

GBIF and data citations

Daniel Noesgaard | Science Communications Officer



Data Citation Community of Practice workshop, 8 June 2021





WHAT IS GBIF?

- International network and data infrastructure funded by the world's governments and aimed at providing anyone, anywhere, open access to data about all types of life on Earth
- 1.7 B records of organisms occurring in time and space, originating from 36,000 different datasets
- Data mediated by GBIF was used and cited in 1000+ scientific publications published in 2020, with citations increasing ~30% every year.

DOWNLOADING DATA FROM GBIF

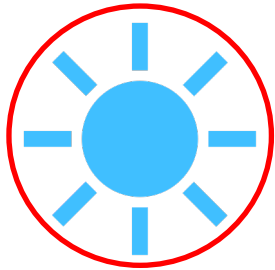


Photo: *Agalychnis callidryas* (Cope, 1862) observed in Mexico by Luis Guillermo (CC BY-NC 4.0)

DATA DOWNLOADS

Number of records

Date

Filters

Contributing datasets

Download

Rerun

Recommended citation

DOI

DOWNLOAD | 7 JUNE 2021

35,120 occurrences included in download

DOI: [10.15468/dl.mavj4d](https://doi.org/10.15468/dl.mavj4d) DOWNLOAD

FILTER APPLIED 7 JUNE 2021 RERUN QUERY

Citation: GBIF.org (07 June 2021) GBIF Occurrence Download <https://doi.org/10.15468/dl.mavj4d>
License: [CC BY-NC 4.0](#)
File: 2 MB Simple
Involved datasets: [126](#)
Make sure to read the [data user agreement](#) and [citation guidelines](#).

And API

- Country or area: Mexico
- Month: March • April • May
- Occurrence status: present
- Scientific name: Anura

INCLUDES RECORDS FROM 126 DATASETS DOWNLOAD AS TSV

UCM Amphibian and Reptile Collection (Arctos)	506
MSB Amphibian and Reptile Collection (Arctos)	133

LINKING USE TO DATA

... eas. *International Journal of Climatology*, 37(12), 4302-4303. <https://doi.org/10.1002/joc.5086>

Forsythe, A., Vogan, A., & Xu, J. (2016). Genetic and environmental influences on the germination of basidiospores in the *Cryptococcus neoformans* species complex. *Scientific Reports*, 6(1), 1–12. <https://doi.org/10.1038/srep33828>

GBIF. (2018). *GBIF occurrence, download*. <https://doi.org/10.15468/dl.ofjxcm> and <https://doi.org/10.15468/dl.gmwftf>

Gorelick, N., Hancher, M., Dixon, M., Fluhmann, S., Kowaluk, D., & Moore, R. (2017). Google earth engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, 202, 18–27. <https://doi.org/10.1016/j.rse.2017.06.031>

Granados, D. P., & Castañeda, E. (2006). Influence of climatic conditions on the isolation of members of the *Cryptococcus neoformans* species complex from trees in Cuba from 1992–2004. *FEMS Yeast Research*, 6(4), 636–641. <https://doi.org/10.1111/j.1567-1364.2006.00090.x>

Wang, F., Colgan, M. P., Smith, D., & ... (2018). ...

455 occurrences included in download

1 CITATION

Includes records from 3 datasets: Westmead Mycology Culture Collection (440)

2,346 OCCURRENCES | 9 CITATIONS

Published by Atlas of Living Australia

Publication date: September 26, 2019

The phylogeny of insects in the data-driven era

Journal article | Peer-reviewed

DOI: 10.15468/dl.ofjxcm

Paper is published citing a GBIF download DOI

The citation is logged against the download, shown on the landing page as 1 citation

Unpacking the citations to the individual datasets and publishers in GBIF

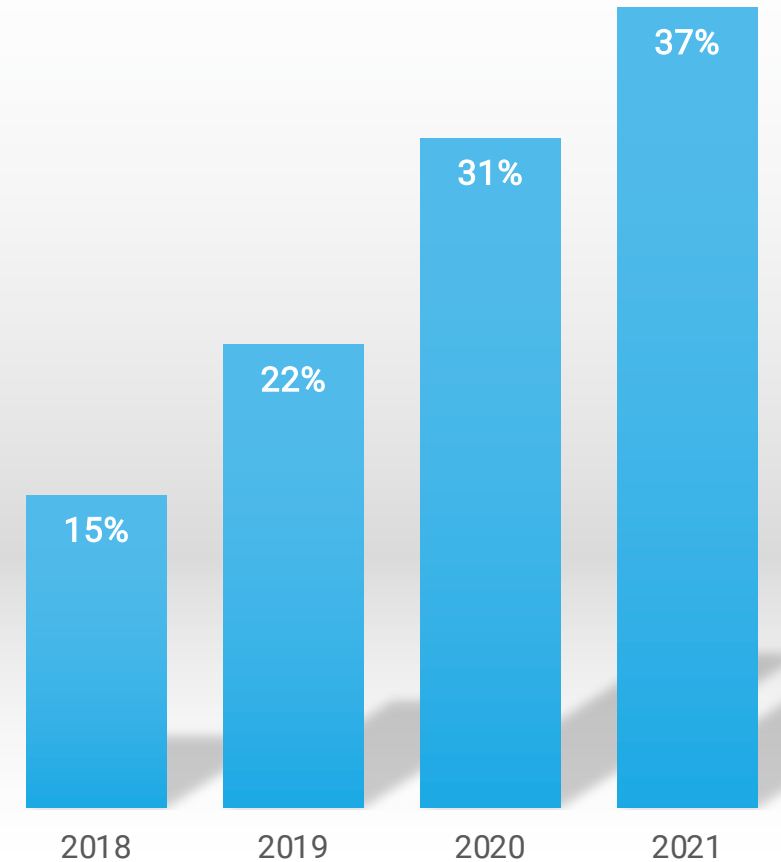
Data publishers can assess how much their data is being used and by which publications

USE OF GBIF DATA



- 1,700 papers logged citing specific data downloads
- In total, ~12,000 specific downloads cited
- More than one in three papers now cite GBIF using DOIs

DOI citations



CHALLENGES

When there's no (suitable) DOI to cite



Photo: *Gorilla beringei* subsp. *beringei* observed in Rwanda by Laura Keene (CC BY-NC 4.0)

NO SUITABLE DOI

- downloads that have been filtered/reduced significantly, or
- data accessed in a cloud computing environment, or
- data obtained by any means for which no DOI was assigned (e.g. third-party tools accessing the GBIF search API)



THE SOLUTION

A new service: derived datasets

To mimick what we do with downloads but based on data that has already been extracted for which no suitable DOI exists

User must extract list of contributing datasets with counts of records and provide a URL of where the dataset can be accessed

In return they get a DOI to cite which can be used to credit data publishers

The screenshot shows the 'Derived dataset' registration page on the GBIF website. At the top, there are navigation links: 'TOOLS', 'MY DERIVED DATASETS', 'REGISTER', and 'ABOUT'. The main heading is 'Derived dataset'. Below this, there is a section titled 'REGISTER NEW DERIVED DATASET'. The form contains the following fields and content:

- Title ***: Filtered dataset of *Cryptococcus neoformans* and *Cryptococcus gattii*
- URL of where derived dataset can be accessed ***: <https://onlinelibrary.wiley.com/action/downloadSupplement?doi=10.1111%2Frisa.13410&file=risa13410-sup-00>
- Or enter dataset keys/DOIs and occurrence counts here**: A table with four rows, each containing a dataset key, an occurrence count, and a trash icon or a plus sign.

Dataset key	Occurrence count	Action
68fc81fa-d63e-4cf1-a0	189	Trash icon
10.15468/giuq7w	12	Trash icon
10.15468/cndomv	5	Trash icon
10.15468/giuq7w	134	Plus sign

Below the table is a rich text editor with a toolbar containing icons for bold, italic, highlight, quote, list, link, image, and other text formatting options. The text in the editor reads: 'Data accessed through rgbif and used in Alaniz AJ, Carvajal JG, Carvajal MA, Cogliati M and Vergara PM (2019) Spatial Quantification of the Population Exposed to *Cryptococcus neoformans* and *Cryptococcus gattii* Species Complexes in Europe: Estimating the Immunocompetent and HIV/AIDS Patients Under Risk. Risk Analysis. Wiley. 40(2): 524-532. Available at: <https://doi.org/10.1111/risa.13410>

#CITETHEDO1

Daniel Noesgaard
dnoesgaard@gbif.org

