

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

Laurence Dedieu

Délégation à l'Information Scientifique et Technique

[gtsodataweb3 : Data Papers : quand ? Comment ? Pourquoi ?](https://gtsodataweb3.org/Data-Papers-quand-Comment-Pourquoi/) - 5 juillet 2022

Concept du *Data paper*



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 BioWoody, 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



Réutilisation des données
Reproductibilité des expériences

- Contexte
- Méthodes
- Potentiel
- + lien vers les données

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 BioWoofEB, 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

Données compréhensibles

Données réutilisables

Documentation associée:

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

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

Data in Brief 15 (2017) 92–96



Data in Brief

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



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 BioWooEE, 34398, Montpellier, France
^c CIRAD, UMR AMAP, 34398, Montpellier, France

Available online 12 September 2017

Keywords:
Morpho-architectural traits
Plant architecture
Wood anatomy
Madagascar

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 information on each specimen is available in the following format:

► **Répor**

* Corresponding author.
E-mail address: loic.brancheriau@cirad.fr (L. Brancheriau).

<http://dx.doi.org/10.1016/j.dib.2017.09.004>
2352-3409/© 2017 The Authors. Published by Elsevier Inc. This is a
(<http://creativecommons.org/licenses/by/4.0/>).



RÉPUBLIQUE
FRANÇAISE

Liberté
Égalité
Fraternité



Data paper + Entrepôt = Bonne pratique



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,*}

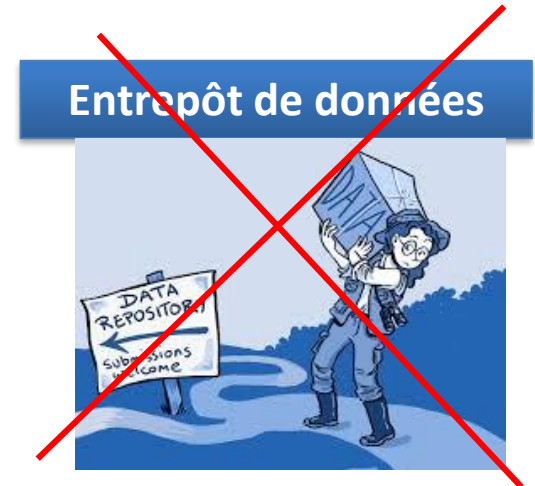
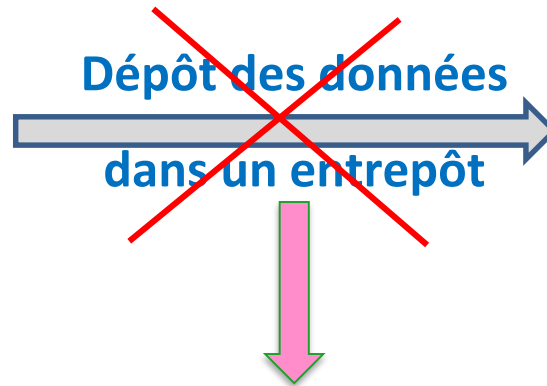
^a University of Montpellier, Dept. of Biodiversity, Ecology and Evolution, Montpellier, France
^b CIRAD, UR BioWoodF, 34398, Montpellier, France
^c CIRAD, UMRI AMAP, 34398, Montpellier, France

Données dans l'article

Keywords:
Morpho-architectural traits
Plant architecture
Wood anatomy
Madagascar

and were also collected during various networks (since 1969). 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 information is indexed from the photographs. The digitized wood slides were deposited at CIRAD.
© 2017 The Authors.

* Corresponding author.
E-mail address: loic.brancheriau@cirad.fr (L. Brancheriau).
<https://doi.org/10.1016/j.dib.2017.09.004>
2352-3409/© 2017 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY 4.0 International license (<http://creativecommons.org/licenses/by/4.0/>).



- 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

Bonnes pratiques

- Un article apportant toutes les infos sur :
 - les données et leur contexte
 - les méthodes d'obtention, équipements, ...
 - le potentiel de réutilisation
 - les fichiers, abréviations, unités, ...
- Des données accessibles dans un entrepôt
 - en format ouvert
 - avec documentation associée complémentaire
 - + description enrichie par métadonnées scientifiques
- Suivant les pratiques disciplinaires courantes pour :
 - les protocoles
 - les formats, normes, standards, métadonnées scientifiques
 - le choix de la revue et de l'entrepôt de données

Enjeux du *Data paper*

- Réutilisation des données
 - Reproductibilité des études
 - Contribuer aux avancées de la science et la société
 - Intégrité scientifique
 - = Fiabilité, honnêteté, rigueur des activités de recherche
 - = Transparence, qualité des recherche
 - Meilleure garantie contre la fraude scientifique
 - Renforcer la confiance des citoyens
- ➡ = enjeux de Science ouverte

Science ouverte

- Enjeu mondial de science ouverte
- Décliné à l'échelle nationale

[Plan d'action national](#), [PNSO](#), [COSO](#), [ANR](#), [Loi 2016](#), [décret 2021](#)

- Appliqué aux projets ANR : PGD, *Data paper* et entrepôt

Recommandation n°6 :

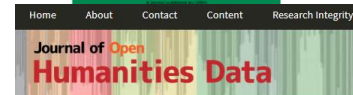
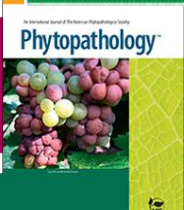
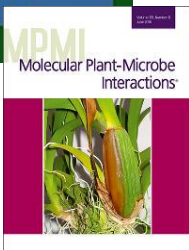
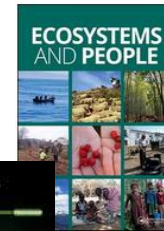
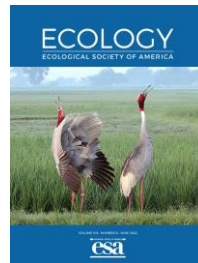
Encourager la publication de data papers (articles de données) qui permettent de décrire les jeux de données publiés, de favoriser leur réutilisation et d'être cités dans des articles scientifiques classiques.

- Enjeu pour la publication scientifique
 - Politique d'accès aux données
 - *Data Availability Statement* à intégrer dans l'article
 - Selon les revues : Dépôt des données recommandé ou obligatoire

Publier un *Data paper*

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
- Tous contributeurs à la production de données peut être co-auteur d'un *Data paper*
Informaticiens, développeurs, documentalistes, analystes,....

Revues publiant des *Data papers*



Revue 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

➤ Relecture par les pairs (*peer-reviewing*)

➤ ≠ noms : *Data paper, Data descriptor, Data article, Datasets, Data Briefs, Resource Announcements, Data Resource Profile*

Critères de choix d'une revue

- **Domaine scientifique / lectorat** (futurs utilisateurs)
- **Exigences de la revue** (Instructions aux auteurs)
 - Echelle du jeu de données : couverture géographique, temporelle ou taxonomique, intérêt pour large communauté, ...
ex : Global Ecology and Biogeography, GigaScience, Plant Journal, Nature Biotechnology, International Journal of Epidemiology
- 
 - Modèle du *Data paper* : simple, complexe, texte libre
 - 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 €

Modèles de *Data paper*

➤ Classique

Data in Brief, Geoscience Data Journal

Modèle classique



Title, Authors, Abstract

<http://www.journals.elsevier.com/data-in-brief/>

Data

Décrire l'étude, les données et les fichiers de données

Methods

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

Value of the data

Pourquoi et à qui ces données sont-elles utiles?

Comment peuvent-elles être réutilisées ?



**Accès aux données déposées dans un entrepôt
ou intégrées dans l'article**

Acknowledgements, References

Modèles de *Data paper*

- Classique

Data in Brief, Geoscience Data Journal

- 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. LIKENS¹

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

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

1 June 1963. This weekly stream water sample from 1 June 1963 to 30 May 1964. The chemistry for precipitation and stream water in the Forest are reported

4 Brook Experimental

in the journal are published

Filename

Word document, 125.5 KB

ecy1894-sup-0001-MetadataS1.doc

Zip archive, 684.2 KB

[ecy1894-sup-0002-DataS1.zip](#)

Class III. Data set status and accessibility

Status

- Latest update
- Metadata status
- Data verification

Accessibility

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

Class IV. Data structural descriptors

A. Data Set Files

- Data set Identity
- Size
- Format

B. Variable information

- Variable definition
- Units of measurement
- Data type

Modèles de *Data paper*

- Classique

Data in Brief, Geoscience Data Journal

- 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,² Gobong Choi,² Jongbum Jeon,³ Kyeongchae Cheong,² Jaeho Ko,² Halbin Xu,^{4,5} and Yong-Hwan Lee^{2,3,5,†}

¹Key Laboratory for Sustainable Forestry in Southern China, College of Forestry, Nanjing Forestry University, Jiangsu 210037, China; ²Department of Agricultural Biotechnology, Seoul National University, Seoul 08826, Korea; ³Genetics and Genomics Program in Agricultural Genomics, Seoul National University; ⁴Department of Biology and the Environment, Nanjing Forestry University; ⁵Center for Fungal Genetic Resources, Plant Genomics and Breeding Institute, and Research Institute of Agriculture 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 71x and 414x. 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

Parameter	<i>C. gloeosporioides</i>		<i>C. fruticola</i>	
	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. fruticola* Nara gc5, and *C. fruticola* 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. fruticola* 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 GFRH000000000, PRJNA471237. The draft genome and SAMN09205517 for BioSample. The genome sequence and gene models are also available from the *Colletotrichum* Fungal Genomic Platform (Choi et al. 2013b) and its sister databases described above.

Literature Cited

Alkan, M., Meng, X., Friedlander, G., Rouven, E., Sukno, S., Shuman, A., Thon, M., Fuhr, 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.

GenBank
Accession N°



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

Modèles de *Data paper*

- Classique

Data in Brief, Geoscience Data Journal

- 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


Localisation des données

- La revue impose 1 entrepôt de données (rare)
- La revue recommande des entrepôts (le + fréquent)
 - Disciplinaires, thématiques
 - Généralistes, institutionnels, ...
 - Editeurs

Localisation des données

SCIENTIFIC DATA

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

nature publishing group 

Data Descriptor

A database of common vampire bat reports

Paige Van de Vuurst¹, M. Mónica Díaz², Annia Rodríguez-San Pedro³, Juan Luis Allendes⁴, Natalie Brown⁵, Juan David Gutiérrez⁶, Heliot Zarza⁷, Stefan V. de Oliveira⁸, Elsa Cárdenas-Canales⁹, Rubén M. Barquez¹⁰ & Luis E. Escobar^{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**

Localisation des données



[http://onlinelibrary.wiley.com/journal/10.1002/\(ISSN\)2049-6060](http://onlinelibrary.wiley.com/journal/10.1002/(ISSN)2049-6060)

DATA PAPER

Geoscience Data Journal RMetS WILEY

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

Vaneeda Allken¹  | Shale Rosen¹ | Nils Olav Handegard¹  | Ketil Malde^{1,2}

¹Institute of Marine Research, Bergen, Norway

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

Correspondence

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

Funding information

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

Abstract

Developing high-performing machine learning algorithms requires large amounts of annotated data. Manual annotation of data is labour-intensive, and the cost and effort needed are an important obstacle to the development and deployment of automated analysis. In a previous work, we have shown that deep learning classifiers can successfully be trained on synthetic images and annotations. Here, we provide a curated set of fish image data and backgrounds, the necessary software tools to generate synthetic images and annotations, and annotated real datasets to test classifier performance. The dataset is constructed from images collected using the Deep Vision system during two surveys from 2017 and 2018 that targeted economically important pelagic species in the Northeast Atlantic Ocean. We annotated a total of 1,879 images, randomly selected across trawl stations from both surveys, comprising 482 images of blue whiting, 456 images of Atlantic herring, 341 images of Atlantic mackerel, 335 images of mesopelagic fishes and 265 images containing a mixture of the four categories.

KEYWORDS

data augmentation, fish dataset, machine learning, synthetic data

Approved thematic data repositories

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

[4TU.ResearchData](#)

[British Atmospheric Data Centre \(BADC\)](#)

[British Oceanographic Data Centre \(BODC\)](#)

[CISL Research Data Archive](#)

[CSIRO Data Access Portal](#)

[Deep Carbon Observatory](#)

[earth₂Observe](#)

[Environmental Information Data Centre \(EIDC\)](#)

[IEDA:EarthChem](#)

[IEDA:MGDS](#)

National Center for Atmospheric Research (NCAR), USA

[Earth Observing Lab \(EOL\)](#),

[Research Data Archive \(RDA\)](#)

[National Geoscience Data Centre \(NGDC\)](#)

[NERC Earth Observation Data Centre \(NEODC\)](#)

[NOAA National Climatic Data Center \(NCDC\)](#)

[NOAA National Oceanographic Data Center \(NODC\)](#)

[NOAA National Geophysical Data Center \(NGDC\)](#)

[PANGAEA](#)

[Polar Data Centre \(PDC\)](#)

[Figshare](#)

[Zenodo](#)

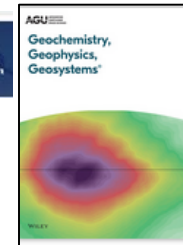
Localisation des données

- La revue impose 1 entrepôt de données (rare)
- La revue recommande des entrepôts (le + fréquent)
 - Disciplinaires, thématiques
 - Généralistes, institutionnels, ...
 - Editeurs
- La revue accepte d'intégrer les données dans l'article
 - Non conseillé
 - données seront moins accessibles, moins visibles et plus difficilement réutilisables

Modalités d'accès aux données

➤ La revue exige l'accès aux données

- au moment de la publication
- dès la soumission au moins pour les évaluateurs



➤ La revue accepte que les données soient accessibles

- après embargo
- sur demande ou collaboration



Modalités de diffusion des données

Data paper



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 BioWooE, 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 xylotèque (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 available on the website <http://www.cirad.fr/mediacentre/archivage>.

© 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/>).

Licence CC-BY

Entrepôt de données



Licence
≠

Données déposées
dans un entrepôt
de données

Licences de diffusion

Licence sur l'article dans la revue

+

Licence sur les données dans l'entrepôt

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

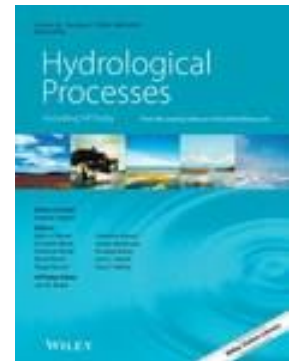
Licence CC0 ou CC-BY

cybergeo

CC0
CC-BY
CC-BY-SA



CC-BY
CC-BY-SA



Licence CC-BY-NC
Pas d'utilisation commerciale

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 !
 - <https://www.cell.com/cell/authors#policies>

Choisir une revue scientifique

- Consulter sa communauté scientifique + revues + littérature citée
- Listes disponibles
 - Liste 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

Où Publier ?

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

The screenshot shows the 'Rechercher une revue' website. A yellow speech bubble labeled 'Thème' points to the 'Thèmes / Sous-thèmes' dropdown menu, which currently shows 'Géographie et espace foncier (6)'. Another yellow speech bubble labeled 'Type d'article' points to the 'Types d'articles' dropdown menu, which currently shows 'Data papers (6)'. The interface includes search filters for 'Titre de la revue', 'Mots ou expression', 'Libre accès', 'Langues', 'Notoriété', and 'Types d'articles'. A list of journals is visible on the right, including Cybergeos, Earth System Science Data, Geoscience Data Journal, Global Ecology and Biogeography, One Ecosystem, and Open Geospatial Data, Software and Standards.

Publier un *Data paper*

1. Anticiper le choix de la revue et de l'entrepôt
 - modèle de *Data paper*
 - exigences à suivre : métadonnées, normes, formats
2. Documenter régulièrement ses protocoles, données, métadonnées, ... au fur et à mesure des recherches
 - selon les pratiques de sa discipline
3. Rédiger le *Data paper* avec suffisamment d'infos pour qu'un lecteur/relecteur puisse
 - comprendre et reproduire l'étude
 - avoir confiance dans la rigueur de l'étude
 - constater la fiabilité et la qualité des données
 - évaluer l'importance et le potentiel des données
 - interpréter et réutiliser les données

Publier un *Data paper*

4. Déposer ses données dans un entrepôt

→ selon format, normes, métadonnées demandés par la revue et l'entrepôt

→ avec la documentation associée

+ licence de diffusion

→ à condition d'avoir le droit de les publier

→ avec l'accord de tous les contributeurs

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

Publier un *Data paper*

6. Si votre Data paper est complet et vos données FAIR
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
- Permet à d'autres de découvrir votre travail et les données que vous produisez
- Vous rend visible dans votre communauté scientifique
- Montre la transparence de vos recherches

Publier un *Data paper*



Valorisation

Crédibilité

Notoriété

Reconnaissance

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



Attractivité

Collaborations

Nouveaux projets

Mes données
sont préservées



Mon *Data paper* est publié. Il montre que:

j'applique des bonnes pratiques de recherche,

je produis des données FAIR

je contribue à la connaissance scientifique

je réponds aux enjeux d'intégrité scientifique

et de science ouverte.

Merci de votre attention

Laurence Dedieu

laurence.dedieu@cirad.fr

Fiche CoopIST: Publier un Data paper. Dedieu Laurence. 2022.

<https://doi.org/10.18167/coopist/0057>