RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

cirad
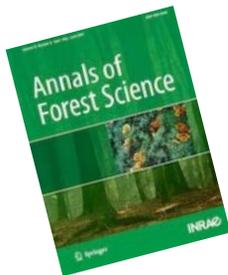LA RECHERCHE AGRONOMIQUE
POUR LE DÉVELOPPEMENT

Formation
Rédiger un *Data paper*

Laurence Dedieu
laurence.dedieu@cirad.fr

Délégation à l'Information Scientifique et Technique (DIST)

Plan de la formation

Concept et objectifs du *Data paper*

Modèles de *Data papers*

    Consultation de *Data papers* publiés

Vers 10h

Rédiger un *Data paper*

Rechercher des données et mise en pratique

Déjeuner entre 12 h 30 et 14 h

Revues publiant des *Data papers* + recherche de revues

Entrepôts de données + recherche d'entrepôts

Outils de rédaction de *Data papers*

A retenir

Pourquoi publier un

Data paper ?

Pourquoi publier un *Data paper*?

- **Vos données ont de la valeur**

  Vous produisez des données de recherche qui peuvent intéresser d'autres scientifiques ….

  mais aussi d'autres types d'acteurs….

Pourquoi publier un *Data paper*?

- Vos données ont de la valeur, un intérêt pour d'autres et peuvent être réutilisées

- Rédiger un *Data paper* peut être simple

- Large choix de revues scientifiques pour publier

  et d'entrepôts pour déposer les données

- Tous contributeurs à la production de données peut être co-auteur d'un *Data paper*

  Informaticiens, développeurs, documentalistes, analystes,….

RÉPUBLIQUE
FRANÇAISE
Liberté
Égalité
Fraternité

cirad

Pourquoi publier un *Data paper* ?

- Publier un *Data paper* permet de le faire savoir
  → paternité et reconnaissance
- Mettre en valeur votre travail → visibilité

- Mettre à disposition ses données permet à d'autres de les utiliser → citations

- S'inscrire dans une démarche d'intégrité scientifique
  → bonnes pratiques de recherche

- Contribuer aux avancées scientifiques ou sociétale
  → valorisation

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

cirad

Pourquoi publier un *Data paper*?

The nationwide 'ZNIEFF' inventory in France: an open dataset of more than one million species data in zones of high ecological value

Fanny Lepareur, Mathieu Manceau, Yorick Reyjol, Julien Touroult, Solène Robert, Frédéric Vest, Arnaud Horellou, Laurent Poncet

A Data Set for US Horror Film Trailers

Arts and Media

Nick Redfern
Independent research
nickredfernres@outlo

nprising audio, colour, motion, and s
sing horror films at the US box office
available for computational film an
film style and is the only existing d
nal analyses. Data is stored in csv file
4.0 International license on Zenodo

Les déterminants naturels et politiques des AOC viticoles de Côte-d'Or

Natural and Political Determinants of Wine-Related Geographical Indications in the Côte-d'Or Département of France

Determinantes naturales y políticas de las AOC vitivinícolas de la Côte-d'Or

JEAN-SAUVEUR AY ET MOHAMED HILAL

DATA NOTE

Data on gender representation in food and beverage print advertisements found in corner stores from Guatemala and Peru

Lucila Rozas[1], Peter Busse[1*], Joaquin Barnoya[2] and Alejandra Garrón[3]

Pourquoi publier un *Data paper*?

- Si votre *Data paper* est complet et vos données FAIR

  → vous avez toutes les chances d'être publié

  ➢ un article supplémentaire publié, évalué par les pairs citable et indexé par les moteurs de recherche.

notoriété

valorisation

attractivité

collaborations

Satisfaction
de contribuer
aux avancées scientifiques
ou sociétales

Crédit: Nos pensées. La valeur de la reconnaissance.
https://nospensees.fr/la-valeur-de-la-reconnaissance/

Pensez-vous que vos données peuvent faire l'objet d'un *Data paper* ?

> Oui

> Non

> Ne sait pas

Concept et objectifs
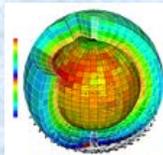
du *Data paper*

Modèle climatique

Loutre

enquêtes

Séquence de gènes

Avez-vous déjà lu un *Data paper* ?

➤ Oui

➤ Non

➤ Ne sait pas

A quoi sert de publier un *Data paper ?*

➢ A - à se faire bien voir

➢ B - à augmenter son taux de publication

➢ C - à rendre disponibles et réutilisables ses données

Concept du *Data paper*: article + entrepôt

Jeu de données

Entrepôt de données

Objectif : que les données soient compréhensibles et réutilisables

Concept du *Data paper*: article + entrepôt

Data in Brief 32 (2020) 106264

Data paper

ELSEVIER

Data Article

Dataset of organic sample near infrared spectra acquired on different spectrometers

Céline Chauvergne[a], Laurent Bonnal[b,c], Denis Bastianelli[b,c], Hélène Carrère[a], Yves Griveau[d], Marie-Pierre Jacquemot[d], Matthieu Reymond[d], Valérie Méchin[d], Virginie Rossard[a], Éric Latrille[a,*]

[a] INRAE, Montpellier University, LBE 102 Avenue des Étangs, 11100 Narbonne, France
[b] CIRAD, UMR SELMET, F-34398 Montpellier, France
[c] Institu Agro, SELMET, University of Montpellier, CIRAD, INRAE, Montpellier, France
[d] Institut Jean-Pierre Bourgin, INRAE, AgroParisTech, Université Paris-Saclay, 78000 Versailles, France

ARTICLE INFO

Article history:
Received 8 July 2020
Revised 18 August 2020
Accepted 27 August 2020
Available online 2 September 2020

Keywords:
Near infrared spectroscopy
Spectrum transfer model
Chemometrics
Organic samples

ABSTRACT

This dataset presents 127 raw near infrared spectra of different organic samples acquired on three different spectrometers in three different labs. An example of data processing is shown to create six spectra transfer models between the three spectrometers (two by two). In order to build and validate these transfer models, the dataset is split into two sets of spectra: a first set was used to compute six spectra transfer models thanks to the Piecewise Direct Standardisation function (PDS). A second set the first one was used to validate treatments and models were

RÉPUBLIQUE FRANÇAISE
Liberté Égalité Fraternité

cirad

Jeu de données

N0410	35.4199982	71.4770966	30.7443008	14.7789001
N0411	2.5599999	4.1999998	96.8622971	1.4170001
N0412	33.5699997	66.5667038	49.2616005	22.1518993
N0413	12.0600004	24.9799995	98.4954987	0.1404
N0414	22.8600006	45.9847984	56.3255997	29.2628994
N0415	33.7099991	61.1310997	31.4731998	16.0713997
N0416	12.8500004	25.7000008	73.9356003	3.9460001
N0417	40.4000015	68.1708984	44.9892998	8.5288
N0418	20.1100006	34.7122002	51.871703	2.8877001
N0419	40.3300018	83.280098	16.1361008	49.5060005
N0420	36.5099983	75.524498	26.6084995	15.3761997
N0421	31.3500004	58.5014992	36.3340988	23.4272995
N0422	43.5999985	90.9889984	18.3896008	35.1469002
N0423	25.2700005	64.7564011	46.0335007	19.7765007
N0424	45.5900002	96.9000015	2.7472999	14.1758003
N0425	35.5400009	66.8264008	39.6702995	19.7801991
N0426	46.1199989	93.8000031	6.4763999	71.2404022
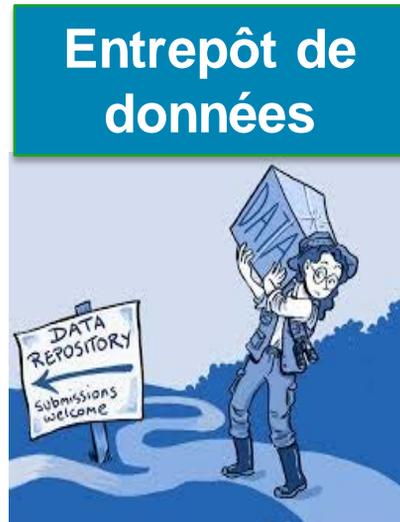
Entrepôt de données

DATA REPOSITORY

Submissions welcome

Documentation complète

Fichiers de données + Documentation associée
- protocoles,
- questionnaire d'enquête, …
- fichier « Read me »
- métadonnées disciplinaires

- Contexte de l'étude
- Méthodes
- Potentiel des données
  + lien vers entrepôt

Données compréhensibles

↓

Données réutilisables

Concept du *Data paper:* article + entrepôt

1. **Décrire un jeu de données et son contexte**
   suffisamment pour permettre de le comprendre et le réutiliser

2. **Décrire les méthodes d'obtention**
   suffisamment pour reproduire l'étude : protocole, méthode d'échantillonnage, équipements, contrôle qualité…

3. **Montrer le potentiel de réutilisation des données**
   suffisamment pour montrer l'originalité des données et leur importance scientifique, sociétale, environnementale,….

   Pas de résultats, ni analyses, ni interprétation

4. **Donner accès au jeu de données**

RÉPUBLIQUE
FRANÇAISE
Liberté
Égalité
Fraternité

cirad

Concept du *Data paper*: article + entrepôt

- Entrepôt: fichiers de données +
                     documentation associée

Exploitations cotonnières et pratiques culturales dans quatre villages de la zone CMDT au Mali, campagne 1997-98

	1 to 4 of 4 Files		Download ▾
☐	**List_variables.tab** — Liste des variables		⬇
	Tabular Data - 9.7 KB - 24 mars 2021 - 0 Downloads		
	3 Variables, 115 Observations - UNF:6:3EbheRvxTye6OB3BuVnABQ==		
	Liste des varibales avec signification en français et en anglais		
☐	**ML_97_parcelles.xlsx**		⬇
	MS Excel Spreadsheet - 155.7 KB - 24 mars 2021 - 2 Downloads	Données	
	MD5: d534a174a61f2cba7533a57589cb3aef		
	Données de suivi des parcelles de culture		
☐	**ML_97_Struc.xlsx**		⬇
	MS Excel Spreadsheet - 19.4 KB - 24 mars 2021 - 0 Downloads		
	MD5: c276e26c851d95a45a44ac4c41cc890c		
	Données sur les caractéristiques structurelles des exploitations		
☐	**Readme.txt** — Read me		⬇
	Plain Text - 550 B - 24 mars 2021 - 0 Downloads		
	MD5: 2bc7f46d64efe883907547d8adb54e23		

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

cirad

Concept du *Data paper* : article + entrepôt

Article

journal homepage: www.elsevier.com/locate/dib

Data Article

Dataset of chemical and near-infrared spectroscopy measurements of fresh and dried poultry and cattle manure

Check for updates

Fabien Gogé[a,*], Laurent Thuriès[b,c], Youssef Fouad[a,*], Nathalie Damay[d], Fabrice Davrieux[e,f], Géraud Moussard[b,c], Caroline Le Roux[d], Séverine Trupin-Maudemain[g], Matthieu Valé[h], Thierry Morvan[a]

[a] INRAE, Institut Agro, UMR SAS, 35000 Rennes, France
[b] CIRAD, UPR Recyclage et Risque, F-97743 Saint-Denis, Réunion, France
[c] Recyclage et Risque, Univ Montpellier, CIRAD, Montpellier, France
[d] LDAR, F-02007 Laon, France
[e] CIRAD, UMR Qualisud, F-97410 Saint-Pierre, Réunion, France
[f] Qualisud, Univ Montpellier, Avignon Université, CIRAD, Institut Agro, Univ Réunion, Montpellier, France
[g] Arvalis-Institut du Végétal, SQV, F-91720 Boigneville, France
[h] Auréa AgroSciences, 45160 Ardon, France

ARTICLE INFO

Article history:
Received 21 October 2020
Revised 2 December 2020
Accepted 8 December 2020
Available online 13 December 2020

Keywords:
Poultry manure
Cattle manure
NIR spectroscopy
Piecewise direct standardization

ABSTRACT

Combined wi...
infrared (NIR)
cise and inexp...
contents of org...
itation of this ...
model may de...
ent spectrome...
tion methods ...
tion (PDS) alg...
The dataset p...
nure samples
located in ma...
land France ar...
for seven chen...
tory methods....

* Corresponding authors.
E-mail addresses: fabien.goge@inrae.fr (F. Gogé), youssef.fouad@a...

https://doi.org/10.1016/j.dib.2020.106647
2352-3409/© 2020 Published by Elsevier Inc. This is an open access ...
(http://creativecommons.org/licenses/by-nc-nd/4.0/)

+

Entrepôt

Recherche Data Gouv > Data INRAE > Experimental - Observation - Simulation Dataverse > T_Morvan_UMR SAS dataverse >

Dataset of chemical and Near-Infrared spectroscopy measurements of fresh and dried poultry and cattle manure

Version 1.2

Gogé, Fabien; Thuriès, Laurent; Fouad, Youssef; Damay, Nathalie; Davrieux, Fabrice; Moussard, Geraud; Le Roux, Caroline; Trupin-Maudemain, Séverine; Valé, Matthieu; Morvan, Thierry, 2020, "Dataset of chemical and Near-Infrared spectroscopy measurements of fresh and dried poultry and cattle manure", https://doi.org/10.15454/JIGO8R, Recherche Data Gouv, V, UNF:6:xLw9kITHRq6tG3QhnE14KA== [fileUNF]

Citer le dataset - Pour en apprendre davantage sur le sujet, consulter le document Data Citation Standards [en].

Modalités d'accès au dataset -
Contact | Partager

Statistiques d'utilisation sur les datasets
1 110 consultations
374 téléchargements
0 citation

Description The dataset consists of 332 manure samples from poultry and cattle origin, sampled in farms located in major regions of livestock production in mainland France and Reunion Island. The samples were analysed for seven chemical properties following conventional laboratory methods : dry matter (DM), total ammonium nitrogen (NH4), total nitrogen (N), phosphorus (P2O5), calcium (CaO), magnesium (MgO) and potassium (K2O). Near-Infrared (NIR) spectra were acquired with three spectrometers on fresh homogenized and dried ground samples and standardized by PDS algorithm. (2020-08-31)

➢ Répond aux attentes des éditeurs

➢ des évaluateurs (reviewers)

➢ Pour les auteurs, double chance d'être :
- visible : Indexés dans bases de données biblio et données
- trouvé : 2 accès par moteurs de recherche
- cité : 2 identifiants numériques (texte et données)

➢ Meilleure mise en valeur des données

Quelles sont selon vous les types de
données concernés ?

Quelles données concernées ?

- ## Tous types de jeux de données

  Séquences, mesures, localisations, photos, enquêtes, presse,…

**Données
d'observation**

télédétection

**Données
expérimentales**

*Séquence de
gènes*

*Résultats
agronomiques*

**Simulations
Modèles**

Modèle climatique

Enquêtes

Données biblio

Définition des données de recherche:

observations, mesures, faits, images, codes, simulations informatiques,….
sur lesquels s'appuient des résultats de recherche.

Exemple 1 *Data paper* sur données d'enquête

Data from an online survey on lentil consumption practices in France in 2022

▼ Aïssétou Lounayo Yabré, Jeanne-Marie Membré

Abstract ▲

Background

In a context of transition towards plant-based protein diet, a survey aiming to collect the lentil consumer practices in France in 2022 was performed. There were 607 responses to the survey, of which a large majority (556) were lentil consumers. Amongst those, 283 people indicated that they currently eat more lentils than 5 years ago.

New information

The questions were related to type of lentil meals, frequency of consumption, type of preparation, storage duration once cooked etc. (Table 1). There were also general questions on age, gender and region. The survey may be used to obtain information on what type of lentils is consumed (and how often) in France, how it is cooked and stored. This information may be then plugged into a food safety risk assessment to refine, for instance, a microbial exposure model.

enquêtes

Table 1.

List of the 21 questions asked in the survey. The response options are provided in the 2nd document (pdf). The responses to the questions are given in the dataset (csv document).

Download as CSV XLSX

Number	Question	Type of responses
1	Do you eat lentils or meals based on lentils?	Single choice
2	If yes, why?	Multiple choice
3	If not, why?	Multiple choice
4	In the last 5 years, your consumption of lentils has increased, remained the same or decreased?	Multiple choice
5	Which types of lentils do you buy?	Multiple choice
6	How often do you eat lentils?	Double array: Single choice per type of lentils
7	If you consume raw lentils, what is the type of preparation?	Multiple choice
8	If hot meals, how do you cook them?	Multiple choices
9	If cold meals, how do you cook them?	Multiple choices
10	If you cook hot meals, what is the time of cooking?	Free response

607 réponses, 12 regions de France

442 ♀ ; 154 ♂ : de 18 à 91 ans

Étude de l'alimentation à base de protéines végétales
→ Pratiques de consommation de lentilles en France

Données dans l'entrepôt
Recherche Data Gouv
https://doi.org/10.57745/KMGODH

Exemple 1 *Data paper* sur données d'enquête

Online survey lentil consumer practices 2022

Version 1.5

Membré, Jeanne-Marie; Yabré, Aïssétou Lounayo, 2022, "Online survey lentil consumer practices 2022", https://doi.org/10.57745/KMGODH, Recherche Data Gouv, V1, UNF:6:yZY1hdpaO4elwozmqHHiMA== [fileUNF]

Citer le jeu de données ▾

Pour en apprendre davantage sur le sujet, consulter le document Data Citation Standards [en].

Description

The dataset includes the 607 responses provided to an online survey on lentil consumer practices in France in 2022. The 2nd file is a document including the questions asked: type of lentil meals, frequency of consumption, type of prepa... general questions on age, gender and ... type of lentils is consumed (and how oft... may be then plugged into a food safety ... exposure model. English (2022-06-16)

1 à 2 de 2 Fichiers

Dataset_English_Final version_underscoreaulieuvirgule.tab
Données tabulaires - 242.9 Ko
Publié 4 juil. 2022
13 téléchargements
30 Variables, 607 Observations UNF:6:yZY1...iMA==
dataset in format csv

Questionnaire lentille Anglais.pdf
Adobe PDF - 420.9 Ko
Publié 4 juil. 2022
10 téléchargements
MD5: 516...421
List of the 21 questions with the possible responses

Exemple 1 *Data paper* sur données d'enquête

Data from an online survey on lentil consumption practices in France in 2022

Aïssétou Lounayo Yabré, Jeanne-Marie Membré

Abstract

Background

In a context of transition towards plant-based protein diet, a survey aiming to colle
consumer practices in France in 2022 was performed. There were 607 responses to th
which a large majority (556) were lentil consumers. Amongst those, 283 people in
they currently eat more lentils than 5 years ago.

New information

The questions were related to type of lentil meals, frequency of consumption, type of
storage duration once cooked etc. (Table 1). There were also general questions on
and region. The survey may be used to obtain information on what type of lentils is consume
(and how often) in France, how it is cooked and stored. This information may be then plugged int
a food safety risk assessm exposure model.

Total views: **417**

RÉPUBLIQUE FRANÇAISE recherche.**data**.gouv.fr

Online survey lentil consumer practices 2022

Version 1.5

Membré, Jeanne-Marie; Yabré, Aïssétou Lounayo, 2022, "Online survey lentil consumer practices 2022", https://d
oi.org/10.57745/KMGODH, Recherche Data Gouv, V1, UNF:6:yZY1hdpaO4elwozmqHHiMA== [fileUNF]

Citer le jeu de données ▾

Pour en apprendre davantage sur le sujet, consulter le document
Data Citation Standards [en].

Description ⓘ

The dataset includes the 607 responses provided to an online survey on lentil consumer practices in
France in 2022. The 2nd file is a document including the questions asked: type of lentil meals,
frequency of consumption, type of preparation, storage duration once cooked, etc. There were also
general questions on age, gender and region. The survey may be used to get information on what
type of lentils is consumed (and how often) in France, how it is cooked and stored. This information
may be then plugged into a food safety risk assessment to refine, for instance, the consumer
exposure model. English (2022-06-16)

Statistiques d'utilisation sur les jeux de données ⓘ

450 consultations ⓘ

20 téléchargements ⓘ

Exemple 2 *Data paper* données biodiversité
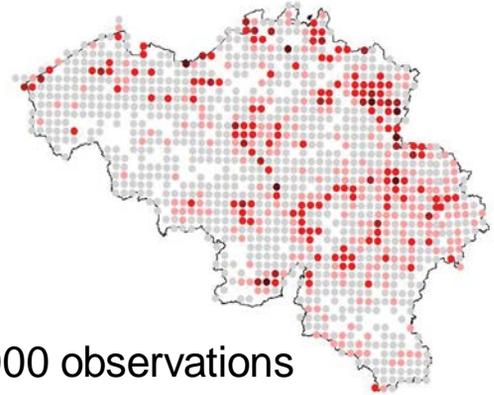
FORMIDABEL: The Belgian Ants Database

Methods

Method step description:

A large portion of the occurrence data have been collected by volunteers, other records originated from several projects and research programs. The data and specimens were sent to the Belgian ant curators, and after validation, the information was incorporated in the database. The collection records "dry specimen" originate from the Gembloux "Ant" collection and the Charles Gaspar collection, the collection of the "Cercle des entomologists Liégeois", the RBINS collection and the private collection "Roland Vannieuwenhuyse". After revision and validation, this information was also included in the database. The literature-based records were retrieved from van Boven 1970; van Boven and Mabelis 1986; Dekoninck et al. 2006 and references therein. How the database evolved is described in the Database history section.

Sampling description: Most occurrence records originate from hand/nest sampling (42, 3% of all records and mainly from Wallonia) and pitfall sampling (36, 7% mainly from Flanders). The followed procedure differs from region to region. This is due to historical reasons. Some very interesting occurrence records were obtained by sifting, coloured water traps and Malaise traps (all less than 3 % of the total sampling). An extensive description of the sampling methods is provided by Schauff (2001).

Quality control description: All the records were validated by the dataset curators before being added to the FORMIDABEL database. The dataset curators also checked the determinations of the collection specimens. If needed, the determination was adapted and made consistent with modern taxonomy; Radchenko and Elmes (2010) for the genus *Myrmica* and Seifert (2007) for the other genera. Before the

27.000 observations

→ 76 espèces de fourmis natives

→ 9 espèces introduites

→ description de chaque microhabitat de fourmis.

Données actualisées 2 fois par an.

Données sur la distribution des fourmis en Belgique

Données dans entrepôt GBIF
https://www.gbif.org/dataset/b528799a-2d52-4023-aa02-9ce081e3ca5f

Exemple 2*Data paper* données biodiversité

Formidabel; Belgian Ants Database
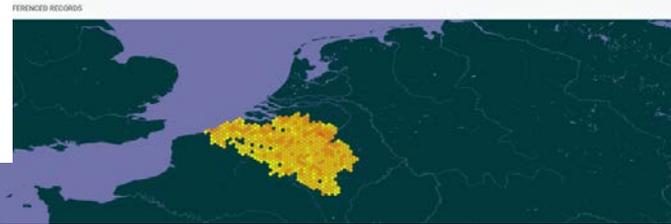
Published by Belgian Biodiversity Platform

DATASET   METRICS   ACTIVITY   ⬇ DOWNLOAD   ⇨ HOME PAGE

FORMIDABEL is a database of Belgian Ants containing more than 27.000 occurrence records. These records originate from collections, field sampling and literature. The database gives information on 76 native and 9 introduced ant species found in Belgium. The collection records originated mainly from the Ants collection in Royal Belgian Institute of Natural Sciences (RBINS), the 'Gaspar' Ants collection in Gembloux and the zoological collection of the University of Liège (ULG). The oldest occurren…   More

Publication date: March 25, 2021
Metadata last modified: March 25, 2021
Hosted by: Belgian Biodiversity Platform
Licence: CC BY 4.0
99 How to cite   DOI 10.15468/xdapub

| 27,264 Occurrences | 100% With taxon match | 99.8% With coordinates | 99.6% With year |

Exemple 2 *Data paper* données biodiversité

FORMIDABEL: The Belgian Ants Database

Dimitri Brosens [1], François Vankerkhoven [2], David Ignace [3], Philippe Wegnez [4], Nicolas Noé [5], André Heughebaert [5], Jeannine Bortels [6], Wouter Dekoninck [7]

Abstract

FORMIDABEL is a database of Belgian Ants containing more than 27.000 occurrence records. These records originate from collections, field sampling and literature. The database gives information on 76 native and 9 introduced ant species found in Belgium. The collection records originated mainly from the ants collection in Royal Belgian Institute of Natural Sciences (RBINS), the 'Gaspar' Ants collection in Gembloux and the zoological collection of the University of Liège (ULG). The oldest occurrences date back from May 1866, the most recent refer to August 2012. FORMIDABEL is a work in progress and the database is updated twice a year.

The latest version of the dataset is publicly and freely accessible through this url: http://ipt.biodiversity.be/resource.do?r=formidabel. The dataset is also retrievable via the GBIF data portal through this link: http://data.gbif.org/datasets/resource/14697

Keywords

Formicidae, Belgium, Flanders, Wallonia, Brussels Capital Region, ecological data, grid mapping, UTM, historical data, literature, collections, observations, trapping, ants

Cité 10 fois

Global Biodiversity Information Facility (GBIF)

Formidabel; Belgian Ants Database

Published by Belgian Biodiversity Platform

Dekoninck W • Brosens D

METRICS   ACTIVITY   ⬇ DOWNLOAD   HOME PAGE   27,264 OCCURRENCES   64 CITATIONS

ABEL is a database of Belgian Ants containing more than 27.000 occurrence records. These records originate lections, field sampling and literature. The database gives information on 76 native and 9 introduced ant found in Belgium. The collection records originated mainly from the Ants collection in Royal Belgian Institute al Sciences (RBINS), the 'Gaspar' Ants collection in Gembloux and the zoological collection of the University of LG). The oldest occurren...   More

Publication date: March 25, 2021
Metadata last modified: March 25, 2021
Hosted by: Belgian Biodiversity Platform
Licence: CC BY 4.0
How to cite   DOI 10.15468/xdapub

| 27,264 Occurrences | 100% With taxon match | 99.8% With coordinates | 99.6% With year |

FERENCED RECORDS

27,264 OCCURRENCES   64 CITATIONS

Bonne pratique

= Publier un *Data paper* + déposer ses données dans un entrepôt

➢ Répond aux attentes des éditeurs et évaluateurs

➢ Pour vous, plus de chances :
- d'être visible, donc trouvé, et peut-etre cité
- de valoriser votre travail et vos données.

➢ Pour la science
- Données plus accessibles et réutilisables
- Meilleure contribution aux avancées scientifiques

Modèles de

Data paper

Publier un *Data paper*

- **Choisir 1 revue**
  - → rédiger selon le modèle proposé de *Data paper*
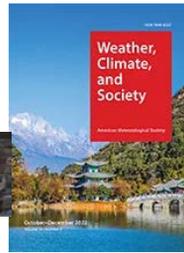
- **Choisir un entrepôt de données**
  - → déposer ses données + documentation associée

Protocoles,
Liste des équipements,
Questionnaire,
Dictionnaire des variables
…..

- **Soumettre le manuscrit**

  contenant le lien vers les données

Modèles de *Data paper*

- **La plupart des revues propose un modèle**
  - ➢ Consulter les instructions aux auteurs

- **Présentation de 2 exemples de modèles classiques**
  - Revue multidisciplinaire *Data in Brief* (Elsevier)
  - Revue disciplinaire *Ecological Research* (Wiley)

Modèle de *Data paper* *Data in Brief*

Title, Authors, Abstract, Keywords

Data description

Décrire les données et les fichiers

Experimental design, materials and methods

Décrire l'experience et les méthodes utilisées pour générer les données

Value of the data

Quel est l'interet de ces données ? à qui sont-elles utiles?

Comment peuvent-elles être réutilisées ?

➡️ **Accès aux données déposées dans un entrepôt**

Acknowledgements, References

Exemple de *Data paper* *Data in Brief*

https://doi.org/10.1016/j.dib.2023.108906

Title, Authors, Abstract, Keywords

Data Article

Image dataset of important grape varieties in the commercial and consumer market

Lafta R. Al-khazraji[a], Mohammed Abdallazez Mohar, Dhafar Hamed Abd[c], Wasiq Khan[d], Bilal Khan[e], Abir Jaafar Hussain[f,d,*]

[a] Ministry of Education, General Directorate of Education, Salah al-Din, Iraq
[b] University of Karbala, College of Computer Science and Information Technology, Informa Department, Iraq
[c] Department of Computer Science Al-Maarif University College, Alanbar, Iraq
[d] Liverpool John Moores University, Liverpool L3 3AF, UK
[e] School of Computer Science and Engineering, California State University San Bernardino, San Bernardino, CA 92407, USA
[f] Department of Electrical Engineering, University of Sharjah, Sharjah, UAE

The proposed dataset is publicly available at Mendeley platform as described below:

Repository name: Mendeley

Data identification number:

Direct URL to data: https://data.mendeley.com/datasets/7n3d6696hz/2

ABSTRACT

This work presents a primary dataset collected from various geographic locations in Iraq for the seedlings of eight varieties of grapes that are used for local consumption and export. Grape types included in the dataset are: deas al-annz, kamali, halawani, thompson seedless, aswud balad, ri-asi, frinsi, shdah. Leaves of each type of the seasoned fruit were photographed with high resolution device. A total of 8000 images (i.e., 1000 images per category) were captured using random sampling approach while maintaining the balance and diversity within grape image data. The proposed dataset is of significant potential impact and usefulness with features including (but not limited to) 8 varieties, that have different tastes and can support various industry in agriculture and food manufactures.

Exemple de *Data paper* *Data in Brief*

1. Objective

To reduce the time, cost, and effort in identifying the desired seedlings of grape cultiva prior to planting, the proposed dataset is configured to build a model capable of classifyin eight cultivars of grapes desired in Iraq.

2. Data Description

The dataset consists of eight folders (corresponding to the included grapes types), each folder contains 1000 images of JPG format. The images within the dataset are all of the resolution of 6000×4000 pixels. Because of the high resolution of the images, the size of the images became 49.8 GB, which is a large size for uploading and downloading the dataset from the Internet, so the resolution of the images was changed using a tool resize pictures in Windows to be the dimensions 1620×1080 pixels, Thus, the data size is 1.95 GB, After compressing with a zip program, the data size became 1.83 GB.Each folder represents one of the eight grape varieties (Deas Al-Annz, Kamali, Halawani, Thompson Seedless, aswud balad , riasi ,frinsi, shdah). The grape leaves have been photographed as shown in Fig. 1.

3. Experimental Design, Materials and Methods

Acquisition of the images for each grape type followed the workflow of based on a random leaf selection approach as illustrated in Fig. 2. In the proposed approach, selection of the con-stituent of the population of each grape type followed the uniform distribution with an equal probability of being selected.

Data diversity was maintained via random leaf selection approach, by which both the leaves to be photographed and tree for each grape type were selected randomly.

In the post image acquisition process for each grape type, images per each grape type were transferred from the camera's memory to an external hard drive and kept in a folder bearing the name of the grape type. The acquisition of the images for the next grape type would then follow after the removal of the images for the transferred grape type, and so on. Leaves containing dust were not photographed at a high rate, as they represent noise, and the leaves that were eaten were not photographed. Photographed with the knowledge and consent of the owners of the gardens.
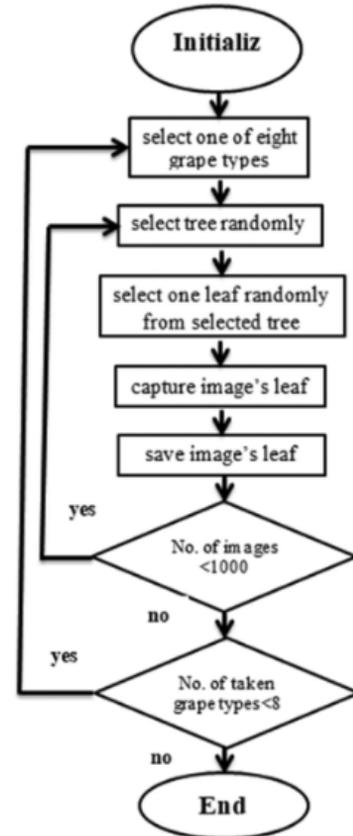
Exemple de *Data paper* *Data in Brief*

Value of the data

- This dataset can be proven useful for the identification of the types of grape seedlings mainly before their plantation and growth [1,2].
- The early identification of the desired grape type is critical due to the long seeding duration (approximately # of years) along with the effort and cost associated with their growth [3].
- Conventional methods for the identification of the type for a seedling requires expert knowledge while at the seeding site where the effort and cost associated with the involved labor can be significant [4].
- The development of a classification model for grape class identification with higher accuracy can be an important undertaking providing significant advantage with respect to the economic factors as well as timely identification.
- The proposed dataset can be proven useful for computer science community, particularly computer vision, machine learning and deep learning to build robust grape classification models that could accurately classify grapes of various types.

Acknowledgements, References

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

cirad

Exemple de Data paper *Data in Brief*

Accès aux données

Grape Varieties Dataset

Published: 24 November 2022 | Version 2 | DOI: 10.17632/7n3d6696hz.2
Contributors: Mohammed Mohammed Abdallazez, Lafta Raheem Ali Al-Khazraji, dhafar hamed, Abir Hussain, Wasiq Khan, Bilal Khan

Description

Type of data   Images , Specific subject area: Comp...   deep learning. This dataset has been
collected for eight grape varieties, where leaves of g...   f Iraq for eight common varieties that are
desirable in the local market and for export. Useful...   ing is provided in the dataset that is
typically used to identify grape variety . Seedlings ...   entify their types with highest confidence.
Dataset for 8 varieties included (deas al-annz, kam...   , frinsi, shdah) comprises of a total of
8000 image, 1000 images for each grape variety.
The dataset consists of eight folder (corresponding...   1000 images of JPG format. The images
within the dataset are all of the resolution of 1620...   ompressing with a zip program, the data
size became 1.83 GB. Each folder represents one o...

Download All 1880 MB ⓘ

Files

📁 Grape Varieties Dataset

Root > Grape Varieties Dataset

📦 Aswud Balad.rar	242 MB ⬇
📦 Deas Al-Annz.rar	278 MB ⬇
📦 Frinsi.rar	224 MB ⬇
📦 Halawani.rar	220 MB ⬇
📦 Kamali.rar	205 MB ⬇
📦 Riasi.rar	255 MB ⬇
📦 Shdah.rar	213 MB ⬇
📦 Thompson Seedl	244 MB ⬇

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

cirad

Exemple de Data paper : *Data in Brief*

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

ELSEVIER

Data Article

Image dataset of important grape varieties in the commercial and consumer market

Lafta R. Al-khazraji[a], Mohammed Abdallazez Mohammed[b], Dhafar Hamed Abd[c], Wasiq Khan[d], Bilal Khan[e], Abir Jaafar Hussain[f,d,*]

[a] Ministry of Education, General Directorate of Education, Salah al-Din, Iraq
[b] University of Karbala, College of Computer Science and Information Technology, Information Technology Department, Iraq
[c] Department of Computer Science Al-Maarif University College, Alanbar, Iraq
[d] Liverpool John Moores University, Liverpool L3 3AF, UK
[e] School of Computer Science and Engineering, California State University San Bernardino, 5500 University Parkway, San Bernardino, CA 92407, USA
[f] Department of Electrical Engineering, University of Sharjah, Sharjah, UAE

ARTICLE INFO

ABSTRACT

This work presents a primary dataset collected fr... ous geographic locations in Iraq for the seedlings varieties of grapes that are used for local consump... export. Grape types included in the dataset are: deas al-annz, kamali, halawani, thompson seedless, aswud balad, ri-asi, frinsi, shdah. Leaves of each type of the seasoned fruit were photographed with high resolution device. A total of 8000 images (i.e., 1000 images per category) were captured using random sampling approach while maintaining the bal-ance and diversity within grape image data. The proposed ... ...otential impact and usefulness with ... ...ot limited to) 8 varieties, that have ... support various industry in agricul-...res.

en ligne depuis 20 janvier 2023

Mendeley Data

Grape Varieties Dataset

Description

Type of data Images, Specific subject area: Computer Vision, Pattern Recognition, machine learning, deep learning. This dataset has been collected for eight grape varieties, where leaves of grape seedlings were captured in different regions of Iraq for eight common varieties that are desirable in the local market and for export. Useful information of the parts of the photographed seedling is provided in the dataset that is typically used to identify grape variety. Seedlings were photographed during the fruiting season to identify their types with highest confidence. Dataset for 8 varieties included (deas al-annz, kamali, halawani, thompson seedless, aswud balad, riasi, frinsi, shdah) comprises of a total of 8000 image, 1000 images for each grape variety.

The dataset consists of eight folder (corresponding to the included grapes types), each folder contains 1000 images of JPG format. The images within the dataset are all of the resolution of 1620 x 1080 pixels, Thus, the data size is 1.95 GB, After compressing with a zip program, the data size became 1.83 GB. Each folder represents one of the eight grape varieties.

Download All 1880 MB (i)

Files

Grape Varieties Dataset

Dataset metrics

Usage	
Views:	82
Downloads:	35

en ligne depuis 8 novembre 2022

Modèle de la revue *EcologicaResearch*

Title, Authors, Abstract, Keywords

Introduction

Contexte et importance des données pour l'écologie.

Data description

Détail des méthodes de production des données :
protocoles, équipements, traitements, contrôle qualité,…

Description des données, tableaux et fichiers :
explication des lignes et colones, unites de mesure, abréviations, valeurs manquantes, abérrantes, …

Fichiers de données déposés dans 1 entrepôt

Acknowledgements, References

Exemple de *Data paper* *Ecological Research*

ECOLOGICAL RESEARCH

DATA ARTICLE | 🔓 **Full Access**

Distribution and functional data of fungal families

Fungi have unique ecosystem functions, such as organic matter decomposition, parasitism, and symbiosis with other organisms in terrestrial and aquatic ecosystems. Their taxonomic and functional diversities are essential factors for predicting ecosystem functions and their responses to environmental changes. With the widespread use of high-throughput sequencing (HTS) in recent years, the detection of fungal DNA sequences in various regions and on different environmental substrates has advanced. HTS-obtained DNA sequences can be compared with those in databases to identify the taxa of organisms from which they are derived. Therefore, a global DNA database containing taxonomic information has been developed. However, functional data on the distribution and function of individual taxa remain scarce. In this study, the ecological information of each fungal family was compiled from review papers and published books. Specifically, the following information was collected: reported distribution (11 categories including information on the presence of fungi in Japan) and function (38 categories related to ecological functional data, such as guilds and habitats) of 553 families included in the literature. These data will provide information on the ecology of specific fungi detected in the field and help estimate the functional group composition and diversity of fungi from their DNA assemblage data. The detailed Metadata for this abstract published in the Data Article section of the journal is available in MetaCat in JaLTER at http://db.cger.nies.go.jp/JaLTER/metacat/metacat/ERDP-2022-02.1/jalter-en.

→ *Abstract*

1. **Introduction**
2. **Data description**
   Methods
   Data structure
Lien d'accès aux données

Acknowledgements
References

Exemple de *Data paper* *Ecological Research*

ECOLOGICAL RESEARCH

DATA ARTICLE   🔓 Full Access

Distribution and functional data of fungal families

1 INTRODUCTION

Fungi have a high species diversity ... ately 3.8 million species (Hawksworth & L...

functions, such as organic matter de... other organisms, such as plants and ... terrestrial (e.g., forests, grasslands, ... (Peay et al., 2016). Therefore, under... diversity of fungi have major ecolog...

With the recent development and d... especially high-throughput sequenc... extensively by detecting fungal DNA ... typical HTS-based studies, such as D... of fungal mark... present in the ... those in a data... Doi, 2019; Nils... of mycology ar...

2 DATA DESCRIPTION

Methods

We summarized the taxonomy, distribution... fungal families from published books (Cann... review papers (Aldossari & Ishii, 2021; Oson... are comprehensive references on the distri... family.

Taxonomic information was recorded by cla... following phyla: Ascomycota, Basidiomycot... Chytridiomycota, and Neocallimastigomyco...

Distribution was first presented in 11 categ...

... bd ...s, ... d ... y f ...es ...ed ...ho ... eae, and Shiralaceae.

TABLE 1. File name and description of the data

Data file name	Description
Taxonomy.csv	The phylum to which each family belongs
Distribution.csv	The distribution of each family
Ecology.csv	Information on ecological functions of each family

TABLE 2. The definition of variables in each data file

Data file name	Variable name	Variable definition
Ecology.csv	Saprobic	A family including saprobic species
	Parasitic	A family including parasitic species
	Heteroecious	A family including heteroecious species
	Autoecious	A family including autoecious species
	Pathogens	A family including pathogenic species
	Commensal	A family including commensal species
	Ectomycorrhizal	A family including ectomycorrhizal species
	Mycorrhizal	A family including mycorrhizal species, excluding ectomycorrhizal fungi
	Lichenized	A family including lichenized species
	Necrotrophic	A family including necrotrophic species

RÉPUBLIQUE
FRANÇAISE
Liberté
Égalité
Fraternité

Exemple de *Data paper* *Ecological Research*

New Search   Japanese Skin   Terms of use

Data Set Citation:

When using this data, please cite the data package:

Matsuoka S , Hatano Y , and Osono T.
Distribution and functional data of fungal families
ERDP-2022-02.1.3 (https://db.cger.nies.go.jp/JaLTER/metacat/metacat/ERDP-2022-02.1.3/jalter-en)

General Information:

Title: **Distribution and functional data of fungal families**

Identifier: ERDP-2022-02.1.3

Abstract: Fungi have unique ecosystem functions, such as organic matter decomposition, parasitism, and symbiosis with other organisms in terrestrial and aquatic ecosystems. Their taxonomic and functional diversities are essential factors for predicting ecosystem functions and their responses to environmental changes. With the widespread use of high-throughput sequencing (HTS) in recent years, the detection of fungal DNA sequences in various regions and on different environmental substrates has advanced. HTS-obtained DNA sequences can be compared with those in databases to identify the taxa of organisms from which they are derived. Therefore, a global DNA database containing taxonomic information has been developed. However, functional data on the distribution and function of individual taxa remain scarce. In this study, the ecological information of each fungal family was compiled from review papers and published books. Specifically, the following information was collected: reported distribution (11 categories including information on the presence of fungi in Japan) and function (38 categories related to ecological functional data, such as guilds and habitats) of 553 families included in the literature. These data will provide information on the ecology of specific fungi detected in the field and help estimate the functional group composition and diversity of fungi from their DNA assemblage data. The detailed Metadata for this abstract published in the Data Paper section of the journal is available in

Data Table, Image, and Other Data Details:

Metadata download:	Ecological Metadata Language (EML) File
Data Table:	Distribution (View Metadata | Download File ⬇)
Data Table:	Ecology (View Metadata | Download File ⬇)
Data Table:	Taxonomy / Taxonomy (View Metadata | Download File ⬇)
Other Data:	data descriptor (View Metadata | Download File ⬇)

A retenir

- **Bonne pratique**
  - Publier 1 *Data paper* + dépot des données dans un entrepôt.
  - Assurer la compréhension des données pour faciliter leur réutilisation.
- **Data paper**
  - Décrit complétement les données, méthodes, fichiers,….
  - Décrit l'intérêt et l'utilité des données
  - Donne le lien d'accès aux données
- **Entrepôt de données**
  - Accès aux données + documentation complémentaire (ex: questionnaire d'enquête, protocole, ….)

RÉPUBLIQUE
FRANÇAISE
Liberté
Égalité
Fraternité

cirad

PAUSE CAFÉ

Consultation de *Data papers* publiés
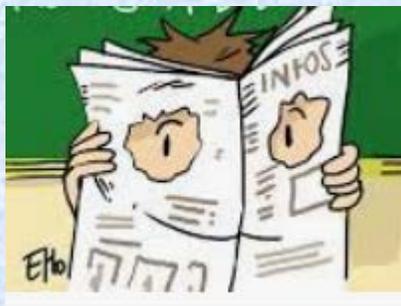
Consultation de *Data papers* publiés

➢ ## Choisir un *Data paper*

▪ À partir de la revue généraliste *Data in Brief*

https://www.sciencedirect.com/journal/data-in-brief/vol/48/suppl/C

▪ À partir de la sélection thématique (diapos suivantes)

- Agronomie, Forêt
- Ecologie, Biodiversité, Faune sauvage
- Climat, Sol, Eau,
- Génomique, Santé,
- Economie, SHS, Archéologie
- *Data papers* à partir de données issues de recherches bibliographiques
- *Data papers* en français
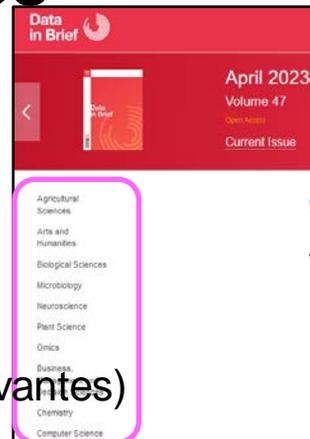
Exemples de *Data papers* publiés

▪ Agronomie

➢ French crop yield, area and production data for ten staple crops from 1900 to 2018 at county resolution. Scientific Data: https://doi.org/10.1038/s41597-022-01145-4

➢ Survey data of a traditional communal water irrigation system in Northern Thailand. Data in Brief: https://doi.org/10.1016/j.dib.2022.108515

➢ A global experimental dataset for assessing grain legume production. Scientific Data: https://doi.org/10.1038/sdata.2016.84

➢ The 'Plantain-Optim' dataset. Data in Brief: https://doi.org/10.1016/j.dib.2018.01.065

➢ Eco-physiological responses of 24 sunflower genotypes to water deficit. Data in Brief: https://doi.org/10.1016/j.dib.2018.10.045

➢ Global database of Hemiptera-Phytoplasma-Plant biological interactions. Biodiversity Data Journal: https://bdj.pensoft.net/article/32910/

➢ Data on the effects of fertilization on growth rates, biomass allocation, carbohydrates and nutrients of nitrogen-fixing and non-nitrogen-fixing tree legumes during tropical forest restoration. BMC Research Notes: https://bmcresnotes.biomedcentral.com/articles/10.1186/s13104-021-05552-5

Consultation de *Data papers* publiés

➢ **Les données sont-elles accessibles ?**

L'information se trouve dans un paragraphe nommé : *Data accessibility, Data availability, Availability of data & materials, Data Records, Data resources*

➢ **Sont-elles déposées dans un entrepôt ?**
- si oui, lequel ?
- si non, sont-elles intégrées dans l'article ? dans ce cas, vous semblent-elles facilement réutilisables ?

➢ **La description vous parait-elle suffisante ?**
- pour reproduire l'étude
- pour comprendre et réutiliser les données
- si « non »: quels éléments ou informations manquent ?

RÉPUBLIQUE
FRANÇAISE
Liberté
Égalité
Fraternité

cirad

Exemples de *Data papers* publiés

- ## Foresterie

  - Diversity of Woody Species in Djamde Wildlife Reserve, Northern Togo, West Africa. Data Science Journal. 2019. http://doi.org/10.5334/dsj-2019-018

  - The GenTree Platform: growth traits and tree-level environmental data in 12 European forest tree species. GigaScience: https://doi.org/10.1093/gigascience/giab010

  - Data on the effects of fertilization on growth rates, biomass allocation, carbohydrates and nutrients of nitrogen-fixing and non-nitrogen-fixing tree legumes during tropical forest restoration. *BMC Research Notes.* https://doi.org/10.1186/s13104-021-05552-5

  - 3 decades of annual growth, mortality, physical condition, and microsite for ten tropical rainforest tree species. Ecology : https://doi.org/10.1002/ecy.2394

  - Data on dendrometric parameters, basic wood density, below- and aboveground biomass of tree species from Mangrove, Miombo, Mopane, and Mecrusse woodlands. Data in Brief. https://doi.org/10.1016/j.dib.2020.105154

Exemples de *Data papers* publiés

- ## Ecologie

  - GIbase1.0: A database of green infrastructure plant species in England and Scotland. *Ecological Solutions and Evidence.* https://doi.org/10.1002/2688-8319.12133

  - Commensal small mammal trapping data in Southern Senegal, 2012–2015: where invasive species meet native ones. *Ecology.* https://esajournals.onlinelibrary.wiley.com/doi/10.1002/ecy.3470

    (Pour accéder à l'article entier (car *Ecology* ne publie que le résumé), il faut ouvrir le fichier zip dans *Supporting information.*)

  - Tundra Trait Team: A database of plant traits spanning the tundra biome. *Global Ecology and Biogeography.* https://doi.org/10.1111/geb.12821

  - 8 million phenological and sky images from 29 ecosystems from the Arctic to the tropics: the Phenological Eyes Network. *Ecological Research* : https://link.springer.com/article/10.1007/s11284-018-1633-x

  - A global spatially explicit database of changes in island palaeo-area and archipelago configuration during the late Quaternary. *Global Ecology and Biogeography* https://doi.org/10.1111/geb.12715

Exemples de *Data papers* publiés

■ Biodiversité

➤ A database of freshwater fish species of the Amazon Basin. *Scientific Data.*
https://www.nature.com/articles/s41597-020-0436-4

➤ AmphiBIO, a global database for amphibian ecological traits. *Scientific Data*.
https://doi.org/10.1038/sdata.2017.123

➤ The data of the Swedish Malaise Trap Project, a countrywide inventory of Sweden's insect fauna. *Biodiversity Data Journal*. https://bdj.pensoft.net/article/56286/list/8/

➤ Aquatic eDNA for monitoring French Guiana biodiversity. *Biodiversity Data Journal*. https://bdj.pensoft.net/article/37518/instance/5252969/

Voir aussi les exemples de *Data papers* sur les pages :

- *Data papers* du GBIF: http://www.gbif.fr/page/ressources/data-papers

- *Data publishing* de l'éditeur Pensoft :
https://natureconservation.pensoft.net/about#DataPublishingGuidelines

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

cirad

Exemples de *Data papers* publiés

- ## Faune sauvage

- Wildlife inventory from camera-trapping surveys in the Azores (Pico and Terceira islands). *Biodiversity Data Journal.* https://doi.org/10.3897/BDJ.8.e47865

- Jaguar movement database: a GPS-based movement dataset of an apex predator in the Neotropics. *Ecology* : https://esajournals.onlinelibrary.wiley.com/doi/10.1002/ecy.2379

- Kakila database: Towards a FAIR community approved database of cetacean presence in the waters of the Guadeloupe Archipelago, based on citizen science. *Biodiversity Data Journal.* https://doi.org/10.3897/BDJ.9.e69022

- The Hummingbird Collection of the Natural History and Science Museum of the University of Porto (MHNC-UP), Portugal. *Biodiversity Data Journal*. https://doi.org/10.3897/BDJ.9.e59913

- A quasi-experimental study of impacts of Tanzania's wildlife management areas on rural livelihoods and wealth. *Scientific data* : https://www.nature.com/articles/sdata201887#t6

Exemples de *Data papers* publiés

▪ Climat - météorologie

➢ A meteorological dataset of the West African monsoon during the 2016 DACCIWA campaign. *Scientific Data*. https://www.nature.com/articles/s41597-022-01277-7#Sec10

➢ Reconstruction of a long-term historical daily maximum and minimum air temperature network dataset for Ireland (1831-1968). *Geoscience Data Journal*. https://doi.org/10.1002/gdj3.92

➢ Data on and methodology for measurements of microclimate and matter dynamics in transition zones between forest and adjacent arable land. *One Ecosystem* : https://oneecosystem.pensoft.net/articles.php?id=24295

➢ Long-term groundwater resource observatory for Southwestern Madagascar. *Hydrological Processes.* https://doi.org/10.1002/hyp.14108

➢ Satellite-based time-series of sea-surface temperature since 1981 for climate applications. *Scientific Data*. https://doi.org/10.1038/s41597-019-0236-x

➢ ClimateEU, scale-free climate normals, historical time series, and future projections for Europe. *Scientific Data*: https://doi.org/10.1038/s41597-020-00763-0

Exemples de *Data papers* publiés

- ## Sol

  - A global database of land management, land-use change and climate change effects on soil organic carbon. Scientific Data: https://doi.org/10.1038/s41597-022-01318-1

  - Soil hydraulic functions of international soils measured with the Extended Evaporation Method (EEM) and the HYPROP device. Open Data Journal for Agricultural Research : https://odjar.org/article/view/15763/15359

  - Dataset on ammonia, nitrous oxide, methane, and carbon dioxide fluxes from 2 soils fertilized amended with treated and non-treated cattle slurry. Data in Brief https://doi.org/10.1016/j.dib.2018.10.124

  - LUCAS Soil Biodiversity and LUCAS Soil Pesticides, new tools for research and policy development. European Journal of Soil Science. 2022. https://doi.org/10.1111/ejss.13299

  - Soil microbial biomass and enzyme data after 6 years of cover crop and compost treatments in organic vegetable production. Data in Brief https://doi.org/10.1016/j.dib.2018.09.013

RÉPUBLIQUE FRANÇAISE
*Liberté
Égalité
Fraternité*

cirad

Exemples de *Data papers* publiés

Eau

- Mapping Flow-Obstructing Structures on Global Rivers. Water Resources Research: https://doi.org/10.1029/2021WR030386

- Sea surface temperature (SST) and SST anomaly (SSTA) datasets over the last four decades (1977–2016) during typhoon season (May to November) in the entire Global Ocean, North Pacific Ocean, Philippine Sea, South China sea, and Eastern China Sea. Data in Brief: https://doi.org/10.1016/j.dib.2022.108646

- Simulating Core Floods in Heterogeneous Sandstone and Carbonate Rocks. Water Resources Research: https://doi.org/10.1029/2021WR030581

- Catchments of German surface water bodies. Hydrological Processes: https://doi.org/10.1002/hyp.14272

- 14 000 years of geochemical and isotopic data from Lake Simcoe, Canada. Data in Brief: https://doi.org/10.1016/j.dib.2022.108541

- Long-term groundwater resource observatory for Southwestern Madagascar. Hydrological Processes: https://doi.org/10.1002/hyp.14108

Exemples de *Data papers* publiés

Génomique

- Genome sequencing of the sweetpotato whitefly *Bemisia tabaci*. *GigaScience*: https://doi.org/10.1093/gigascience/gix018

- Transcriptome data from 3 endemic Myrtaceae species from New Caledonia displaying contrasting responses to myrtle rust (*Austropuccinia psidii*). *Data in Brief* https://doi.org/10.1016/j.dib.2018.12.080

- Draft Genome Resources of 2 Strains of *Xylella fastidiosa* associated with Almond Leaf Scorch Disease in Alicante, Spain. *Phytopathology* https://doi.org/10.1094/PHYTO-09-18-0328-A

- 72-h diurnal RNA-seq analysis of fully expanded third leaves from maize, sorghum, and foxtail millet at 3-h resolution. *BMC Research Notes*: https://bmcresnotes.biomedcentral.com/articles/10.1186/s13104-020-05431-5

- De novo assembly and annotation of the mangrove cricket genome. *BMC Research Notes:* https://doi.org/10.1186/s13104-021-05798-z

Exemples de *Data papers* publiés

■ Santé

➤ VectorNet Data Series 3: *Culicoides* Abundance Distribution Models for Europe and Surrounding Regions. Open Health Data: http://doi.org/10.5334/ohd.33

➤ Coccidioidomycosis (Valley Fever) Case Data for the Southwestern United States. Open Health Data: http://doi.org/10.5334/ohd.31.

➤ Data on the physical function of children with cerebral malaria. Data in Brief: https://doi.org/10.1016/j.dib.2021.106961

➤ Analysis of Chagas disease vectors occurrence data: the Argentinean triatomine species database. Biodiversity Data Journal: https://doi.org/10.3897/BDJ.8.e58076

➤ Response2covid19, a dataset of governments' responses to COVID-19 all around the world. Scientific Data: https://doi.org/10.1038/s41597-020-00757-y

➤ Data Resource Profile: COVerAGE-DB: a global demographic database of COVID-19 cases and deaths. International Journal of Epidemiology. https://doi.org/10.1093/ije/dyab027

Exemples de *Data papers* publiés

- ## Economie

  - IABSE-ADIAB – IAB Job Vacancy Survey Data Linked to Administrative Data. Journal of Economics and Statistics: https://doi.org/10.1515/jbnst-2023-0004

  - Forecasted data of prices for the most common households' fuels utilized in Nigeria during the period 2010–2024. Data in Brief: https://doi.org/10.1016/j.dib.2022.108561

  - Dataset on retail outlet product prices for Botswana, Lesotho and South Africa. Data in Brief: https://doi.org/10.1016/j.dib.2018.05.006

  - Household economy, forest dependency & opportunity costs of conservation in eastern rainforests of Madagascar. Scientific Data: https://doi.org/10.1038/sdata.2018.225

  - World carbon pricing database: sources and methods. Scientific Data: https://doi.org/10.1038/s41597-022-01659-x

  - An extensive data set on energy, economy, environmental pollution and institutional quality in the petroleum-reliant developing and transition economies. Data in Brief: https://doi.org/10.1016/j.dib.2021.106766

  - The Household, Income and Labour Dynamics in Australia (HILDA) Survey. Journal of Economics and Statistics: https://doi.org/10.1515/jbnst-2020-0029.

RÉPUBLIQUE FRANÇAISE
*Liberté
Égalité
Fraternité*

cirad

Exemples de *Data papers* publiés

- ## Sciences humaine et sociale

➤ Dataset on farmers' perception of commodity futures market. Data in Brief. 2022. https://doi.org/10.1016/j.dib.2022.108429

➤ Survey data on income, food security, and dietary behavior among women and children from households of differing socio-economic status in urban and peri-urban areas of Nairobi, Kenya. Data in Brief : https://doi.org/10.1016/j.dib.2020.106542

➤ Using Crowd-Sourced Data to Explore Police-Related-Deaths in the United States (2000–2017): The Case of Fatal Encounters. Open Health Data: http://doi.org/10.5334/ohd.30

➤ An integrated dataset for stakeholder perceptions of environmental change and instrumented measures of change. Data in Brief: https://doi.org/10.1016/j.dib.2018.10.112

➤ Voir aussi les exemples de *Data papers publiés dans* Journal of Open Humanities Data : https://openhumanitiesdata.metajnl.com/articles?section=1

Exemples de *Data papers* publiés

En français

➢ Les communes nouvelles françaises (2012-2022) : une méthode pour l'analyse de données à l'échelon municipal selon des limites évolutives. Cybergeo.
https://doi.org/10.4000/cybergeo.39387

➢ Une base de données pour étudier vingt années de dynamiques du marché immobilier résidentiel en Île-de-France. *Cybergeo*. https://doi.org/10.4000/cybergeo.37430

➢ Les déterminants naturels et politiques des AOC viticoles de Côte-d'Or. Cybergeo.
https://doi.org/10.4000/cybergeo.36443

➢ Données d'enquêtes socioéconomiques sur les ménages agricoles dans les pays du Sud. Cahiers Agricultures. Numéro thématique contenant 4 *Data paper.*
https://www.cahiersagricultures.fr/fr/component/toc/?task=topic&id=889

- Enquêtes sur la consommation, la perception et les utilisations de l'huile de palme rouge chez les ménagères et restauratrices de Yaoundé, Cameroun

- Explorer les liens entre agriculture, migration et sécurité alimentaire : une enquête auprès de ménages agricoles diversifiés et multilocalisés du nord-ouest du Nicaragua

- L'informel et le non-marchand dans les systèmes d'activités : enquête représentative sur les ménages kanak en tribus de Nouvelle-Calédonie

- Explorer les liens entre agriculture et sécurité alimentaire : une enquête auprès des femmes du gouvernorat de Sidi-Bouzid en Tunisie

Rédiger un

Data paper

Etapes à suivre

1. Choisir la revue et voir le modèle de *Data paper*

2. Choisir l'entrepôt de données et voir ses exigences

3. Rédiger le *Data paper* selon le modèle de la revue

4. Préparer les données selon les infos (métadonnées) et le format demandés par la revue et l'entrepôt à condition d'avoir le droit de les publier

5. Déposer les données dans l'entrepôt avec la documentation

6. Soumettre le Data paper à la revue avec le lien vers l'entrepôt où est déposé le jeu de données

Structure classique d'un *Data paper*

- Décrire les données

- Décrire les méthodes d'obtention

- Décrire le potentiel des données

➡️ **Donner le lien d'accès aux données**

Objectif du *Data paper*

➢ Faciliter la réutilisation des données

➢ Le *Data paper* doit décrire complètement :

  • le jeu de données

  • tout le protocole + méthodes + équipements

  • + informations nécessaires pour que le futur utilisateur puisse comprendre les données

➢ Donner le lien pour accéder au jeu de données dans un entrepôt de données

1. Décrire les données

- **Quelles informations peuvent-être utiles ?**

Quel **objet d'étude ?**
variétés ? souches ?
populations ?
écosystème ?

Où s'est déroulée
l'observation ?

Sur quelle **période**
s'est déroulée la collecte ?

Qui a collecté et traité
les données ?

Quel est le **protocole**
expérimental ?

Quels **équipements** ?
contrôles ?
logiciels ? versions ?

Fichiers
de
données

Quels **paramètres**
ont été mesurés ?
avec quelles **unités** ?

Explications des **fichiers**:
Noms, colonnes/lignes
abréviations, variables, unités

++ toutes infos utiles
spécifiques du **domaine**

Toutes les informations sont pertinentes
vous ne savez pas quels sont les besoins d'un futur utilisateur.

RÉPUBLIQUE
FRANÇAISE
Liberté
Égalité
Fraternité

cirad

Inspiré de https://doranum.fr/enjeux-benefices/principes-fair/

1. Décrire les données

➢ Quel est l'objet d'étude et son contexte ?

    plante, animal, écosystème, population, microbes, …
    taxonomie: espèce, genre,…variété, souches,…
    organe étudié, clone, plasmide, …

➢ Quand a eu lieu l'étude ?

    Date, période : en suivant la norme internationale

➢ Où a eu lieu l'étude ?

    Géolocalisation : Geoname, ISO 19115
    Contexte: habitat, climat, type de population, …

➡ Suffisamment pour comprendre l'étude et les données, avec standards disciplinaires si possible

Quelle est la norme internationale pour écrire une date ?

1. **AAAA-MM-JJ** (année-mois-jour)
   2023-04-25

2. **MM-JJ-AAAA** (mois-jour-année)
   04-25-2023

3. **JJ-MM-AAAA** (jour-mois-année)
   25-04-2023

Norme internationale pour écrire une date

- **Norme ISO-8601**: **AAAA-MM-JJ** https://www.iso.org/fr/iso-8601-date-and-time-format.html

  - Pour lever l'ambiguïté quand les dates sont exprimées en chiffres

  - Permet d'exprimer la date de manière claire et compréhensible par les personnes et les machines.

Vol Abidjan – New York

Départ le 05-01-23

5 janvier 2023
DD-MM-YYYY

Arrivée le 05-01-23

1 mai 2023
MM-DD-YYYY

Départ le 2023-01-05

Guides de standards de métadonnées

Standards de métadonnées = Règles de description des données définies par une communauté scientifique pour une discipline.

Digital Curation Centre

http://www.dcc.ac.uk/resources/metadata-standards

Research Data Alliance

http://rd-alliance.github.io/metadata-directory/

FAIRsharing.org
standards, databases, policies

https://fairsharing.org/

elixir

https://www.elixir-europe.org/communities

https://www.iso.org/fr/home.htm

Utiliser un standard est recommandé pour faciliter la compréhension des données

2. Décrire les méthodes

- **Méthodes de collecte sur le terrain**
  données d'observation, inventaire de biodiversité, …
  méthode d'échantillonnage
  enquêtes: face/face, tél, en ligne, nombre personnes interrogées

- **Protocole expérimental (labo, serre, animalerie, …)**
  mode de culture (champs, *in vitro*, …), traitements, intrants,
  variables et paramètres étudiés, unités de mesure
  Contrôle qualité,…

- **Type d'équipements** de mesures, de séquençage, …

→ Suffisamment pour reproduire l'étude,
avec protocole classique dans la discipline si possible

3. Décrire le potentiel des données

Objectif: convaincre l'éditeur que vos données sont dignes d'intérêt, crédibles, fiables, originales et même indispensables.

➤ Montrez qu'elles ont une portée scientifique
- Contribution à un défi majeur: *changement climatique, biodiversité, santé,…*
- Collectées sur plusieurs années, plusieurs espèces, toute une filière, …
- Données de référence, généralisables à plusieurs espèces,
- Utiles pour développer des modèles, outils d'intelligence artificielle, …

➤ Une valeur économique, sociétale, environnementale, …
- Développement de tests, de vaccins, d'outils, de services,…
- Impact bénéfique sur la société ou sur les politiques publiques
- Valeur historique, patrimoniale, …

RÉPUBLIQUE
FRANÇAISE
Liberté
Égalité
Fraternité

cirad

3. Décrire le potentiel des données

➢ **Montrez que vos données sont originales**

- Données rares ou uniques :
  expérimentations impossibles à répéter ou très couteuses, phénomènes rares,…

- Données jamais publiées

➢ **Evaluez l'originalité de vos données** en cherchant si des données similaires ont déjà été publiées:

- Outils de recherche de données:
  Datacite Search, Google Dataset Search

- Voir aussi « Explorer les moteurs de recherche scientifiques » :
  https://www.datacc.org/vos-besoins/trouver-des-donnees/moteurs-de-recherche-scientifiques-et-data-journals/

➡ Description suffisante pour convaincre l'éditeur du potentiel, de la valeur et de la qualité de vos données

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

cirad

4. Préparer les données

> Identifier le plus tôt possible la revue et l'entrepôt pour voir les recommandations
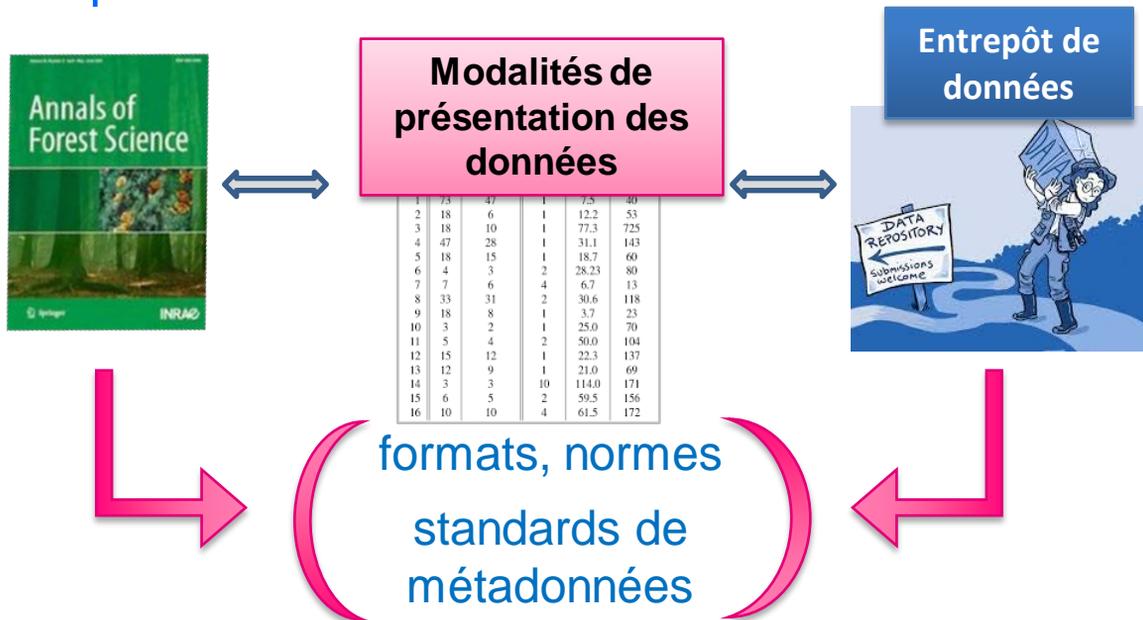
4. Préparer les données

➤ Vérifier que vous répondez aux attentes de la revue et de l'entrepôt
  - Format recommandé
  - Standard et métadonnées requises
  - Documentation associée

➤ Expliquer vos fichiers et tableaux
  - Noms des fichiers,
  - Abréviations,
  - Lignes / colonnes ,….

→ Suffisamment pour que le lecteur puisse comprendre, interpréter et accéder aux données

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

cirad

Selon vous quelles informations manquent pour comprendre ce tableau ??

	A	B	C	D	E
1		GTT date	GTT weight	Time	Glucose mg/dl
2	321	2/9/15	24.5	0	99.2
3				5	349.3
4				15	286.1
5				30	312
6				60	99.9
7				120	217.9
8	322	2/9/15	18.9	0	185.8

Expliquer les fichiers de données

> Expliquez vos tableaux

	A	B	C	D	E
1		GTT date	GTT weight	Time	Glucose mg/dl
2	321	2/9/15	24.5	0	99.2
3				5	349.3
4				15	286.1
5				30	312
6				60	99.9
7					
8	322	2/9/15			

Signification des lignes et des colonnes ? abréviations ?

Date: JJ/MM/AA ? MM/JJ/AA ?

Unités de mesure : Temps et Poids

Pourquoi certaines cases sont vides ?

Critères d'évaluation (peer reviewing)

➢ **Les Data papers sont tous évalués sur:**

- Originalité et valeur des données
- Intérêt des données dans la discipline et pour des recherches futures
- Rigueur de la méthode, Qualité et fiabilité des données
- Adéquation avec les pratiques courantes de la discipline
- Description suffisante pour permettre à d'autres de comprendre, interpréter et réutiliser les données

➢ **Contrôle de l'accès aux données**

dès la soumission pour certaines revues !!

A retenir

➢ 1ère étape : choix de la revue et de l'entrepôt
  → Détermine le modèle de *Data paper*
  → organisation et description des données (format, standards,…)

➢ Rédiger le *Data paper* avec assez d'infos pour qu'1 lecteur puisse:
  - comprendre et reproduire l'étude
  - interpréter et réutiliser les données
  - avoir confiance dans la rigueur de la recherche
  - évaluer l'importance et le potentiel des données

➢ Un *Data paper* compte comme un vrai article
  - validé par une relecture par les pairs
  - citable

Rechercher des données

Rechercher grâce aux moteurs de recherc

- **DataciteCommons :** https://commons.datacite.org/

- **Google Dataset Search :** https://datasetsearch.research.google.com/

- **BASE :** https://www.base-search.net/Search/Advanced

- Voir aussi : Explorer les moteurs de recherche scientifiques
  https://www.datacc.org/vos-besoins/trouver-des-donnees/moteurs-de-recherche-scientifiques-et-data-journals/

Rechercher grâce aux moteurs de recher

- **DataciteCommons :** https://commons.datacite.org/

- Datacite : fournisseur mondial de DOI pour les données scientifiques : → ~ 20 millions de jeux de données issus de près de 2000 entrepôts.

- Donc le meilleur outil pour rechercher des données.

Mots clés

Rechercher grâce aux moteurs de recherc

- **DataciteCommons :** https://commons.datacite.org/

sugarcane

Sélection
« Dataset »

Autres critères :
Publication year
Field of science

- **DataciteCommons :** https://commons.datacite.org/

A vous avec
vos mots-clés

DataCite
Commons

Type to search...   🔍   Pages ▾   Support   🔑 Sign In

📧 Works   👤 People   🏛 Organizations   📦 Repositories

Search works by keyword(s) and/or identifier. Documentation is available in **DataCite Support.**

RÉPUBLIQUE
FRANÇAISE
Liberté
Égalité
Fraternité

cirad

Pause déjeuner

cirad

Revues scientifiques publiant des *Data papers*

Connaissez-vous des revues publiant des *Data paper* ?

➢ Oui

➢ Non

➢ Ne sait pas

Revues publiant des *Data papers*

Revues publiant des *Data papers*

- ### Data journaux
  - publient seulement des *Data papers*
  - généralement en libre accès : gratuit ou coût variable
- ### Revues classiques
  - publient ≠ types d'articles dont des Data papers
  - gratuite ou non, libre accès ou en option payante
- ### multidisciplinaires, disciplinaires ou thématiques
- ### avec ou sans facteur d'impact
- ### ≠ noms : *Data paper, Data descriptor, Data article, Datasets, Data Briefs, Resource Announcements, Data Resource Profile*

Critères de choix d'une revue

- **Domaine scientifique / lectorat** (futurs utilisateurs)

- **Exigences de la revue** (Instructions aux auteurs)
  - Echelle du jeu de données : couverture géographique, temporelle ou taxonomique, intérêt pour large communauté, …
    ex : *Global Ecology and Biogeography, GigaScience, Plant Journal, Nature Biotechnology, International Journal of Epidemiology*
  - Modèle du *Data paper* : simple, complexe, texte libre
  - Modalités d'accès aux données
  - Accès aux données et entrepôts recommandés
  - Modalités de diffusion des données (licences)

- **Libre accès à l'article**

- **Coût de publication** : Varie de gratuit à + de 3 000 €

Critère 1 – Modèle de *Data paper*

- Selon la revue, le modèle de *Data paper* diffère entre:
- Modèle classique
  Data in Brief, Geoscience Data Journal, Ecological research

Modèle classique de *Data paper*

Title, Authors, Abstract, Keywords

Objective : *Décrire le contexte de l'étude*

Data description : *Décrire les données et les fichiers*

Methods : *Décrire les méthodes pour générer les données*

Value of the data : *Décrire le potentiel de reutilisation des données*

➡️ **Lien d'accès aux données déposées dans un entrepôt**

Acknowledgements, References

Critère 1– Modèle de *Data paper*

- Selon la revue, le modèle de *Data paper* diffère entre:

➤ Modèle classique
  Data in Brief, Geoscience Data Journal, Ecological research

➤ Modèle classique + table de métadonnées
  Annals of Forest Science

Modèle avec table de métadonnées

Spatial coverage	**Minimum longitude (x min in decimal degre**
	Maximum longitude (x max in decimal degr
	Minimum latitude (y min in decimal degre
	Maximum latitude (y max in decimal degre
Temporal coverage of	**Start date**
the collected data	**End date**
Accessibility	**Online ressource URL**
	Online ressource name
	Constraint/License
Metadata provider(s)	**Name and first name**
	Organization / Institute
	Telephone (optional)
	Address
	City
	Postal code
	Country
	E-mail address
	ORCID identifier (optional)

Type of data	Measurements
	LFMC raw measurements and different statistics (mean, robust mean estimate and robust standard error) are given for a given date, a given site, and a given species. In the raw data file flags are provided (for each replicate and the mean value) allowing to filter outliers. The procedure to filter outlier is explained in the data paper (Martin-StPaul et al)
	Rainfall measurements are provided in a separated file. The measurement date and the Day Of Year (DOY) of measurements and the previous DOY when a reliable measurements was performed are given. See Martin-StPaul et al
	The data come from measurements taken in 35 different sites. Coordinates of each site are given in the table **InfoSite_ReseauHydrique**
Protocols	Apical and lateral shoots of branches fully exposed to the sun are sampled from different individuals of a given species within circular plots with a radius between five and twenty meters. All sampled individuals are representative of the average status of each chosen species. Samples are mixed together and separated in five subsamples of ten to fifteen grams of living vegetation. Samples are enclosed in small containers sealed with paper tape and weighted fresh in laboratory, oven-dried at 60°C during twenty four hours and weighted dry. The sampling operations take place at ca. 12:00 UT. LFMC is computed and released on fresh mass basis. The conversion to a dry mass basis is given in Martin-StPaul et al
Equipment / software	Oven
	Balance (0.001g precision)
	Rainfall gauge

	1	2	3	4	5	6	7
	Variable name	**Access**	**Additional access information**	**Description**	**Type**	**Unit**	**Value range**
	SiteCode	http://doi.org/10.5281/zenodo.162978	spread sheet "1_LFMC_final_Table"	Code allowing to identify the site where measurements are performed. The site labels are «DmSn» where m is the «département» number and n the site number within a «département». In the database presented below, site identifiers are unique and static (contrary to the operational database). Data from non-geolocalized (NG) sites are labelled «DmSNGn» with n the site number within département and m the index of the non-geolocalized site.	character		
	SiteName	http://doi.org/10.5281/zenodo.162978	spread sheet "1_LFMC_final_Table"	Common name of the site (i.e. "place called")	character		
	Species	http://doi.org/10.5281/zenodo.162978	spread sheet "1_LFMC_final_Table"	latin binomial name (Genus species)	character		
	SiteXSpecies	http://doi.org/10.5281/zenodo.162978	spread sheet "1_LFMC_final_Table"	unique identifier for the site and species	character		
	Date	http://doi.org/10.5281/zenodo.162978	spread sheet "1_LFMC_final_Table"	sampling date	character		
	Year	http://doi.org/10.5281/zenodo.162978	spread sheet "1_LFMC_final_Table"	sampling year	numeric integer		1996 – 2016
	Month	http://doi.org/10.5281/zenodo.162978	spread sheet "1_LFMC_final_Table"	sampling month	numeric integer		1 – 12
	Doy	http://doi.org/10.5281/zenodo.162978	spread sheet "1_LFMC_final_Table"	sampling day of year	numeric integer		1 – 366
	RobustLFMC	http://doi.org/10.5281/zenodo.162978	spread sheet "1_LFMC_final_Table"	Robust estimates of mean LFMC. Detailed of computation are provided in Martin-StPaul et al 2017	numeric float	%	10 -- 320

Critère 1 – Modèle de *Data paper*

- Selon la revue, le modèle de *Data paper* diffère entre:

➢ Modèle classique
Data in Brief, Geoscience Data Journal, Ecological research

➢ Modèle classique + table de métadonnées
Annals of Forest Science

➢ Modèle structuré par des métadonnées disciplinaires
Ecology, Open Health Data, Freshwater Metadata Journal

Modèle structuré par métadonnées disciplinaires

ECOLOGY

Data Papers

Fifty years of continuous precipitation and stream chemistry data from the Hubbard Brook ecosystem study (1963–2013)

GENE E. LIKENS[1]

Cary Institute of Ecosystem Studies, 2801 Sharon Turnpike, Millbrook, New York 12545 USA

1 June 1963. This veekly stream water ine 1963 to 30 May chemistry for preci-Forest are reported

Brook Experimental

Class II. Research origin descriptors
Site Description
- Site type
- Geography (location, size)
- Habitat
- Geology, Landform
- Climate

Experimental or sampling design
- Design characteristics
- Variables included
- Species sampled
- Data collection period, frequency

Research methods
- Field/Laboratory
- Instrumentation

Filename
Word document, 125.5 KB
ecy1894-sup-0001-MetadataS1.doc
Zip archive, 684.2 KB
ecy1894-sup-0002-DataS1.zip

Class III. Data set status and accessibility
Status
- Latest update
- Metadata status
- Data verification

Accessibility
- Storage location and medium
- Contact person(s)
- Copyright restriction
- Costs

Class IV. Data structural descriptors
A. Data Set Files
- Data set Identity
- Size
- Format

B. Variable information
- Variable definition
- Units of measurement
- Data type

RÉPUBLIQUE FRANÇAISE
Liberté Égalité Fraternité

cirad

Modèle structuré par métadonnées disciplinaires

Class II. Research origin descriptors
Site Description

Class II. Research origin descriptors
Sampling design

Class IV. Data structural descriptors
B. Variable Information

ECOLOGY
ECOLOGICAL SOCIETY OF AMERICA

1. Site description-

a. Site type: The site is an experimental forest located in the [White Mountains, New] Hampshire, operated under the supervision of the Northern R[esearch Station of the United] States Forest Service.

b. Geography: The Hubbard Brook Experimental Forest is a[t] altitude from 222 to 1,015 m.

> *"The HBEF is located within the townships of North*
> *Ellsworth (~40 %), and Warren (<1 %), NH, within [the White Mountain National]*
> *Forest of north central New Hampshire. Coordinate[s]*
> *the area. The Atlantic Ocean is about 116 km to the*
> *West Thornton, NH."*

(Likens 2013: 5)

c. Habitat: The ecosystem is a mixed temperate hardwood [forest (...) sugar maple]
(Acer saccharum), American beech *(Fagus grandifolia)* and [(...) yellow birch (Betula]
alleghaniensis).|

d. Geology: Bedrock geology is composed of quartz-mica s[chist (...) of the Silurian]
Rangeley Formation. Significant glacial deposits overlay m[uch of the (...) in the]
spodosol and inceptosol orders.

e. Watersheds/hydrology:

> *"The experimental watershed-ecosystems range in*
> *altitude from 500 to 910 m. These headwater waters[heds have (...)]*
> *slope of 20–30 %), with well-incised channels and re[latively (...)]*
> *divides. The height of the land surrounding each wa[tershed (...)]*
> *and the topography all have been determined from g[round surveys (...) aerial]*
> *photography and most recently augmented by Laser [Imaging Detection and]*
> *Ranging (LIDAR). Experimental Watersheds 1–6 fa[ce (...) watersheds 7-9]*
> *face N to NE."*

(Likens 2013: 7)

f. Site history: Secondary forest, logged ca. 1910. Extensiv[e (...) may be]
found in Likens and Bormann (1979), Holmes and Likens ([2016]) and at
http://www.hubbardbrook.org/overview/sitedescription.shtm[l]

g. Climate: mean annual precipitation - 1434 mm, mean an[nual]

2. Sampling design: The sampling design is exhaustively described i[n]
Buso et al. (2000). Perhaps the most important aspect to explain here [in the]
precipitation dataset is curated. Contamination issues always plague r[ain (...) the]
policy for the HBEF long-term record has been to curate the dataset b[ut (...) raw]
sample information is retained. To help explain this we provide the fo[llowing]
approach to QA/QC of precipitation samples

Precipitation Chemistry records within the dataset were compiled fro[m (...) order]
to represent the "best" sample (read "cleanest") for each weekly colle[ction. (...) At]
the HBES, the issue of precipitation contamination by coarse particle[s (...)]
clearly seen as a potential problem in open collectors. While not excl[usively a summer]
problem, the worst incidents tended to occur in the summer and fall. [(...) To]
eliminate foreign objects, such as screens, wads of poly-wool, and us[ing bottle]
collectors, but these were problematic in their own way. Screens had [(...) wherein]
precipitation was 'filtered' through trapped debris. It was eventually d[ecided that]
a bulk (continuously open) collector was the most appropriate choice. [(...) used]
several collectors so that at least one sample collected from at least on[e was]
reasonably clean.

Although occasional grossly contaminated samples were not analyzed [(at time of]
collection), most precipitation samples were sent to the analytical lab[. (...)]
the presence of amorphous fine particulates. It turned out that some fi[ne (...)]
and influence the analysis (probably bird feces), while others appeare[d to be]
fine pieces of bark or grit). There was no evidence, in the early days o[f (...)]
chemistry from different sites around HBEF would result in significan[t (...)]
particulate contamination was considered a more important factor in i[nfluencing]
difference in chemistry across the valley. Long-term data since then b[ear (...) Of]
course, the resulting input flux values were always calculated using th[e (...)]
or a pattern of rain gauges in the case of a whole watershed, since the [(...)]
is significant.

Thus, precipitation data reported here represents a 'mosaic' of the hig[hest (...) samples]
from several sites, picked based on obvious field observations of entr[ained debris (...)]
chemistry returned from the laboratory. The basic intention of this ap[proach is (...) best]
sample for each period (week) in which precipitation occurred. For ea[ch (...)]
a single rain chemistry estimate for the south facing watersheds (Site [(...)]

1. Variable name	2. Variable definition	3. Unit	4.a. Storage type	4.c. Range	4.e. Precision
W(ws)	Each watershed is numbered, see Table 3 below.	N/A	integer		N/A
datetime	Date and time of sampling, YYYY-MM-DD hh:mm:ss format	N/A	string		1 minute
date	date of sampling, YYYY-MM-DD format (included for samples where there exists no timestamp)	N/A	string		1 day
Ca	calcium concentration	mg/L	floating point	[0.005, 11.8]	0.001
Mg	magnesium concentration	mg/L	floating point	[0.005, 4.62]	0.01
K	potassium concentration	mg/L	floating point	[0.003, 8.35]	0.01
Na	sodium concentration	mg/L	floating point	[.005, 2.75]	0.01
Al	total aluminum concentration	mg/L	floating point	[0. 3.4]	0.001
NH4	ammonium concentration	mg/L as NH4, NOT NH4-N	floating point	[.001, 1.0]	0.001
pH	pH	unitless	floating point	[3.70, 6.93]	0.01
SO4	sulfate concentration	mg/L as SO4	floating point	[0.05, 11.8]	0.01

# Critère 1 – Modèle de *Data paper*

- **Selon la revue, le modèle de *Data paper* diffère entre:**

➤ Modèle classique
Data in Brief, Geoscience Data Journal, Ecological research

➤ Modèle classique + table de métadonnées
Annals of Forest Science

➤ Modèle structuré par des métadonnées disciplinaires
Ecology, Open Health Data, Freshwater Metadata Journal

➤ Texte libre mais limité en taille (2 pages, 1000 mots, …)
Phytopathology, Plant Phenome Journal, Hydrological Processes

Modèle en texte libre et court

https://apsjournals.apsnet.org/journal/mpmi

MPMI
Molecular Plant-Microbe Interactions

RESOURCE ANNOUN

A High-Quality Draft Genome Sequence of *Colletotrichum gloeosporioides* sensu stricto SMCG1#C, a Causal Agent of Anthracnose on *Cunninghamia lanceolata* in China

Lin Huang,[1] Ki-Tae Kim,[1] Ji-Yun Yang,[1] Hyeunjeong Song,[3] Gobong Choi,[3] Jongbum Jeon,[3] Kyeongchae Cheong,[3] Jaeho Ko,[2] Haibin Xu,[4,5] and Yong-Hwan Lee[2,3,5,†]

[1] Co-Innovation ... China, College of Forestry, Nanjing Forestry University ...
[2] Department ... University, Seoul 08826, Korea;
[3] Interdisciplinary Program ... Seoul National University;
[4] College of Biology ... Environment, Nanjing Forestry University;
[5] Center for Fungal ... ic Resources, Plant Genomics and Breeding Institute, and Research Institute of Agriculture and ... ce Sciences, Seoul National University

Résumé

Abstract

Colletotrichum has a broad host range and causes major yield losses of crops. The fungus *Colletotrichum gloeosporioides* is associated with anthracnose on Chinese fir. In this study, we present a high-quality draft genome sequence of *C. gloeosporioides* sensu stricto SMCG1#C, providing a reference genomic data for further research on anthracnose of Chinese fir and other hosts.

Colletotrichum is one of the most important groups of phytopathogenic fungi in the world because of its scientific and economic importance (Dean et al. 2012). *Colletotrichum gloeosporioides* is a ubiquitous plant pathogen that infects a wide range of plant species (Weir et al. 2012). According to recent advances in taxonomy, *C. gloeosporioides* is considered a species complex and is segregated into 22 species and one subspecies (Weir et al. 2012). Chinese fir (*Cunninghamia lanceolata*) has been cultivated for over 3,000 years and contributes about 40% of timber in southern China (Huang et al. 2018; Shi et al. 2010). Anthracnose caused by *C. gloeosporioides* is one of the most serious fungal diseases on Chinese fir, which is widely distributed in the cultivated areas of Chinese fir, and causes enormous economic losses (Lan et al. 2015). *C. gloeosporioides* SMCG1#C was isolated from the infected leaves of Chinese fir in Nanjing, China. Based on the phylogenetic tree calculated from the alignment of concatenated sequences of ITS, ACT, CAL, CHS-1, and APDH, strain SMCG1#C was identified as *C. gloeosporioides* sensu stricto (*unpublished data*).

The genome of *C. gloeosporioides* SMCG1#C was sequenced, using both PacBio Sequel System (Tianjin Biochip Corporation, Tianjin, China) and Illumina Hiseq X Ten System (Novogene Corporation, Beijing, China). A total of 519,294 reads and 171,464,766 paired-end 150-bp Illumina reads were generated, with respective coverages of 71× and 414×. De novo assemblies were performed using Velvet version 1.2.10 (Zerbino and Birney 2008). We obtained 28 contigs with an average length of 2,210,112 bp, an N50 of 4,696,547 bp, and L50 of 5. Finally, a draft genome of 18 scaffolds was produced by using BLASR and BLASTn algorithms (Camacho et al. 2009; Chaisson and Tesler 2012), a total of 61.9 Mb, a G+C content of 50.3%, N50 of 5,209,244 bp, and L50 of 5 (Table 1). The validation of assembly

Table 1. Genome assembly statistics of *Colletotrichum gloeosporioides* sensu stricto SMCG1#C and the other *C. gloeospori...* species

Parameter	*C. gloeosporioides*		*C. fructicola*	
	SMCG1#C	Cg-14	Nara gc5	1104-7
Host	Chinese fir	Avocado	Strawberry	Apple
Total assembly length (bp)	61,916,549	53,209,944	55,607,143	57,082,694
Number of scaffolds	18	4,537	1,241	684
G+C (%)	50.3	53.4	53.4	53.2
N50 (bp)	5,209,244	25,337	112,809	339,165
L50	5	656	152	50
Number of genes	16,287	16,538	15,381	17,827
Number of secreted proteins	1,830	1,648	1,657	1,913
BUSCO completeness	99.3%	91.7%	94.8%	99.3%
Reference	This study	Alkan et al. 2013	Gan et al. 2013	Liang et al. 2018

was achieved by BUSCO v3.0.2, using the fungi dataset (Waterhouse et al. 2018), and it showed 99.3% completeness for the assembled genome. A whole-genome alignment analysis using MUMmer v3.23 with species in the *C. gloeosporioides* species complex (Delcher et al. 2002), including *C. gloeosporioides* Cg-14, *C. fructicola* Nara gc5, and *C. fructicola* 1104-7 (Alkan et al. 2013; Gan et al. 2013; Liang et al. 2018), revealed that strain SMCG1#C was close to the other *C. gloeosporioides* (83% coverage) rather than the *C. fructicola* strains (69 and 71% coverage for Nara gc5 and 1104-7, respectively).

Structural annotation of the genome was performed using the MAKER v2.31.8 pipeline (Holt and Yandell 2011), and 16,287 protein-coding genes were identified. Among them, functions of 14,269 proteins (87.6% of proteome) were annotated by InterProScan 5.21-60.0 (Jones et al. 2014). According to the gene family pipelines previously described, 23 laccases, 48 peroxidases, 137 plant cell wall–degrading enzymes, 707 transcription factors, 281 Cytochrome P450, and 1,830 secretory protein–coding genes were predicted (Choi et al. 2010, 2013a, 2014; Park et al. 2008a and b). Among the secretome, 750 proteins were identified as small secreted proteins (<300 amino acids) that might function as effectors (Kim et al. 2016). In addition, 1,076 CAZymes, 930 peptidases, and 246 lipases were predicted by dbCAN release 6.0, MEROPS release 12.0, and LED release 3.0 pipelines, respectively (Fischer and Pleiss 2003; Rawlings et al. 2018; Yin et al. 2012). The ortholog clustering analysis of *C. gloeosporioides* SMCG1#C with the species complex and with *C. orbiculare*, *C. graminicola*, and *C. higginsianum* as outgroup (Dallery et al. 2017; Gan et al. 2013; O'Connell et al. 2012), using OrthoFinder v2.2.6 revealed 2,947 orthogroups specific to the species complex (Emms and Kelly 2015). Among them, 1,438 orthogroups were shared by all four strains and 547 orthogroups were only shared between the two *C. gloeosporioides* strains (557 genes in SMCG1#C and 550 genes in Cg-14). Lastly, the strain SMCG1#C had 407 orphan genes and 55 genes were functionally annotated as cation binding, transport, and integral component of membrane for the top three gene ontology terms.

The genome of *C. gloeosporioides* sensu stricto SMCG1#C is, so far, the best quality genome within the published genomes of *C. gloeosporioides* species complex (Table 1), and it will be ...

genome data has been deposited in the National Center for Biotechnology Information NCBI GenBank database under accession number QFRH00000000, PRJNA471237 for Bio-Project, and SAMN09205517 for BioSample. The genome sequence and gene models are also available from the Comparative Fungal Genomics Platform 2.0 (Choi et al. 2013b) and its sister databases described above.

GenBank Accession N°

Literature Cited

Alkan, N., Meng, X., Friedlander, G., Reuveni, E., Sukno, S., Sherman, A., Thon, M., Fluhr, R., and Prusky, D. 2013. Global aspects of pacC regulation of pathogenicity genes in *Colletotrichum gloeosporioides* as revealed by transcriptome analysis. Mol. Plant-Microbe Interact. 26:1345-1358.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T. L. 2009. BLAST+: Architecture and applications. BMC Bioinformatics 10:421.

Chaisson, M. J., and Tesler, G. 2012. Mapping single molecule sequencing reads using basic local alignment with successive refinement (BLASR): Application and theory. BMC Bioinformatics 13:238.

Choi, J., Cheong, K., Jung, K., Jeon, J., Lee, G. W., Kang, S., Kim, S., Lee, Y. W., and Lee, Y. H. 2013b. CFGP 2.0: A versatile web-based platform for supporting comparative and evolutionary genomics of fungi and Oomycetes. Nucleic Acids Res. 41 (D1):D714-D719.

Critère 1– Modèle de *Data paper*

- Selon la revue, le modèle de *Data paper* diffère entre:

➢ Modèle classique
  Data in Brief, Geoscience Data Journal, Ecological research

➢ Modèle classique + table de métadonnées
  Annals of Forest Science

➢ Modèle structuré par des métadonnées disciplinaires
  Ecology, Open Health Data, Freshwater Metadata Journal

➢ Texte libre mais limité en taille (2 pages, 1000 mots, …)
  Phytopathology, Plant Phenome Journal, Hydrological Processes

➢ Modèle avec Résultats et Discussion
  Earth System Science Data, Ethnobiology Letters, Plant Journal

➡ Bien choisir sa revue

Critère 2 – Modalités d'accès aux données

- **La revue exige l'accès aux données**

  - au moment de la publication

  - dès la soumission au moins pour les évaluateurs
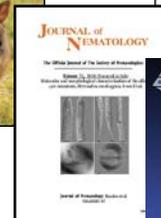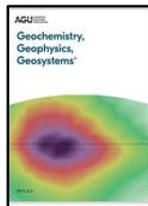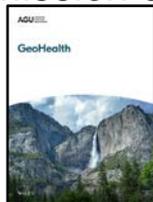
- **La revue accepte que les données soient accessibles**

  - après embargo

  - sur demande ou collaboration
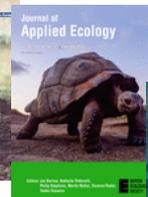
Critère 3 – accès aux données et entrepôts

- **La revue recommande des entrepôts** (le + fréquent)
    - Disciplinaires, thématiques, généralistes, éditeurs

Accès aux données et entrepôt

Seventy years of data from the world's longest grazed and irrigated pasture trials

Rich. W. McDowell ✉, R. A. Moss, C. W. Gray, L. C. Smith & G. Sneath

If a discipline-specific repository does not exist,
data should be submitted to a generalist repository.

Abstract

Pastures are the most widespread land use, globally. The Winchmore trial
established in 1948–1949 in Canterbury, New Zealand and examined eith
rates of phosphorus (P) fertiliser on the same irrigation schedule (Fertiliser
different irrigation scheduling at the same rate of P application (Irrigation
96,000 records of soil chemistry and physical data and pasture yield and l
composition are available along with nearly 7000 soil samples. These data
used in 475 publications that have explored topics as diverse as: improvei
sheep, dairy and deer production; the efficacy and scheduling of irrigation
improvements in pasture and crop production; agronomic and environme
water research; and entomology. In addition to above topics, these data a
invaluable for calibrating models to predict long-term issues like the accur
soil carbon or contaminants like cadmium and informing policy on climate
and agricultural practices. The data and soil samples are available for use i
yield discoveries, unforeseen 70 years ago.

View data repositories

- **Biological sciences:** Nucleic acid sequence; Protein sequence; Molecular & supramolecular structure; Neuroscience; Omics; Taxonomy & species diversity; Mathematical & modelling resources; Cytometry and Immunology; Imaging; Organism-focused resources
- **Health sciences**
- **Chemistry and Chemical biology**
- **Earth, Environmental and Space sciences:** Broad scope Earth & environmental sciences; Astronomy & planetary sciences; Biogeochemistry and Geochemistry; Climate sciences; Ecology; Geomagnetism & Palaeomagnetism; Ocean sciences; Solid Earth sciences
- **Physics**
- **Materials science**
- **Social sciences**
- **Generalist repositories**

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

cirad

Accès aux données et entrepôts

RMetS
Geoscience
Data Journal

DATA PAPER

Geoscience Data Journal

A real-world dataset and data simulation algorithm for automated fish species identification

Vaneeda Allken[1] | Shale Rosen[1] | Nils Olav Handegard[1] | Ketil Malde

[1]Institute of Marine Research, Bergen, Norway

[2]Department of Informatics, University of Bergen, Bergen, Norway

Correspondence
Vaneeda Allken and Shale Rosen, Institute of Marine Research, P. O. Box 1870, Nordnes, N-5817 Bergen, Norway.
Email: vaneeda.allken@hi.no; shale.rosen@hi.no

Funding information
Norges Forskningsråd, Grant/Award Number: 203477 and 270966/O70; Norwegian Ministry of Trade, Industry and Fisheries

Abstract
Developing high-performing machine learning algorithms requi of annotated data. Manual annotation of data is labour-intensive effort needed are an important obstacle to the development and tomated analysis. In a previous work, we have shown that deep can successfully be trained on synthetic images and annotations. a curated set of fish image data and backgrounds, the necessary generate synthetic images and annotations, and annotated real da sifier performance. The dataset is constructed from images collec Vision system during two surveys from 2017 and 2018 that targ important pelagic species in the Northeast Atlantic Ocean. We a 1,879 images, randomly selected across trawl stations from both s 482 images of blue whiting, 456 images of Atlantic herring, 341 mackerel, 335 images of mesopelagic fishes and 265 images conta the four categories.

KEYWORDS
data augmentation, fish dataset, machine learning, synthetic data

Approved thematic data repositories commonly used by the scientific community it supports, formal data management policy in place, provide a stable URL and unique identifier for the dataset.

4TU.ResearchData
British Atmospheric Data Centre (BADC)
British Oceanographic Data Centre (BODC)
CSIRO Data Access Portal
Deep Carbon Observatory
eartH$_2$Observe
Environmental Information Data Centre (EIDC)
IEDA:EarthChem
IEDA:MGDS
Earth Observing Lab (EOL),
National Geoscience Data Centre (NGDC)
NERC Earth Observation Data Centre (NEODC)
NOAA National Climatic Data Center (NCDC)
NOAA National Oceanographic Data Center
NOAA National Geophysical Data Center
PANGAEA
Polar Data Centre (PDC)
Figshare
Zenodo

Accès aux données et entrepôts

Phytoplankton species abundance in Tokyo Bay (Japan) from 1998 to 2019

Takashi Nakada, Toshiya Katano ✉ Keigo Hashimoto, Maiko Kagami ✉

Read the full text >   📄 PDF   🔧 TOOLS   ❮ SHARE

Abstract

Tokyo Bay is a semienclosed coastal system located in the center of Japan. Eutrophication has progressed since 1950, resulting in red tides and blue tides appearing frequently. This Data Paper reports the abundance of phytoplankton species sampled monthly from April 1998 to March 2019 at 17 stations in Tokyo Bay. Monitoring has been carried out by Chiba Prefectural Government. Phytoplankters were enumerated by microscopy as numbers of cells or colonies. The abundance of each species was expressed as cells per milliliter of seawater. The total cell density ranged from 10^2 to 10^6 cells·ml^{-1}. The dominant taxa were diatoms, such as *Chaetoceros* spp., *Skeletonema* spp., and *Thalassiosira* spp. in most of the months. The dominant diatom species reached concentrations of 10^6 cells·ml^{-1}. Microflagellates or Cryptomonadaceae were also abundant, especially in winter, reaching 10^6 cells·ml^{-1}. These data can be used to appreciate how anthropogenic disturbances such as eutrophication and global warming affect the density and community composition of phytoplankton. The detailed Metadata for this abstract published in the Data Paper section of the journal is available in MetaCat in JaLTER at http://db.cger.nies.go.jp/JaLTER/metacat/metacat/ERDP-2021-03.1/jalter-en.

https://esj-journals.onlinelibrary.wiley.com/journal/14401703

LTER
International Long-Term
Ecological
Research Network
(Entrepôt thématique)

or other open-access depositories (e.g. FigShare, Dryad, GBIF).

Critère 3 – accès aux données et entrepôts

- **La revue recommande des entrepôts** (le + fréquent)
  - Disciplinaires, thématiques, généralistes, éditeurs

- **La revue impose un entrepôt de données** (rare)

  Studies in Mycology (MycoBank)

  Chemosphere (Pangaea)

  Systematic Botany (DRYAD, GenBank)

- **La revue accepte d'intégrer les données dans l'article**
  - Non conseillé : les données seront moins accessibles, moins visibles et plus difficilement réutilisables

Accès aux données et entrepôts

Data in Brief
Volume 42, June 2022, 107998

ELSEVIER

Data Article

Data on growth, uptake and N_2 fixation of grass-clover leys fertilized with mineral N fertilizer and cattle slurry

Abstract

This article presents the data obtained from a field experiment in which grass clover leys were fertilized with increasing N rates applied in either mineral N fertilizer and/or cattle slurry forms. The leys were composed of a 2–species m of white clover (*Trifolium repens L.*) and ryegrass (*Lolium perenne L.*) and a 4–species mixture of white clover, red clover (*Trifolium pratense L.*), festulolium (*Festuloli braunii*) and ryegrass. In total, eighty fields were established at two farm sites western part of Denmark on sandy soils and monitored for two herbage seaso (2018–2019). Dry matter yield, botanical composition, N concentration and th proportion of N derived from the atmosphere using the 15 N dilution metho recorded in the harvestable biomass after each cut. Furthermore, the specific growth, N uptake and quantitative biological N fixation of the species were determined.

Specifications Table

Subject	Agronomy and Crop Science
Specific subject area	N fertilization of grass–clover leys
Type of data	Tables, Excel workbook
How data were acquired	Field experiments performed during two consecutive years at two farmer sites in Western Jutland (Denmark).
Data format	Raw data
Description of data collection	Biomass was harvested and sampled five times during the herbage season. An included dry matter yields, botanical composition, N concentration and %Nd
Data source location	South-western Jutland (SW) and Mid-western Jutland (MW); Denmark SW coordinates: 55°32′N, 8°29′E MW coordinates: 56°10′N, 8°46′E
Data accessibility	With the article

Critère 4 – Modalités de diffusion des données

DATA NOTE   **Open Access**

Non-structural carbohydrates and morphological traits of leaves, stems and roots from tree species in different climates

Guangqi Zhang[1*], Pascale Maillard[2], Zhun Mao[1], Loic Brancheriau[3], Julien Engel, Bas
Claire Fortunel[1], Jean-Luc Maeght[1], Jordi Martínez-Vilalta[4,5], Merlin Ramel[1], Sophie Nou
Stéphane Fourtier[1] and Alexia Stokes[1]

Données déposées dans un entrepôt de données

Licence ≠

Abstract

Objectives: Carbon fixed during photosynthesis is exported from leaves towards sink organs as non-structural carbohydrates (NSC), that are a key energy source for metabolic processes in trees. In xylem, NSC are mostly stored as soluble sugars and starch in radial and axial parenchyma. The multi-functional nature of xylem means that cells possess several functions, including water transport, storage and mechanical support. Little is known about how NSC impacts xylem multi-functionality, nor how NSC vary among species and climates. We collected leaves, stem and root xylem from tree species growing in three climates and estimated NSC in each organ. We also measured xylem traits linked to hydraulic and mechanical functioning.

Data description: The paper describes functional traits in leaves, stems and roots, including NSC, carbon, nitrogen, specific leaf area, stem and root wood density and xylem traits. Data are provided for up to 90 angiosperm species from temperate, Mediterranean and tropical climates. These data are useful for understanding the trade-offs in resource allocation from a whole-plant perspective, and to better quantify xylem structure and function related to water transportation, mechanical support and storage. Data will also give researchers keys to understanding the ability of trees to adjust to a changing climate.

Keywords: Angiosperms, Non-structural carbohydrates, Fibers, Functional traits, Mediterranean, Parenchyma, Root, Temperate, Tropical, Vessels

Objective

Non-structural carbohydrates (NSC) are essential substrates for metabolic processes in trees, including respiration, osmoregulation, growth, reproduction and defense [1–4], as well as having major consequences for downstream processes such as microbial activity in the rhizosphere [5]. NSC is a product of photosynthesis and comprises mainly soluble sugars involved in transport or in diate functions, and starch stored in different plant organ future use and maintaining functionality when carbon is higher than supply (e.g., under severe drought str Therefore, understanding how patterns of NSC v will enable us to better evaluate the role of N siological processes and ecological strategies, s a broad range of

➢ **Licences de diffusion**
Licence sur l'article dans la revue imposée par l'éditeur
+
Licence sur les données dans l'entrepôt que vous choisissez sauf si imposée par la revue !

Licence CC-BY

Modalités de diffusion des données

- Revue impose une licence de diffusion des données
- Revue propose un choix de licences +/- ouvertes

Licence CC0 ou CC-BY

Licence CC0
*la + ouverte
aucune restriction
possible*

CC0
CC-BY
CC-BY-SA
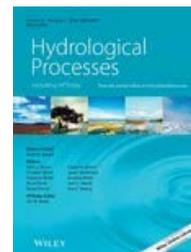
CC-BY
CC-BY-SA

Licence CC-BY-NC
*Pas d'utilisation
commerciale*

Modalités de diffusion des données

FLOTROP: *base de données botaniques d'Afrique tropicale*

Global Biodiversity Information facility

Licence CC-BY-NC

Licence CC-BY

Licence autorisée : CC-0 ou CC-BY
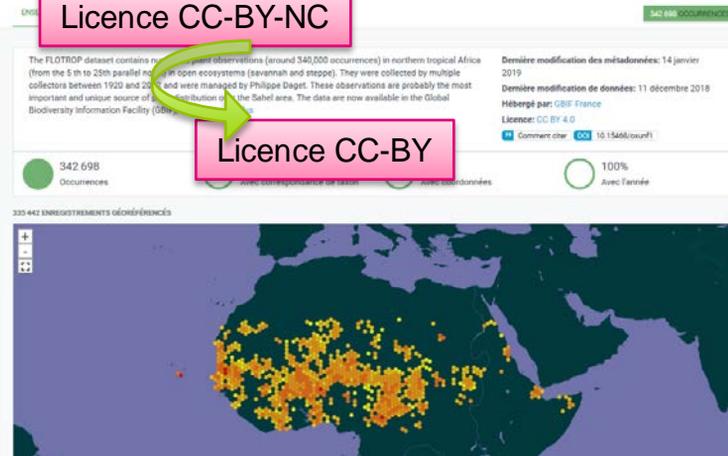
Critère 5- *Data paper* Article de recherche

- *Data paper* complémentaire du *Research paper*
  Mais vérifiez, selon la revue choisie pour l'article de recherche, que publier le *Data paper* ne sera pas bloquant

- Certaines revues l'acceptent

  Voir liste : https://open-research-europe.ec.europa.eu/data-policies

  ou la note des revues du groupe Nature ou des éditeurs BMC, PLoS, Elsevier, SAGE, PNAS, Science

- Pour d'autres, la publication du *Data paper* entraine une perte d'originalité pour l'article de recherche.

  Toutes les revues de l'éditeur Cell Press : https://www.cell.com/cell/authors#policies

Choisir une revue scientifique

➢ Consulter sa communauté scientifique

▪ Listes disponibles

- **Liste Cirad de revues publiant des *Data papers*** : https://doi.org/10.18167/coopist/0057
- Forschungsdaten : https://www.forschungsdaten.org/index.php/Data_Journals
- University of Edinburgh : https://www.wiki.ed.ac.uk/display/datashare/Sources+of+dataset+peer+review
- GBIF (biodiversité): http://www.gbif.fr/page/contrib/publier-un-datapaper

▪ **Où publier** (site internet)

- 2300 revues
- Critères de recherche :

  thèmes / sous-thèmes
  types d'articles
  libre accès
  notoriété

Exercice de recherche de revues

20 minutes

Exercice: choisir une revue scientifique

■ A partir du site Où publier :

https://ou-publier.cirad.fr/

■ A partir des listes disponibles

- **Liste Cirad de revues publiant des *Data papers*** : https://doi.org/10.18167/coopist/0057

- Forschungsdaten : https://www.forschungsdaten.org/index.php/Data_Journals

- University of Edinburgh : https://www.wiki.ed.ac.uk/display/datashare/Sources+of+dataset+peer+review

- GBIF (biodiversité): http://www.gbif.fr/page/contrib/publier-un-datapaper

Critères à retenir pour choisir une revue

➢ Domaine scientifique et lectorat visé

▪ Cohérence entre vos données et les exigences de la revue telles que:

- Portée des données: large échelle, fort intérêt pour la communauté,…
- Contraintes en termes de :
  - modèle du *Data paper*
  - métadonnées imposées ou non
  - choix d'entrepôt
  - modalités d'accès et de diffusion des données

▪ Libre accès au *Data paper*

▪ Coût de publication

Entrepôts de donné

Pourquoi déposer vos jeux de données dans un entrepôt ?

➢ A - pour les conserver dans un environnement réfrigéré

➢ B - pour les cacher avec d'autres données

➢ C - pour les faire connaitre et faciliter leur réutilisation

RÉPUBLIQUE
FRANÇAISE
Liberté
Égalité
Fraternité

cirad

Dépôserses données dans un entrepôt

➢ Optimise les possibilités de réutilisation

- + de visibilité : 2 sources indépendantes
- + d'opportunités que vos données soient trouvées
- + de chances d'être cité : 2 identifiants numériques

- Meilleure préservation des données
- Valorisation des données (Ex: SEANOE, GBIF)
- Suivi des citations de vos données

➢ Répond aux attentes des éditeurs et des évaluateurs

Concept du *Data paper* : article + entrepôt

DATA NOTE

Open Access

Non-structural carbohydrates and morphological traits of leaves, stems and roots from tree species in different climates

Guangqi Zhang[1*], Pascale Maillard[2], Zhun Mao[1], Loïc Brancheriau[3], Julien Engel[1], Bas, Claire Fortunel[1], Jean-Luc Maeght[1], Jordi Martínez-Vilalta[4,5], Merlin Ramel[1], Sophie Nou Stéphane Fourtier[1] and Alexia Stokes[1]

Abstract

Objectives: Carbon fixed during photosynthesis is exported from leaves towards sink organs as non-structural carbohydrates (NSC), that are a key energy source for metabolic processes in trees. In xylem, NSC are mostly stored as soluble sugars and starch in radial and axial parenchyma. The multi-functional nature of xylem means that cells possess several functions, including water transport, storage and mechanical support. Little is known about how NSC impacts xylem multi-functionality, nor how NSC vary among species and climates. We collected leaves, stem and root xylem from tree species growing in three climates and estimated NSC in each organ. We also measured xylem traits linked to hydraulic and mechanical functioning.

Data description: The paper describes functional traits in leaves, stems and roots, including NSC, carbon, nitrogen, specific leaf area, stem and root wood density and xylem traits. Data are provided for up to 90 angiosperm species from temperate, Mediterranean and tropical climates. These data are useful for understanding the trade-offs in resource allocation from a whole-plant perspective, and to better quantify xylem structure and function related to water transportation, mechanical support and storage. Data will also give researchers keys to understanding the ability of trees to adjust to a changing climate.

Keywords: Angiosperms, Non-structural carbohydrates, Fibers, Functional traits, Mediterranean, Parenchyma, Root, Temperate, Tropical, Vessels

Objective

Non-structural carbohydrates (NSC) are essential substrates for metabolic processes in trees, including respiration, osmoregulation, growth, reproduction and defense [1–4], as well as having major consequences for downstream processes such as microbial activity in the rhizosphere [5]. NSC is a product of photosynthesis and comprises mainly soluble sugars involved in transport or immediate functions, and starch stored in different plant organs for future use and maintaining functionality when carbon demand is higher than supply (e.g., under severe drought stress) [6–9]. Therefore, understanding how patterns of NSC vary in trees will enable us to better evaluate the role of NSC in tree physiological processes and ecological strategies, especially across a broad range of

Données déposées dans un entrepôt de données

Bien choisir son entrepôt

Connaissez-vous des entrepôts de données dans votre domaine ?

➢ Oui

➢ Non

➢ Ne sait pas

Utilisez-vous des données issues d'entrepôts tels que FAOStat, WorldClim, WorldPop, GBIF, WorldWide Antimalarial Resistance Network, ...

➤ Oui

➤ Non

➤ Ne sait pas

Les types d'entrepôts de données

- **Un entrepôt de données est une infrastructure permettant de préserver et de mettre à disposition des données.**

- **Il existe de nombreux types d'entrepôts:**
  - Certains sont portés par des organismes internationaux (ex: FAOStat, WorldClim,…)

    ils donnent accès à des données de référence
    mais ne permettent pas d'y déposer des données

  - D'autres accueillent les données de recherche.

    Vous pouvez déposer vos données
    et télécharger d'autres données.

- **Re3data : répertoire d'entrepôts** https://www.re3data.org/

Les types d'entrepôts de données

- **Entrepôts généralistes**

    Dryad, FigShare,

    Dataverse, Open Science Framework
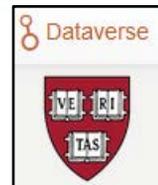
- **Institutionnels, Nationaux, Européens**

    Dataverse (Universités, Cirad, IRD, CEA, ….)

    Recherche Data Gouv (entrepôt national)

    Zenodo, Eudat, B2Share

- **Portés par des éditeurs de revues**

    Oxford Univ Press (GigaDB), Ubiquity Press (Dataverse), Elsevier (Mendeley Data)

- **Thématiques** ou **disciplinaires**: recommandés par les revues !

Les entrepôts thématiques et disciplinaires

- **Nombreux thèmes et disciplines**

Biodiversité : Global Biodiversity Information Facility

Phenotypage de plantes:

Sciences de la terre :

Microbio alimentaire:

OMICs :
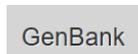
Pastoralisme – Mobilité animale :

Biologie

Enquêtes:

SHS:

Codes, Logiciels :

L'entrepôt Dataverse du Cirad

- **Entrepôt de données institutionnel**
  - Accepte tous types de données + documentation associée
  - Dépôt gratuit
  - Ouvert aux UR, aux projets et aux partenaires de projets
  - Appui : Céline Barthelemy et Gwenaël Doux (Dist)

- **Entrepôt reconnu par les éditeurs**
  - Vous pouvez y déposer vos données
  - Choisir une licence de diffusion, parfois un embargo
  - Un DOI sera attribué aux jeu de données pour insérer dans le manuscrit du *Data paper*
  - Ou une URL privée pour transmettre de manière sécurisée aux reviewers l'accès à vos données.

Critères de choix d'un entrepôt

- **Entrepôt reconnu dans la discipline**
  - Porté par une communauté scientifique
  - Correspondant aux publics scientifiques visés
  - Valorisant vos données

- **Visible dans les moteurs de recherche**

- **Délivrant un identifiant numérique pérenne et unique**

- **Adapté à la taille de vos fichiers**

- **Certifié** si possible

- **Gratuit** : la plupart sauf par ex : PANGAEA, DRYAD, TreeBase

Critères de choix d'un entrepôt

- En accord avec : le bailleur, l'institution, les partenaires, la revue où vous publiez

- Autorisant les modalités de diffusion que vous souhaitez
  - souvent l'entrepôt propose un choix de licences
  - Mais attention parfois la licence est imposée
    - licence CC0
    = domaine public, aucune restriction d'usage
    et sans obligation de citer les créateurs des données

- Voir la fiche Déposer des données de recherche dans un entrepôt

Pour choisir un entrepôt de données

➢ Consultez les pratiques de votre communauté scientifique / vos collègues

➢ Regardez les entrepôts recommandés par la revue ciblée

➢ Consultez les moteurs de recherche de données pour identifier les entrepôts utilisés pour des données similaires: https://commons.datacite.org/ ; Google Dataset search; ....

➢ Consultez les répertoires d'entrepôts

Répertoires d'entrepôt de données

https://www.re3data.org/search

3152 entrepôts

https://fairsharing.org/search?fairsharingRegistry=Database

2043 entrepôts

Informe aussi sur des entrepôts
portés par des organismes internationaux
(ex: FAOStat, WorldClim, WorldPop, ISRIC…)
donnant accès à des données de référence.
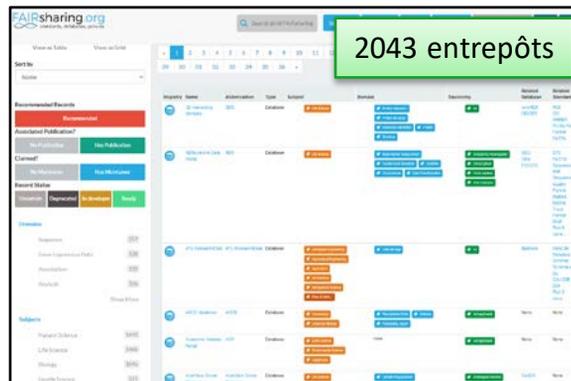
Entrepôts en biologie moléculaire

https://elixir-europe.org/platforms/data/elixir-deposition-databases

Database Commons
a catalog of worldwide biological databases

https://ngdc.cncb.ac.cn/databasecommons/

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

cirad

Répertoire re3data

re3data.org

Sea

re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

https://www.re3data.org/search

Filter

Subjects ⊞
Content Types ⊞
Countries ⊞
AID systems ⊞
API ⊞
Data access ⊞
Data access restrictions ⊞
Database access ⊞
Database access restrictions ⊞
Database licenses ⊞
Data licenses ⊞
Data upload ⊞
Data upload restrictions ⊞
Enhanced publication ⊞
Institution responsibility type ⊞
Institution type ⊞
Keywords ⊞
Metadata standards ⊞
PID systems ⊞
Provider types ⊞
Quality management ⊞
Repository languages ⊞
Software ⊞
Syndications ⊞
Repository types ⊞
Versioning ⊞

Mot-clé

Nombre de résultats

pathogen

Previous | 1 | Next →

Found 18 result(s)

Pathogen Portal
the Bioinformatics Resource Centers Portal

Subject(s)	Basic Biological and Medical Research Microbiology, Virology and Immunology Bioinformatics and Theoretical Biology
Content type(s)	Standard office documents Scientific and statistical data formats Raw data Structured text Software applications
Country	United States

Pathogen Portal is a repository linking to the Bioinformatics Resource Centers (BRCs) sponsored by the National Institute of Allergy and Infectious Dise
Virginia Bioinformatics Institute. The BRCs are providing web-based resources to scientific community conducting basic and applied research on organis
biowarfare or bioterrorism or causing emerging or re-emerging diseases. The Pathogen Portal supports and links to five Bioinformatics Resource Centers
different group of pathogens, focusing on, but not limited to, pathogens causing (Re-)Emerging Infectious Diseases, and those in the NIAID Category A-l
research. The scope of the BRCs also includes Invertebrate Vectors of Human Disease. Pathogen Portal covers EuPathDB, IRD, PATRIC, VectorBase a

Host - Pathogen Interaction Database
HPIDB

Subject(s)	Basic Biological and Medical Research Microbiology, Virology and Immunology Biology Life Sciences Medicine
Content type(s)	Standard office documents Raw data
Country	United States

HPIDB is a public resource, which integrates experimental PPIs from various databases into a single database. The Host-Pathogen Interaction Databas
to understanding molecular interactions between key organisms and the pathogens to which they are susceptible.

SilkPathDB
Silkworm Pathogen Database

Subject(s)	Biology Life Sciences Basic Biological and Medical Research Pathology and Forensic Medicine Microbiology, Virolog
Content type(s)	Standard office documents Scientific and statistical data formats Images Raw data Software applications other P
Country	China

Silkworm Pathogen Database (SilkPathDB) is a comprehensive resource for studying on pathogens of silkworm, including microsporidia, fungi, bacteria
not only genomic data including functional annotation of genes and gene products, but also extensive biological information for gene expression data an
will be help with researches on pathogens of silkworm as well as other Lepidoptera insects.

*Liberté
Égalité
Fraternité*

Répertoire re3data

SEANOE

| General | Institutions | Terms | Standards |

Name of repository	**SEANOE**
Additional name(s)	Sea scientific open data publication Publication de données scientifiques marines
Repository URL	https://www.seanoe.org/
Subject(s)	Geosciences (including Geography) · Life Sciences · Natural Sciences · Biology · Agriculture, Forestry, Horticulture and Veterinary Medicine · Chemistry · Physics
Description	"Seanoe (SEA scieNtific Open data Edition) is a publisher of scientific data in the field of marine sciences. It is operated by Ifremer (http://wwz.ifremer.fr/). Data published by SEANOE are available free. They can be used in accordance with the terms of the Creative Commons license selected by the author of data. Seance contributes to Open Access / Open Science movement for a free acces for everyone to all scientific data financed by public funds for the benefit of research. An embargo limited to 2 years on a set of data is possible; for example to restrict access to data of a publication under scientific review. Each data set published by SEANOE has a DOI which enables it to be cited in a publication in a reliable and sustainable way. The long-term preservation of data filed in SEANOE is ensured by Ifremer infrastructure. "
Contact	frederic.merceur@ifremer.fr data@seanoe.org
Content type(s)	Raw data · other · Standard office documents · Structured graphics
Keyword(s)	marine sciences
Persistent identifier(s) of the repository	ISSN 2491-1836 FAIRsharing_doi:10.25504/FAIRsharing.bb5rna
Repository type(s)	disciplinary
Mission statement for designated community	https://www.seanoe.org/html/about.htm

modalités et licences

embargo

Répertoire re3data

https://www.re3data.org/search

SEANOE

General | Institutions | **Terms** | Standards

modalités
et licences

Policies (1)

Policy Name — About SEANOE

Database access

Type of access to research
data repository — open

Modalités d'accès
aux données

Data access (3)

Type of access to data — open

Type of access to data — embargoed

Type of access to data — restricted

Data access restriction
type(s) — other

Data licenses (1)

DataLicense — CC

Modalités de dépôt
de données

Data upload (1)

Type of data upload — open

Data upload license (1)

Data upload license name — Publish your marine data

A retenir

➢ Déposer ses données dans 1 entrepôt de confiance est la bonne pratique pour publier un *Data paper*

- gain de visibilité et de valorisation
- optimise les possibilités de réutilisation
- conservation des données assurées

■ Nombreux entrepôts: institutionnels, généralistes, thématiques (le plus recommandé)

■ Critères de choix d'un entrepôt

- le + utilisé dans votre discipline
- recommandé par: éditeur, financeur, partenaires, institution…
- adapté à vos données
- proposant la licence souhaitée
- gratuit ou coût raisonnables

Exercice de recherche

d'entrepôts de données

20 minutes

Exercice: choisir un entrepôt de données

▪ A partir de la revue ciblée pour publier votre *Data paper*

▪ A partir d'un répertoire :

- re3data : https://www.re3data.org/search

- FairSharing: https://fairsharing.org/search?fairsharingRegistry=Database

- Pour les données de biologie moléculaire : ELIXIR
  https://elixir-europe.org/platforms/data/elixir-deposition-databases
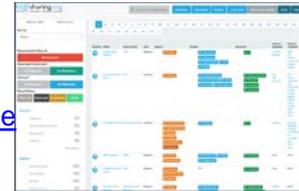
➤ Vérifiez :

- s'il est adapté à vos données: domaine, taille fichiers, format, …

- quelles sont les licences disponibles

- S'il y a un coût de dépôt des données.

Outils de rédaction d'un
Data paper

Outils de rédaction en ligne de *Data paper*

- Générateur de *Data paper* de Research Data Gouv : l'entrepôt national de données :
  https://entrepot.recherche.data.gouv.fr/datapartage-datapapers-web/
  - permet de créer une ébauche de *Data paper* à partir du DOI d'un jeu de données déposé dans l'entrepôt.
  - en cours d'évolution

- Arpha Writing Tool (AWT) de l'éditeur Pensoft
  - adapté aux revues en écologie et biodiversité

- Outil utilisé par Freshwater Metadata Journal:
  http://www.freshwaterjournal.eu/

- Authorea utilisé par Geoscience Data Journal:
  https://www.authorea.com/users/616189/articles/642279-geoscience-data-journal-template?mode=edit

Outil de rédaction AWT

- Arpha Writing Tool (AWT) de l'éditeur Pensoft
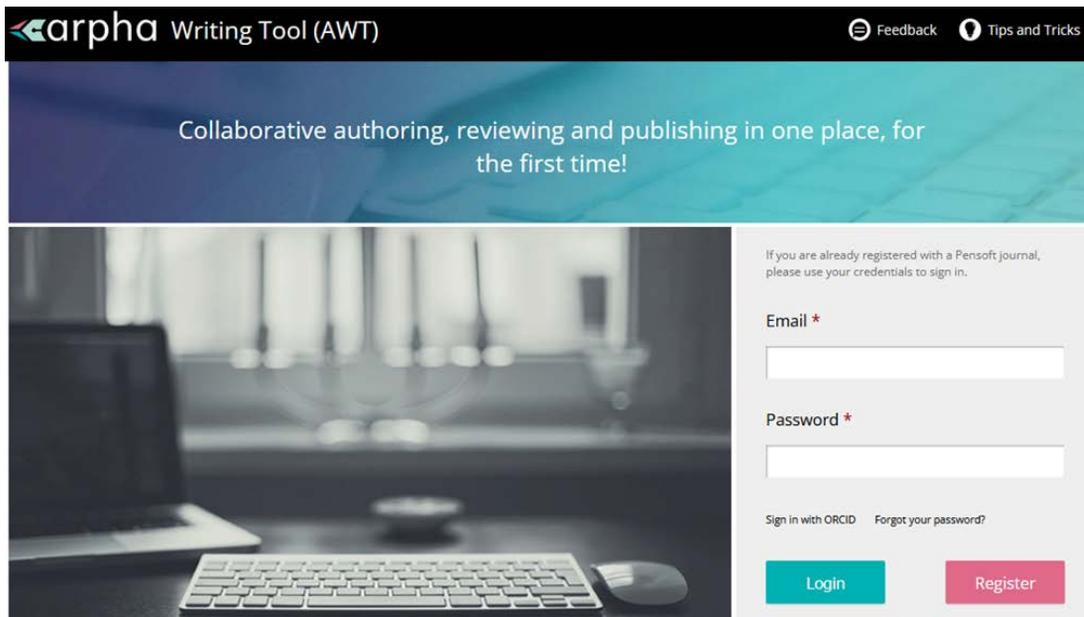
Outil de rédaction AWT

To start a manuscript select a journal and an article type

Choix de la revue

- Biodiversity Data Journal
- Research Ideas and Outcomes
- One Ecosystem
- BioDiscovery
- Biodiversity Information Science and Standards
- Food and Ecological Systems Modelling Journal
- ARPHA Conference Abstracts
- Viticulture Data Journal
- Biosystematics and Ecology

Choix du type d'article

Research ideas	Early research outcomes	Brief research outcomes
Data Management Plan	Applied Study	Commentary
Grant Proposal	Case Study	Conference Abstract
PhD Project Plan	Clinical Case Studies	Correspondence
PostDoc Project Plan	Data analytics	Ecosystem Accounting Table
Research Idea	Data Paper (Biosciences)	Ecosystem Inventory
Software Management Plan	Data Paper (Generic)	Ecosystem Service Mapping
	Emerging Technique	Ecosystem Service Models
	Formal Model Article Format	Institutional/Society announcement
	Forum Paper	Interdisciplinary Perspectives
	FSKX (Food Safety Knowledge)	Monitoring Schema
	Methods	Opinions
	Model validation	Research Poster
	OMICS Data Paper	Research Presentation
	Project Report	Short Communication
	Questionnaire	Single-media Publication
	R Package	
	Software Description	
	Workshop Report	

Research outcomes	PhD theses	Editorial matters
Alien Species Profile	PhD Thesis	Biography
Bibliography		Book Review
Catalogues/Checklist		Corrigendum
Guidelines		Data Review

RÉPUBLIQUE FRANÇAISE
*Liberté
Égalité
Fraternité*

cirad

Outil de rédaction AWT

Outil de rédaction AWT

Choix de la revue

- Biodiversity Data Journal
- Research Ideas and Outcomes
- One Ecosystem
- BioDiscovery
- Biodiversity Information Science and Standards
- Food and Ecological Systems Modelling Journal
- ARPHA Conference Abstracts
- Viticulture Data Journal
- Biosystematics and Ecology

Research ideas	Early research outcomes	Brief research outcomes
Data Management Plan	Applied Study	Commentary
Grant Proposal	Case Study	Conference Abstract
PhD Project Plan	Clinical Case Studies	Correspondence
PostDoc Project Plan	Data analytics	Ecosystem Accounting Table
Research Idea	Data Paper (Biosciences)	Ecosystem Inventory
Software Management Plan	Data Paper (Generic)	Ecosystem Service Mapping
	Emerging Technique	Ecosystem Service Models
	Formal Model Article Format	Institutional/Society announcement
	Forum Paper	Interdisciplinary Perspectives
	FSKX (Food Safety Knowledge)	Monitoring Schema
	Methods	Opinions
	Model validation	Research Poster
	OMICS Data Paper	Research Presentation
	Project Report	Short Communication
	Questionnaire	Single-media Publication
	R Package	
	Software Description	
	Workshop Report	

Research outcomes	PhD theses	Editorial matters
Alien Species Profile	PhD Thesis	Biography
Bibliography		Book Review
Catalogues/Checklist		Corrigendum
Guidelines		Data Review
Interactive Key		Editorial
Policy Brief		Obituary
Replication Study		Software Review
Research Article		
Review Article		
Species Conservation Profiles		
Standards		
Taxonomy & Inventories		
Wikipedia Article		

Import d'un fichier à partir d'1 entrepôt

Import a Data Paper Manuscript
From GBIF IPT Metadata (EML)
From DataONE Metadata (EML)
From LTER Metadata (EML)

Import Occurrence Records
From Excel Spreadsheet
Manually
From GBIF
From BOLD
From iDigBio
From Plutof

Reset selection

Create manuscript

OR

Import a manuscript

Outil de rédaction AWT / import fichier

Outil de rédaction AWT / import fichier

Data paper automatiquement généré

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

cirad

A retenir
sur les *Data papers*

Publier un *Data paper* est : une bonne pratique de recherche

- Répond aux enjeux de transparence, rigueur et intégrité scientifique de la recherche
- Témoigne de bonnes pratiques de gestion des données
- Répond aux recommandations faites à l'ANR d'inciter à la publication de *Data papers*
- S'inscrit dans le mouvement mondial de science ouverte
- Garantit la réplication et validation des recherches
- Contribue aux avancées scientifiques et sociétales.

Publier un Data paper est : à la portée de tous

- Rédiger un *Data paper* peut être simple à condition de:
  - bien traiter ses données et d'être rigoureux dans ses méthodes
  - les documenter au fur et à mesure de la collecte ou de la production
  - appliquer les pratiques de sa discipline
  - choisir une revue et un entrepôt adaptés et suivre les recommandations
  - faire une biblio régulière pour bien évaluer l'originalité et l'intérêt de ses données.

Publier un *Data paper* est : bénéfique pour vous

- Si votre *Data paper* est complet et vos données FAIR
- Votre manuscrit sera accepté → 1 article publié, évalué par les pairs, citable, complémentaire d'un article de recherche
- Valorisation de votre travail et de vos données
- Pour vous : visibilité, reconnaissance, citations,

notoriété

paternité

attractivité

collaborations

Satisfaction
de contribuer
aux avancées scientifiques
ou sociétales

Crédit: Nos pensées. La valeur de la reconnaissance.
https://nospensees.fr/la-valeur-de-la-reconnaissance/

Laurence Dedieu

laurence.dedieu@cirad.fr

Liens utiles

➢ Divers documents

- Guide de Bonnes pratiques. CNRS. 2020. https://mi-gt-donnees.pages.math.unistra.fr/guide/07-publier.html

- *Data Paper* : émergence d'une nouvelle donne scientifique. Revue française des sciences de l'information et de la communication. 2022. https://doi.org/10.4000/rfsic.12219

- Guide pratique à l'usage des doctorants. 2021. https://www.ouvrirlascience.fr/passeport-pour-la-science-ouverte-guide-pratique-a-lusage-des-doctorants/

- Les cours issus du projet FOSTER sur la science ouverte : https://callisto-formation.fr/pluginfile.php/13539/mod_resource/content/1/01_FOSTER_Science%20ouverte_bonnes_pratiques_recherche_FRADAPT_Inist-CNRS.pdf

- Entrepôts de données de recherche : mesurer l'impact de l'Open Science à l'aune de la consultation des jeux de données déposés. V Rebouillat. 2020. ⟨hal-02928817⟩

- Vidéos de EOSC. 2022. https://www.eosc-pillar.eu/node/397?utm_campaign=Data%20veille&utm_medium=email&utm_source=Revue%20newsletter

Liens utiles

➢ Produits par le CIRAD

- Publier un *Data paper* (2022) - avec la liste des revues publiant des *Data papers*. https://doi.org/10.18167/coopist/0057
- Wébinaire: Data Papers : quand ? Comment ? Pourquoi ? Laurence Dedieu. 05/07/2022. https://www.youtube.com/watch?v=SCW3xNxK4PU
- Déposer des données de recherche dans un entrepôt (2020). https://doi.org/10.18167/coopist/0070

➢ Produits par le GBIF (plateforme mondiale de biodiversité)

- *Data papers* et biodiversité : https://www.gbif.org/data-papers
- Wébinaire sur les *Data papers* (2018): https://www.gbif.org/fr/event/7j0sFhaug80CIMqs6aWegK/webinar-introduction-to-data-papers

Liens utiles

➤ Ressources pédagogiques

- Webinaire « Data paper - Une incitation à la qualification et à la réutilisation des jeux de données ». DoRANum. 2020.
Vidéos: https://doranum.fr/2020/11/20/supports-du-webinaire-data-paper-une-incitation-a-la-qualification-et-a-la-reutilisation-des-jeux-de-donnees/
Synthèse: https://mi-gt-donnees.pages.math.unistra.fr/site/download/Christine%20Hadrossek%20-%20Synth%C3%A8se%20webinaire%20du%205%20novembre%20Data%20Paper.pdf

- *Ressources sur les Data papers.* DoRANum: https://doranum.fr/data-paper-data-journal/minute-publication-data-papers/ ; https://doranum.fr/data-paper-data-journal/contenu-data-paper/

- Dépôt et Entrepôts » : https://doranum.fr/depot-entrepots/fiche-synthetique/

- Choisir un entrepôt. Infographie interactive : https://doranum.fr/depot-entrepots/choix-entrepot-depot-donnees/

- Les critères d'évaluation des data papers: https://doranum.fr/data-paper-data-journal/criteres-evaluation-data-papers/

- Ressources du groupe de travail « Données » du groupe de travail Science Ouverte (GTSO). Couperin. https://www.couperin.org/science-ouverte/ressources-du-gtso/groupe-donnees

- Vidéo-Déposer ses données: https://doranum.fr/depot-entrepots/deposer-ses-donnees/

Liens utiles

➢ Quiz sur les données

- Data paper et Data journals : https://doranum.fr/data-paper-data-journal/

- Format ouvert ou fermé : https://doranum.fr/stockage-archivage/quiz-format-ouvert-ou-ferme/ propose une liste interactive des formats

- Quiz : jeu sur les entrepôts de données : https://doranum.fr/depot-entrepots/deposez-un-jeu-sur-les-entrepots-de-donnees_10_13143_aabd-hs57/

- Licences Creative Commons : https://uniofcam.libwizard.com/f/CreativeCommons

- Étude de cas de gestion des données en Économie : https://doranum.fr/economie-gestion/etude-de-cas-gestion-donnees-economie/ar-introduction-to-data-papers