondant à des normes et standards internationaux reconnus, avec un certain niveau de description (métadonnées).
- Les données sont publiées sous la forme de flux structuré et lisible par une machine (XML par exemple) sur une plateforme qui permet de manipuler les jeux de données (ex : IPOL <http://www.ipol.im/>).

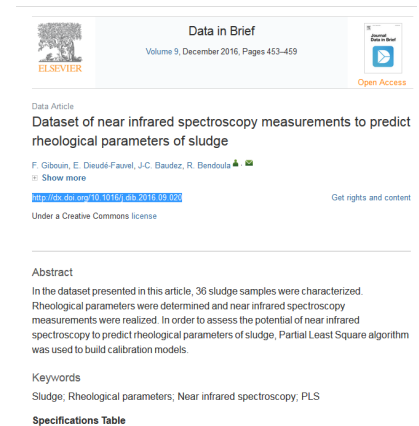
Comment procéder : soumettre son "data paper" en 5 étapes

Lire attentivement les instructions aux auteurs

En effet, elles comportent toutes les indications de publication du jeu de données (base de dépôts, métadonnées recommandées...).

Identifier la plateforme de dépôt et déposer ses jeux de données

En général, le dépôt dans les entrepôts de données est payant mais le tarif reste modeste (quelques centaines d'euros). Les bases de dépôt de données sont en général d'accès libre (« Open repositories »). De plus, les entrepôts de données proposent un lien entre article et donnée en attribuant un identifiant (en général un DOI). Un schéma de métadonnées est à respecter pour chaque entrepôt, il comporte les instructions de dépôts, et les formats et les standards utilisés. De manière générale, le standard de description de jeux de données est le Dublin Core. Voir fiche sur Partager ses données.



The screenshot shows a 'Data in Brief' article on the Elsevier platform. The article title is 'Dataset of near infrared spectroscopy measurements to predict rheological parameters of sludge'. The authors listed are F. Giboun, E. Druod-Favez, J.C. Bauder, R. Bendoua. The article is under a Creative Commons license. The abstract states: 'In the dataset presented in this article, 36 sludge samples were characterized. Rheological parameters were determined and near infrared spectroscopy measurements were realized. In order to assess the potential of near infrared spectroscopy to predict rheological parameters of sludge, Partial Least Square algorithm was used to build calibration models.' The keywords are 'Sludge; Rheological parameters; Near infrared spectroscopy; PLS'. There is a 'Specifications Table' link at the bottom.

En savoir plus

Trouver un data journal

Trac open source project : <http://proj.badc.rl.ac.uk/prepade/bl/og/DataJournalsList>

Exemple de data journals

[Data in brief](#) (Elsevier)
[Geoscience Data Journal](#) (Wiley)
[Scientific Data](#) (Nature)
[Biodiversity Data Journal](#) (Pensoft)

Soumettre à la revue (les fichiers à préparer peuvent diverger d'une revue à l'autre, il est absolument nécessaire de lire les instructions aux auteurs) :

- le fichier de son article ;
- le fichier de description de ses jeux de données dans un « Data Descriptor Template » qui comprend de manière générale les métadonnées suivantes : titre / auteurs, présentation des jeux de données, la description des méthodes d'enregistrements, versions... ;
- les fichiers des données (vérifier les formats recommandés).

Le processus classique de « peer reviewing »

Il est déclenché jusqu'à la validation ou le refus de l'article dans les mêmes conditions de publication qu'un article scientifique classique. Dans le cas d'un refus, les jeux de données déposés ne sont pas publiés.

La question des droits

Il faut distinguer les droits de cession de droits concernant :

- le data paper (contrat d'édition classique c'est le droit d'auteur qui s'applique) ;
- les jeux de données publiés en général en licence CC0 : La licence CC 0 définit une réutilisation libre des données mais elle ne fait pas renoncer à la législation sur le droit d'auteur donc l'auteur conservera ses droits sur les données produites même si elles sont sous licence CC0.

Pour en savoir plus :

<http://www.ist.agropolis.fr/istfiles/pdf/OAWeek15/L-DEDIEU-Publier-un-Data-paper-23-octobre-2015-OA-Montpellier.pdf>

Un exemple de data journal : Data in Brief

Deux méthodes pour soumettre un article : directement dans Data in brief ou bien la possibilité de soumettre un data article en même temps qu'un papier de recherche dans certaines revues (voir liste ci-après) : l'article pointe alors sur les données du data paper publié conjointement. Pour en savoir plus :

<http://www.journals.elsevier.com/data-in-brief/policies-and-guidelines/participating-elsevier-journals-data-in-brief/>

Les données sont jointes soit à l'article (si taille de fichiers pas trop élevée), soit dans des entrepôts (voir liste disponible) , soit dans Mendeley Data. En cas de publications de plusieurs data papers, les auteurs payent une fois pour 10 data papers.

La FAQ sur Data in brief est ici :

<http://www.journals.elsevier.com/data-in-brief/policies-and-guidelines/data-in-brief-faq>

Voir par exemple un data paper publié par des agents Irstea :

Dataset of near infrared spectroscopy measurements to predict rheological parameters of sludge ,F. Gibouin, E. Dieudé-Fauvel, J-C. Baudez, R. Bendoula, Data in Brief, Volume 9, December 2016, Pages 453-459. <http://dx.doi.org/10.1016/j.dib.2016.09.020>