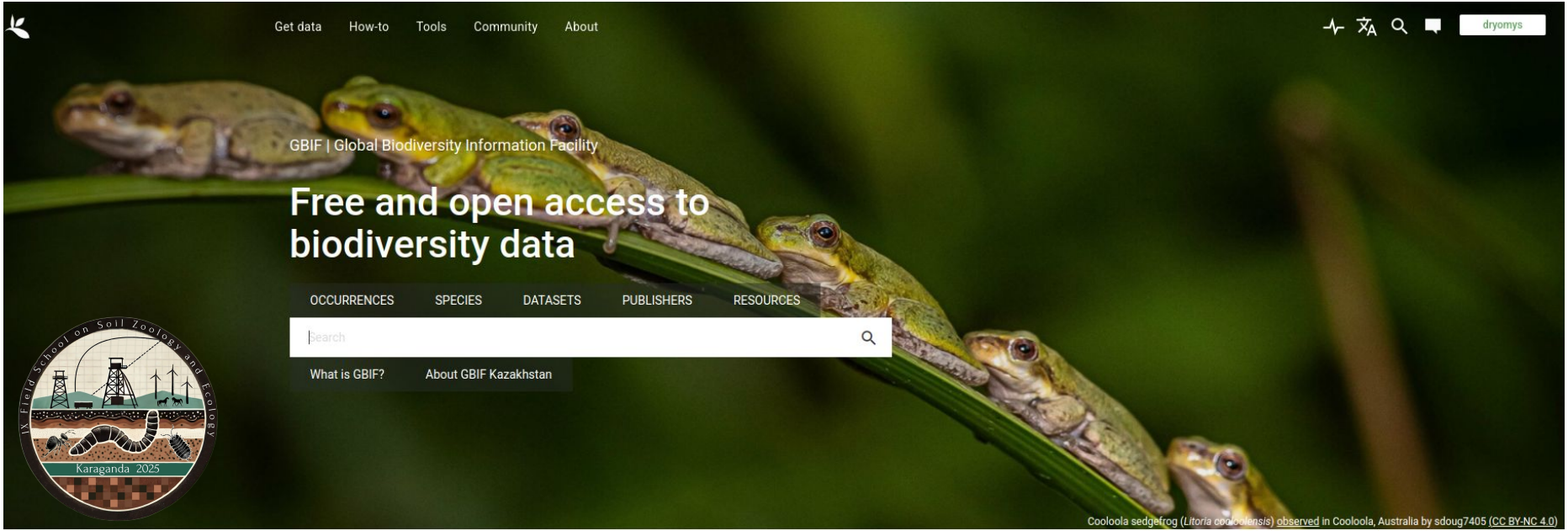
# Global Biodiversity Information Facility




# Outside and inside

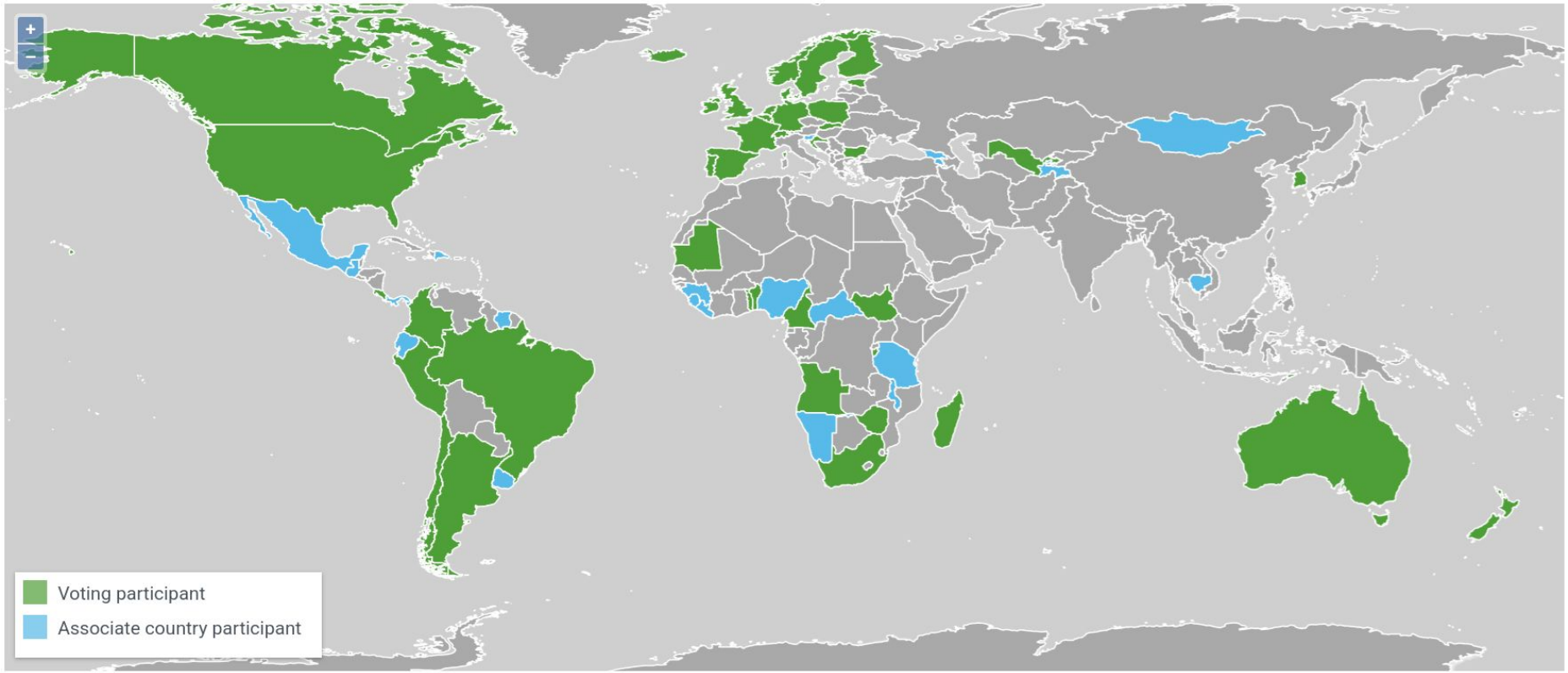
## Снаружи и внутри

Natalya Ivanova

# Part 1. OUTSIDE

**GBIF is an 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.**

# The GBIF Network




45 VOTING PARTICIPANTS | 21 ASSOCIATE COUNTRY PARTICIPANTS | 42 OTHER ASSOCIATE PARTICIPANTS | 2,596 PUBLISHERS

https://www.gbif.org/the-gbif-network

PUBLISHER | SINCE JANUARY 26, 2023

# Karaganda Buketov University

ABOUT METRICS HOME PAGE

15,438 OCCURRENCES | 78,026 HOSTED OCCURRENCES | 21 DATASETS | 85 CITATIONS

**Description:** The University provides educational and scientific services, releases research in the field of assessment of biological diversity of Kazakhstan: both of vascular and non-vascular plants, vertebrates and invertebrates animals; curated collections of seed bank, herbarium, ex situ; holds a museum of nature with collection of invertebrate and vertebrate animals.

**Endorsed by:** Participant Node Managers Committee

**Installations:** Karaganda Biodiversity Information Network

**Administrative contact:** Serikbolat Talzhanov

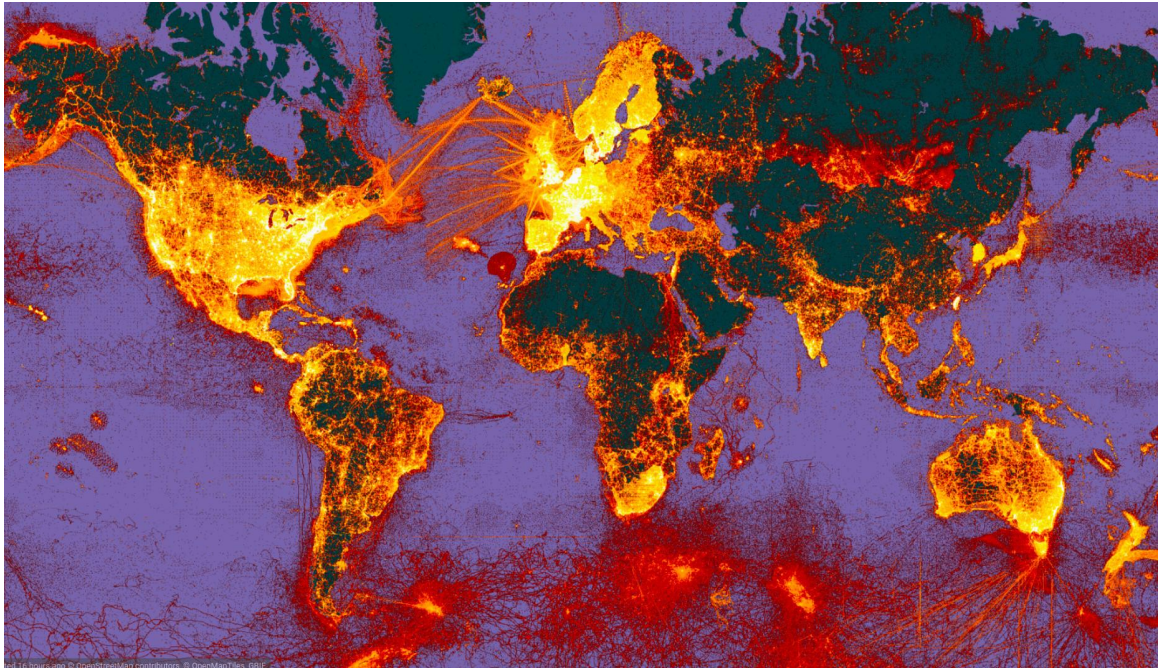
**Technical contact:** Maxim Shashkov

**Country or area:** Kazakhstan

**Hosting:** 26 datasets ( 6 publishers• 2 countries )

Download activity report

# The Global Biodiversity Information Facility



**3,144,592,728**
species occurrences

**115,669**
datasets

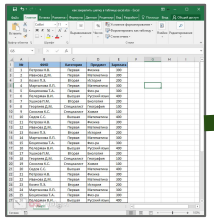
**2,537**
publishers

**23,254** peer-reviewed papers using data

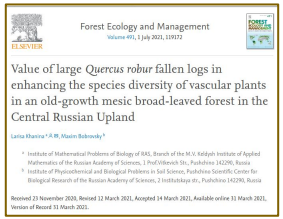
# Main GBIF data sources

- Natural-history collections
- Leaving collections
- Citizen science
- Field surveys
- Camera-trap data
- Species occurrences extracted from articles (books, reports, PhD theses, etc.)
- DNA data

# Darwin Core: how to prepare table with my data for publishing through GBIF

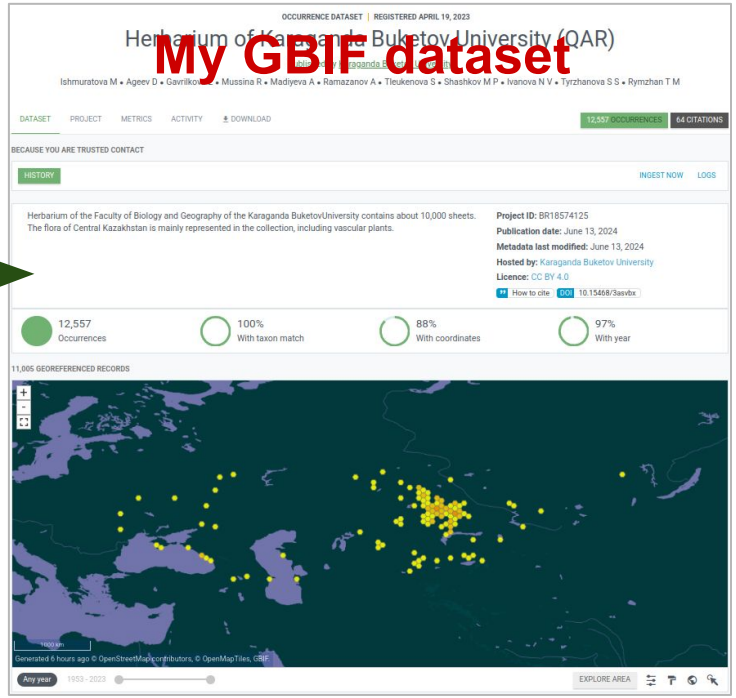



My data



Forest Ecology and Management

Value of large *Quercus robur* fallen logs in enhancing the species diversity of vascular plants in an old-growth mesic broad-leaved forest in the Central Russian Upland




doi:10.1371/journal.pone.0029715

## Required information about the occurrence

- A unique identifier for the occurrence (occurrenceID)
- Basis of the record (basisOfRecord)
- The full scientific name of the organism (scientificName)
- The date or date interval (eventDate)

## Strongly recommended information: taxonomy

- The taxonomic rank of the supplied scientific name (taxonRank)
- The full scientific name specifying the kingdom (kingdom)

## Strongly recommended information: abundance

- To record the quantity of a species occurrence, e.g. as the number of individuals, percentage of vegetation coverage, or the biomass (individualCount, organismQuantity & organismQuantityType)

## Strongly recommended information: location

- A two-letter standard abbreviation for the country of the occurrence locality (countryCode)
- The geographic latitude or longitude, respectively, in decimal degrees (decimalLatitude & decimalLongitude)
- The coordinate system (geodeticDatum)
- The horizontal distance from the given decimalLatitude and decimalLongitude in meters, describing the smallest circle containing the whole of the Location (coordinateUncertaintyInMeters)

OCCURRENCE | 31 AUGUST 1969

# *Carabus granulatus* Linnaeus, 1758

**Granulate Ground Beetle** In English **Collected in Luxembourg**

Animalia › Arthropoda › Insecta › Coleoptera › Carabidae › *Carabus*

DETAILS

**Species:** Carabus granulatus Linnaeus, 1758
**Location:** Europe › Luxembourg
**Basis of record:** Preserved specimen

**Event ID:** DSS0043900001ZJP
**Dataset:** Collections and observation data National Museum of Natural History Luxembourg
**Publisher:** National Museum of Natural History, Luxembourg
**Issues:** Country derived from coordinates | Continent derived from coordinates


https://www.gbif.org/occurrence/5209638409

Get data How-to Tools Community About Login

OCCURRENCE DATASET | REGISTERED JANUARY 20, 2025

# Records of White-headed Duck (Oxyura leucocephala) in Central Asia

Published by Karaganda Buketov University

Shishkina Y

DATASET METRICS ACTIVITY DOWNLOAD 153 OCCURRENCES

This dataset compiles records of White-headed Duck (Oxyura leucocephala) observations across Central Asia, with a primary focus on Kazakhstan and neighboring regions. It provides information on sighting locations, population estimates, sex and age distribution, and references to scientific publications supporting these records. The data are derived from literature sources. This dataset is valuable for monitoring population trends, tracking changes in abundance and distribution, and assessing habitat conditions for this species.

**Publication date:** January 20, 2025

**Metadata last modified:** January 20, 2025

**Hosted by:** Karaganda Buketov University

**Licence:** CC BY 4.0

How to cite DOI 10.15468/ruvr34

153 Occurrences

100% With taxon match

69% With coordinates

90% With year

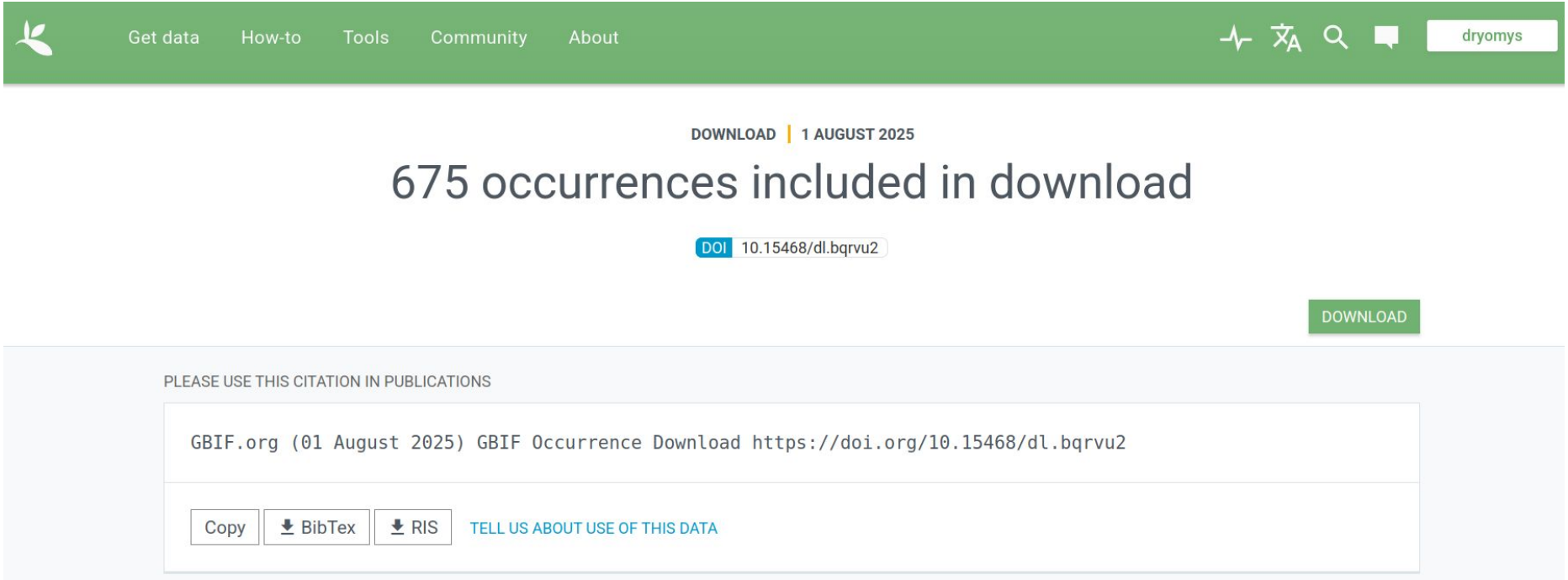
https://www.gbif.org/dataset/78f7dbd8-859a-4d46-a278-85439d28db49

## White-headed Duck: occurrences from Kazakhstan

https://doi.org/10.15468/dl.bqrvu2

## White-headed Duck in Kazakhstan: GBIF data sources

| | | |
|---|---|---|
| ☐ | EOD – eBird Observation Dataset | 353 |
| ☐ | Records of White-headed Duck (Oxyura leucoc... | 127 |
| ☐ | iNaturalist Research-grade Observations | 79 |
| ☐ | Hatikka.fi observations | 45 |
| ☐ | Observation.org, Nature data from around the ... | 39 |
| ☐ | RU-BIRDS.RU, Birds observations database fro... | 20 |
| ☐ | INSDC Sequences | 5 |
| ☐ | Museum of Comparative Zoology, Harvard Unive... | 3 |
| ☐ | Birds and Mammals Collections of the Zoologica... | 2 |
| ☐ | NABU\|naturgucker | 1 |




**Citation guidelines: https://www.gbif.org/citation-guidelines**

OCCURRENCE DATASET | REGISTERED OCTOBER 13, 2022

# Distribution of marsh frogs (Pelophylax ridibundus complex) in Kazakhstan

Published by Institute of Zoology of the Republic of Kazakhstan

Dujsebayeva T • Kaptyonkina A • Arifulova I • Ualiyeva D • Akhmedenov K • Ivanov A • Khromov V • Krainyuk V • Sarzhanov F • Tarasovskaya N • Titov S • Timoshenko A • Ermakov O • Malakhov D • Starikov S



The presented data are the result of generalization and reconciliation of literary, museum and archival information on the distribution of lake frogs of the P. ridibundus complex in Kazakhstan, and new data were obtained during field work in 2021-2022. Based on the collected material, a database has been compiled for all the frog finds known today for the period from the end of the XX century to the present.

**Publication date:** April 9, 2023

**Metadata last modified:** April 9, 2023

**Hosted by:** GBIF Secretariat

**Licence:** CC BY 4.0

How to cite DOI 10.15468/et4dus

838 Occurrences

100% With taxon match

99.6% With coordinates

94% With

Global Change Biology

RESEARCH ARTICLE | Open Access

## Forecasting potential invaders to prevent future biological invasions worldwide

Arman N. Pili, Boris Leroy, John G. Measey, Jules E. Farquhar, Adam Toomes, Phillip Cassey, Sebastian Chekunov, Matthias Grenié, Dylan van Winkel, Lisa Maria, Mae Lowe L. Diesmos, Arvin C. Diesmos, Damaris Zurell, Franck Courchamp, David G. Chapple

First published: 15 July 2024 | https://doi.org/10.1111/gcb.17399 |

ELSEVIER

Biological Conservation

Volume 302, February 2025, 110887

## Vertebrates in trade that pose high invasion risk to the United States

Wesley M. Daniel [a 1], Helen R. Sofaer [b 1], Catherine S. Jarnevich [c 1], Richard A. Erickson [d 1], Brett A. DeGregorio [e], Peder S. Engelstad [f], Jonathan Freedman [g 2], Susan Canavan [h], E.M. Dean [a], Michael J. Adams [i], Charmayne L. Anderson [j], Mindy Barnett [k], Marybeth K. Brey [d], Kyle J. Brumm [l], Matthew S. Bunting [m], Emily Caffrey [n], Laura Cardador [o], Jacoby Carter [p], Phillip Cassey [q], Duane C. Chapman [r]...Deah Lieurance [ar]

Show more

Add to Mendeley | Share | Cite

https://doi.org/10.1016/j.biocon.2024.110887 Get rights and content

nature communications

Explore content | About the journal | Publish with us

nature > nature communications > articles > article

Article | Open access | Published: 30 April 2025

## Modelling the species-area relationship using extreme value theory

Luís Borda-de-Água, M. Manuela Neves, Luise Quoss, Stephen P. Hubbell, Filipe S. Dias & Henrique M. Pereira

*Nature Communications* **16**, Article number: 4045 (2025) | Cite this article

**5113** Accesses | **52** Altmetric | Metrics

MOLECULAR ECOLOGY

ORIGINAL ARTICLE | Open Access

## A Latitudinal Gradient of Reference Genomes

Ethan B. Linck, Carlos Daniel Cadena

First published: 14 October 2024 | https://doi.org/10.1111/mec.17551 | Citations: 6

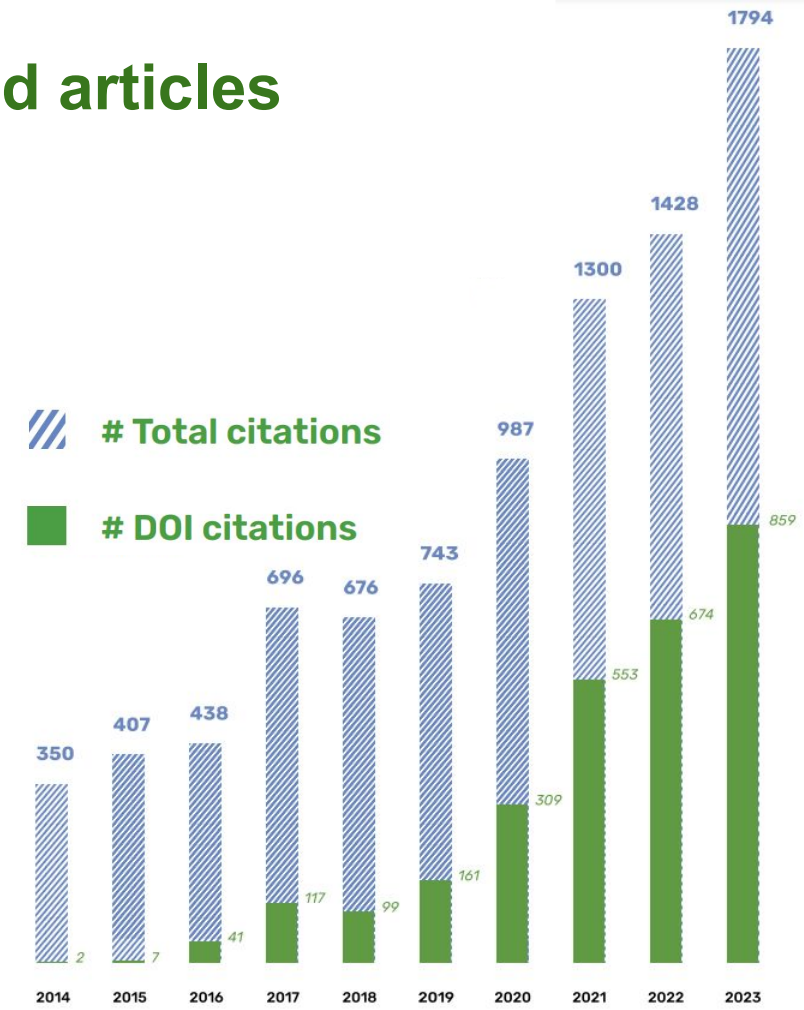
**Handling Editor:** Martin Kardos

**Funding:** The authors received no specific funding for this work.

# Annual number of peer-reviewed articles using GBIF-mediated data

| | |
|---|---|
| Agriculture | 744 |
| Biodiversity science | 560 |
| Biogeography | 1,012 |
| Citizen science | 170 |
| Climate change | 2,941 |
| Conservation | 1,905 |
| Data management | 72 |
| Data paper | 96 |
| Ecology | 3,748 |
| Ecosystem services | 111 |
| Evolution | 1,279 |
| Freshwater | 179 |
| Human health | 553 |
| Invasives | 1,836 |
| Marine | 379 |
| Phylogenetics | 725 |
| Species distributions | 1,091 |
| Taxonomy | 765 |


https://doi.org/10.35035/d9pk-1162



# Outcomes

- GBIF is the largest source of species occurrences for distribution modelling.
- The data's origin, authorship, and licences are clear.
- Data have global identifiers.
- All data are available free of charge.
- GBIF supports stable access to data.

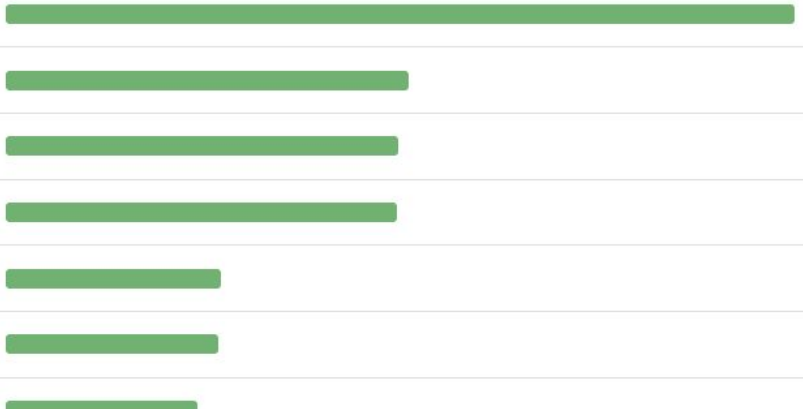
# Part 1. INSIDE

***Let's start using GBIF data***


- Taxonomic coverage
- Data quality
  ~ Gaps in GBIF Backbone
  ~ GBIF Issues & Flags

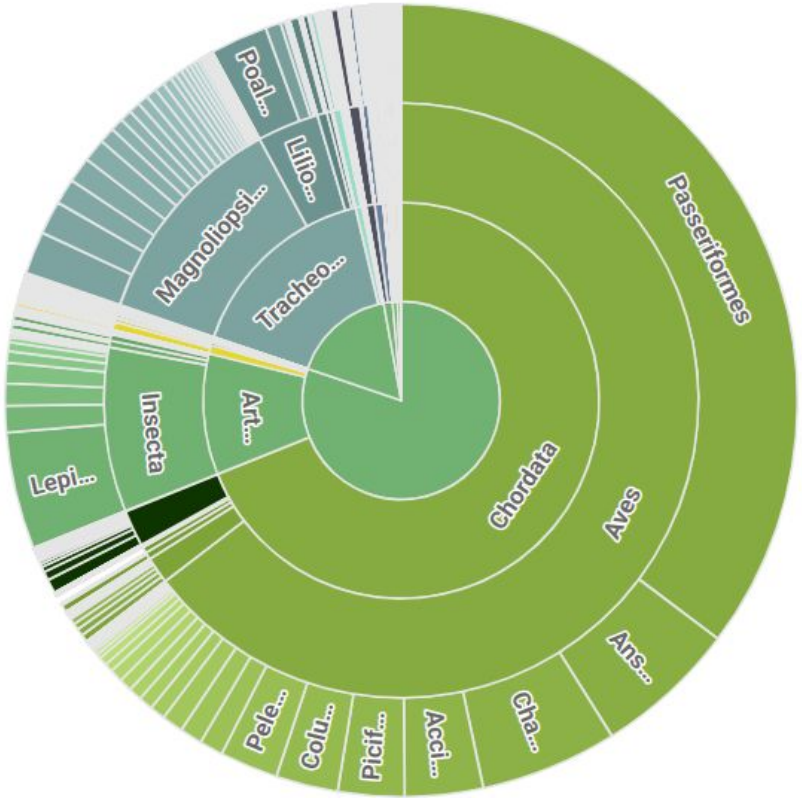
# 10 largest GBIF datasets

| Dataset | Count |
|---|---|
| EOD – eBird Observation Dataset | 1,512,208,412 |
| iNaturalist Research-grade Observations | 122,346,432 |
| Observation.org, Nature data from around the World | 114,762,164 |
| Artportalen | 113,397,914 |
| DOF/BirdLife Denmark - Observations from DOFbasen | 35,948,083 |
| Norwegian Species Observation Service | 35,279,706 |
| Macro-moth distribution records for the UK from Bu... | 30,822,537 |
| Southern African Bird Atlas Project 2 | 21,979,397 |
| INPN - Données flore des CBN agrégées par la FCBN | 20,999,334 |
| NABU\|naturgucker | 16,762,198 |



**Citizen science:**
**~2/3 of GBIF data**

# Taxonomic scope of the GBIF data: world of birds

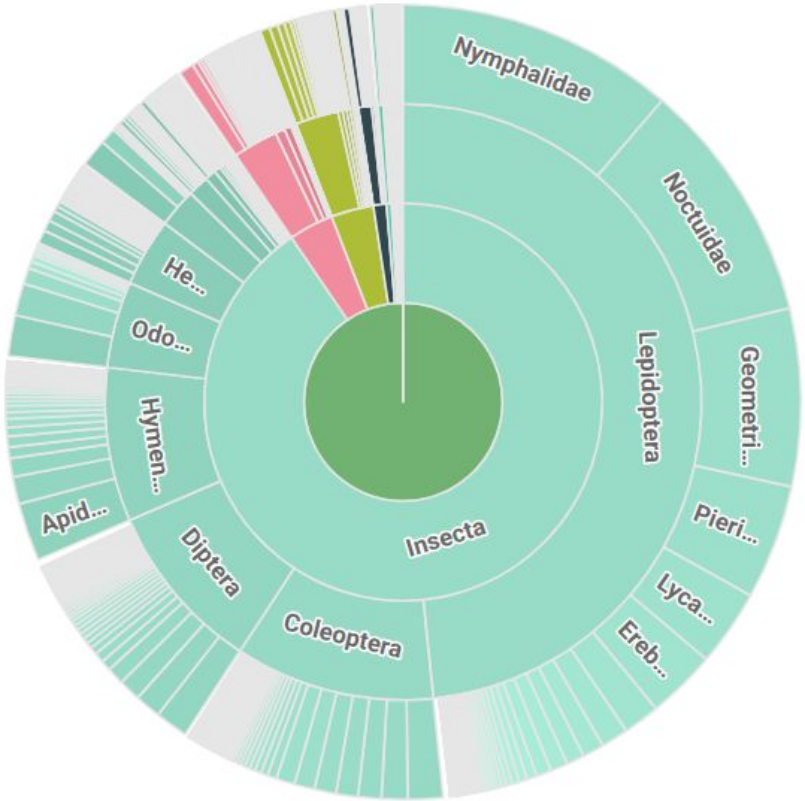



GBIF:
3,145,943,737 occurrences

Aves:
2,026,326,060 occurrences

Arthropoda:
309,646,862 occurrences (9.8%)

# Taxonomic scope of insects: ~50% Lepidoptera

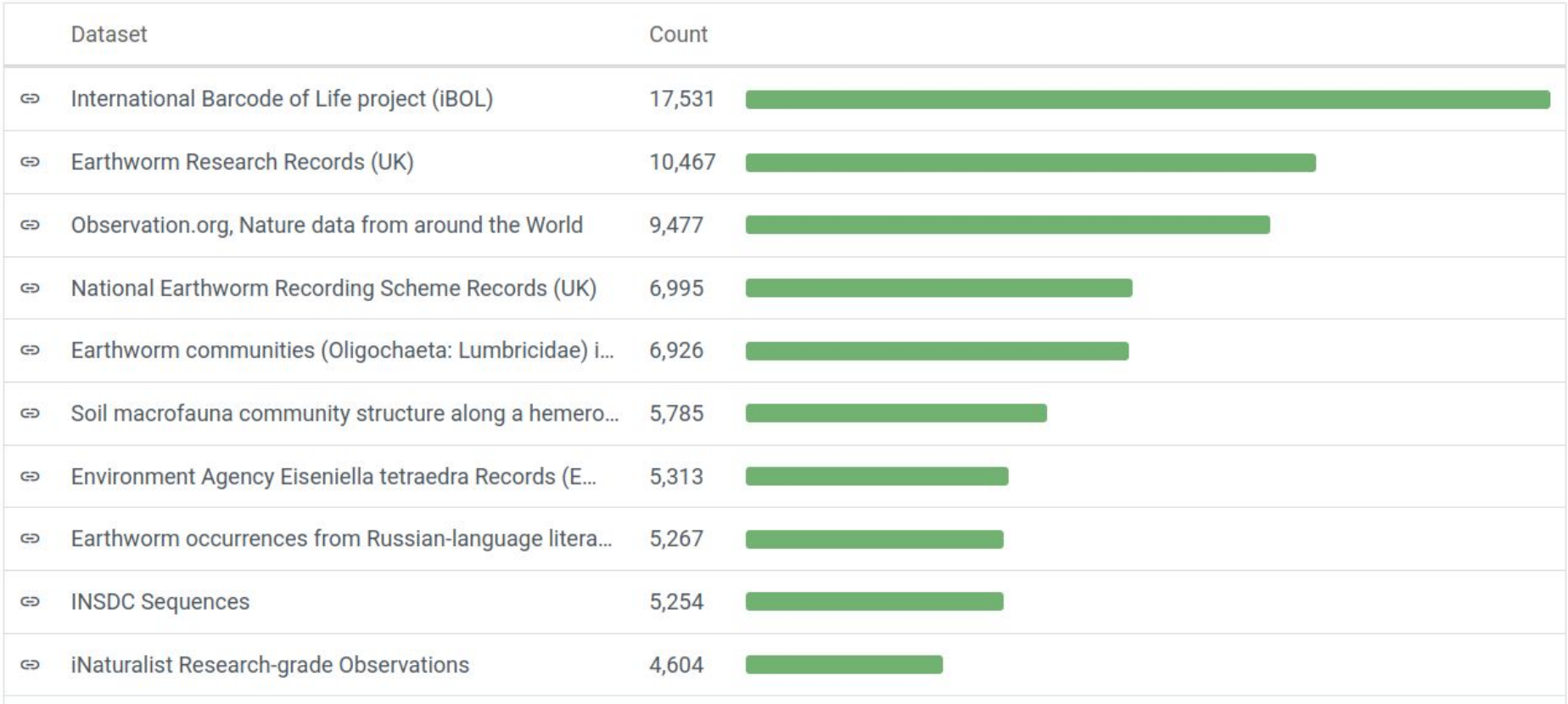



Insects:
280,594,300 occurrences

Lepidoptera:
149,995,720 occurrences

Coleoptera:
33,615,422 occurrences

# Family Lumbricidae: 166,379 occurrences



| Dataset | Count |
|---|---|
| International Barcode of Life project (iBOL) | 17,531 |
| Earthworm Research Records (UK) | 10,467 |
| Observation.org, Nature data from around the World | 9,477 |
| National Earthworm Recording Scheme Records (UK) | 6,995 |
| Earthworm communities (Oligochaeta: Lumbricidae) i... | 6,926 |
| Soil macrofauna community structure along a hemero... | 5,785 |
| Environment Agency Eiseniella tetraedra Records (E... | 5,313 |
| Earthworm occurrences from Russian-language litera... | 5,267 |
| INSDC Sequences | 5,254 |
| iNaturalist Research-grade Observations | 4,604 |

# Family Carabidae: 4,419,760 occurrences

| Dataset | Count |
|---|---|
| Ground Beetle Recording Scheme - data verified via... | 423,207 |
| Swiss National Coleoptera Databank | 386,051 |
| iNaturalist Research-grade Observations | 316,201 |
| Observation.org, Nature data from around the World | 194,503 |
| Artportalen | 185,318 |
| University of Alberta E. H. Strickland Entomologic... | 170,860 |
| VIT-Coleopterotheca (The Natural History Museum of... | 82,129 |
| NEON Biorepository Carabid Collection (Pinned Vouc... | 80,302 |
| CAS Entomology (ENT) | 76,782 |
| International Barcode of Life project (iBOL) | 66,459 |

# Order Araneae: 7,250,317 occurrences

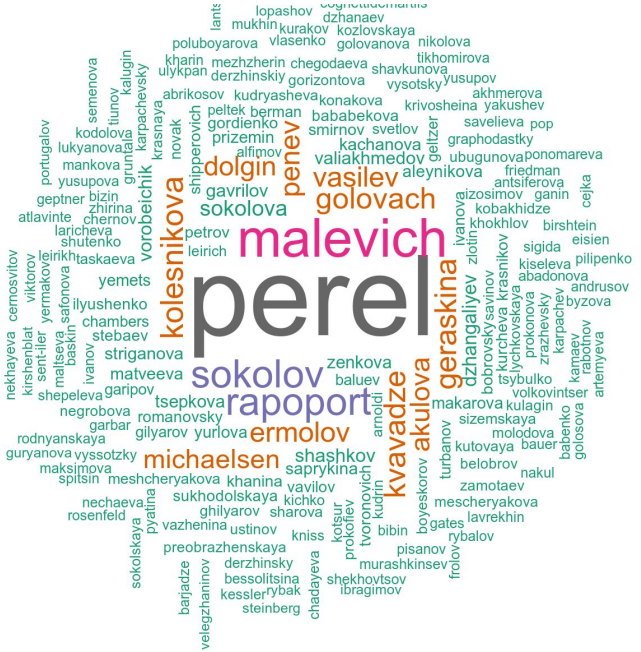
| Dataset | Count |
|---|---|
| iNaturalist Research-grade Observations | 1,969,924 |
| Observation.org, Nature data from around the World | 566,116 |
| ARABEL - Arachnologia Belgica | 291,402 |
| Swiss National Spider Databank | 253,932 |
| International Barcode of Life project (iBOL) | 192,953 |
| Artportalen | 189,432 |
| The Environmental Sample Collection of Spiders at ... | 159,152 |
| NABU\|naturgucker | 149,839 |
| Arachnida and Myriapoda (Luomus) | 129,463 |
| Inventaire national des araignées de France métrop... | 122,792 |

OCCURRENCE DATASET | REGISTERED MAY 4, 2023

# Earthworm occurrences from Russian-language literature

Published by Institute of Mathematical Problems of Biology RAS – the Branch of Keldysh Institute of Applied Mathematics of Russian Academy of Sciences

Shashkov M P • Ivanova N V • Ermolov S A



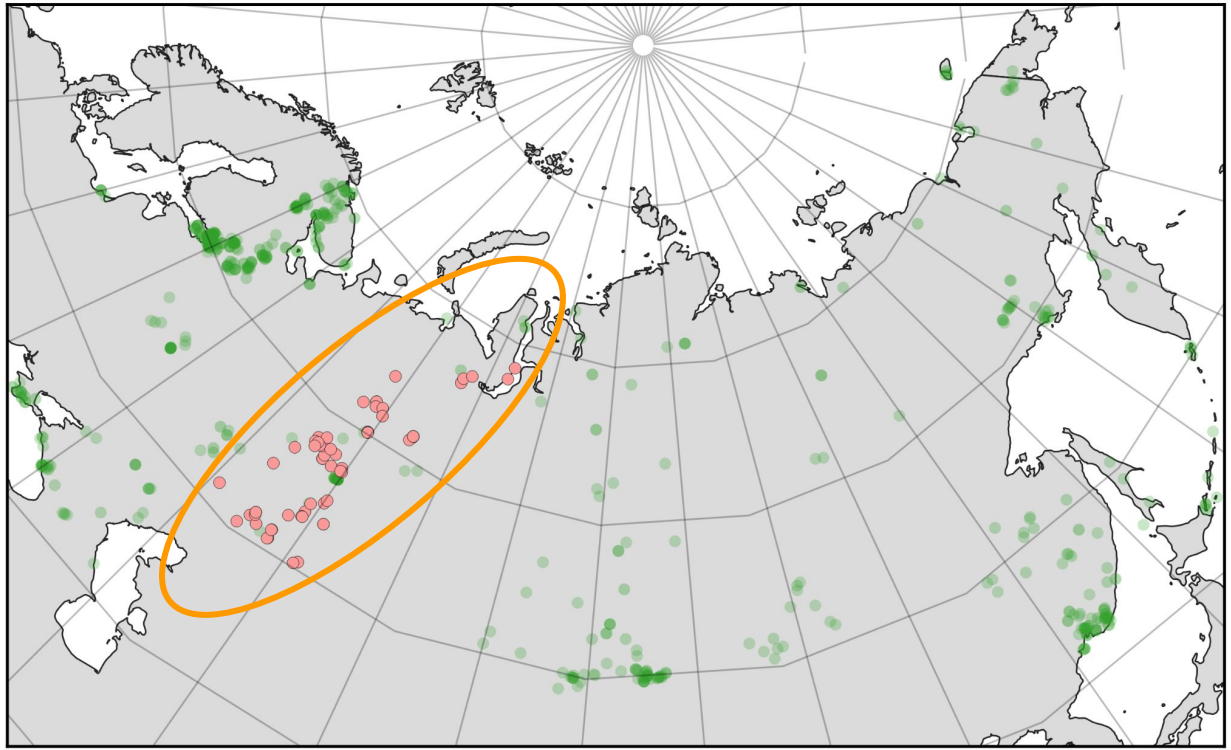

- 5304 species occurrences from 159 Russian-language scientific publications
- Data collected from 1868 to 2022
- Data cover territory of 27 countries

Российский научный фонд
#23-24-00112

See dataset

# The mobilization techniques for primary data on biodiversity: from literature legacy to digital lake

**August 2025:**
7415 mined occurrences
71 digitised articles
291 volunteers



РНФ | Российский научный фонд
проект № 24-24-00460

# Outcomes

- Most likely, GBIF will be incomplete for your species' occurrences.
- You will need additional data sources (non-digitised collections, literature, etc.).

# GBIF Backbone Taxonomy

It is a single, synthetic management classification with the goal of covering all names GBIF is dealing with. It's the taxonomic backbone that allows GBIF to integrate name-based information from different resources.


https://doi.org/10.15468/39omei

# Earthworm project: GBIF Backbone Update

Tamara Vsevolodova-Perel (1930-2018)

Range and Regularities in the Distribution of Earthworms of the USSR Fauna. Perel, 1979 (DOI: 10.15468/qp8aqx) **+ 32 names to GBIF Backbone**

The Earthworms of the Fauna of Russia. Perel, 1997 (DOI: 10.15468/d3jg8d) **+ 5 names to GBIF Backbone**

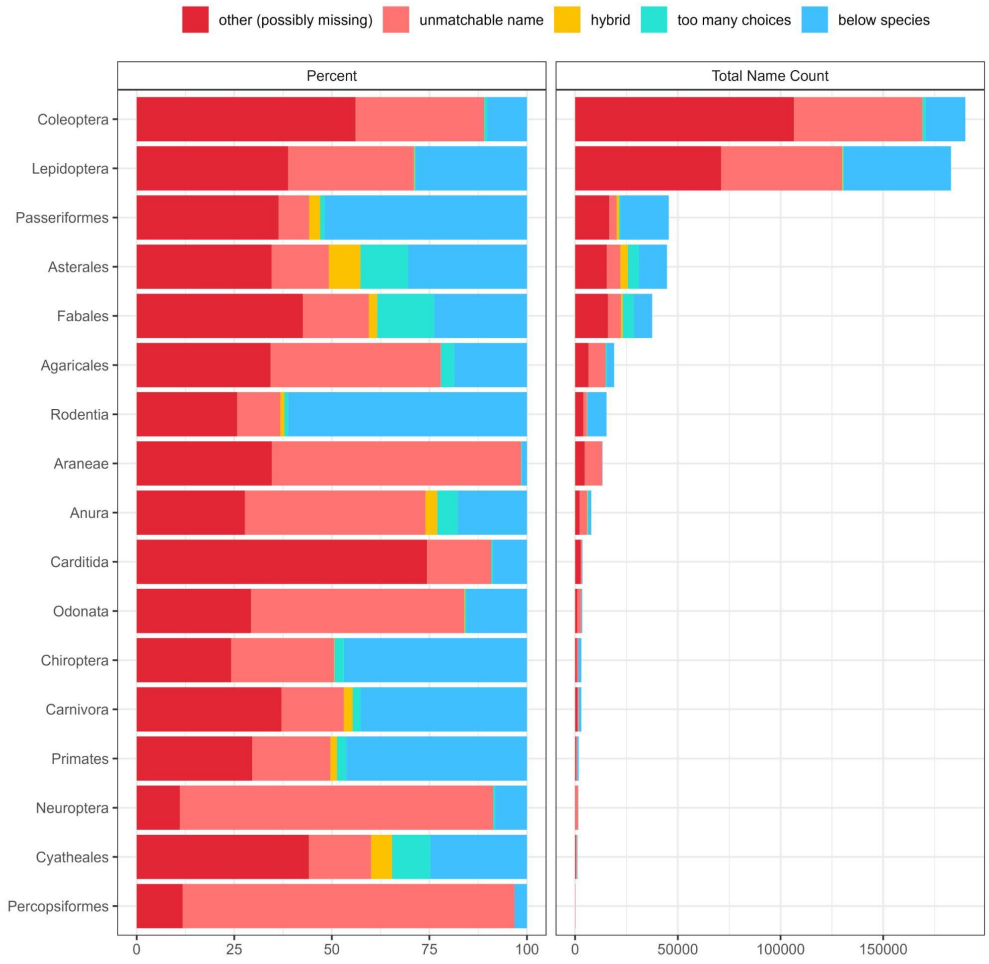
Synonyms for the Earthworms of the USSR Fauna. Perel, 1979 (DOI: 10.15468/wwanyr)

# GBIF Backbone vs. checklist (Misirlioğlu et al. 2023): mismatches found

**Missed in GBIF Backbone:**
*Dendrobaena tellermanica* Perel, 1966
*Eisenia pallida* Malevič, 1956
*Eisenia bashkirica* (Malevič, 1950)
*Cernosvitovia sturanyi* Rosa, 1895

**Different spellings of the same species are considered separate accepted names:**
*Dendrobaena mariupolienis* Wyssotzky, 1898
*Dendrobaena mariupoliensis*

https://doi.org/10.3897/BDJ.12.e130897



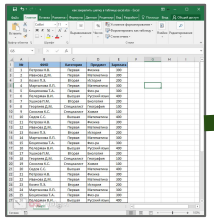

# Data Gaps in the GBIF Backbone Taxonomy

*Part of:* Waller JT (2022) Finding Data Gaps in the GBIF Backbone Taxonomy. Biodiversity Information Science and Standards 6: e91312. https://doi.org/10.3897/biss.6.91312

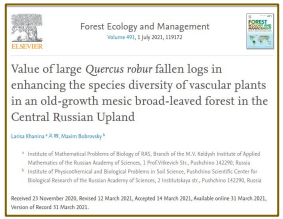
# Outcomes

- Carefully check the taxonomic data

# Darwin Core: how to prepare table with my data for publishing through GBIF

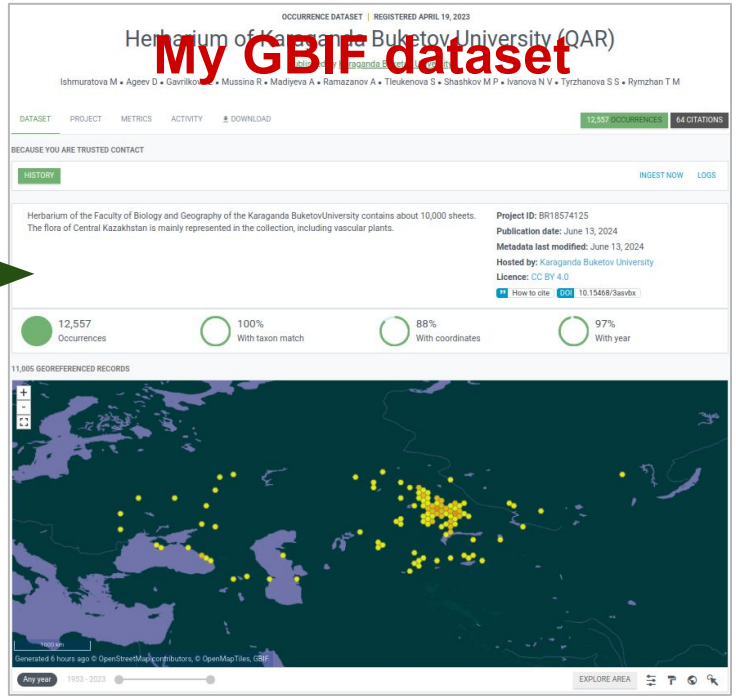



My data



Forest Ecology and Management

Value of large *Quercus robur* fallen logs in enhancing the species diversity of vascular plants in an old-growth mesic broad-leaved forest in the Central Russian Upland




doi:10.1371/journal.pone.0029715

# GBIF Data Processing Pipelines

Every single occurrence record that gets published to GBIF goes through a series of three processing steps until it becomes available on GBIF.org.

1. source downloading
2. parsing into verbatim occurrences
3. interpreting verbatim values

Waller J et al., 2021 doi: 10.3897/biss.5.75686

Get data How-to Tools Community About

OCCURRENCE | 25 JUNE 1980

# Anthocopa fyrneri Tkalcu

**Collected in Uzbekistan**

Animalia › Arthropoda › Insecta › Hymenoptera › Megachilidae

DETAILS

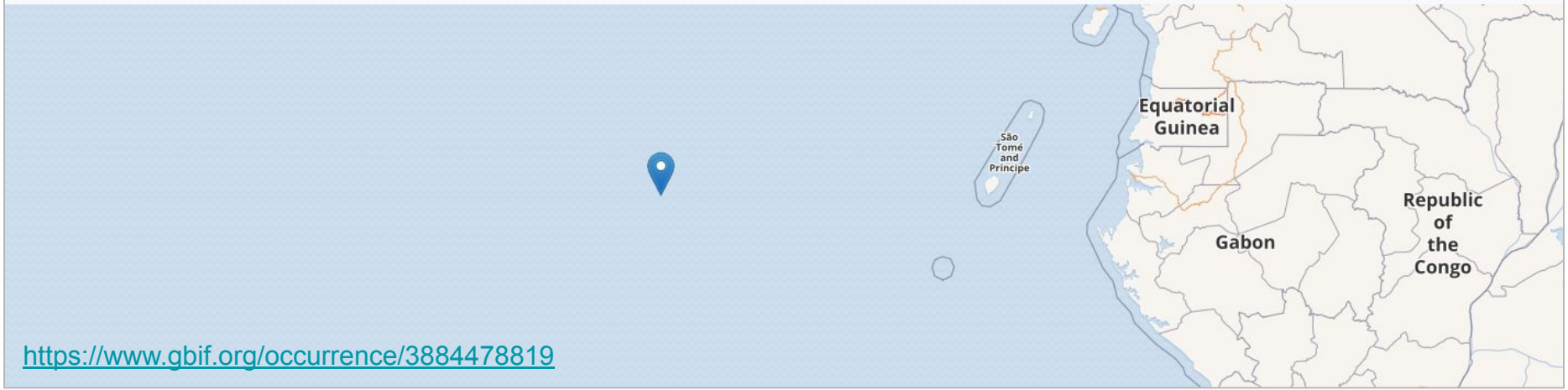
**GBIF Taxon interpretation:** Anthocopa Lepeletier & Audinet-Serville, 1825 **Dataset:** ZOBODAT (Zoological Botanical Database)

**Location:** Uzbekistan **Publisher:** Biologiezentrum Linz Oberoesterreich

**Basis of record:** Preserved specimen

**Issues:** Zero coordinate | Coordinate uncertainty metres invalid | Country coordinate mismatch | Taxon match higherrank

10




https://www.gbif.org/occurrence/3884478819

# GBIF Issues & Flags

**Location**

| Term | Interpreted | Original | Remarks |
|---|---|---|---|
| Coordinate uncertainty in metres | | 9.999999e+06 | Coordinate uncertainty metres invalid |
| Country or area | Uzbekistan | | Country coordinate mismatch |
| Country code | UZ | UZ | Country coordinate mismatch |
| Decimal latitude | 0 | 0 | Zero coordinate<br>Country coordinate mismatch |
| Decimal longitude | 0 | 0 | Zero coordinate<br>Country coordinate mismatch |
| Geodetic datum | WGS84 | WGS84 | Country coordinate mismatch |
| Locality | Taschkent env., Galvasai | Taschkent env., Galvasai | |

**See details:** https://www.researchgate.net/publication/373218947_GBIF_Issues_Flags

https://www.gbif.org/occurrence/3884478819

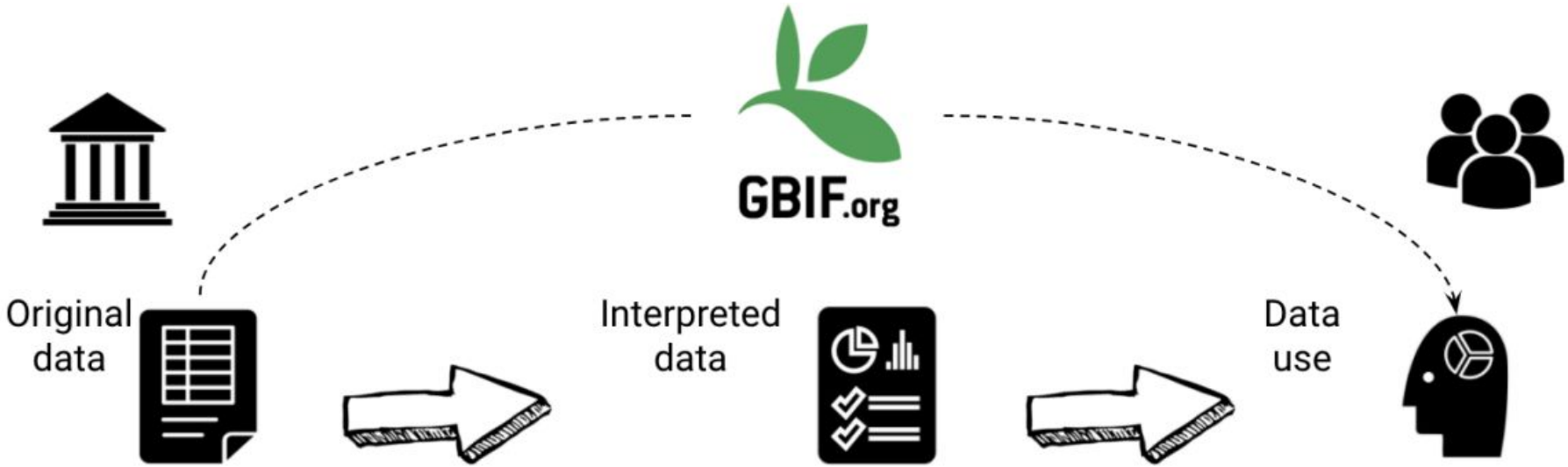
# GBIF Issues & Flags

**Taxon**

| Term | Interpreted | Original | Remarks |
|---|---|---|---|
| Kingdom | Animalia | Metazoa | Taxon match higherrank |
| Phylum | Arthropoda | Arthropoda | Taxon match higherrank |
| Class | Insecta | Insecta | Taxon match higherrank |
| Order | Hymenoptera | Hymenoptera | Taxon match higherrank |
| Family | Megachilidae | Apidae | Taxon match higherrank |
| Genus | Anthocopa | Anthocopa | Taxon match higherrank |
| Specific epithet | | fyrneri | Taxon match higherrank |
| Scientific name | Anthocopa Lepeletier & Audinet-Serville, 1825 | Anthocopa fyrneri Tkalcu | Taxon match higherrank |
| Scientific name authorship | Lepeletier & Audinet-Serville, 1825 | Tkalcu | Taxon match higherrank |
| Rank | Genus | | Inferred |
| Taxonomic status | taxonomicStatus.PROVISIONALLY_ACCEPTED | | Inferred |

https://www.gbif.org/occurrence/3884478819

# You download interpreted data through GBIF




**See details:** https://www.researchgate.net/publication/373218947_GBIF_Issues_Flags

# Outcomes

- Most GBIF data contain technical issues.
- Check out the list of issues and flags corrected in your downloaded GBIF data.

# Outcomes

GBIF benefits

- All data are available free of charge and have global identifiers.
- The data's origin, authorship, and licences are clear.

GBIF challenges

- Data on soil invertebrates are incomplete.
- Data cleaning required.


Thank you for your attention!