RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

cirad
LA RECHERCHE AGRONOMIQUE POUR LE DÉVELOPPEMENT

Publier un *Data Paper* : enjeux de bonnes pratiques, d'intégrité scientifique et de science ouverte

Laurence Dedieu
Laurence.dedieu@cirad.fr

Délégation à l'Information Scientifique et à la science Ouverte

Pourquoi publier un *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* les rend visibles et accessibles
- Affirme votre paternité
- Expose la transparence et l'intégrité de vos recherches

- Rédiger un *Data paper* peut être simple
- Large choix de revues et d'entrepôts de données

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

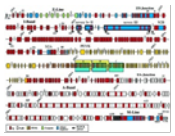
cirad

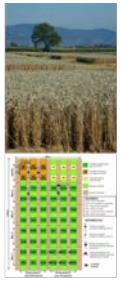
Quelles données concernées ?

- Tous types de jeux de données

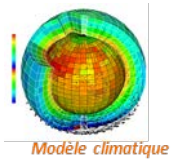
  Séquences, mesures, localisations, photos, enquêtes, presse,…

Données d'observation



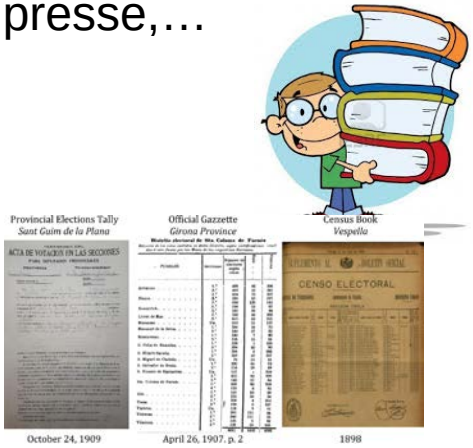


Données expérimentales



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.

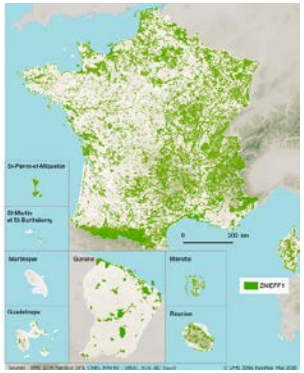
Quelles données concernées ?

Data Paper

Biodiversity Data Journal 10: e71222
https://doi.org/10.3897/BDJ.10.e71222 (04 Mar 2022)

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 Ves
Arnaud Horellou, Laurent Poncet



RESEARCH DATA JOURNAL FOR THE HUMANITIES AND SOCIAL SCIENCES (2021) 1–12
BRILL

A Data Set for US Horror Film Trailers

Arts and Media

Nick Redfern
Independent researc
nickredfernres@outlo

mprising audio, colour, motion, and sho
ssing horror films at the US box office fr
v available for computational film analy
f film style and is the only existing data
mal analyses. Data is stored in csv files a
n 4.0 International license on Zenodo: w

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]

Concept et objectif du *Data paper*

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

cirad

Data paper = article + entrepôt

Contents lists available at ScienceDirect

Data in Brief

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

Data Article

An annotated dataset for event-based surveillance of antimicrobial resistance

Nejat Arınık[a,c], Wim Van Bortel[d], Bahdja Boudoua[a,c], Luca Busani[e], Rémy Decoupes[a,c], Roberto Interdonato[b,c], Rodrique Kafando[a,c], Esther van Kleef[f], Mathieu Roche[b,c], Mehtab Alam Syed[b,c], Maguelonne Teisseire[a,c,*]

[a] INRAE, Montpellier F-34398, France
[b] CIRAD, Montpellier F-34398, France
[c] TETIS, Univ. Montpellier, AgroParisTech, CIRAD, CNRS, INRAE, Montpellier 34090, France
[d] ITM, Institute of Tropical Medicine, Department of Biomedical Sciences, Antwerp, Belgium
[e] Center for Gender-Specific Medicine, Istituto Superiore di Sanità Viale Regina Elena 299, 00161 Rome, Italy
[f] ITM, Institute of Tropical Medicine, Department of Public Health, Outbreak Research Team, Antwerp, Belgium

ARTICLE INFO

Article history:
Received 16 November 2022
Revised 15 December 2022
Accepted 27 December 2022
Available online 1 January 2023

Dataset link: MOOD - News AMR dataset - Hackathon 2022 (Original data)

Keywords:
Antimicrobial resistance (AMR)
Text mining
Annotation
Epidemiology

ABSTRACT

This paper presents an annotated dataset used in the MOOD Antimicrobial Resistance (AMR) hackathon, hosted in Montpellier, June 2022. The collected data concerns unstructured data from news items, scientific publications and national or international reports, collected from four event-based surveillance (EBS) Systems, i.e. ProMED, PADI-web, HealthMap and MedISys. Data was annotated by relevance for epidemic intelligence (EI) purposes with the help of AMR experts and an annotation guideline. Extracted data were intended to include relevant events on the emergence and spread of AMR such as reports on AMR trends, discovery of new drug-bug resistances, or new AMR genes in human, animal or environmental reservoirs. This dataset can be used to train or evaluate classification approaches to automatically identify written text on AMR events across the different reservoirs and sectors of One Health (i.e. human, animal,

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	51.8717003	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

Objectif : que les données soient compréhensibles et réutilisables

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

cirad

Data paper = article + entrepôt

1. Décrire un jeu de données et son contexte

   suffisamment pour assurer sa compréhension

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 convaincre l'éditeur de l'originalité des données et de leur importance scientifique, sociétale, environnementale,….

4. Donner le lien d'accès au jeu de données

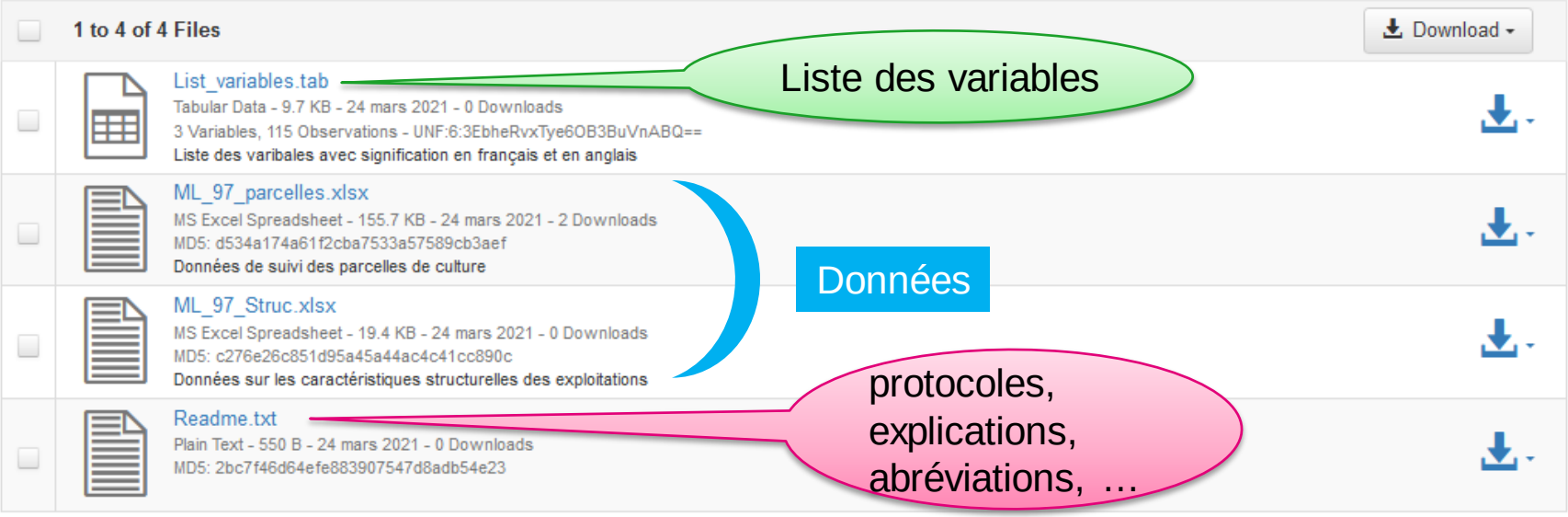
Pas de résultats, ni analyses, ni interprétation

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

cirad

Data paper = article + entrepôt

- Entrepôt: accès aux 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

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

cirad

Ex 1: *Data paper* sur données d'enquête

Food and Ecological Systems Modelling Journal

Data Paper (Biosciences)

Food and Ecological Systems Modelling Journal 3: e91025
https://doi.org/10.3897/fmj.3.91025 (05 Oct 2022)

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.

Pratiques de consommation de lentilles en France

→ étude de l'alimentation à base de protéines végétales

Méthodes

Table 1. Download as CSV XLSX

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).

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 👤; 11 ? : de 18 à 91 ans

Lien vers les données
dans l'entrepôt Recherche Data Gouv
https://doi.org/10.57745/KMGODH

Ex 1: *Data paper* sur données d'enquête

Lien vers les données
dans l'entrepôt Recherche Data Gouv
https://doi.org/10.57745/KMGODH

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

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://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 60
France in 2022. The 2nd file
frequency of consumption, t
general questions on age, g
type of lentils is consumed (
may be then plugged into a
exposure model. English (20

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

Ex 2: *Data paper* - données biodiversité

ZooKeys Home Articles About About Pen

Data Paper

ZooKeys 306: 59-70
https://doi.org/10.3897/zookeys.306.4898 (03 Jun 2013)

FORMIDABEL: The Belgian Ants Database

Methods

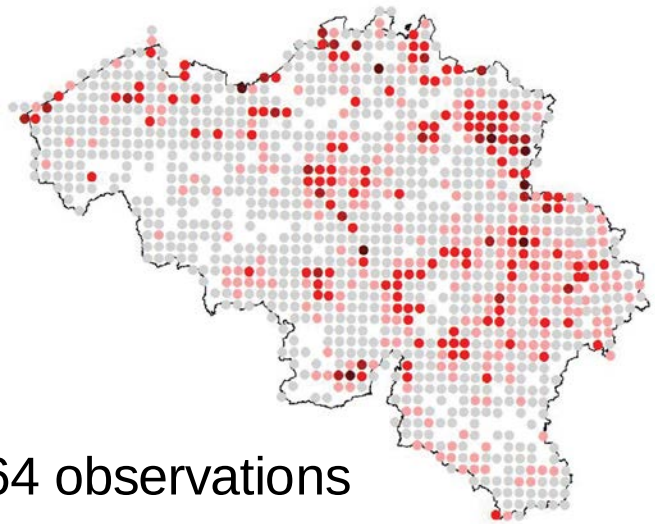
Method step description:

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) andpitfall 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 final publication of the database all the records were tested on geographical consistency by the Belgian

Base de données sur la distribution des fourmis en Belgique



27 264 observations

→ 76 espèces de fourmis natives

→ 9 espèces introduites

→ description de chaque microhabitat de fourmis.

Données actualisées 2 fois par an.

Lien vers les données dans l'entrepôt GBIF
https://www.gbif.org/dataset/b528799a-2d52-4023-aa02-9ce081e3ca5f

Ex 2: *Data paper* - données biodiversité

Lien vers les données dans l'entrepôt GBIF
Global Biodiversity Information Facility
https://www.gbif.org/dataset/b528799a-2d52-4023-aa02-9ce081e3ca5f

Formidabel; Belgian Ants Database

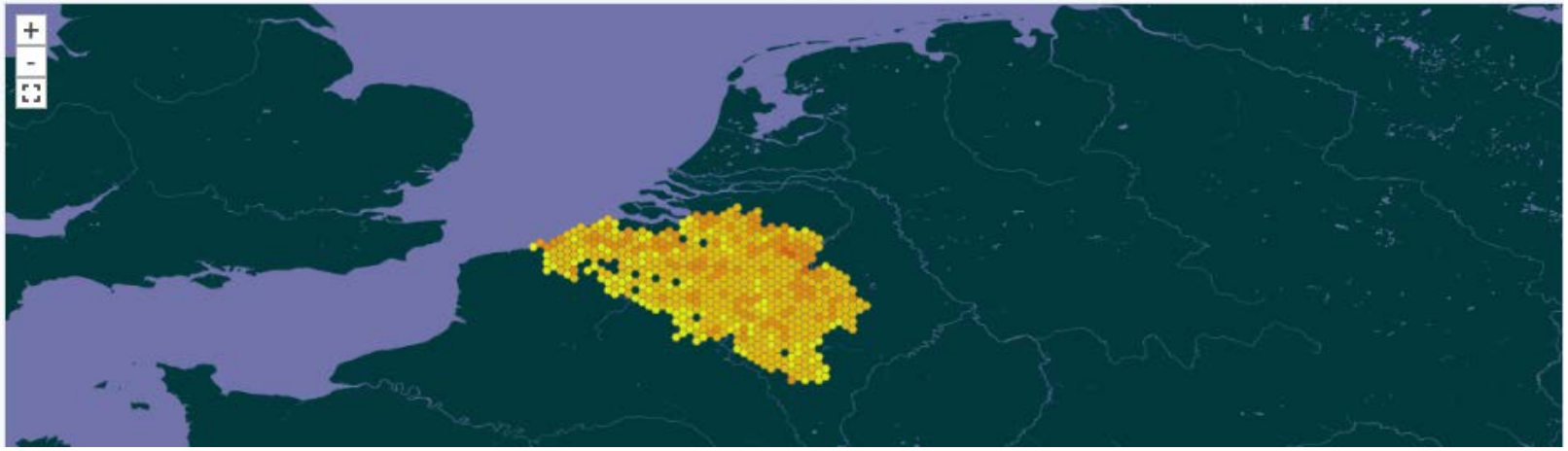
DATASET METRICS ACTIVITY DOWNLOAD HOME PAGE

27,264 OCCURRENCES 64 CITATIONS

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
How to cite DOI 10.15468/xdapub

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



Data paper = article + entrepôt

ZooKeys Home Articles About About Pen

Data Paper ZooKeys 306: 59-70
https://doi.org/10.3897/zookeys.306.4898 (03 Jun 2013)

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) andpitfall 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 final publication of the database all the records were tested on geographical consistency by the Belgian Biodiv

Moteurs de recherche biblio
Web of Science
Scopus
Google scholar

4 citations

Global Biodiversity Information Facility
GBIF

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

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
How to cite DOI 10.15468/xdapub

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

27,221 GEOREFERENCED RECORDS

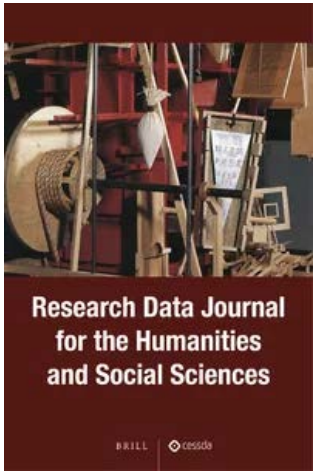
Moteurs de recherche de données
Datacite search
Google Dataset search
Mendeley Data

85 citations

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

cirad

Publier un *Data paper*



Les étapes

1. Choisir la revue et consulter 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 le format et les informations (métadonnées) 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 associée
6. Soumettre le *Data paper* à la revue

   avec le lien vers l'entrepôt où est déposé le jeu de données

Critères d'évaluation

- Tous les *Data papers* sont évalués :
  - Originalité des données et intérêt pour la discipline
  - Description suffisante pour permettre à d'autres de comprendre, interpréter et réutiliser les données
  - Adéquation des méthodes avec les pratiques disciplinaires
  - Qualité et fiabilité des données
  - Potentiel de réutilisation des données.
- Exigences différentes selon la revue
  - Evaluation du texte +/- organisation des données, formats, entrepôt, parfois jusqu'aux données elles-mêmes.

Les reviewers peuvent vous demander vos données dès la soumission du manuscrit !

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

cirad

Choisir la revue et l'entrepôt de données

Revues publiant des *Data papers*

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

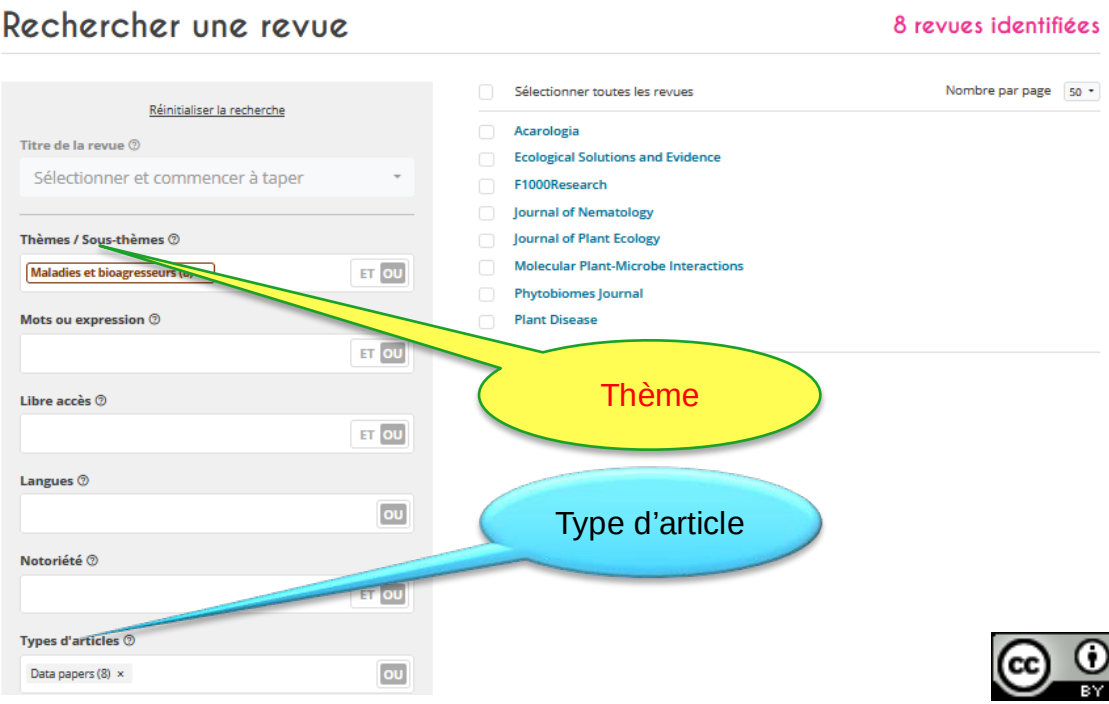
cirad

Revues publiant des *Data papers*

- Choisir en consultant sa communauté scientifique
- Listes disponibles
  - **Liste Cirad de revues publiant des *Data papers*** : https://doi.org/10.18167/coopist/0057
  - GBIF (biodiversité): http://www.gbif.fr/page/contrib/publier-un-datapaper
  - 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
- Où publier

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

  - 2300 revues
  - Plusieurs critères de recherche



RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

cirad

Revues publiant des *Data papers*

- Data journaux publient uniquement des *Data papers*
- Revues classiques publient ≠ types d'articles dont *Data papers*

- Revues en libre accès ou non
- Publication gratuite ou payante

- Multidisciplinaires, disciplinaires ou thématiques
- avec ou sans facteur d'impact

- ≠ noms : *Data paper, Data note, Data article, Data Briefs, Data descriptor, 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 ou plus complexe
  - Modalités d'accès aux données lors de la soumission du manuscrit
  - Localisation des 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 €

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

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

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

cirad

Modèle de *Data paper*

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

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

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

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

cirad

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

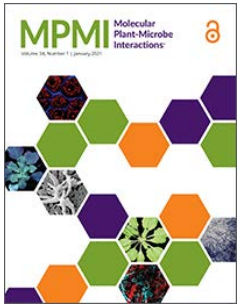
Alkan, N., Meng, X., Friedlander, G., Reuveni, E., Sukno, S., Sherman, A., Thon, M., Fluhr, R., and Prusky, D. 2013. Global aspects of pacC regulation of pathogenicity genes in *Colletotrichum gloeosporioides* as revealed by transcriptome analysis. Mol. Plant-Microbe Interact. 26:1345-1358.

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

Modèle de *Data paper*

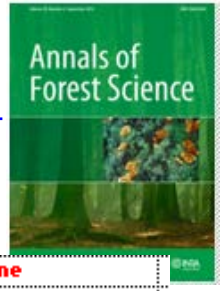
- Selon la revue, le modèle de *Data paper* diffère entre:

- Modèle classique
  Data in Brief, Geoscience Data Journal, Ecological research
- Texte libre mais limité en taille (2 pages, 1000 mots, …)
  Phytopathology, Plant Phenome Journal, Hydrological Processes
- Modèle classique + table de métadonnées
  Annals of Forest Science

Modèle avec table de métadonnées

Family	Value
Species	
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
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

https://annforsci.biomedcentral.com/



Data provider(s)	Name and first name
	Organization / Institute
	Telephone (optional)
	Address
	City
	Postal code
	Country
	E-mail address
	ORCID identifier (optional)
Spatial coverage	Minimum longitude (x min in decimal degree)
	Maximum longitude (x max in decimal degree)
	Minimum latitude (y min in decimal degree)
	Maximum latitude (y max in decimal degree)
Temporal coverage of the collected data	Start date
	End date

Variable name	Description	Type	Unit	Value range
SiteCode	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 fhe non-geolocalized site.	character		
SiteName	Common name of the site (i.e. "place called")	character		
Species	latin binomial name (Genus species)	character		
SiteXSpecies	unique identifier for the site and species	character		
Date	sampling date	character		
Year	sampling year	numeric integer		1996 – 2016
Month	sampling month	numeric integer		1 – 12
Doy	sampling day of year	numeric integer		1 – 366
RobustLFMC	Robust estimates of mean LFMC. Detailed of computation are provided in Martin-StPaul et al 2017	numeric float	%	10 -- 320
RobustStandErrLFM	Robust estimates of the standard error of LFMC. Detailed of computation are provided in Martin-StPaul et al 2017	numeric float	%	0,6 -- 11
RobustNval	Number of robust values used to compute the robust estimates of the mean and the standard error	numeric integer		1 -- 5

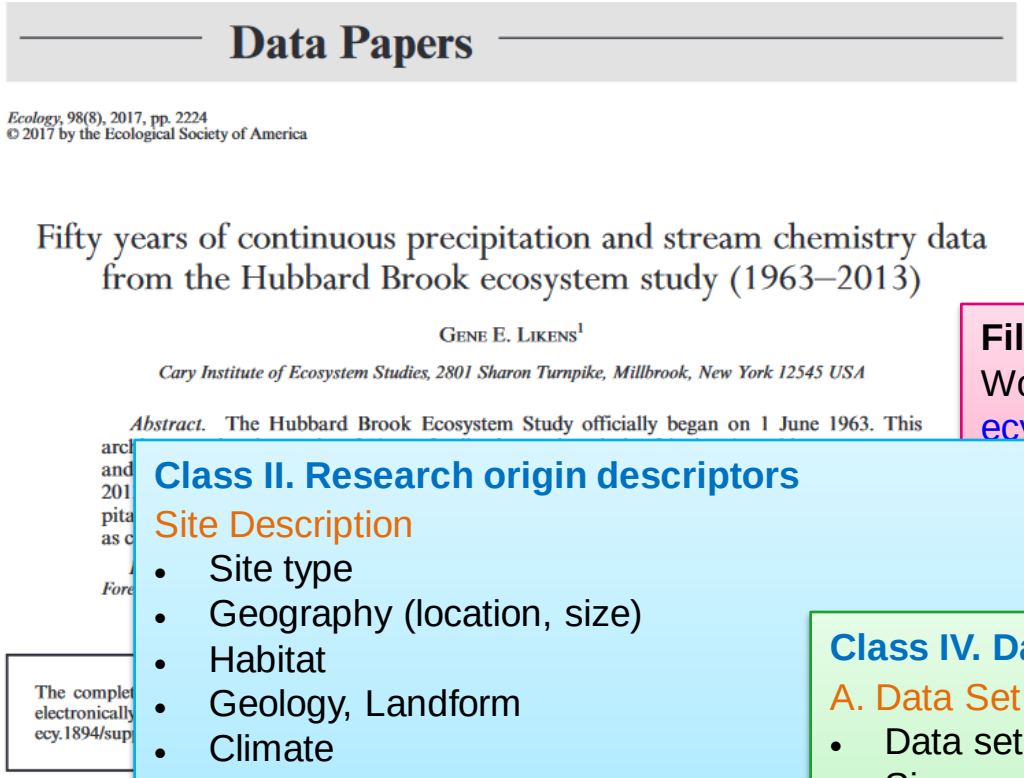
Modèle de *Data paper*

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

- Modèle classique
  Data in Brief, Geoscience Data Journal, Ecological research
- Texte libre mais limité en taille (2 pages, 1000 mots, …)
  Phytopathology, Plant Phenome Journal, Hydrological Processes
- 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

Exemple du **standard EML** = Ecological metadata language



Data Papers

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

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

Abstract. The Hubbard Brook Ecosystem Study officially began on 1 June 1963. This

The complet
electronically
ecy.1894/sup

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

Filename
Word document, 125.5 KB
ecy1894-sup-0001-MetadataS1.doc
34.2 KB
02-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 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

Modèle structuré par métadonnées disciplinaires

Exemple du **standard EML** = Ecological metadata language

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

Class II. Research origin descriptors
Site Description

1. Site description-

a. Site type: The site is an experimental forest locat
Hampshire, operated under the supervision of the N
States Forest Service.

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

"The HBEF is located within the townships o
Ellsworth (~40 %), and Warren (<1 %), NH,
Forest of north central New Hampshire. Coo
the area. The Atlantic Ocean is about 116 km
West Thornton, NH."

(Likens 2013: 5)

c. Habitat: The ecosystem is a mixed temperate harc
(*Acer saccharum*), American beech (*Fagus grandifo*
alleghaniensis).

d. Geology: Bedrock geology is composed of quartz
Rangeley Formation. Significant glacial deposits ove
spodosol and inceptosol orders.

e. Watersheds/hydrology:

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

(Likens 2013: 7)

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

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

Class II. Research origin descriptors
Sampling design

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

Precipitation Chemistry records within the dataset were compi
to represent the "best" sample (read "cleanest") for each weekl
the HBES, the issue of precipitation contamination by coarse p
clearly seen as a potential problem in open collectors. While n
problem, the worst incidents tended to occur in the summer an

ted

Class II. Research origin descriptors
Research Methods

3. Research methods: The analytical methods prior to 200
technical report by Buso et al. (2000). Samples were analyz
Studies (Millbrook, NY)

This data record concludes in May 2013, As of June 2013,
chemical analyses were transferred from the Cary Institute
Service Lab in Durham, NH.

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
SO4	sulfate concentration	mg/L as SO_4	floating point	[0.05, 11.8]	0.01

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

cirad

Modèle de *Data paper*

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

- Modèle classique
  Data in Brief, Geoscience Data Journal, Ecological research
- Texte libre mais limité en taille (2 pages, 1000 mots, …)
  Phytopathology, Plant Phenome Journal, Hydrological Processes
- Modèle classique + table de métadonnées
  Annals of Forest Science
- Modèle structuré par des métadonnées disciplinaires
  Ecology, Open Health Data, Freshwater Metadata Journal
- Modèle avec Résultats et Discussion
  Earth System Science Data, Ethnobiology Letters, Plant Journal

→ **Bien choisir sa revue**

Soumission et 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

Dépôt des données dans un entrepôt

- La revue recommande des entrepôts (le + fréquent)

Entrepôts de données

- Infrastructure qui préserve et diffuse des données de recherche

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

- Différents types d'entrepôts de données

  Institutionnels, européens, nationaux

Recherche Data Gouv

(Recherche Data gouv)

Génération datapaper

Statistiques 806 515 téléchargements

Contact Partager

L'entrepôt pluridisciplinaire *Recherche Data Gouv* est une solution souveraine pour le partage et l'ouverture des données de recherche produites par les communautés qui ne disposent pas d'un entrepôt disciplinaire reconnu. Il est basé sur le logiciel libre Dataverse. Le dépôt des données doit se faire dans l'espace institutionnel attribué à l'établissement dont relève un des contributeurs. Un espace générique est dédié aux données produites par les établissements ne disposant pas d'espace dédié. Les tests sont à effectuer dans le bac à sable.
Retrouver les actualités et événements de la plateforme *Recherche Data Gouv*.

The *Recherche Data Gouv* multidisciplinary repository is a sovereign solution for sharing and opening up data produced by communities that do not have a recognized disciplinary repository. It is based on the Dataverse software. Data should be deposited in the space assigned to an institution which one of the contributor belongs to. A generic space is available for data produced by institutions which do not such a space yet themselves. Tests should be performed in the sandbox.
The latest news and events on the *Recherche Data Gouv* platform.

Data Repository Grenoble Alpes

Data Univ. Gustave Eiffel

Data_UMontpellier

Data UPPA

Entrepôts de données

- Infrastructure qui préserve et diffuse des données de recherche
- Différents types d'entrepôts de données

  Institutionnels, européens, nationaux

  Généralistes

  Disciplinaires ou thématiques (recommandés)

- Répertoire d'entrepôts de données

  Re3data https://www.re3data.org FairSharing https://fairsharing.org

> 3000

2075

- Critères de choix d'un entrepôt

  Recommandé par la revue

  Le + utilisé dans votre discipline

  Délivrant un identifiant numérique à votre jeu de données

  Services: visibilité, aide au dépôt, suivi de citations, embargo,…

  Cout de dépôt : Gratuit (la plupart); certains payants

Ni un site web de projet

Ni Research Gate !!!!!!

Dépôt des données dans un entrepôt

➢ Recommandation : Entrepôt disciplinaire de préférence

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**

SCIENTIFIC DATA
nature publishing group npg

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

Ecology

TERN Data Discovery Portal
Environmental Data Initiative (formerly LTER Network Information System Data Portal)
Global Biodiversity Information Facility (GBIF)
KNB: The Knowledge Network for Biocomplexity

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

cirad

Dépôt des données dans un entrepôt

- La revue recommande des entrepôts (le + fréquent)

  Disciplinaires, thématiques, généralistes

- Dans certaines disciplines: entrepôts obligatoires

  OMICS : GenBank, ENA, GEO, UniProt, EVA, EGA,

  Cristallographie : Cambridge Structural Database, Crystallography Open Database

  https://www.springernature.com/gp/authors/research-data-policy/repositories-mandates/19540364

- Certains éditeurs proposent leur entrepôt de données

  mais pas d'obligation

  ex : Mendeley (Elsevier), Dataverse (Ubiquity Press)

- La revue accepte d'intégrer les données dans l'article

  N'est pas une bonne pratique :
  Données moins accessibles, moins visibles, moins réutilisables
  Pas de choix de licence de diffusion de vos données

Modalités de diffusion des données

scientific **data**

Data Descriptor | Open access | Published: 25 November 2020

Response2covid19, a dataset of governments' responses to COVID-19 all around the world

Simon Porcher

Following the COVID-19 outbreak, governments all around the world have implemented public health and economic measures to contain the spread of the virus and to support the economy. Public health measures include domestic lockdown, school closures and bans on mass gatherings among others. Economic measures cover wage support, cash transfers, interest rates cuts, tax cuts and delays, and support to exporters or importers. This paper introduces 'Response2covid19', a living dataset of governments' responses to COVID-19. The dataset codes the various policy interventions with their dates at the country-level for more than 200 countries from January 1 to October 1, 2020 and is updated every month. The production of detailed data on the measures taken by governments can help generate robust evidence to support public health and economic decision making.

Background & Summary

In December 2019, a new coronavirus appeared in Wuhan, China and spread to nearly every country in the first quarter of 2020. In the beginning of November 2020, according to the Johns Hopkins Coronavirus Resource Center (https://coronavirus.jhu.edu/), there were more than 50.25 million confirmed cases and over 1.25 million deaths linked to the virus. The pandemic forced governments all around the world to adopt diverse public health policies and economic measures that are quite unique in history. Public policies data is needed in pandemics to best monitor the spread of infection, but also to understa the diversity in governments' responses. In order to provide accurate and openly available data, we collected d from other datasets or countries case studies on thirteen public health policies and seven economic measure the country-level and on a daily basis, and merged it with the data on daily cases and deaths from the Europe enter for Disease Prevention and Control (https://www.ecdc.europa.eu/en/geographical-distribution-2019 ases). The resulting dataset, Response2covid19, tracks the implementation and intensity of 20 governme ures taken during the spread of the pandemic, and is updated on a monthly basis.

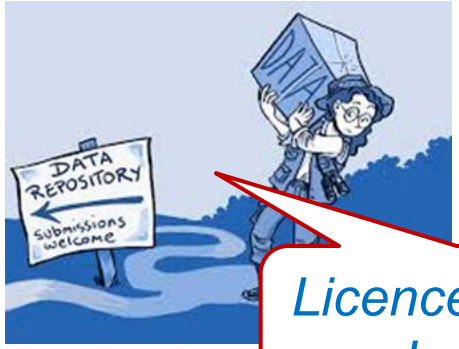
We make sense of the coded measures by creating tw dices of governments' interventions against COVID-19. The first index measures the rigidity of gove sponses to COVID-19 and is based on the implementation of thirteen public health measures. Th ex quantifies the economic responses to COVID-19 based on the coding of seven types of econo ons to face the economic downturn following the various public health measures. The indices dity of governments' responses to the pandemic between January 1, 2020, and October 1, 2020 y basis. In its current version, the final dataset is made of 62,700 country-day observations. 228 ded in the database. The two indices allow cross-country comparisons and the documentation countries.

The dataset is of interest for epidemiologists wishing measures worldwide with the evolution of the number of cases[1]. Several studies already as el restrictions[2], human mobility restrictions[3] or various transmission control measures[4]. T rest for social scientists wishing to study the impact of other factors, e.g. democracy or ins and the timing of the measures tak with economic ou

Licence CC-BY

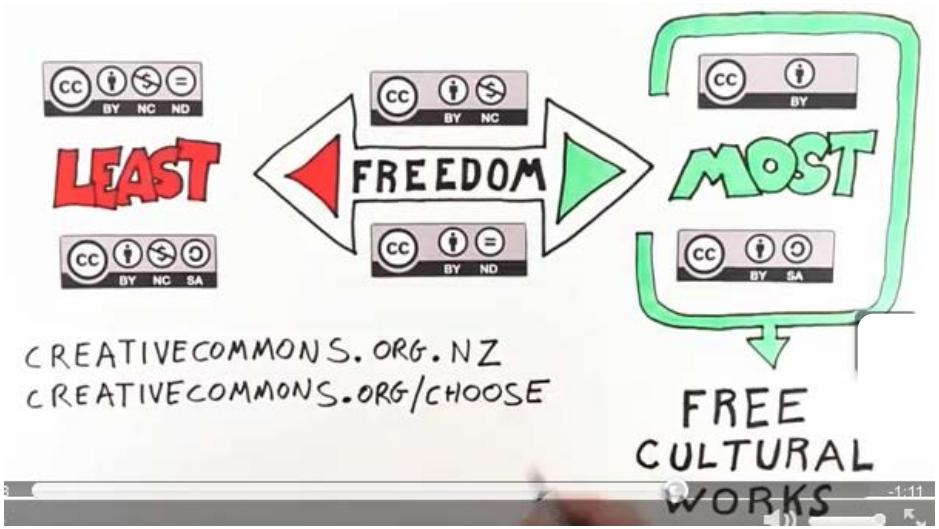
décidée par l'éditeur

Données déposées dans un entrepôt de données



Licence ≠ choisie par le déposant

La licence de diffusion sert à indiquer précisément les droits des utilisateurs



https://www.youtube.com/watch?v=gftrFKqtUlU

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

cirad

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

A retenir

Enjeux de partage, valorisation, intégrité

- Vos données ont de la valeur, après les avoir exploitées :

  Publiez un *Data paper* pour le faire savoir

  Partagez-les pour que d'autres puissent les utiliser…
- *Le Data paper* : meilleur moyen de valoriser ses données
  - c'est un article complémentaire d'un article de recherche
  - évalué par les pairs, indexé par les moteurs de recherche
  - citable (DOI), comptant comme un vrai article publié.
- Publier un *Data paper* est une bonne pratique de recherche
  - Répond aux enjeux de transparence et d'intégrité scientifique
  - S'inscrit dans le mouvement mondial de science ouverte.
  - Atout majeur pour reconnaître la paternité des données
  - Dans un *Data paper* : tous les contributeurs peuvent être auteurs.

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

cirad

Les bénéfices pour vous

- Réussir à publier un *Data Paper*, c'est:
  1. Bien choisir revue et entrepôt et suivre les instructions
  2. Rédiger suffisamment le *Data paper* pour qu'1 lecteur puisse
     - comprendre, interpréter et réutiliser les données
     - avoir confiance dans la rigueur de la recherche
     - mesurer l'importance et le potentiel des données.
  3. Déposer le jeu de données dans un entrepôt où :
     - vous serez identifiés en tant que créateurs des données
     - vos données seront préservées, accessibles et réutilisables.

Visibilité

Paternité

Citations

Reconnaissance

transparence
Intégrité

Attractivité

Collaborations
Projets

Data paper publié

Satisfaction de contribuer
à la science ouverte
et
aux avancées scientifiques
ou sociétales

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

cirad

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

Laurence Dedieu

laurence.dedieu@cirad.fr

Documentation et liens utiles

CoopIST – Fiches pratiques (site internet du Cirad) : https://coop-ist.cirad.fr/

- Publier un *Data paper*. Dedieu L. 2022. https://doi.org/10.18167/coopist/0057
- Déposer des données dans un entrepôt. Dedieu L. Barale M. 2020. https://doi.org/10.18167/coopist/0070
- Trouver des jeux de données. Deboin MC. 2020. https://doi.org/10.18167/coopist/0071

Choisir une revue scientifique

- Où Publier (site internet du Cirad): https://ou-publier.cirad.fr/
- Liste de revues publiant des *Data papers.* Cirad. Accès par « Publier un *Data paper* », à gauche dans rubrique « Voir aussi » : https://doi.org/10.18167/coopist/0057

Choisir un entrepôts de données

- Répertoire re3data : https://www.re3data.org/
- Sélectionner un entrepôt thématique de confiance pour la diffusion des données de recherche : note méthodologique. COSO. 2023. https://www.ouvrirlascience.fr/selectionner-un-entrepot-thematique-de-confiance-pour-la-diffusion-des-donnees-de-recherche-note-methodologique/

Rechercher des données

- Outils de recherche: Datacite Search, Google Data Search, BASE, search.dataone.org, ….

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

cirad

Exemples de *Data papers*

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

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

Exemples de *Data papers*

- 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*

- 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
- *Biodiversity Data Journal* https://bdj.pensoft.net/browse_journal_articles.php?form_name=filter_articles&sortby=0&journal_id=1&search_in_=0&section_type%5B%5D=8

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

cirad

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

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

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

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

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

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

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.

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.

Exemples de *Data papers*

- En français
  - Maillages sanitaires en France : « Camailleu-santé », un outil de cartographie en ligne pour suivre leurs évolutions dans le temps. Revue francophone sur la santé et les territoires. https://doi.org/10.4000/rfst.1939
  - 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. 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