RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

cirad
LA RECHERCHE AGRONOMIQUE POUR LE DÉVELOPPEMENT

Rédiger et publier un *Data paper*

Laurence Dedieu

Délégation à l'Information Scientifique et Technique

Laurence.dedieu@cirad.fr

Support mis à disposition selon les termes de la licence Creative Commons Attribution - Pas d'Utilisation Commerciale 4.0 International.

Pourquoi publier 1 Data paper ?

- Vous avez produit des données de recherche
- Vos données ont de la valeur

  → elles ont un intérêt pour d'autres

  → elles peuvent être réutilisées

- Publier un *Data paper* informe la communauté scientifique de l'existence de vos données
- Rédiger un *Data paper* peut être simple
- Large choix de revues scientifiques et d'entrepôts pour déposer les données
- Avoir fait un PGD vous aide à publier un *Data paper*.

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

cirad

CC BY NC

Plan de la formation

- Concept et objectifs du *Data paper*
- Rédiger un *Data paper*

  → Consultation de *Data papers* publiés

- Revues scientifiques et critères de choix

  → Recherche de revues

- Entrepôt de données et critères de choix

  → Recherche d'entrepôts

- A retenir

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

cirad

Concept et objectifs du *Data paper*

Ce support Rédiger un *Data paper* de L. Dedieu est mis à disposition selon les termes de la licence Creative Commons Attribution - Pas d'Utilisation Commerciale 4.0 International.

Concept du *Data paper*

Data in Brief 15 (2017) 92–96

Contents lists available at ScienceDirect

Data in Brief

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

Data Article

Integrated dataset of anatomical, morphological, and architectural traits for plant species in Madagascar

Amira Azizan [a], Emma Guillon [a], Yves Caraglio [c], Patrick Langbour [b], Sébastien Paradis [b], Pierre Bonnet [c], Yannick Brohard [c], Christine Heinz [c], Nabila Boutahar [b], Loïc Brancheriau [b,*]

[a] University of Montpellier, Dept. of Biodiversity, Ecology and Evolution, Montpellier, France
[b] CIRAD, UR BioWooEB, 34398, Montpellier, France
[c] CIRAD, UMR AMAP, 34398, Montpellier, France

ARTICLE INFO

Article history:
Received 7 July 2017
Accepted 5 September 2017
Available online 12 September 2017

Keywords:
Morpho-architectural traits
Plant architecture
Wood anatomy
Madagascar

ABSTRACT

In this work, we present a dataset, which provides information on the structural diversity of some endemic tropical species in Madagascar. The data were from CIRAD xylotheque (since 1937), and were also collected during various fieldworks (since 1964). The field notes and photographs were provided by French botanists; particularly by Francis Hallé. The dataset covers 250 plant species with anatomical, morphological, and architectural traits indexed from digitized wood slides and fieldwork documents. The digitized wood slides were constituted by the transverse, tangential, and radial sections with three optical magnifications. The main specific anatomical traits can be found within the digitized area. Information on morphological and architectural traits were indexed from digitized field drawings including notes and photographs. The data are hosted in the website ArchiWood (http://archiwood.cirad.fr).

© 2017 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Jeu de données

c	# of objs	# of objs +3 frames	min frames	ave frames	max frames
1	73	47	1	7.5	40
2	18	6	1	12.2	53
3	18	10	1	77.3	725
4	47	28	1	31.1	143
5	18	15	1	18.7	60
6	4	3	2	28.23	80
7	7	6	4	6.7	13
8	33	31	2	30.6	118
9	18	8	1	3.7	23
10	3	2	1	25.0	70
11	5	4	2	50.0	104
12	15	12	1	22.3	137
13	12	9	1	21.0	69
14	3	3	10	114.0	171
15	6	5	2	59.5	156
16	10	10	4	61.5	172

Entrepôt de données

- Contexte
- Méthodes
- Potentiel

+ lien vers les données

Réutilisation des données
Reproductibilité des expériences

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

cirad

CC BY NC

Concept du *Data paper*

Data in Brief 15 (2017) 92–96

Contents lists available at ScienceDirect

Data in Brief

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

Data Article

Integrated dataset of anatomical, morphological, and architectural traits for plant species in Madagascar

Amira Azizan [a], Emma Guillon [a], Yves Caraglio [c], Patrick Langbour [b], Sébastien Paradis [b], Pierre Bonnet [c], Yannick Brohard [c], Christine Heinz [c], Nabila Boutahar [b], Loïc Brancheriau [b,*]

[a] University of Montpellier, Dept. of Biodiversity, Ecology and Evolution, Montpellier, France
[b] CIRAD, UR BioWooEB, 34398, Montpellier, France
[c] CIRAD, UMR AMAP, 34398, Montpellier, France

ARTICLE INFO

Article history:
Received 7 July 2017
Accepted 5 September 2017
Available online 12 September 2017

Keywords:
Morpho-architectural traits
Plant architecture
Wood anatomy
Madagascar

ABSTRACT

In this work, we present a dataset, which provides information on the structural diversity of some endemic tropical species in Madagascar. The data were from CIRAD xylotheque (since 1937), and were also collected during various fieldworks (since 1964). The field notes and photographs were provided by French botanists; particularly by Francis Hallé. The dataset covers 250 plant species with anatomical, morphological, and architectural traits indexed from digitized wood slides and fieldwork documents. The digitized wood slides were constituted by the transverse, tangential, and radial sections with three optical magnifications. The main specific anatomical traits can be found within the digitized area. Information on morphological and architectural traits were indexed from digitized field drawings including notes and photographs. The data are hosted in the website ArchiWood (http://archiwood.cirad.fr).

© 2017 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Jeu de données

c	# of objs	# of objs +3 frames	min frames	ave frames	max frames
1	73	47	1	7.5	40
2	18	6	1	12.2	53
3	18	10	1	77.3	725
4	47	28	1	31.1	143
5	18	15	1	18.7	60
6	4	3	2	28.23	80
7	7	6	4	6.7	13
8	33	31	2	30.6	118
9	18	8	1	3.7	23
10	3	2	1	25.0	70
11	5	4	2	50.0	104
12	15	12	1	22.3	137
13	12	9	1	21.0	69
14	3	3	10	114.0	171
15	6	5	2	59.5	156
16	10	10	4	61.5	172

Entrepôt de données

- Contexte
- Méthodes
- Potentiel

+ lien vers les données

Si documentation complète

Documentation associée:

- protocoles, questionnaire d'enquête,
- fichier « Read me »
- métadonnées disciplinaires
- ….

Données compréhensibles

Données réutilisables

RÉPUBLIQUE FRANÇAISE
Liberté Égalité Fraternité

cirad

CC BY NC

Concept du *Data paper*

1. Décrire un jeu de données et son contexte
   - suffisamment pour assurer sa compréhension
2. Décrire les méthodes d'obtention
   - suffisamment pour permettre de reproduire l'étude : protocole, méthode d'échantillonnage, procédures de contrôle qualité…
   - en montrant la rigueur scientifique et la qualité des données
3. Montrer le potentiel de réutilisation des données
   - suffisamment pour convaincre l'éditeur de l'originalité des données et de leur portée scientifique, sociétale, environnementale, historique, ….

Pas de résultats, ni analyses, ni interprétation

→ Donner accès au jeu de données

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité
cirad
CC BY NC

Entrepôt et documentation

- Entrepôt: fichiers de données + documentation associée

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

https://doi.org/10.18167/DVN1/PKCW2S

1 to 4 of 4 Files — Download

List_variables.tab
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
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
Plain Text - 550 B - 24 mars 2021 - 0 Downloads
MD5: 2bc7f46d64efe883907547d8adb54e23

Liste des variables

Données

Read me

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

cirad

CC BY NC

Data paper + Entrepôt = Bonne pratique

- 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

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

cirad

CC BY NC

Exemple de *Data paper*

- Dataset on farmers' perception of commodity futures market. Data in Brief: https://www.sciencedirect.com/science/article/pii/S2352340922006266

ELSEVIER

Data in Brief

Volume 43, August 2022, 108429

Data Article

Dataset on farmers' perception of commodity futures market

S. Srinivasan [a], M. Babu [b], P.S. Shabi Sh

Show more

+ Add to Mendeley Share Cit

Data accessibility	**Repository name:** Mendeley Data**Data identification number:** 10.17632/pw339snbvs.2**Direct URL to Data:** https://data.mendeley.com/datasets/pw339snbvs/2

https://doi.org/10.1016/j.dib.2022.108429

Get rights and content

Under a Creative Commons license

Open access

Abstract

The commodity futures market plays a major role in reducing the price risk for the participants. Unfortunately, the farmers' participation in the futures market particularly from the Tamil Nadu region is very less. A survey was conducted using the interview method to identify the information sources used by farmers for taking pricing decisions, the awareness and perception of farmers towards the futures market, and its effect on preferred marketing alternatives. The data cleaning process was done using content validity, confirmatory factor analysis, and reliability test using Cronbach's alpha, and the assumptions of normality and multicollinearity were examined. The data will be of potential use to researchers who wish to explore farmers' behavior towards hedging in the commodity futures market.

CC BY NC

Exemple de *Data paper*

1. Data Description

The interview schedule includes demographic data which were recorded on a nominal scale, training and farming details noted on an ordinal scale, the cost details were collected
recorded on a three-p
preferred. The farmer
Likert scale which inc
Similarly, the percept
Likert scale responses
The data collection in
information sources f
and "5" Always. The r
confirmatory factor a
towards futures mark
decisions are present
normality of the data
and the examination
Tolerance and Varian

odity futures market. Data in

2. Experimental Design, Materials and Methods

Initially, it was decided to use the questionnaire method for collecting the data. During the pilot study, the farmers were contacted individually at the regulated markets where auctioning of turmeric, cardamom, and chilly are conducted. The informal discussions, with farmers', revealed that the majority were not able to read and understand the questions. Some of the choices, with respect to information sources used by farmers, did not find a place in the questionnaire. All the above issues were addressed after the first pilot study. During the second pilot study, with 125 farmers', additional marketing alternatives namely selling to the traders, co-operative societies, etc., were added. It was noticed during the interaction, that the farmers were not aware of diffe
The only price risk instrument
majority of the farmers were no
outcome of the pilot study reve
for data collection as the respon
Additional explanations in the
farmers to get reliable answers.
to use the Structured Interview
commodity futures helps the fa
future periods. Hence, the data
awareness and perception of fa
auction at the regulated market
include other marketing altern
societies, farmers' associations,
the data on farmers' choice of s

Value of the Data

- The data on Tamil Nadu farmers' participation in futures markets is unknown through any of the sources.
- It will answer the following questions: How many farmers are using the futures market? What other marketing tools are used by farmers to sell their produce?
- This data will provide useful insights into farmers' understanding of various aspects of commodity futures trading.
- This data will add value to the policymakers in developing strategic measures for protecting the farmers from price risk and exploitation by the traders.
- The researchers can make use of the data to develop models on farmers' hedging behavior in the futures market.

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

cir

Exemple de *Data paper*

Mendeley Data

Create account | Sign in

Survey Data on Farmers Awareness, Perception level towards futures market

Published: 20 June 2022 | Version 2 | DOI: 10.17632/pw339snbvs.2
Contributor: S Srinivasan Sri Ramachandra Institute of Higher Education and Research

Description

Price risk is one of the major source of risk faced by all farmers. This data set captures the information sources used by the farmers for taking pricing decisions. A survey was carried out with 314 Chillies farmers at Ramanathapuram District of TamilNadu, 383 Turmeric Cultivators in the Erode District of TamilNadu and 221 Cardamom growers in Theni District of TamilNadu. The awareness level and perception towards commodity futures market and farmers preference towards usage of marketing alternatives in selling their produce were recorded in the dataset. The dataset is in MS-Excel format.

Download All 616 KB

Files

Questionnaire upload Mendeley 14.04.2022.pdf	553 KB
Survey Data on Farmers Awareness towards futures market.xls	740 KB

Steps to reproduce

The data were collected from the farmers through Interview Method. Top three villages by area of cultivation were identified in each of the district and the farmers were contacted personally by the researcher to record the details of farming, mode of selling the produce, interest, awareness, perception towards futures and other preferred marketing alternatives. Future researchers can add other risk management techniques for farmers for data collection. The data set will be helpful for any researcher interested in studying farmers welfare and hedging intention of farmers in commodity futures and options market.

Categories

Agricultural Economics, Financial Market, Financial Econometrics, Financial Instrument

License

CC BY 4.0 — Learn more

Dataset metrics

Usage

Views:	52
Downloads:	21

PLUMX — View details

Latest version

Version 2
Published: 20 Jun 2022
DOI: 10.17632/pw339snbvs.2

Cite this dataset

Sri Ramachandra Institute of Higher Education and Research, S Srinivasan (2022), "Survey Data on Farmers Awareness, Perception level towards futures market", Mendeley Data, V2, doi: 10.17632/pw339snbvs.2

Copy to clipboard

Previous versions

Version 1 — 11 February 2022

Compare to version

Version 1

CC BY NC

Exemple de *Data paper*

Mendeley Data

Create account | Sign in

Survey Data on Farmers Awareness, Perception level towards futures market

ELSEVIER

Data in Brief

Volume 43, August 2022, 108429

Data Article

Dataset on farmers' perception of commodity futures market

S. Srinivasan [a], M. Babu [b], P.S. Shabi Shimny [c], C. Hariharan [d], J. Gayathri [e], G. Indhumathi [f]

Show more

+ Add to Mendeley Share Cite

https://doi.org/10.1016/j.dib.2022.108429

Under a Creative Commons license

Get rights and content

Open access

Abstract

The commodity futures market plays a major role in reducing the price risk for the participants. Unfortunately, the farmers' participation in the futures market particularly from the Tamil Nadu region is very less. A survey was conducted using the interview method to identify the information sources used by farmers for taking pricing decisions, the awareness and perception of farmers towards the futures market, and its effect on preferred marketing alternatives. The data cleaning process was done using content validity, confirmatory factor analysis, and reliability test using Cronbach's alpha, and the assumptions of normality and multicollinearity were examined. The data will be of potential use to researchers who wish to explore farmers' behavior towards hedging in the commodity futures market.

Polychaete dataset of the southwestern G...

Data in Brief, Volume 43, 2022, Article 108370

Download PDF View d

Datasets of productivity and vegetation co

Data in Brief, Volume 43, 2022, Article 108387

Download PDF Vie

Chemical and biochemical qua orga...

Data in Brief, Volume 4 2022, Arti 8

Download PDF View details

Article Metrics

Captures

Readers: 3

PLUMX View details

Dataset metrics

Usage

Views: 52

Downloads: 21

PLUMX View details

Latest version

Version 2

Published: 20 Jun 2022

DOI: 10.17632/pw339snbvs.2

Cite this dataset

Sri Ramachandra Institute of Higher Education and Research, S Srinivasan (2022), "Survey Data on Farmers Awareness, Perception level towards futures market", Mendeley Data, V2, doi: 10.17632/pw339snbvs.2

Copy to clipboard

Previous versions

Version 1 11 February 2022

Compare to version

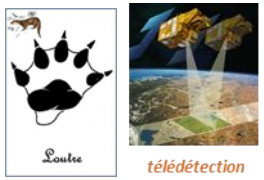
Version 1

CC BY NC

Quelles données concernées ?

- Tous types de jeux de données

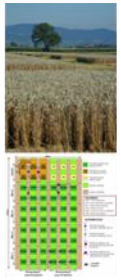
  Séquences, mesures, localisations, photos, enquêtes, presse,…



télédétection

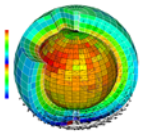
Données d'observation

Séquence de gènes



Résultats agronomiques

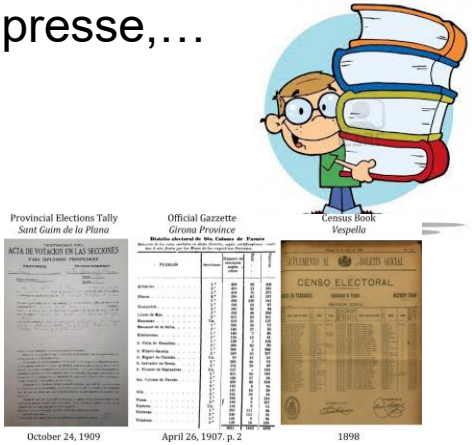
Données expérimentales



Modèle climatique

Simulations Modèles

Enquêtes



Données biblio

Définition des données de recherche:

observations, mesures, faits, images, codes, simulations informatiques,….

sur lesquels s'appuient des résultats de recherche.

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

cirad

CC BY NC

Bénéfices pour vous

- Mise en valeur de vos données et de votre travail
  - article évalué par les pairs, citable, indexé par les moteurs de recherche, comptant comme un article de recherche
  - visibilité des données dans l'entrepôt + citations si réutilisation
- Nombreuses revues + entrepôts (Dataverse institutionnel)
- Contrainte: préparation des données

  → Documenter régulièrement protocoles, données, métadonnées, au fur et à mesure des recherches et selon pratiques disciplinaires

  → intérêt du PGD !!!!
- Tous les contributeurs peuvent être auteurs
- Permet à d'autres de découvrir votre travail et les données que vous produisez→ vous rend visible dans votre communauté
- Montre la transparence de vos recherches.

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

cirad

CC BY NC

Rédiger un *Data paper*

Ce support Rédiger un *Data paper* de L. Dedieu est mis à disposition selon les termes de la licence Creative Commons Attribution - Pas d'Utilisation Commerciale 4.0 International.

Etapes à suivre

a) Choisir la revue et consulter le modèle de *Data paper*

b) Choisir l'entrepôt de données et voir ses exigences

c) Rédiger le *Data paper* selon le modèle de la revue

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

e) Déposer les données dans l'entrepôt
avec documentation

f) Soumettre le *Data paper* à la revue
avec le lien vers l'entrepôt où est déposé le jeu de données

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité
cirad
CC BY NC

Rédiger un *Data paper*

- Modèle classique du *Data paper*
  - Décrire les données
  - Décrire les méthodes d'obtention
  - Décrire le potentiel des données

→ **Préparer et déposer les données dans l'entrepôt**
Donner le lien d'accès aux données

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

cirad

cc BY NC

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, …

- Où a eu lieu l'étude ?

  Géolocalisation : Geoname, ISO 19115

  Contexte: habitat, climat, type de population, …

- Quand a eu lieu l'étude ?

  Date, période : ISO 8601 = AAAA-MM-JJ

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

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

cirad

CC BY NC

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*, …), intrants, traitements,…

  variables, paramètres, 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

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

cirad

CC BY NC

3. Décrire le potentiel des données

- Originalité de vos données
  - Données rares ou uniques, données jamais publiées

    expérimentations impossibles à répéter ou très couteuses

    groupes difficilement accessibles

    phénomènes rares
- Valeur et portée scientifique de vos données
  - Données sur plusieurs années, plusieurs espèces, filière, projet
  - généralisables à plusieurs espèces, données de référence
  - contribuant à un thème de recherche majeur:

    changement climatique, gestion de ressources naturelles, santé publique,…

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité
cirad

CC BY NC

3. Décrire le potentiel des données

- Valeur économique, sociétale, environnementale, culturelle, …
  - Développement de tests, vaccins, outils, modèles, services,…
  - impact sur la société ou sur les politiques publiques
  - avantages culturels ou écologiques

- Evaluer l'originalité des données
  - Datacite Search, Google Data Search, BASE…
  - S'il existe déjà des données similaires, précisez l'apport de vos données dans votre domaine et dans d'autres domaines

→ Description suffisante pour convaincre l'éditeur de l'intérêt et de la qualité de vos données

cc BY NC

4. Préparer et déposer les données dans l'entrepôt

- Selon les recommandations de l'entrepôt et de la revue
  - Format de fichier ouvert
  - Avec documentation associée
    - ✓ protocoles, questionnaire d'enquête,
    - ✓ dictionnaire des variables, …
    - ✓ explication des noms des fichiers, abréviations, colonnes, lignes, etc.
    - ✓ fichier « Read me »

cc BY NC

Importance de la documentation

- Exemple de tableau peu interprétable

	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	
4				15	28
5				30	312
6				60	99.9
7				120	217.9
8	322	2/9/15	18.9	0	185.8

Manque qqs explications :

colonne A ? GTT ?

unité de temps ?

Jours, heures ou minutes?

format Date ?

JJ/MM/AA ou MM/JJ/AA ?

Cases vides ??

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

cirad

CC BY NC

4. Préparer et déposer les données dans l'entrepôt

- Selon les recommandations de l'entrepôt et de la revue
  - Format de fichier ouvert
  - Avec documentation associée
  - Enrichir la description des données en ajoutant des métadonnées scientifiques proposées par l'entrepôt

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

- Dans le manuscrit, préciser :
  - Le nom de l'entrepôt
  - L'identifiant numérique du jeu de données
  - La licence de diffusion.

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

cirad

CC BY NC

5. Soumission et Evaluation

- Tous les *Data papers* sont évalués
- Selon la revue : évaluation du texte et parfois plus : métadonnées, fichiers, formats, entrepôt, organisation des données, +/- données elles-mêmes

  → accès aux données dès la soumission pour les reviewers !

- Critères
  - Valeur scientifique et originalité des données
  - Potentiel de réutilisation des données
  - Rigueur et adéquation de la méthode
  - Description suffisante pour permettre à d'autres de comprendre, interpréter et réutiliser les données
  - Qualité et fiabilité des données

cc BY NC

Mise en pratique: Consultation de *Data papers* publiés

Ce support Rédiger un *Data paper* de L. Dedieu est mis à disposition selon les termes de la licence Creative Commons Attribution - Pas d'Utilisation Commerciale 4.0 International.

Exercice

- Choisir un *Data paper* à partir d'une revue généraliste

Data in Brief : https://www.sciencedirect.com/journal/data-in-brief/issues

Data in Brief | *Open access*

Submit your article

Articles & Issues | About | Publish | Search in this journal | Guide for authors

Actions for selected articles

Select all / Deselect all

Download PDFs

Export citations

Show all article previews

Contents

- Agricultural Sciences
- Arts and Humanities
- Biological Sciences
- Omics
- Business, Management and Decision Sciences
- Computer Science
- Data Science
- Earth and Planetary Sciences
- Economics, Econometrics and Finance
- Energy
- Engineering
- Environmental Science

Agricultural Sciences

Data article • *Open access*

Data on growth performance, glucose concentration and testosterone level of Asian seabass, *Lates calcarifer* juveniles fed with exogenous melatonin at different concentration

Nur Farihah Rani, Muhamad Yazed Abduh, Nor Hakim Norazmi-Lokman, Siti Ariza Aripin

Article 108495

Download PDF | Article preview

Data article • *Open access*

Dataset on effect of decolourisation on metabolomic profile of *Moringa oleifera* leaf powder

Adewumi Toyin Oyeyinka, Oluwafemi Ayodeji Adebo, Muthulisi Siwela, Kirthee Pillay

Article 108508

Download PDF | Article preview

Data article • *Open access*

Draft genome sequence data of *Vibrio harveyi* VH1 isolated from a diseased tiger grouper, *Epinephelus fuscoguttatus*, cultured in Malaysia

Md. Ali Amatul-Samahah, Aslah Mohamad, Nurhidayu Al-saari, Mohd Zamri-Saad, ... Tomoo Sawabe

Article 108533

Download PDF | Article preview

Data article • *Open access*

Survey data of a traditional communal water irrigation system in Northern Thailand

Arriya Mungsunti, Kevin A. Parton, Arief Anshory Yusuf, Tossapond Kewprasopsak

Article 108515

Download PDF | Article preview

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

CC BY NC

Exercice

- Choisir un *Data paper* à partir de revues généralistes

  Data in Brief : https://www.sciencedirect.com/journal/data-in-brief/issues

- À partir des listes thématiques dans diapos suivantes
  - *Agronomie, Forêt*
  - *Ecologie, Biodiversité, Faune sauvage*
  - *Climat, Sol, Eau,*
  - *Génomique, Santé,*
  - *Economie, SHS*
  - *Data papers en français*
  - *Disponibles en format papier pour l'atelier*

cc BY NC

Exemples de *Data papers*

Ce support Rédiger un *Data paper* de L. Dedieu est mis à disposition selon les termes de la licence Creative Commons Attribution - Pas d'Utilisation Commerciale 4.0 International.

Exemples de *Data papers*

- Agronomie
  - 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

cc BY NC

Exemples de *Data papers*

- 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

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

cirad

CC BY NC

Exemples de *Data papers*

- Ecologie

  - 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

cc BY NC

Exemples de *Data papers*

- 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

Exemples de *Data papers*

- 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

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

cirad

CC BY NC

Exemples de *Data papers*

- 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

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

cirad

CC BY NC

Exemples de *Data papers*

- 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

cc BY NC

Exemples de *Data papers*

- 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

cc BY NC

Exemples de *Data papers*

- 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

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité
cirad
CC BY NC

Exemples de *Data papers*

- 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

cc BY NC

Exemples de *Data papers*

- Economie

  - 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

CC BY NC

Exemples de *Data papers*

- Sciences humaines et sociales
  - 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
  - Dependency Treebanks of Ancient Greek Prose. Journal of Open Humanities Data: http://doi.org/10.5334/johd.13.

cc BY NC

Exemples de *Data papers*

- En français
  - 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. Jean-Sauveur Ay et Mohamed Hilal. 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

cc BY NC

Exemples de *Data papers*

- Disponibles en format papier pour l'atelier

  - Aprisco Field Station: the spatial structure of a new experimental site focused on agroecology. *Journal of Plant Ecology*. 2022. https://doi.org/10.1093/jpe/rtac061
  - Long-term groundwater resource observatory for Southwestern Madagascar. *Hydrological Processes*. 2022. https://doi.org/10.1002/hyp.14108
  - ~14 000 years of geochemical and isotopic data from Lake Simcoe, Canada. *Data in Brief*. 2022. https://www.sciencedirect.com/science/article/pii/S235234092200748X
  - Diversity of Woody Species in Djamde Wildlife Reserve, Northern Togo, West Africa. Data Science Journal. 2019. http://doi.org/10.5334/dsj-2019-018
  - Dataset on farmers' perception of commodity futures market. *Data in Brief*. 2022. https://doi.org/10.1016/j.dib.2022.108429

cc BY NC

Choisir la revue scientifique

Ce support Rédiger un *Data paper* de L. Dedieu est mis à disposition selon les termes de la licence Creative Commons Attribution - Pas d'Utilisation Commerciale 4.0 International.

Revues publiant des Data Papers

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

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

cirad

Où Publier ? https://ou-publier.cirad.fr/

Une sélection de revues et d'éditeurs d'ouvrages en sciences appliquées à l'agriculture

CC BY NC

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*

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

cirad

CC BY NC

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
  - Localisation des données et entrepôts recommandés
  - Modalités d'accès et de diffusion des données (licences)
- Libre accès à l'article
- Coût de publication
  - Varie de gratuit à + de 3 000 €

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

cirad

CC BY NC

Modèles de *Data paper*

- Classique

  Data in Brief, Geoscience Data Journal, Ecological research

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

cirad

cc BY NC

Modèle de la revue *Data in Brief*

http://www.journals.elsevier.com/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

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

cirad

CC BY NC

Modèle de *Ecological Research*

Title, Authors, Abstract, Keywords

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

- **Background** and **ecological significance of the data**.
- **Methods:** detailed information about how data were collected and how the dataset was created: design of experiment(s), measurement procedures, names of chemicals & instruments, quality control...
- **Data structure:** physical structure of the data file (e. g. description of data columns), units of values, symbols indicating anomalies, outliers, or missing values, and any additional information about the data.

→ **Data file(s) déposés dans 1 entrepôt**

Acknowledgements, References

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

cirad

CC BY NC

Modèles de *Data paper*

- Classique

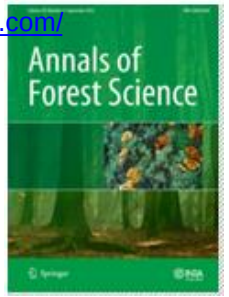
  Data in Brief, Geoscience Data Journal, Ecological research

- Classique + table de métadonnées scientifiques

  Annals of Forest Science

Modèle avec table de métadonnées

https://annforsci.biomedcentral.com/



Title, Authors, Abstract

Background : Contexte de l'étude et des données

Methods : description détaillée des méthodes d'acquisition

Data access & Metadata description :

- Lien vers les données dans un entrepôt
- Fichier Excel obligatoire pour les métadonnées :

  type de données, protocoles, équipements, logiciel, couvertures spatiale et temporelle, couverture taxonomique, variables, unités, …

Technical validation : justifier la qualité des données

Reuse potential and limits : conseils pour réutilisation des données

Acknowledgements, References

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

cirad

cc BY NC

Modèles de *Data paper*

- Classique

  Data in Brief, Geoscience Data Journal, Ecological research

- Classique + table de métadonnées scientifiques

  Annals of Forest Science

- Structuré par métadonnées disciplinaires

  Ecology, Open Health Data, Freshwater Metadata Journal

Modèle structuré par métadonnées disciplinaires

Data Papers

Ecology, 98(8), 2017, pp. 2224
© 2017 by the Ecological Society of America

https://esajournals.onlinelibrary.wiley.com/journal/19399170

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
weekly stream water
une 1963 to 30 May
chemistry for preci-
Forest are reported

l Brook Experimental

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

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

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

CC BY NC

Modèle structuré par métadonnées disciplinaires

Class II. Research origin descriptors
Site Description

Data Papers

1. Site description-

a. Site type: The site is an experimental forest located in the
Hampshire, operated under the supervision of the Northern Research Station of the United
States Forest Service.

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

*"The HBEF is located within the townships of Norti
Ellsworth (~40 %), and Warren (<1 %), NH, within
Forest of north central New Hampshire. Coordinate
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
(*Acer saccharum*), American beech (*Fagus grandifolia*) an
alleghaniensis).

d. Geology: Bedrock geology is composed of quartz-mica
Rangeley Formation. Significant glacial deposits overlay m
spodosol and inceptosol orders.

e. Watersheds/hydrology:

*"The experimental watershed-ecosystems range in
altitude from 500 to 910 m. These headwater water
slope of 20–30 %), with well-incised channels and
divides. The height of the land surrounding each wa
and the topography all have been determined from
photography and most recently augmented by Laser
Ranging (LIDAR). Experimental Watersheds 1–6 fa
face N to NE."*

(Likens 2013: 7)

f. Site history: Secondary forest, logged ca. 1910. Extensiv
found in Likens and Bormann (1979), Holmes and Likens (
http://www.hubbardbrook.org/overview/sitedescription.sht

g. Climate: mean annual precipitation - 1434 mm, mean ar

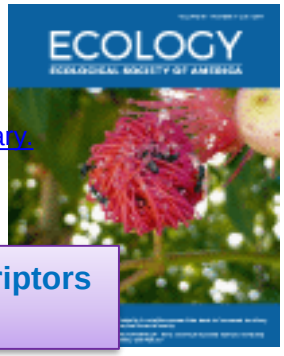
Class II. Research origin descriptors
Sampling design

2. Sampling design: The sampling design is exhaustively describ
Buso et al. (2000). Perhaps the most important aspect to explain h
precipitation dataset is curated. Contamination issues always plag
policy for the HBEF long-term record has been to curate the datas
sample information is retained. To help explain this we provide th
approach to QA/QC of precipitation samples

Precipitation Chemistry records within the dataset were compiled
to represent the "best" sample (read "cleanest") for each weekly c
the HBES, the issue of precipitation contamination by coarse part
clearly seen as a potential problem in open collectors. While not e
problem, the worst incidents tended to occur in the summer and fa
eliminate foreign objects, such as screens, wads of poly-wool, and
collectors, but these were problematic in their own way. Screens h
precipitation was 'filtered' through trapped debris. It was eventual
a bulk (continuously open) collector was the most appropriate cho
several collectors so that at least one sample collected from at leas
reasonably clean.

Although occasional grossly contaminated samples were not analy
collection), most precipitation samples were sent to the analytical
the presence of amorphous fine particulates. It turned out that som
and influence the analysis (probably bird feces), while others appe
fine pieces of bark or grit). There was no evidence, in the early da
chemistry from different sites around HBEF would result in signif
particulate contamination was considered a more important factor
difference in chemistry across the valley. Long-term data since th
course, the resulting input flux values were always calculated usin
or a pattern of rain gauges in the case of a whole watershed, since
is significant.

Thus, precipitation data reported here represents a 'mosaic' of the
from several sites, picked based on obvious field observations of
chemistry returned from the laboratory. The basic intention of this
sample for each period (week) in which precipitation occurred. Fo
a single rain chemistry estimate for the south facing watersheds (S



https://esajournals.onlinelibrary.wiley.com/journal/19399170

Class IV. Data structural descriptors
B. Variable Information

B. Variable information

Table 2A. Variable information for HBEF_stream_chemistry_1963-2013.csv

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 NH_4, NOT NH_4-N	floating point	[.001, 1.0]	0.001
pH	pH	unitless	floating point	[3.70, 6.93]	0.01

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

cirad

Modèles de *Data paper*

- Classique

  Data in Brief, Geoscience Data Journal, Ecological research

- Classique + table de métadonnées scientifiques

  Annals of Forest Science

- Structuré par métadonnées disciplinaires

  Ecology, Open Health Data

- Texte libre mais limité en taille (2 pages, 1000 mots, …)

  Phytopathology, Plant Phenome Journal, Hydrological Processes

Modèle en texte libre et court

MPMI Vol. 32, No. 2, 2019, pp. 139–141. https://doi.org/10.1094/MPMI-05-18-0144-A

RESOURCE ANNOUNCEMENT

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

[1]...ion Center for Sustainable Forestry in Southern China, College of Forestry, Nanjing Forestry ...g, Jiangsu 210037, China;
[2]...icultural Biotechnology, Seoul National University, Seoul 08826, Korea;
[3]...Program in Agricultural Genomics, Seoul National University;
[4]...ge of Biology and the Environment, Nanjing Forestry University;
[5]...r for Fungal Genetic Resources, Plant Genomics and Breeding Institute, and Research Institute of A...ulture and Life 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 GAPDH, 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. gloeosporioides* complex species

	C. gloeosporioides		*C. fructicola*	
Parameter	**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 able to serve as a reference genome for comparative analysis of the species complex. The genome data has been deposited in the National Center for Biotechnology Information NCBI GenBank database under accession number QFRH00000000, PRJNA471237 for the 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

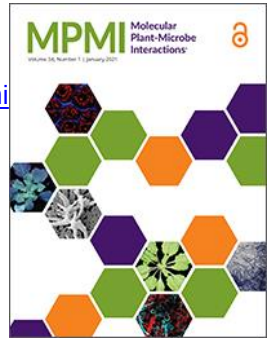
Alkan, N., Meng, X., Friedlander, G., Reuveni, E., Sukno, S., Sherman, A., Thon, M., Fluhr, R., and Prusky, D. 2013. Global aspects of pacC regulation of pathogenicity genes in *Colletotrichum gloeosporioides* as revealed by transcriptome analysis. Mol. Plant-Microbe Interact. 26:1345-1358.

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.

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



RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

cirad

CC BY NC

Modèles de *Data paper*

- Classique

  Data in Brief, Geoscience Data Journal, Ecological research

- Classique + table de métadonnées scientifiques

  Annals of Forest Science

- Structuré par métadonnées disciplinaires

  Ecology, Open Health Data

- Texte libre mais limité en taille (2 pages, 1000 mots, …)

  Phytopathology, Plant Phenome Journal, Hydrological Processes

- Structure avec Résultats et Discussion

  Earth System Science Data, Ethnobiology Letters, Plant Journal

cc BY NC

Localisation des données / entrepôts

- La revue impose un entrepôt de données (rare)

  ex : Studies in Mycology (MycoBank, GenBank/EMBL/DDBJ)

  Chemosphere (Pangaea)

  Journal of Economic History (openICPSR)

  Systematic Botany (DRYAD, GenBank)

- La revue recommande des entrepôts (le + fréquent)
  - Thématiques, généralistes, …
  - Editeurs

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

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

cirad

cc BY NC

Localisation des données / entrepôts

SCIENTIFIC DATA

nature publishing group

https://www.nature.com/sdata/

SCIENTIFIC DATA

Data Descriptor

A database of common vampire bat reports

Paige Van de Vuurst[1], M. Mónica Díaz[2], Annia Rodríguez-San Pedro[3], Juan Luis Allendes[4], Natalie Brown[5], Juan David Gutiérrez[6], Heliot Zarza[7], Stefan V. de Oliveira[8], Elsa Cárdenas-Canales[9], Rubén M. Barquez[10] & Luis E. Escobar[1,11,12]

The common vampire bat (*Desmodus rotundus*) is a sanguivorous (i.e., blood-eating) bat species distributed in the Americas from northern Mexico southwards to central Chile and Argentina. *Desmodus rotundus* is one of only three mammal species known to feed exclusively on blood, mainly from domestic mammals, although large wildlife and occasionally humans can also serve as a food source. Blood feeding makes *D. rotundus* an effective transmissor of pathogens to its prey. Consequently, this species is a common target of culling efforts by various individuals and organizations. Nevertheless, little is known about the historical distribution of *D. rotundus*. Detailed occurrence data are critical for the accurate assessment of past and current distributions of *D. rotundus* as part of ecological, biogeographical, and epidemiological research. This article presents a dataset of *D. rotundus* historical occurrence reports, including >39,000 locality reports across the Americas to facilitate the development of spatiotemporal studies of the species. Data are available at https://doi.org/10.6084/m9.figshare.15025296.

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

CC BY NC

Localisation des données / entrepôts

ELSEVIER

Data in Brief
Volume 44, October 2022, 108544

Data Article

Dataset about the adoption of winter cover crops at the municipality level for mainland France

Benjamin Nowak [a], Audrey Michaud [b], Gaëlle Marliac [c]

1. Generalist repositories are recommended only when specialized repositories (see below) are not available:

Repository name	Size limit	Costs
Mendeley Data	10 GB per dataset	Free
Dryad Digital Repository	300 GB per dataset	Fees applicable; consult with the official repository webpage for the latest information.
Harvard Dataverse	2 GB per file	Free
Open Science Framework	5 GB per file	Free
Zenodo	50 GB per file	Free
Science Data Bank	8 GB per file	Free

Earth and Planetary Sciences	Ecological/environmental research data	Environmental Data Initiative	Free
	Ecological data	ESS-DIVE archive	Free
		KNB: The Knowledge Network for Biocomplexity	Up to 5 GB for free
	Hydrologic data	HydroShare (CUAHSI)	Up to 20 GB free
	Atmospheric radiation data (submission restricted)	Atmospheric Radiation Measurement (ARM) Data Archive	Check with the repository.
	Marine geoscience data	Marine Geoscience Data System (MGDS)	Check with the repository.
	Marine data	Sea scientific open data publication (Seanoe)	Free

Biological Sciences	Proteomics	PRIDE	Free
		Panorama Public	Check with repository.
		ProteomeXchange	Free
	Metabolomics	MetaboLights	Free
		Protein Circular Dichroism Data Bank (PCDDB)	Free
		Biological Magnetic Resonance Data Bank (BMRB)	Free
	Molecular structure and interactions	Worldwide Protein Data Bank (wwPDB)	Free
		Structural Biology Data Grid	Free
		Cambridge Crystallographic Data Centre (CCDC)	Check with the repository.
		Electron Microscopy Data	Free

Egalité
Fraternité

Localisation des données / entrepôts

DATA PAPER

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

Takashi Nakada, Toshiya Katano, Keigo Hashimoto, Maiko Kagami

First published: 11 August 2021 | https://doi.org/10.1111/1440-1703.12254

Funding information: Asahi Glass Foundation, Environmental Field Research; JSPS KAKENHI, Grant/Award Number: JP19H05667

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).

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

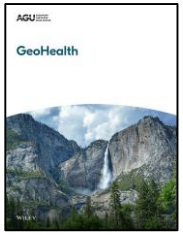
cirad

CC BY NC

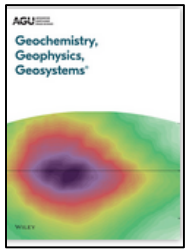
Modalités d'accès aux données

- La revue exige l'accès aux données
  - au moment de la publication
  - dès la soumission au moins pour les évaluateurs

BMC Part of Springer Nature
BMC Research Notes



SCIENTIFIC DATA



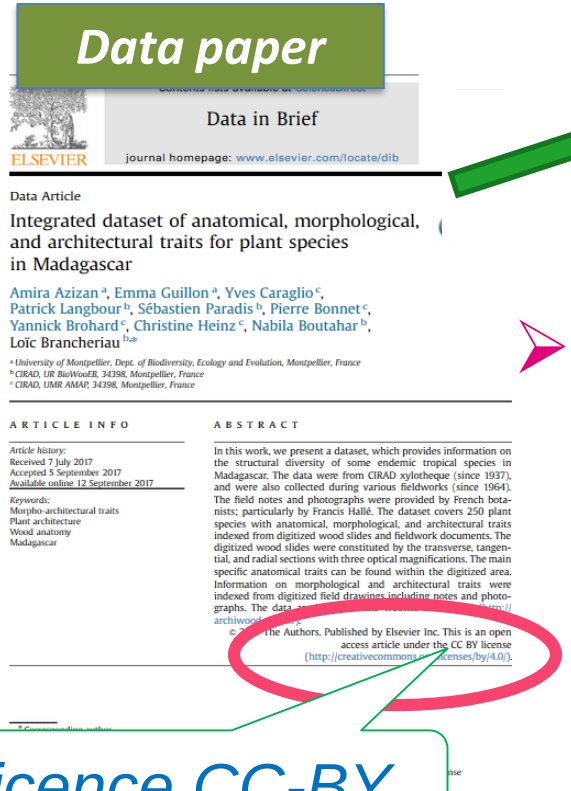
- La revue accepte que les données soient accessibles
  - après embargo
  - sur demande ou collaboration

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

cirad

CC BY NC

Modalités de diffusion des données



Data paper

Data in Brief

ELSEVIER

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

Data Article

Integrated dataset of anatomical, morphological, and architectural traits for plant species in Madagascar

Amira Azizan [a], Emma Guillon [a], Yves Caraglio [c], Patrick Langbour [b], Sébastien Paradis [b], Pierre Bonnet [c], Yannick Brohard [c], Christine Heinz [c], Nabila Boutahar [b], Loïc Brancheriau [b,*]

[a] University of Montpellier, Dept. of Biodiversity, Ecology and Evolution, Montpellier, France
[b] CIRAD, UR BioWooEB, 34398, Montpellier, France
[c] CIRAD, UMR AMAP, 34398, Montpellier, France

ARTICLE INFO

Article history:
Received 7 July 2017
Accepted 5 September 2017
Available online 12 September 2017

Keywords:
Morpho-architectural traits
Plant architecture
Wood anatomy
Madagascar

ABSTRACT

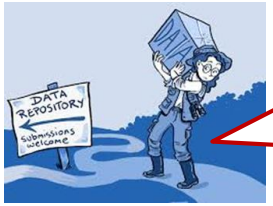
In this work, we present a dataset, which provides information on the structural diversity of some endemic tropical species in Madagascar. The data were from CIRAD xylotheque (since 1937), and were also collected during various fieldworks (since 1964). The field notes and photographs were provided by French botanists; particularly by Francis Hallé. The dataset covers 250 plant species with anatomical, morphological, and architectural traits indexed from digitized wood slides and fieldwork documents. The digitized wood slides were constituted by the transverse, tangential, and radial sections with three optical magnifications. The main specific anatomical traits can be found within the digitized area. Information on morphological and architectural traits were indexed from digitized field drawings including notes and photographs. The data

© The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Licence CC-BY

Données déposées

dans un entrepôt de données



Licence ≠

➢ Licences de diffusion

- Licence sur l'article dans la revue

+

- Licence sur les données dans l'entrepôt

RÉPUBLIQUE FRANÇAISE
Liberté Égalité Fraternité

cirad

cc BY NC

Modalités de diffusion des données

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

(GIGA)n SCIENCE

F1000Research

BMC Research Notes

Licence CC0
la + ouverte
aucune restriction
possible

SCIENTIFIC DATA

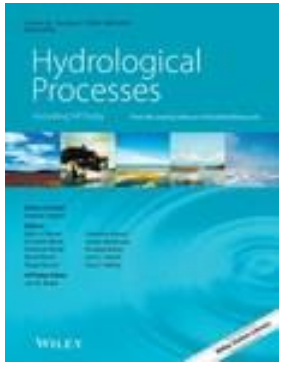
Earth System Science Data

Licence CC0 ou CC-BY

cybergeo

CC0
CC-BY
CC-BY-SA

CC-BY
CC-BY-SA



Journal of open health data

Licence CC-BY-NC
Pas d'utilisation commerciale

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

cirad

CC BY NC

Data paper / Article de recherche

- Vérifiez que publier le *Data paper* ne nuira pas à la publication d'un futur article de recherche

- Certaines revues l'acceptent

  Scientific Data : OK pour publier dans les revues du groupe Nature

  F1000Research : OK pour BMC, PLoS, Elsevier, SAGE, PNAS, Science

- D'autres non.

  ex: l'éditeur Cell Press : environ 50 revues !

  "As a matter of publishing ethics, we cannot consider any paper that contains data that have been published or submitted for publication elsewhere".

  https://www.cell.com/cell/authors#policies

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

cirad

CC BY NC

Choisir une revue scientifique

- Consulter sa communauté scientifique + revues + littérature citée
- 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
  - 2200 revues
  - Critères de recherche :

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

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

cirad

CC BY NC

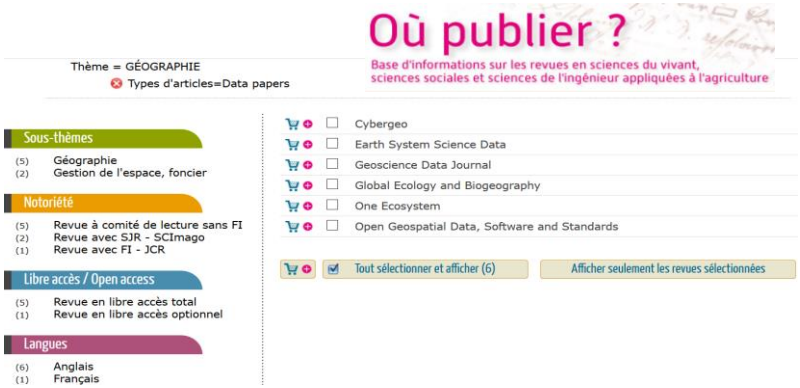
Mise en pratique: Recherche de revues

Ce support Rédiger un *Data paper* de L. Dedieu est mis à disposition selon les termes de la licence Creative Commons Attribution - Pas d'Utilisation Commerciale 4.0 International.

Exercice: choisir 1 revue scientifique

- A partir des listes de revues publiant des *Data papers*
  - Liste Cirad : 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

- A partir du site Où publier :



- Vérifier :
  - si elle est adaptée à vos données: domaine, échelle, portée, …
  - le modèle de *Data paper*
  - si elle propose un choix d'entrepôts
  - et un choix de licences de diffusion des données.

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

cirad

CC BY NC

Choisir l’entrepôt de données

Ce support Rédiger un *Data paper* de L. Dedieu est mis à disposition selon les termes de la licence Creative Commons Attribution - Pas d’Utilisation Commerciale 4.0 International.

Les entrepôts de données

- Nombreux entrepôts de données
  - Généralistes

    Dryad, FigShare, Dataverse
  - Institutionnels, Nationaux, Européens

    Dataverse, Zenodo, Eudat, B2Share
  - Portés par des éditeurs de revues

    Oxford Univ Press (GigaDB), Ubiquity Press (Dataverse), Mendeley (Elsevier)
  - Thématiques ou disciplinaires

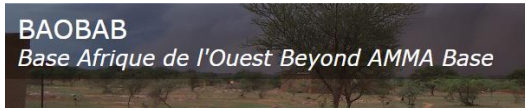
Les entrepôts de données

- **Thématiques**
  - GBIF (Global Biodiversity Information Facility)
  - EDI, DataOne
  - Pangaea, SEANOE
  - Movebank, WormBase, MycoBank, ComBase
  - GenBank, Barcode of Life Data Systems, UniProt, Intact
  - Dataverse, ICPSR, Quetelet, beQuali
  - Software Heritage

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité
cirad
CC BY NC

Les entrepôts de données

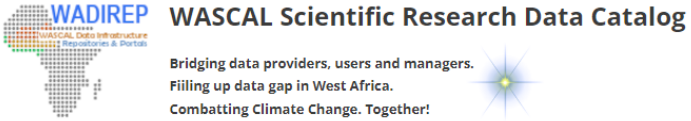
- **En Afrique**



  - BAOBAB : Base Afrique de l'Ouest Beyond AMMA Base
  - West African Vegetation

  - West African Plants

  - WASCAL Data Discovery Portal



    West African Science Service Center
    on Climate Change and Adapted Land Use Data Discovery Portal

  - DataFirst

  - AfDB Statistical Data Portal

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

cirad

CC BY NC

Critères de choix d'un entrepôt

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

  → *"deposit data with a community-recognized repositories"*

- Privilégiez un entrepôt reconnu dans la discipline
  - Porté par une communauté scientifique
  - Correspondant aux publics scientifiques visés
  - Valorisant vos données
- Délivrant un identifiant numérique pérenne et unique
- Adapté à la taille de vos fichiers
- Coût du dépôt : la plupart sont gratuits

  sauf par ex : PANGAEA, DRYAD, TreeBase

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

cirad

cc BY NC

Critères de choix d'un entrepôt

- Autorisant les modalités de diffusion que vous souhaitez
  - souvent l'entrepôt offre un choix de licences
  - mais parfois licence imposée

    ex : Dryad et MoveBank applique la licence CC0 (domaine public)

    DRYAD MOVEBANK For Animal Tracking Data
  - + la licence est ouverte, + les possibilités de réutilisation

    CC0 >> CC-BY >> CC-BY-NC

Voir la fiche Déposer des données de recherche dans un entrepôt

cc BY NC

Choisir un entrepôt

- Consultez d'abord:
  - Votre communauté scientifique / vos collègues
  - Entrepôts recommandés par la revue ciblée

    Scientific Data

    Springer

  - Sites de recherche de données:

    DataCite,

    Google Dataset search,

    WoS (Data Citation Index)…

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

cirad

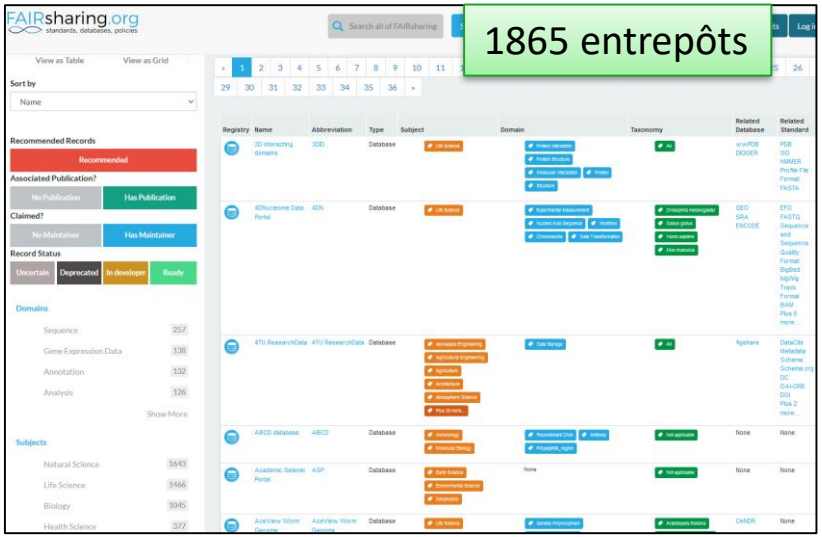
CC BY NC

Choisir un entrepôt

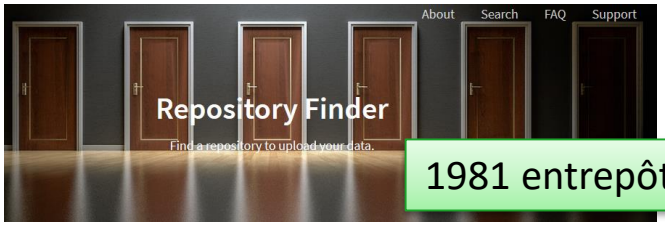
https://www.re3data.org/search

2855 entrepôts

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



1865 entrepôts



1981 entrepôts

https://repositoryfinder.datacite.org/

Entrepôts en biologie moléculaire

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

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

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

cirad

CC BY NC

Mise en pratique: Recherche d’entrepôts

Ce support Rédiger un *Data paper* de L. Dedieu est mis à disposition selon les termes de la licence Creative Commons Attribution - Pas d’Utilisation Commerciale 4.0 International.

A retenir

Ce support Rédiger un *Data paper* de L. Dedieu est mis à disposition selon les termes de la licence Creative Commons Attribution - Pas d'Utilisation Commerciale 4.0 International.

A retenir

- Vos données ont de la valeur et peuvent intéresser d'autres scientifiques → partagez vos données
- Bénéfices: visibilité, notoriété, citations, collaborations

Valorisation

Crédibilité

Notoriété

Reconnaissance

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

Attractivité

Collaborations

Nouveaux projets

Mes données sont préservées

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité
cirad
CC BY NC

A retenir

- Bonne pratique :
  - → Déposez vos données dans 1 entrepôt + documentation associée (protocole, liste des variables, explication abréviations, etc.)
  - → Publiez un *Data paper* pour informer
- Objectif : Réutilisation des données et Reproductibilité des études
- 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
- Pour réussir:
  - → Documenter régulièrement vos expériences, protocoles, données
  - → Avoir fait 1 Plan de gestion des données vous facilite la tâche.

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

cirad

CC BY NC

Publier un *Data paper*

- Si votre *Data paper* est complet
- Et si vos données sont compréhensibles et accessibles dans un entrepôt:

  Vous avez toutes les chances d'être publié, … lu,….cité

- Mise en valeur de votre travail
  - article évalué par les pairs, citable, indexé par les moteurs de recherche, comptant comme un article de recherche
  - visibilité des données dans un entrepôt
- Montre la transparence de vos recherches
- Permet à d'autres de découvrir votre travail et les données que vous produisez
- Vous rend visible dans votre communauté scientifique.

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

cirad

Merci de votre attention

Laurence Dedieu

laurence.dedieu@cirad.fr

Publier un Data paper. Dedieu Laurence. Fiche CoopIST. 2022.
https://doi.org/10.18167/coopist/0057

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

cirad

Support mis à disposition selon les termes de la licence Creative Commons Attribution - Pas d'Utilisation Commerciale 4.0 International.