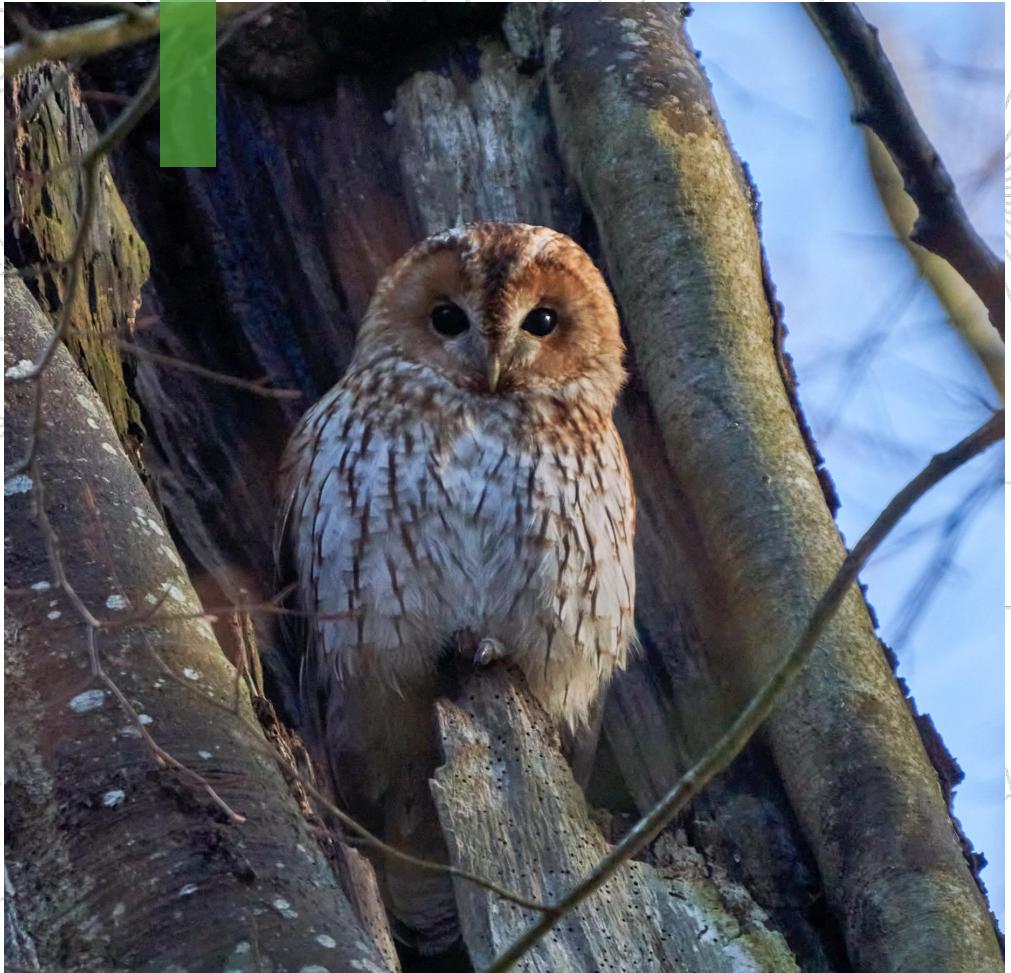


# GBIF - Global Biodiversity Information Facility

Michal Torma, NHM, UIO



UNIVERSITETET  
I OSLO

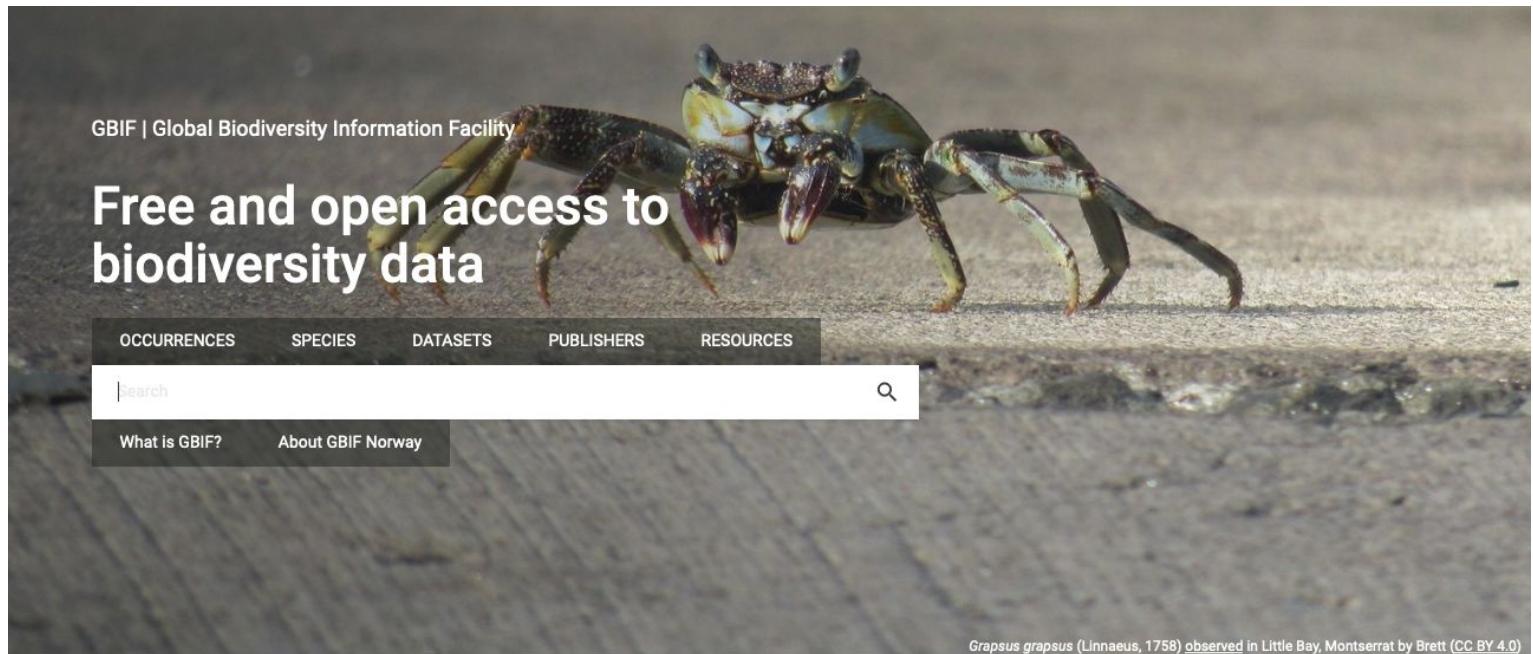


Global Biodiversity  
Information Facility

## What is GBIF?

# What is GBIF?

- **Intergovernmental** network and research infrastructure
- Provides anyone, anywhere, **free and open access to data** about all types of life on Earth
- **Voluntary collaboration** through Memorandum of Understanding (MoU)
- **Participant nodes**, Secretariat in Copenhagen, Denmark



2,309,952,319

Occurrence records



85,218

Datasets



2,037

Publishing institutions



8,703

Peer-reviewed papers  
using data

# BY THE NUMBERS | 10<sup>th</sup> November 2025

Species occurrence records

**3 567 323 913**



Country Participants

**68**

Organizational Participants

**42**

Datasets

**118 721**



Peer-review papers using data

**13 763**



Publishers

**2 645**

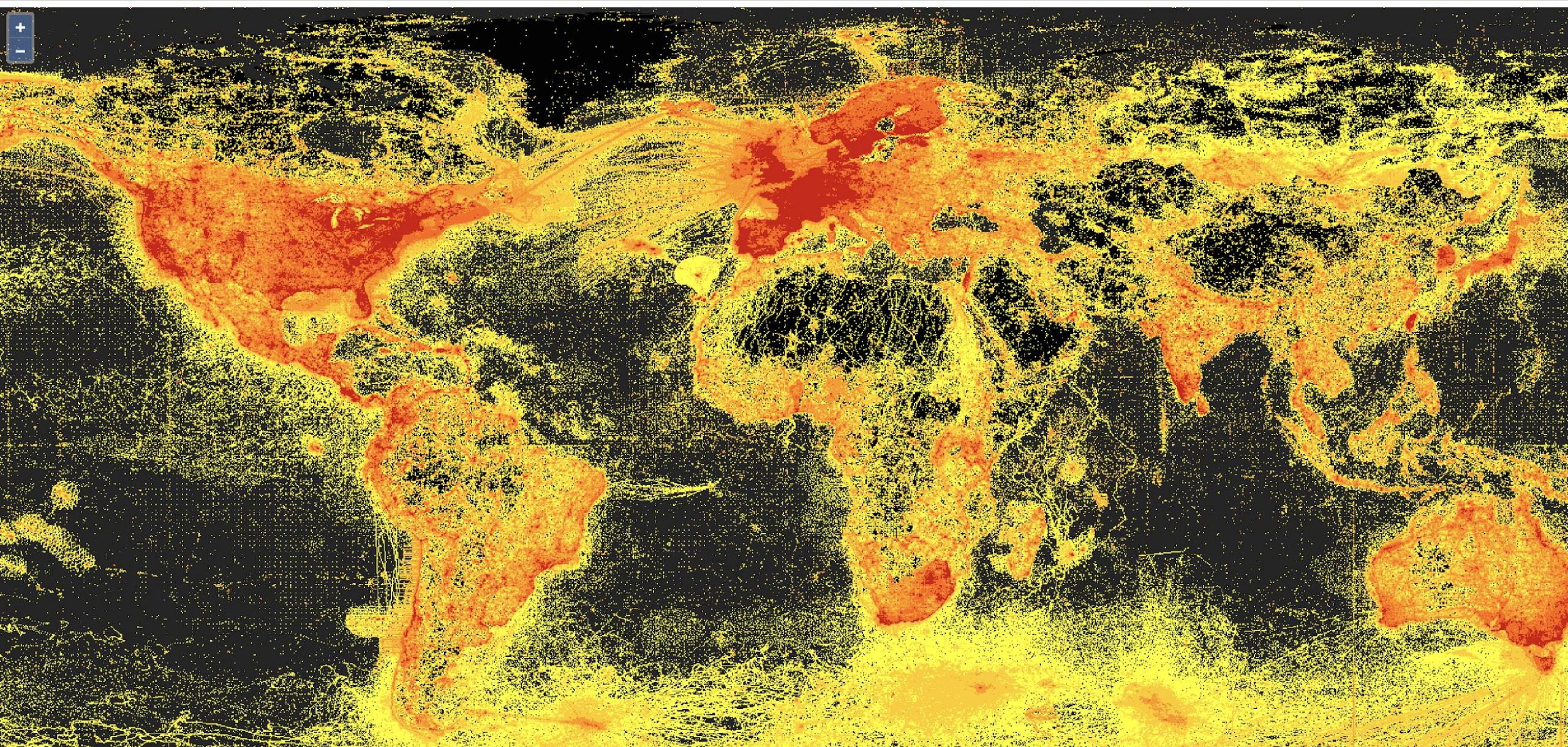


Records downloaded per month in 2022

**>119 billion**



# Data From the GBIF Network | 14<sup>th</sup> November 2023



# BY THE NUMBERS | 10<sup>th</sup> November 2025

Species occurrence records

**3 567 323 913**



Country Participants

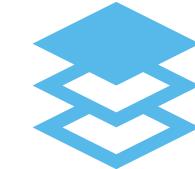
**68**

Organizational Participants

**42**

Datasets

**118 721**



Records downloaded per month in 2022

**>119 billion**



Publishers

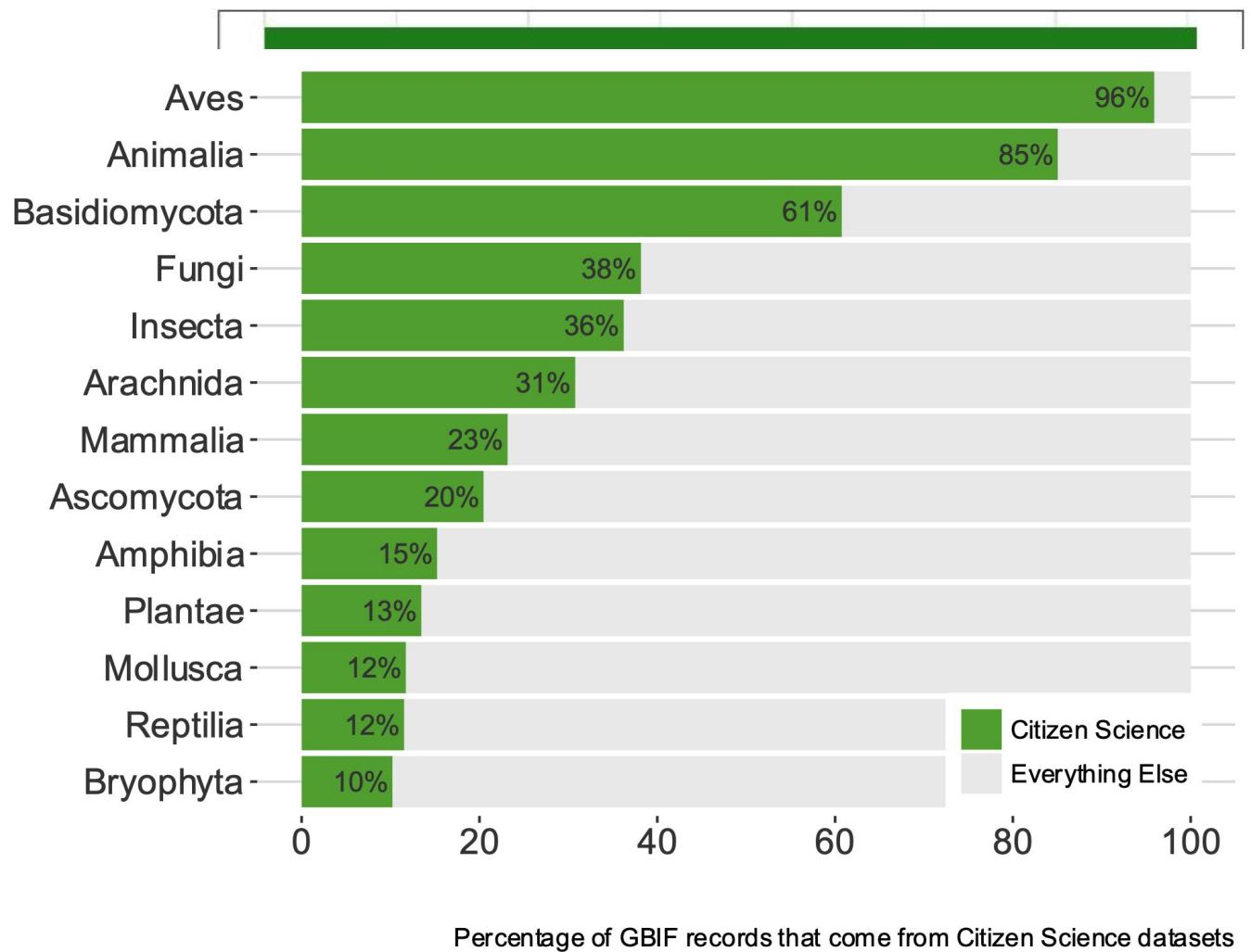
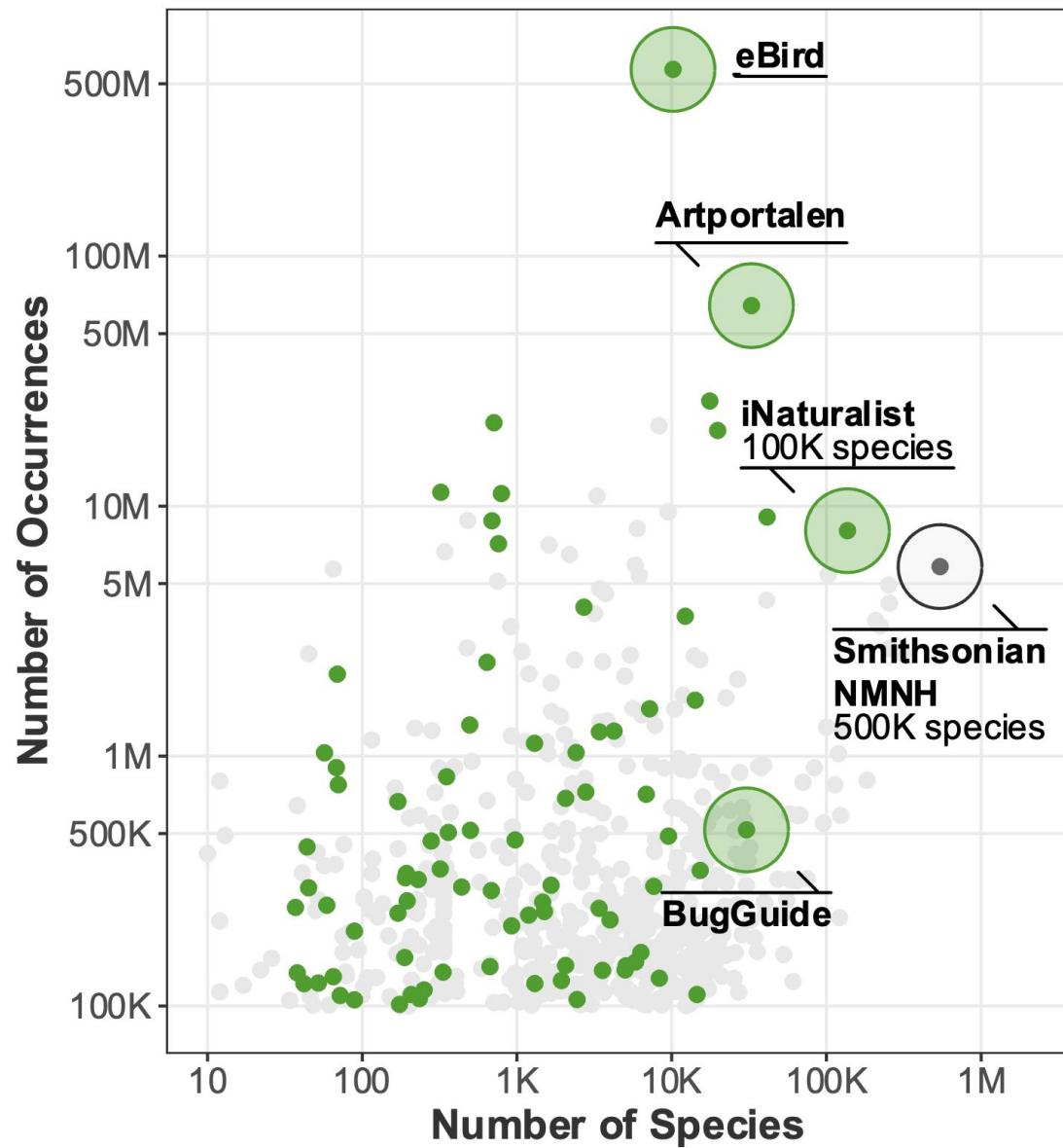
**2 645**



Peer-review papers using data

**13 763**

# Data Sets in GBIF



# BY THE NUMBERS | 10<sup>th</sup> November 2025

Species occurrence records

**3 567 323 913**



Country Participants

**68**

Organizational Participants

**42**

Datasets

**118 721**



Publishers

**2 645**



Peer-review papers using data

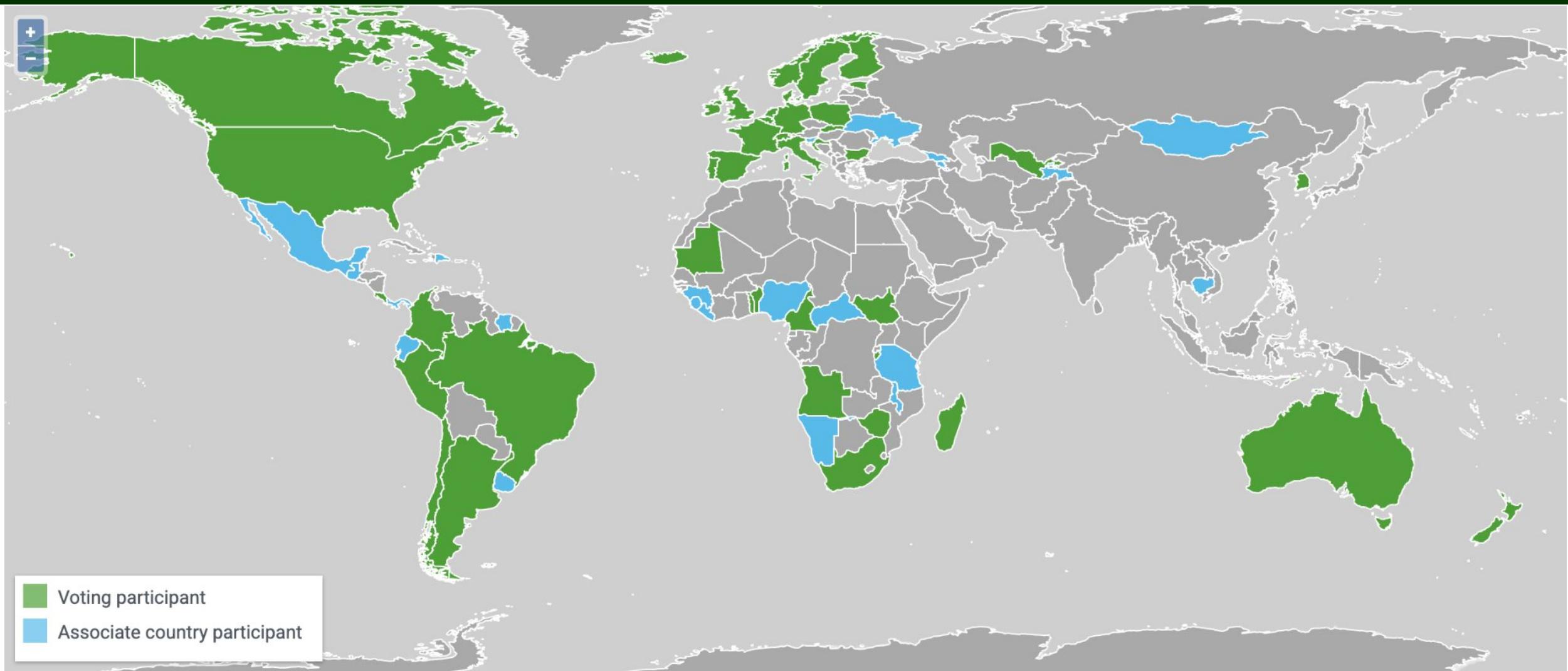
**13 763**



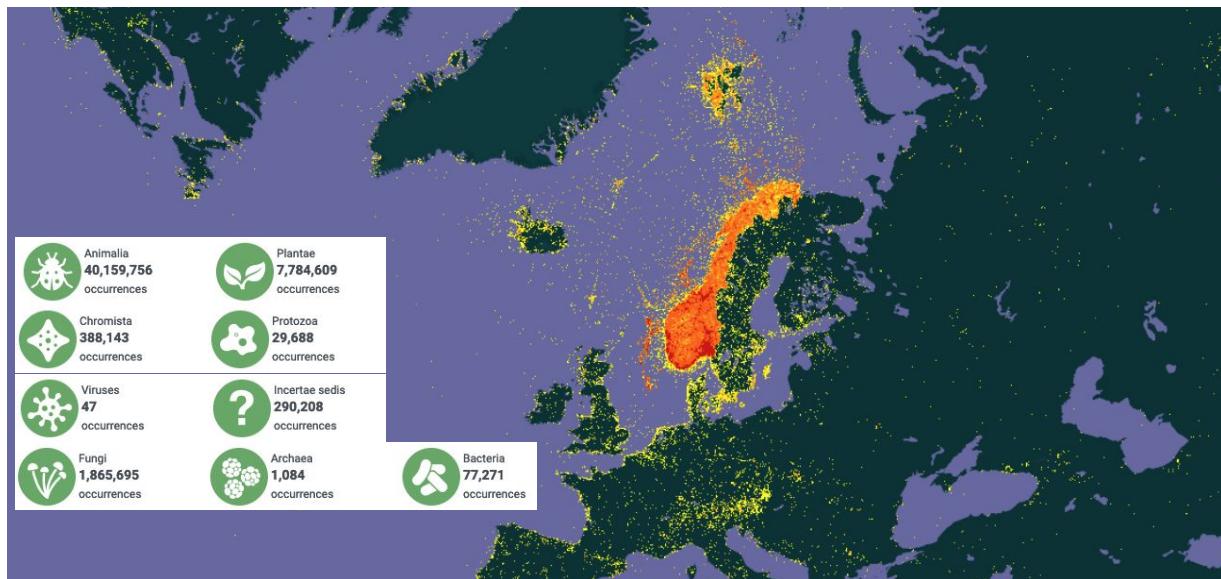
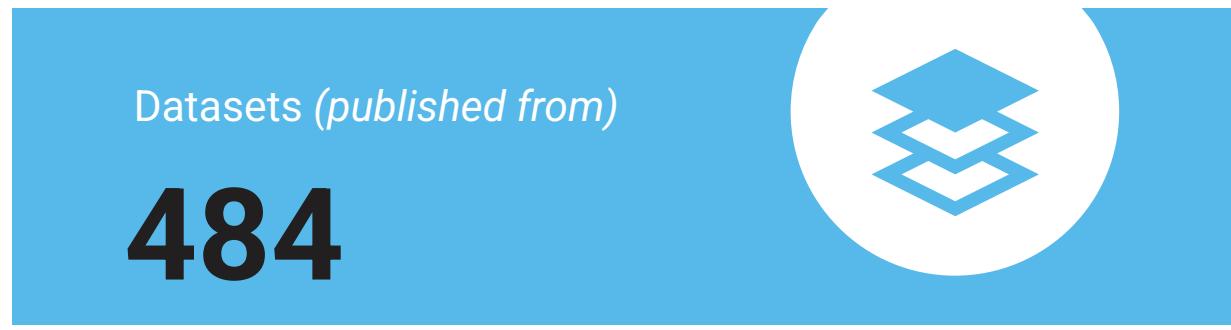
Records downloaded per month in 2022

**>119 billion**

# Data From the GBIF Network | 10<sup>th</sup> November 2025



# BY THE NUMBERS | 10<sup>th</sup> November 2025 - Norway

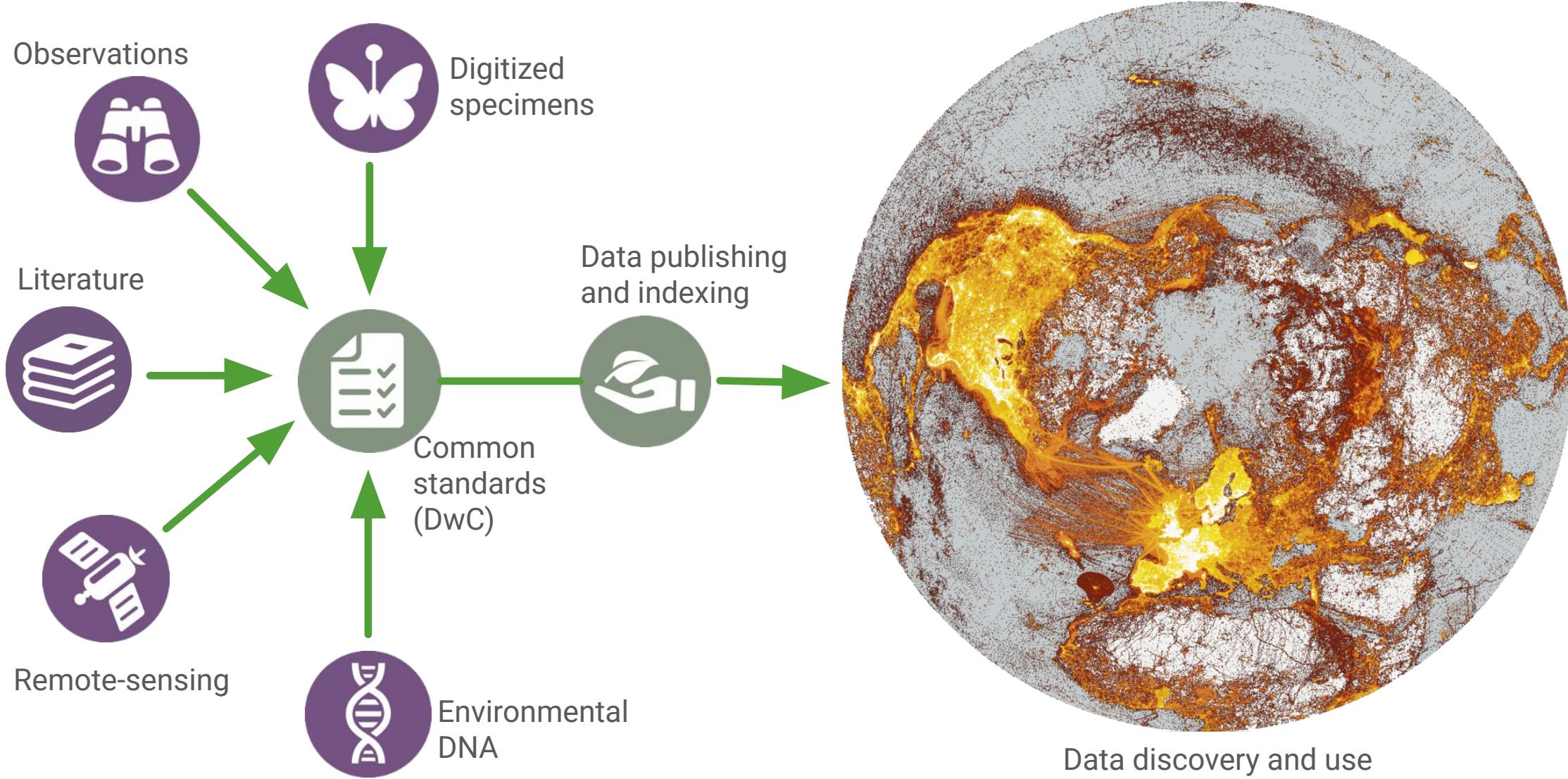




Global Biodiversity  
Information Facility

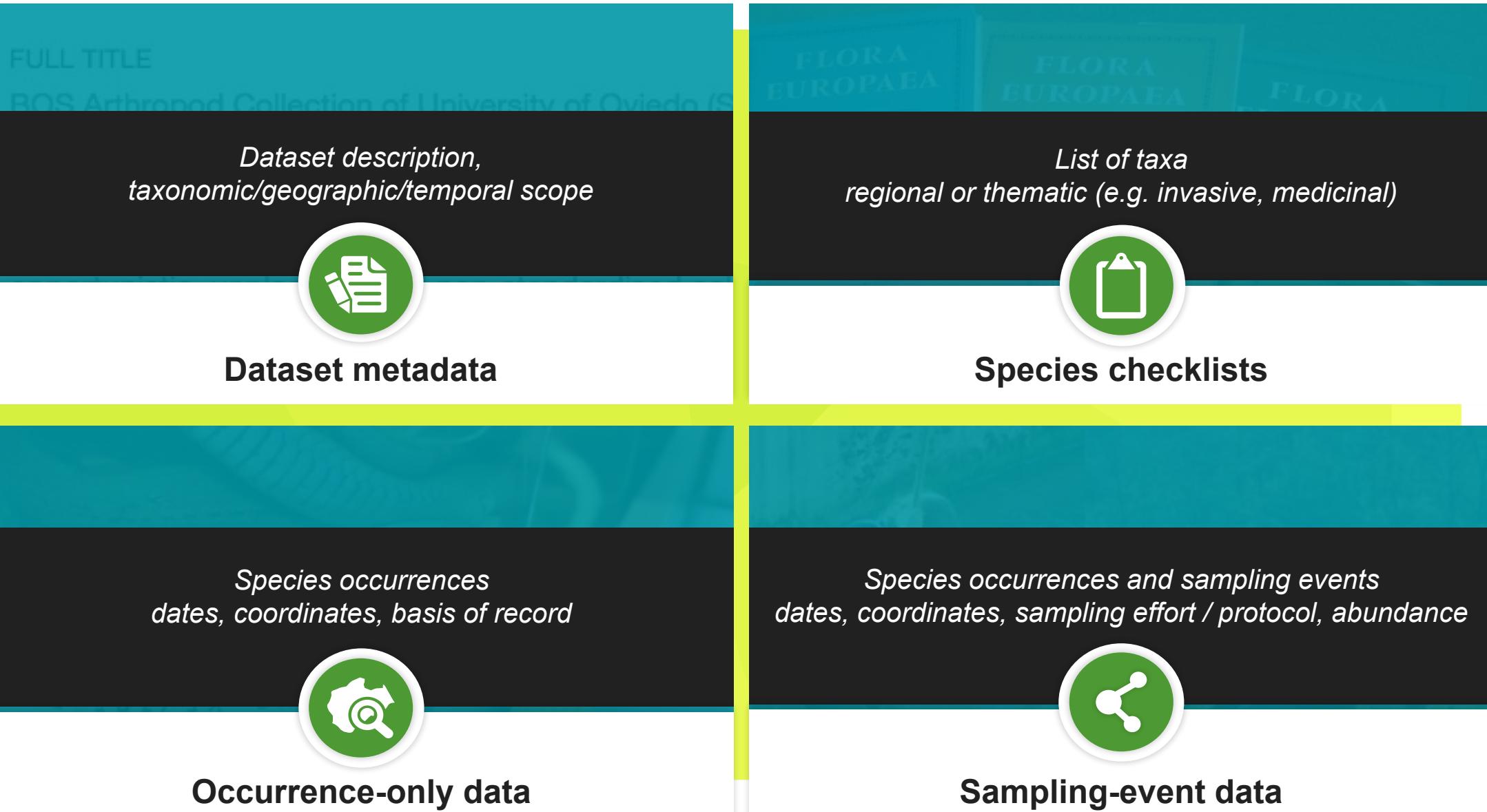
## What Data is in GBIF?

# Evidence About Where Species HAVE Lived & When



# Data Models in GBIF

Data richness levels supported by Gbif



# Sources of GBIF Data: Digitized Museum Collections

OCCURRENCE DATASET | REGISTERED JULY 30, 2012

## Mycology herbarium, Oslo (0)

Published by [Natural History Museum, University of Oslo](#)

DATASET METRICS ACTIVITY DOWNLOAD HOME PAGE

215,107 OCCURRENCES 33 CITATIONS

Fungi in the Natural History Museum, University of Oslo

215,107 Occurrences

193,748 GEOREFERENCED RECORDS

Generated 2 hours ago © OpenStreetMap

Any year 1811–2020

4,098 OCCURRENCES WITH IMAGES

Metadata last modified: October 7, 2020  
by: GBIF.no  
License: CC BY 4.0  
How to cite DOI: 10.15468/6xjrdn

98% With year

EXPLORE

SEE GALLERY

OCCURRENCE DATASET | REGISTERED JULY 30, 2012

## Vascular plant herbarium TRH, NTNU University Museum

Published by [NTNU University Museum](#)

Mika Bendiksby

DATASET METRICS ACTIVITY DOWNLOAD

231,447 OCCURRENCES 188 CITATIONS

This is the largest herbarium at NTNU University Museum, containing around 240 000 specimens from Norway and the Nordic countries. Most of the specimens were collected in Central Norway. The vascular plant herbarium also contains around 32 000 specimens from the rest of the world, most of which is not yet digitized. The arctic collection contains professor Olav Gjærevoll's material from Alaska and Canada - about 6 000 digitized specimens.

231,447 Occurrences

197,938 GEOREFERENCED RECORDS

Generated 2 hours ago © OpenStreetMap

Any year 1762–2020

227,288 OCCURRENCES WITH IMAGES

Metadata last modified: October 7, 2020  
by: GBIF.no  
License: CC BY 4.0  
How to cite DOI: 10.15468/zrlqok

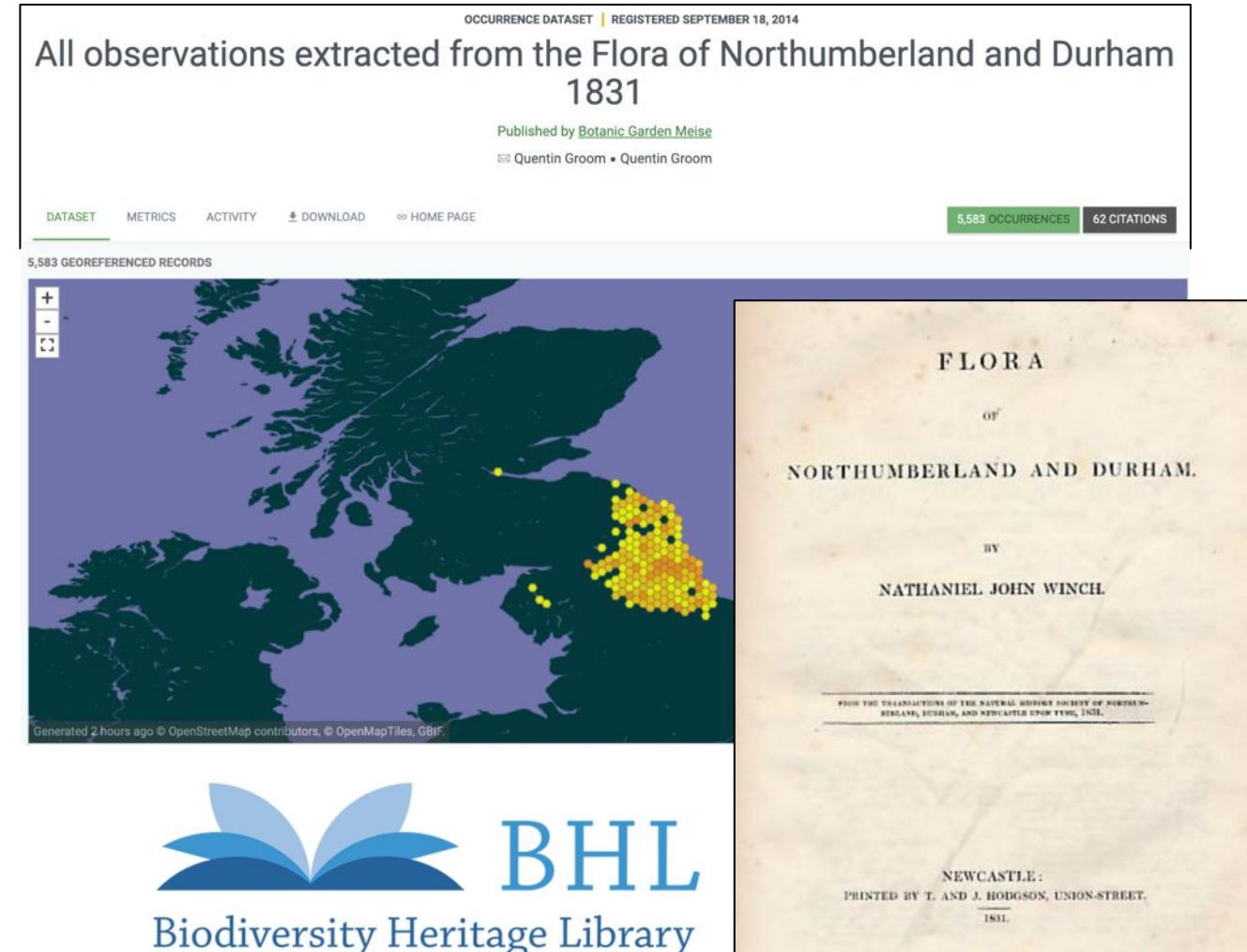
97% With year

EXPLORE

SEE GALLERY

Barcode: V-055822  
Species: Pinus cembra L.  
Common name: Swiss pine, Swiss mountain pine, Swiss stone pine, Cembra pine  
Source: Herbarium  
Digitizing project: Trondheim herbarium  
Coordinates: 63°41'07"N 10°28'00"E Alt: 25 m

# Sources of GBIF Data: Taxonomic Literature



# Sources of GBIF Data: DNA Derived Occurrences

Get data How-to Tools Community About

SAMPLING EVENT | REGISTERED MARCH 7, 2019

## Atlantic salmon microbiota

Published by MGnify

DATASET METRICS ACTIVITY DOWNLOAD

4,702 OCCURRENCES 4 CITATIONS

Atlantic salmon skin mucus and surrounding water have been sampled at production facilities both in freshwater and after transfer to seawater. Extracted DNA is sequenced using barcoded 16S primers using PacBio SCC.

4,702 Occurrences 94% With taxon match 100% With coordinates 100% With year

Metadata last modified: March 13, 2019  
Hosted by: GBIF Secretariat  
License: CC BY 4.0  
[How to cite](#) DOI: 10.15468/ox86z5

4,702 GEOREFERENCED RECORDS

+ - ⌂



Generated 2 hours ago © OpenStreetMap contributors, © OpenMapTiles, GBIF.

Year 2017

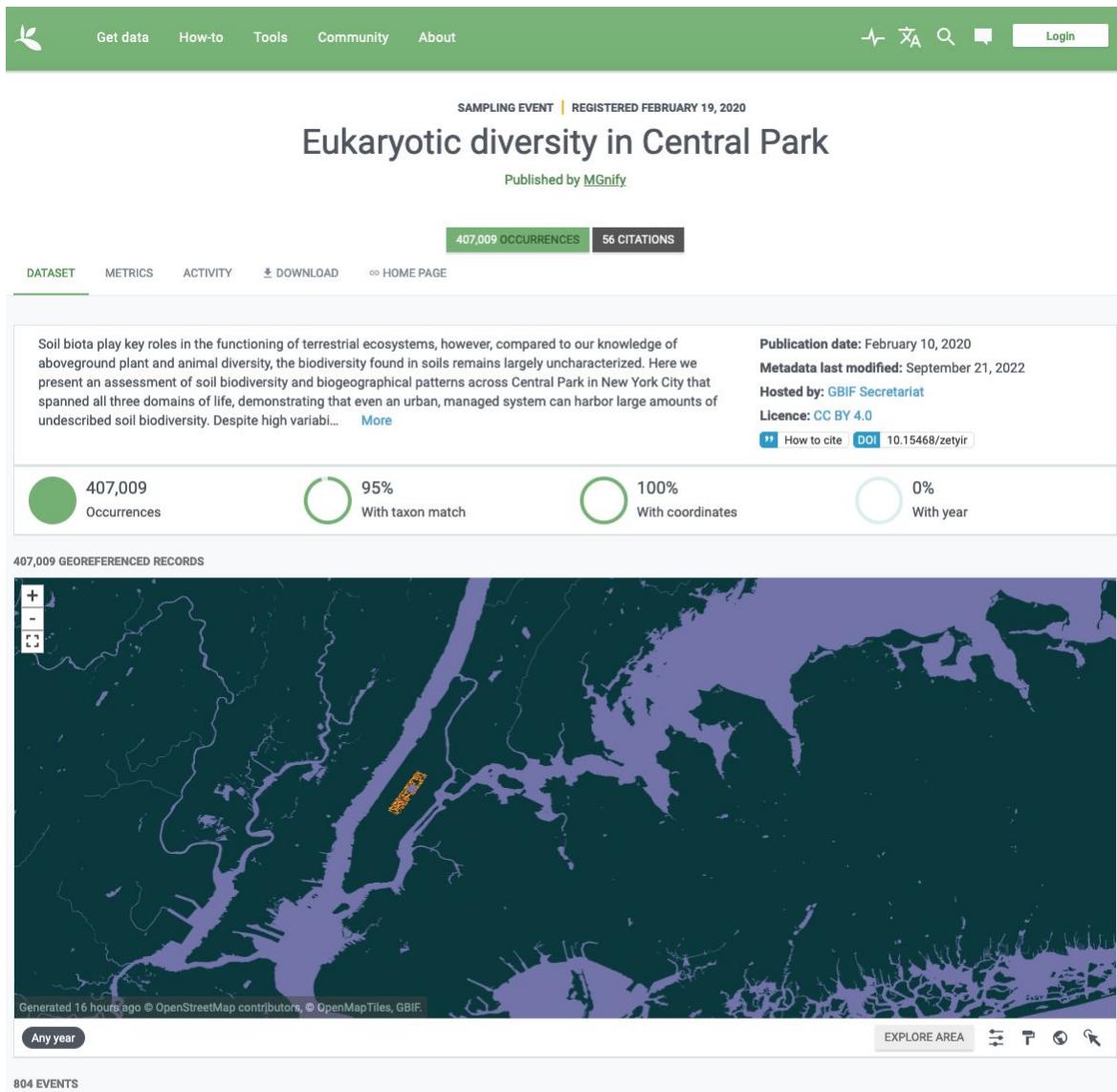
EXPLORE AREA

66 EVENTS

Event ID	Event date	Sampling protocol	Occurrence count
mgya00210126	1 January 2017		297
mgya00210073	1 January 2017		209
mgya00210106	1 January 2017		203



# Sources of GBIF Data: Peer-Reviewed Publications



[rsb.royalsocietypublishing.org](http://rsb.royalsocietypublishing.org)



## Research

**Cite this article:** Ramirez KS *et al.* 2014 Biogeographic patterns in below-ground diversity in New York City's Central Park are similar to those observed globally. *Proc. R. Soc. B* **281:** 20141988.  
<http://dx.doi.org/10.1098/rspb.2014.1988>

Received: 11 August 2014

Accepted: 29 August 2014

### Subject Areas:

ecology

### Keywords:

Bacteria, Archaea, Eukarya, soil biodiversity, 16S rRNA gene, 18S rRNA gene

Biogeographic patterns in below-ground diversity in New York City's Central Park are similar to those observed globally

Kelly S. Ramirez<sup>1</sup>, Jonathan W. Leff<sup>4,5</sup>, Albert Barberán<sup>5</sup>, Scott Thomas Bates<sup>6</sup>, Jason Betley<sup>7</sup>, Thomas W. Crowther<sup>8</sup>, Eugene F. Kelly<sup>2</sup>, Emily E. Oldfield<sup>8</sup>, E. Ashley Shaw<sup>3</sup>, Christopher Steenbock<sup>4</sup>, Mark A. Bradford<sup>8</sup>, Diana H. Wall<sup>1,3</sup> and Noah Fierer<sup>4,5</sup>

<sup>1</sup>School of Global Environmental Sustainability, <sup>2</sup>Department of Soil and Crop Sciences, and <sup>3</sup>Department of Biology, Colorado State University, Fort Collins, CO 80523, USA

<sup>4</sup>Department of Ecology and Evolutionary Biology, and <sup>5</sup>Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, CO 80309, USA

<sup>6</sup>Department of Plant Pathology, University of Minnesota, Saint Paul, MN 55108, USA

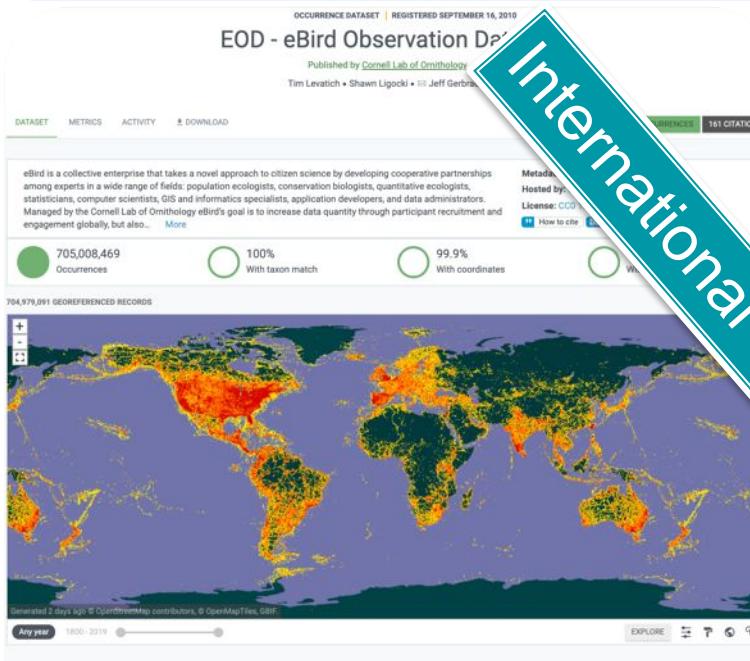
<sup>7</sup>Illumina UK, Chesterford Research Park, Little Chesterford, Saffron Walden, Essex CB10 1XL, UK

<sup>8</sup>School of Forestry and Environmental Studies, Yale University, New Haven, CT 06511, USA

Soil biota play key roles in the functioning of terrestrial ecosystems, however, compared to our knowledge of above-ground plant and animal diversity, the biodiversity found in soils remains largely uncharacterized. Here, we present an assessment of soil biodiversity and biogeographic patterns across Central Park in New York City that spanned all three domains of life, demonstrating that even an urban, managed system harbours large amounts of undescribed soil biodiversity. Despite high variability across the Park, below-ground diversity patterns were predictable based on soil characteristics, with prokaryotic and eukaryotic communities exhibiting overlapping biogeographic patterns. Further, Central Park soils harboured nearly as many distinct soil microbial phylotypes and types of soil communities as we found in biomes across the globe (including arctic, tropical and desert soils). This integrated cross-domain investigation highlights that the amount and patterning of novel and uncharacterized diversity at a single urban location matches that observed across natural ecosystems spanning multiple biomes and continents.

# Sources of GBIF Data: Citizen Science Observations

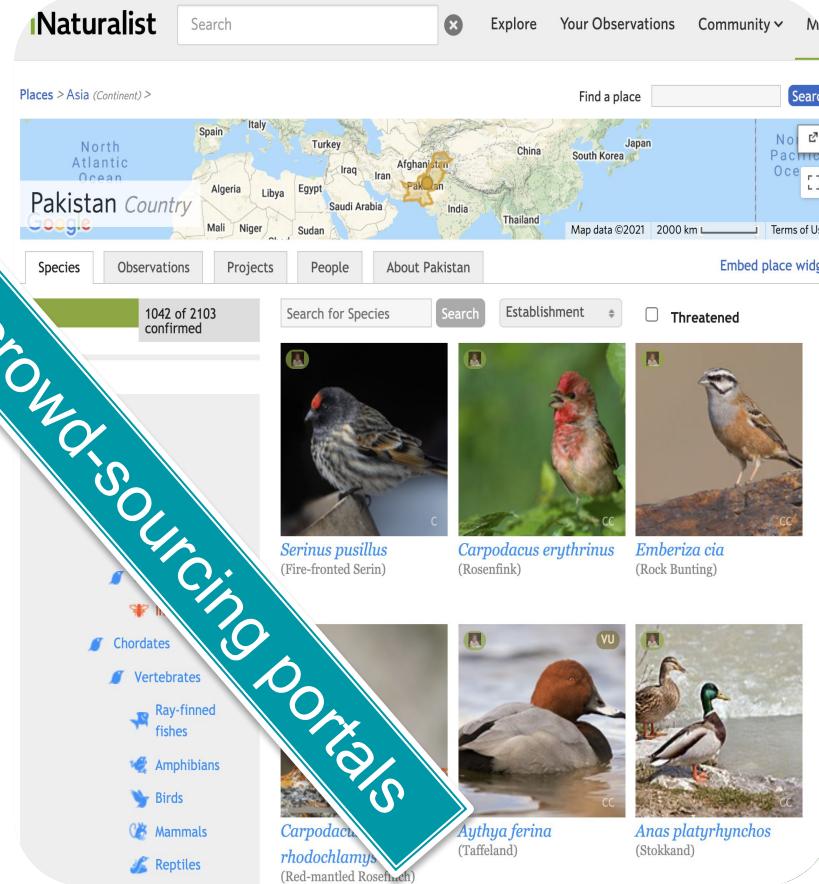
# eBird



## Description

eBird is a collective enterprise that takes a novel approach to citizen science by developing cooperative partnerships among experts in a wide range of fields: population ecologists, conservation biologists, quantitative ecologists, statisticians, computer scientists, GIS and informatics specialists, application developers, and data administrators. Managed by the Cornell Lab of Ornithology eBird's goal is to increase data quantity through participant recruitment and engagement globally, but also to quantify and control for data quality issues such as observer variability, imperfect detection of species, and both spatial and temporal bias in data collection. eBird data are openly available and used by a broad spectrum of students, teachers, scientists, NGOs, government agencies, land managers, and policy makers. The result is that eBird has become a major source of biodiversity data, increasing our knowledge of the dynamics of species distributions, and having a direct impact on the conservation of birds and their habitats.

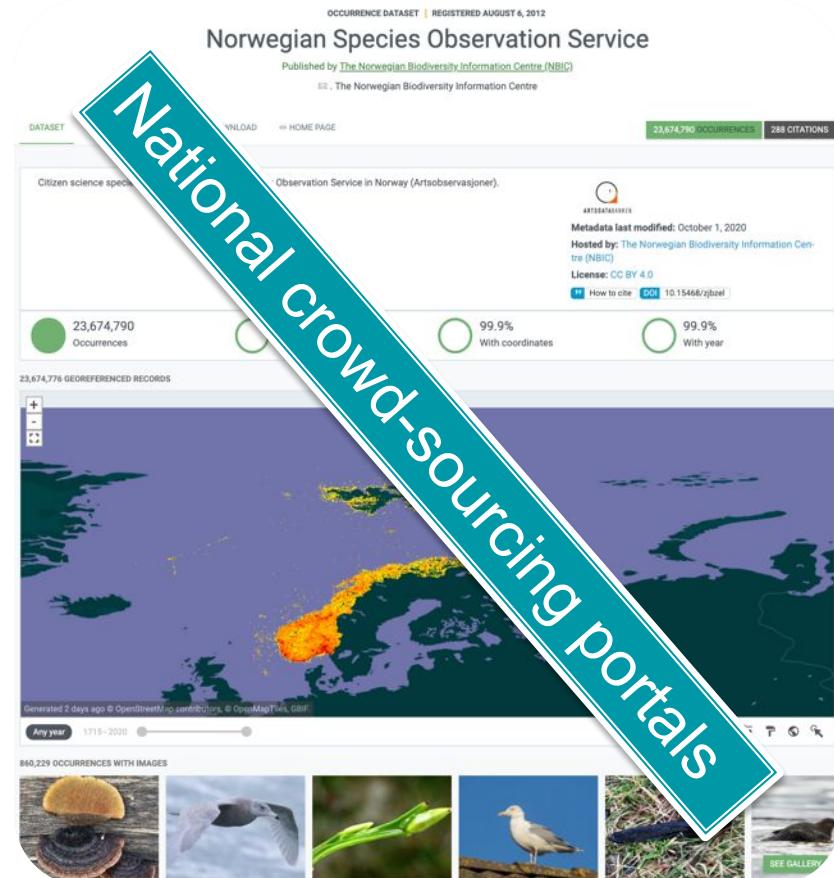
# iNaturalist



## International crowd-sourcing portals



# ARTSDATABANKEN

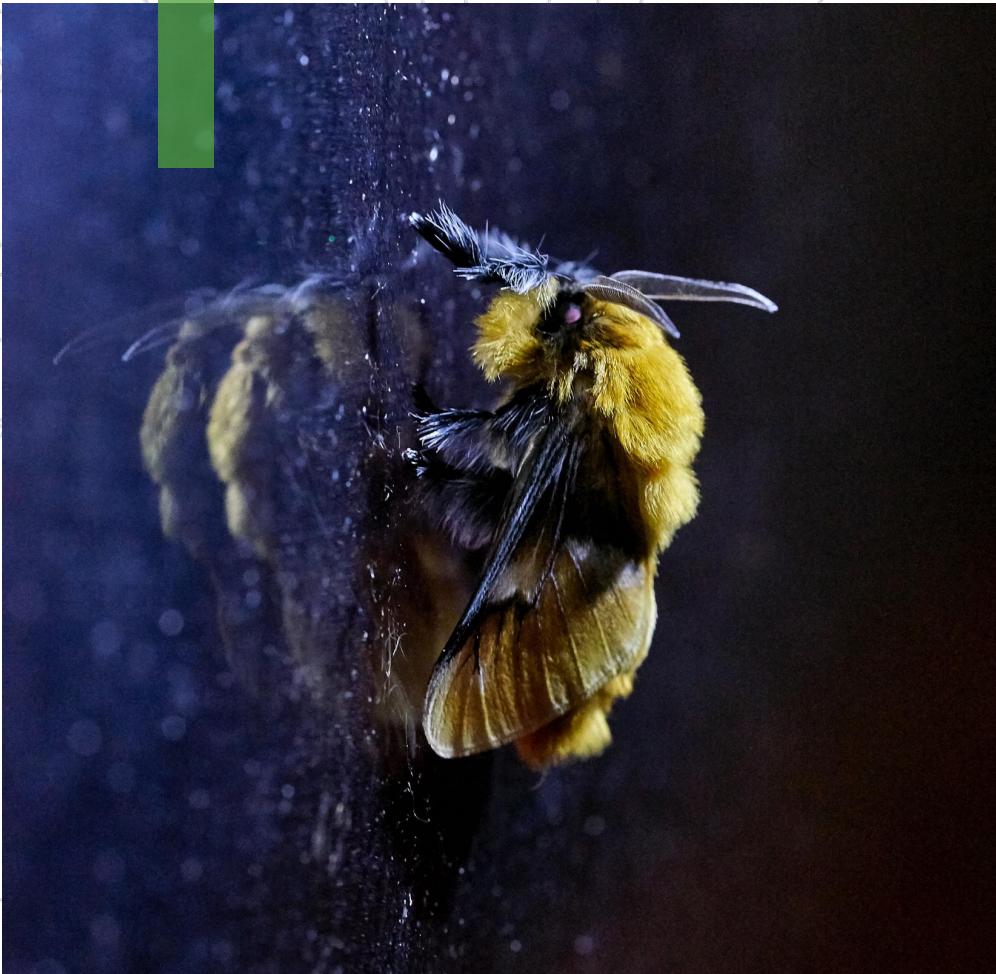


## National crowd-sourcing portals

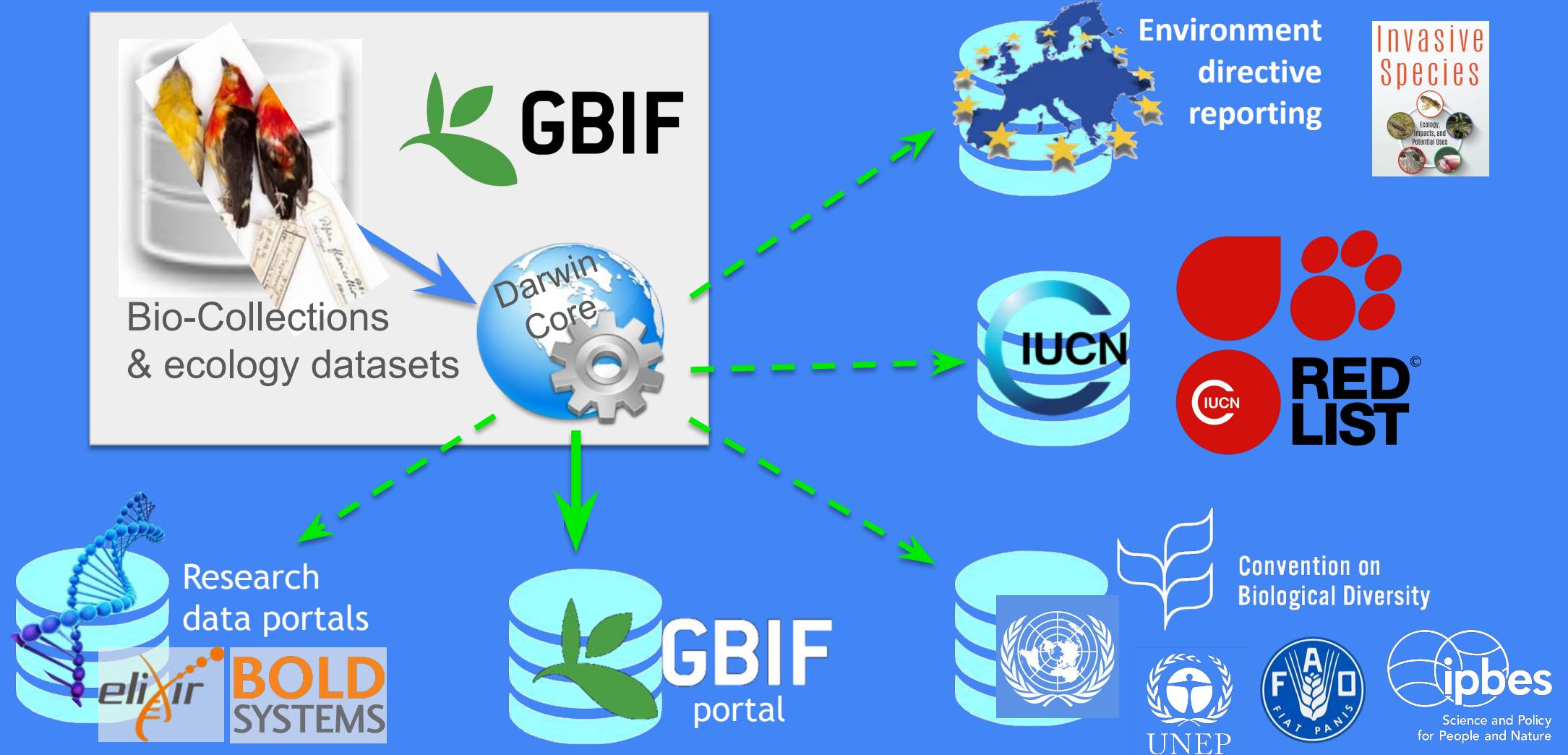


Global Biodiversity  
Information Facility

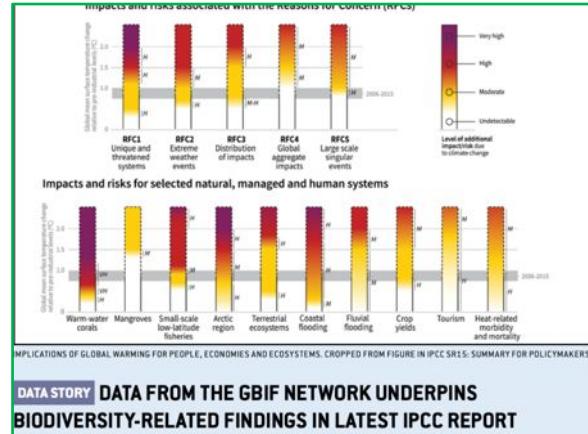
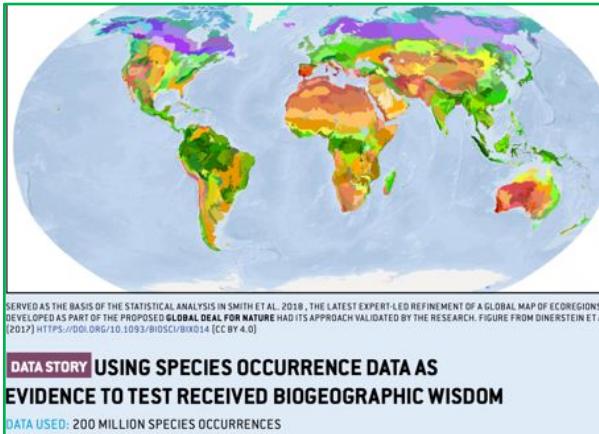
## Using GBIF Data



# GBIF: Multiple-Purpose Data Publishing Services



# Supporting Research and Sustainable Development



## ▪ Conservation

- Protected areas
- Threatened species
- Invasive species risk

## ▪ Food Security

- Crop wild relatives
- In situ, ex situ* conservation of genetic diversity
- Fisheries planning

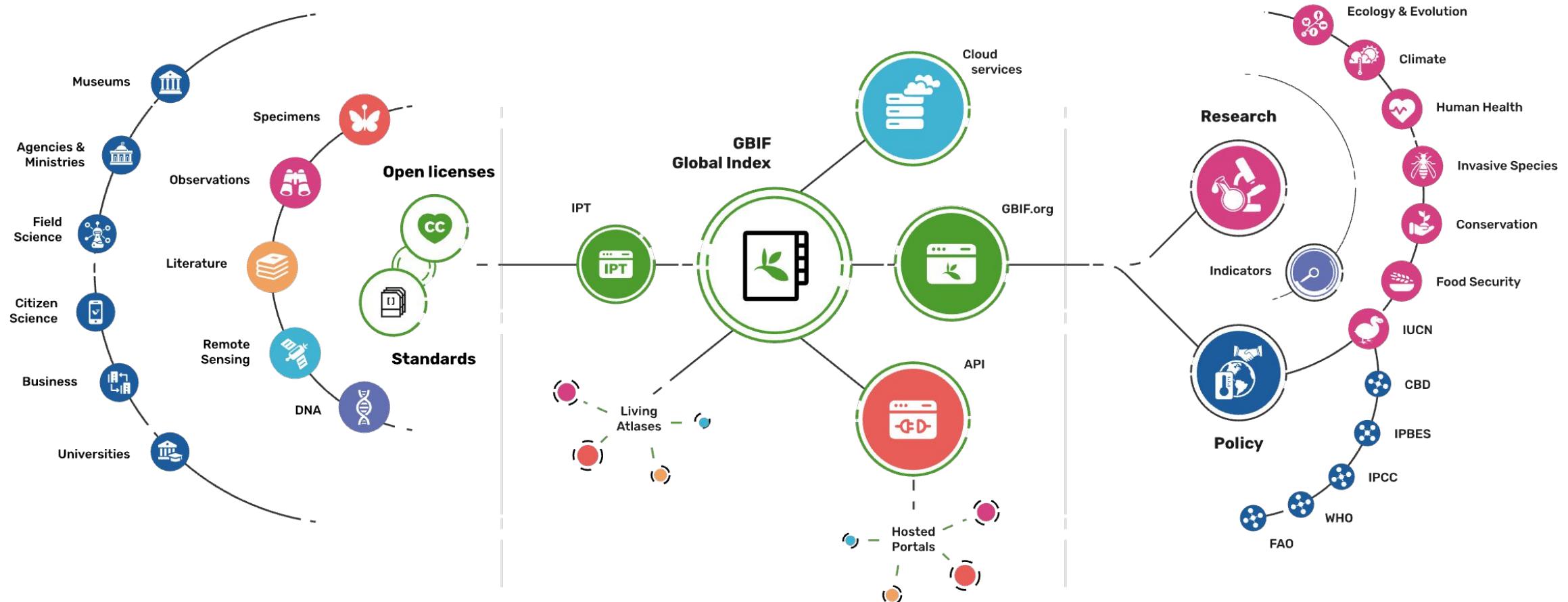
## ▪ Climate change

- Modelling impacts on species ranges
- Adaptation strategies
- Mitigation benefits, risks

## ▪ Human health

- Disease risk based on occurrence of vectors, hosts, reservoirs
- Medicinal plants
- Hazards e.g. snakebite

# Biodiversity Evidence for Research and Policy





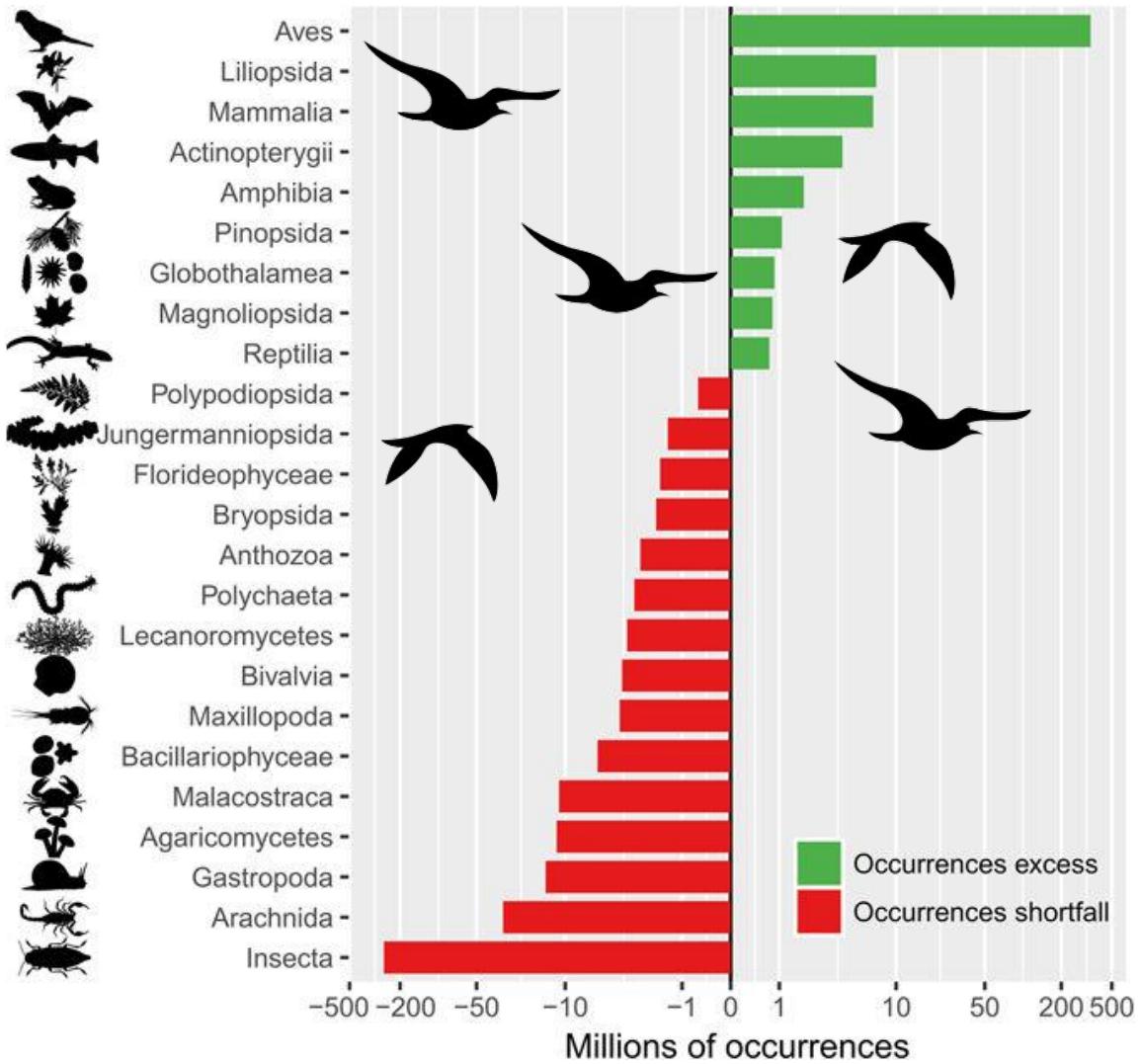
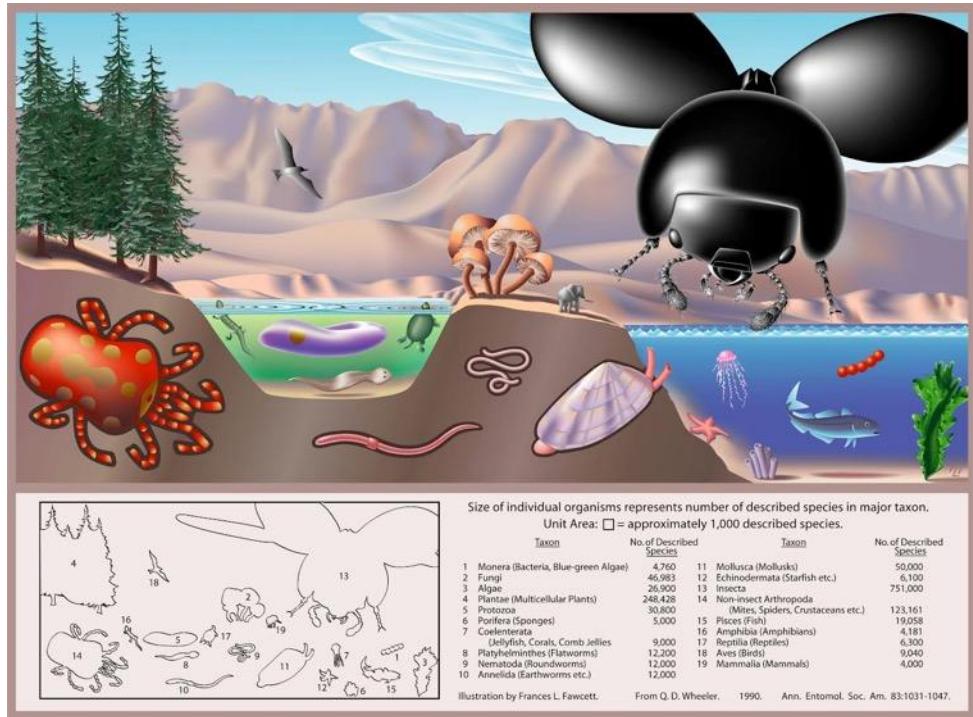
Global Biodiversity  
Information Facility

## GBIF Data Considerations

# Taxonomic Biases



Image: FL Fawcett in Wheller Ann. Entomol. Soc. Am. 1990

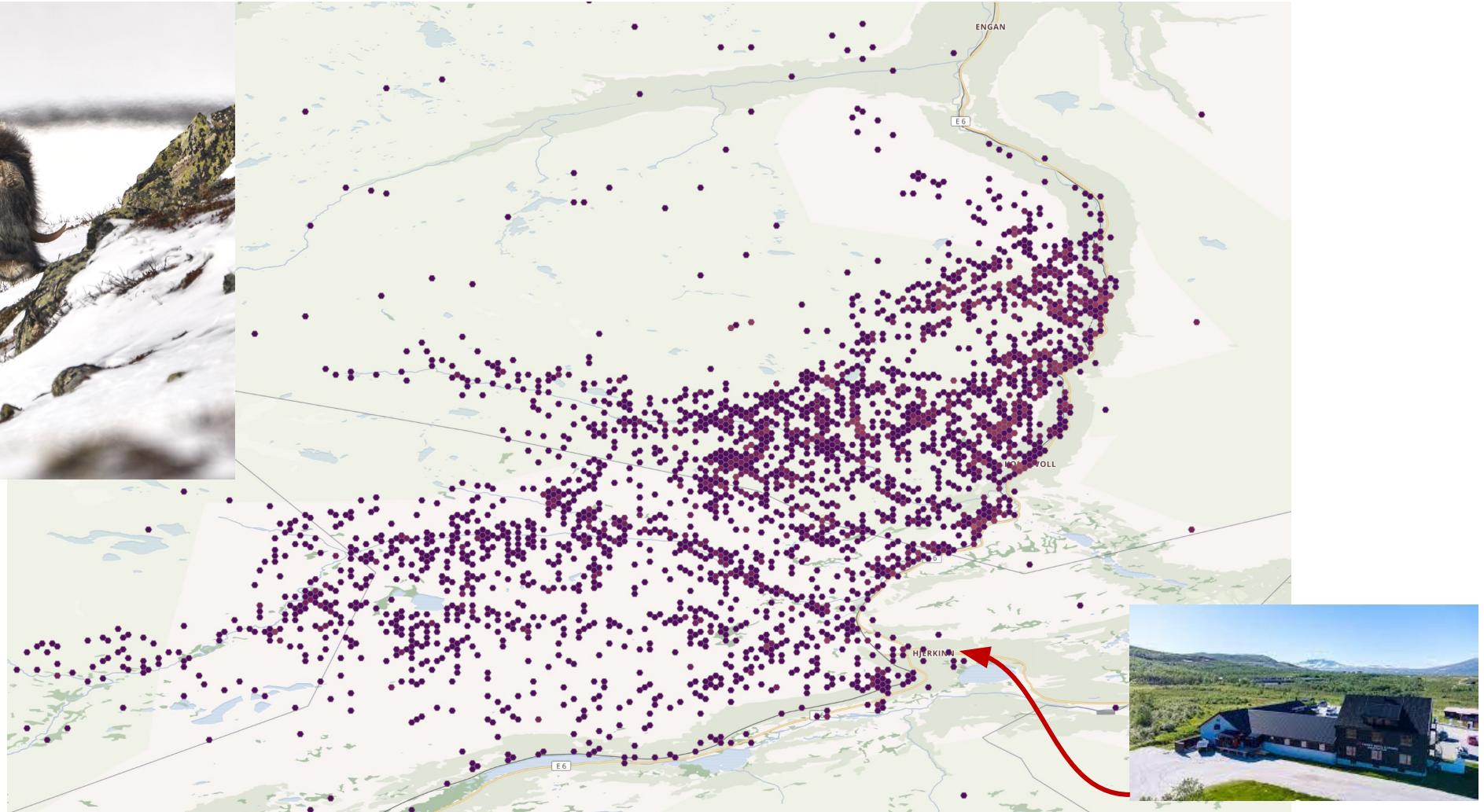


# Spatial Bias



Observations are:

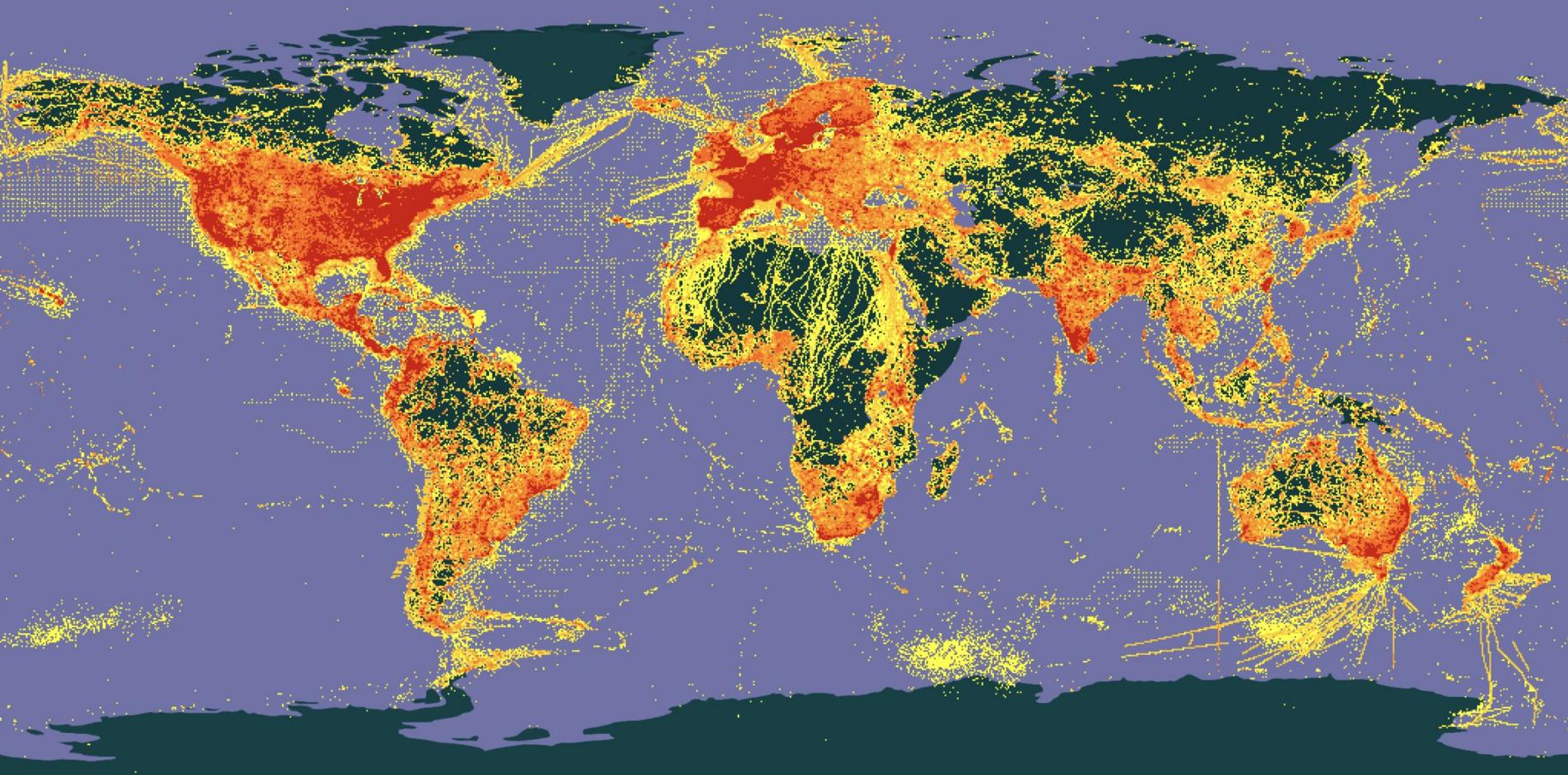
- where people go
- driven by visibility



# Temporal Bias



**2032-2032**



# Data Issues and Flags

- Suite of flags for determining data quality
- The ones to look for especially
  - Taxon match fuzzy
  - Taxon match higherrank

 Get data How-to Tools Community About

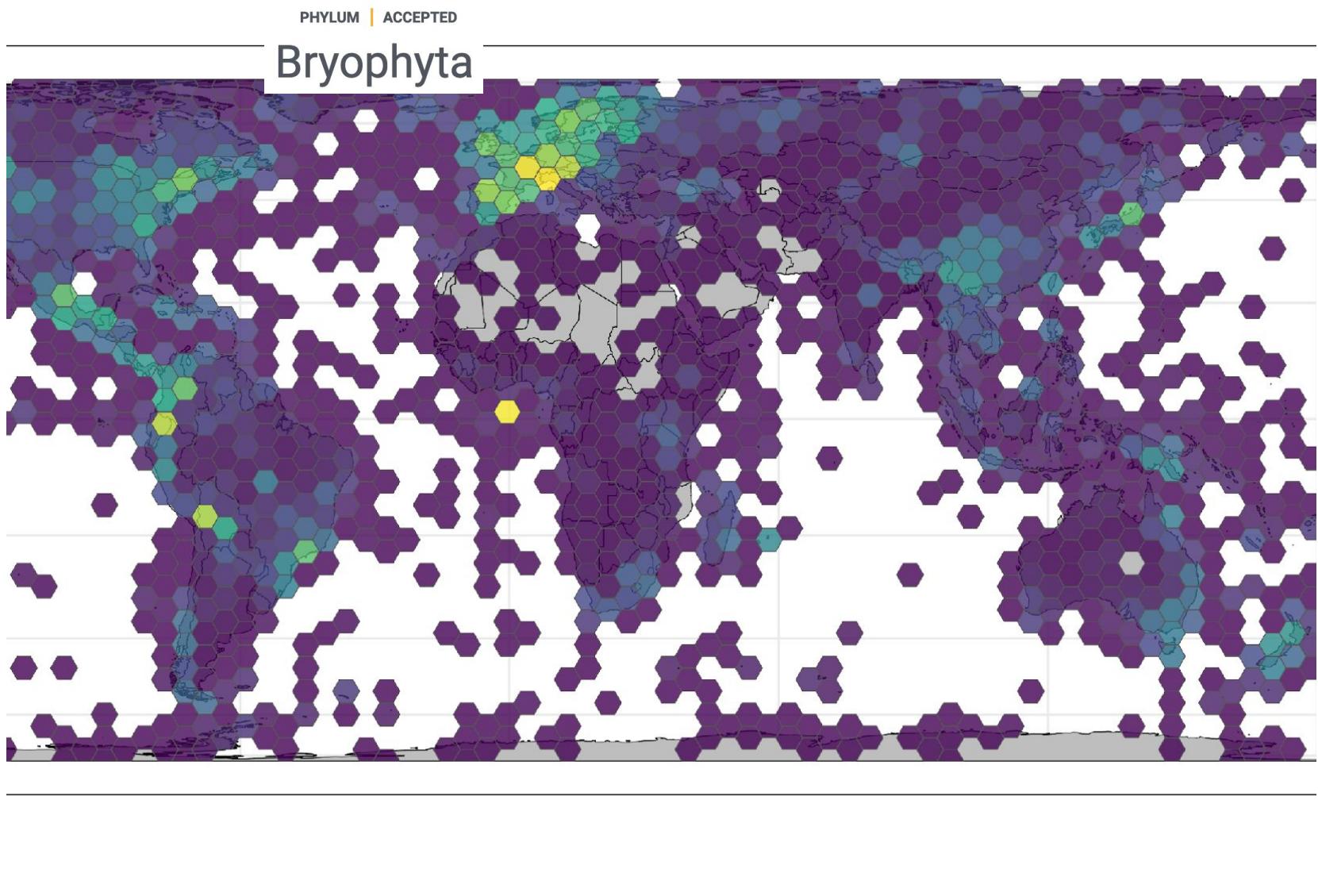
< Occurrences 1

Country or area

Continent

Issues and flags

Flag	Description	Count
<input type="checkbox"/>	Zero coordinate	1,396,989
<input type="checkbox"/>	Coordinate out of range	326,055
<input type="checkbox"/>	Coordinate invalid	6,245,672
<input type="checkbox"/>	Coordinate rounded	33,464,792
<input type="checkbox"/>	Geodetic datum invalid	24,881,806
<input type="checkbox"/>	Geodetic datum assumed WGS84	324,397,323
<input type="checkbox"/>	Coordinate reprojected	5,080,972
<input type="checkbox"/>	Coordinate reprojection failed	272
<input type="checkbox"/>	Coordinate reprojection suspicious	153,545
<input type="checkbox"/>	Coordinate accuracy invalid	0
<input type="checkbox"/>	Coordinate precision invalid	6,708,139
<input type="checkbox"/>	Coordinate uncertainty metres invalid	3,897,046
<input type="checkbox"/>	Coordinate precision uncertainty mismatch	0
<input type="checkbox"/>	Footprint SRS invalid	2,488,256
<input type="checkbox"/>	Footprint WKT mismatch	2,107,702

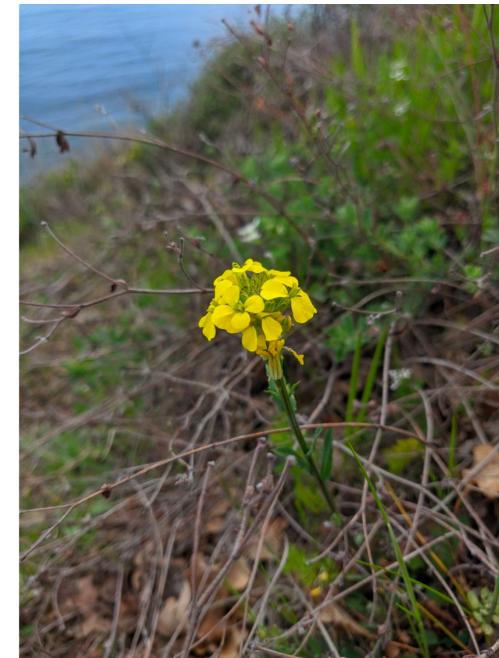


# [Art and Science] of Taxonomy

**Taxonomic homonyms:** when filtering based on scientific name, check if all the results are in the same part of taxonomic tree (at least kingdom)



*Cuspidaria cuspidata* (Olivi, 1792)



*Cuspidaria cuspidata* (M. Bieb.) Takht.

# Data Size

The size of the dataset you download can be substantial, depending on the extent of your query. Ensure you have the necessary storage and computational resources to handle the data.

Consider using [BigQuerry](#) or [Apache Spark](#) for interactions with really big data.

Only query what you need.



THE LD<sub>50</sub> OF TOXICITY DATA IS  
2 KILOGRAMS PER KILOGRAM.



Global Biodiversity  
Information Facility

## Accessing GBIF Data

# BY THE NUMBERS | 10<sup>th</sup> November 2025

Species occurrence records

**3 567 323 913**



Country Participants

**68**

Organizational Participants

**42**

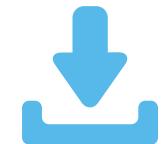
Datasets

**118 721**



Peer-review papers using data

**13 763**



Records downloaded per month in 2022

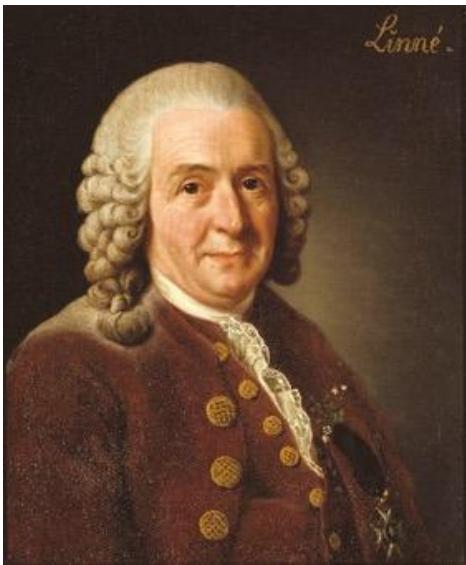
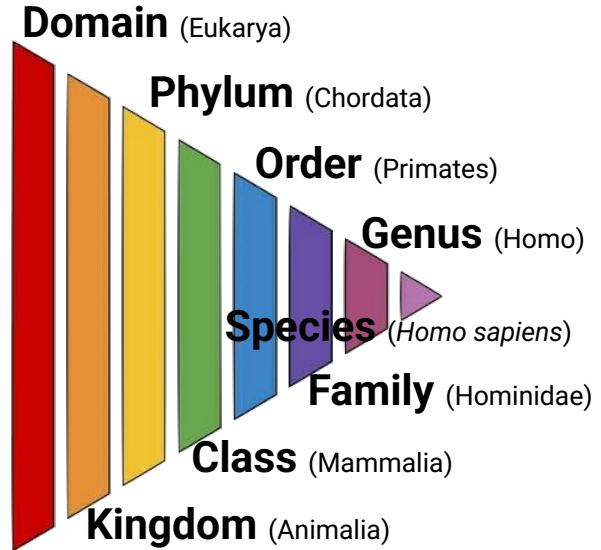
**>119 billion**

Publishers

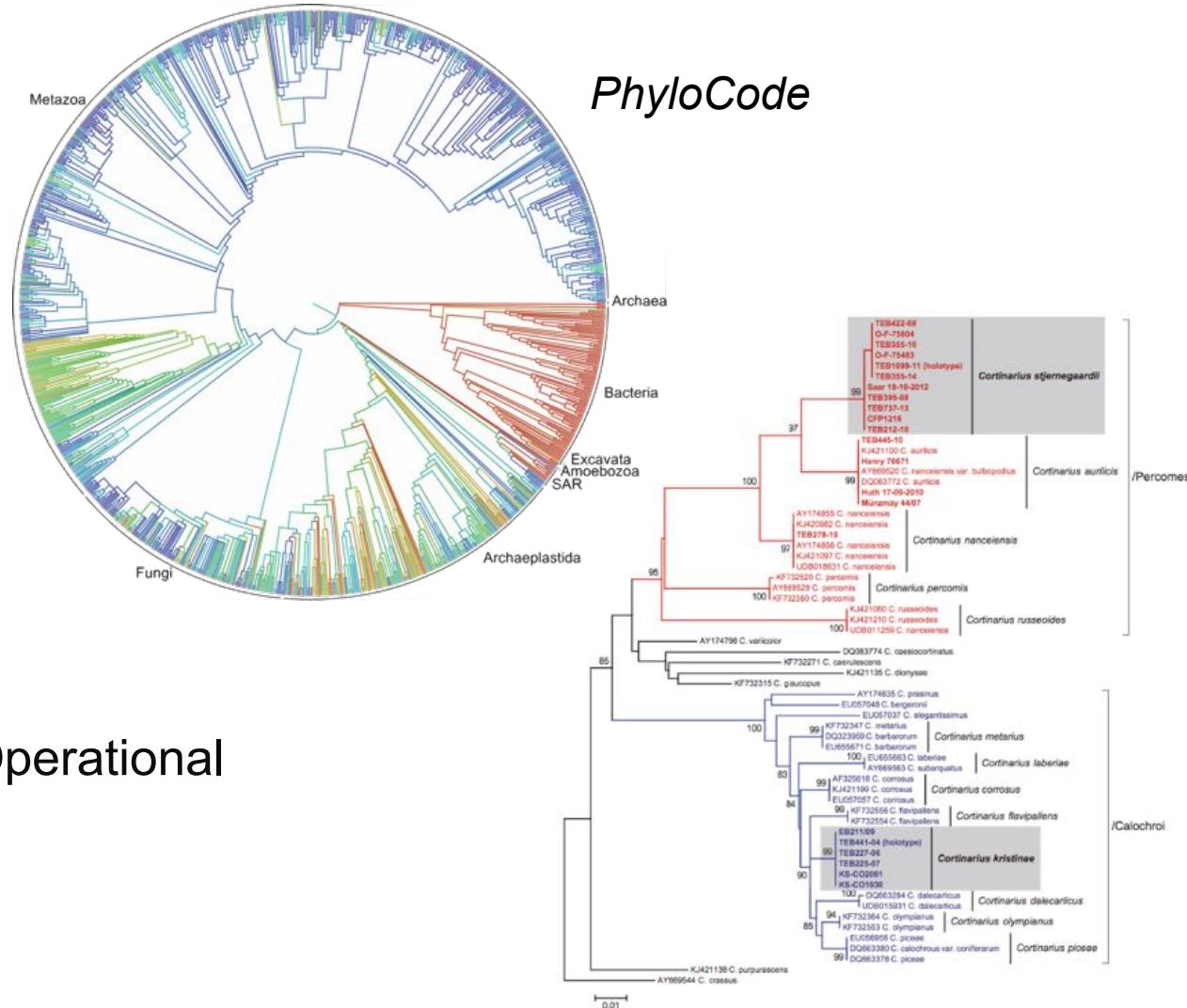
**2 645**



# How are Data Indexed in GBIF?



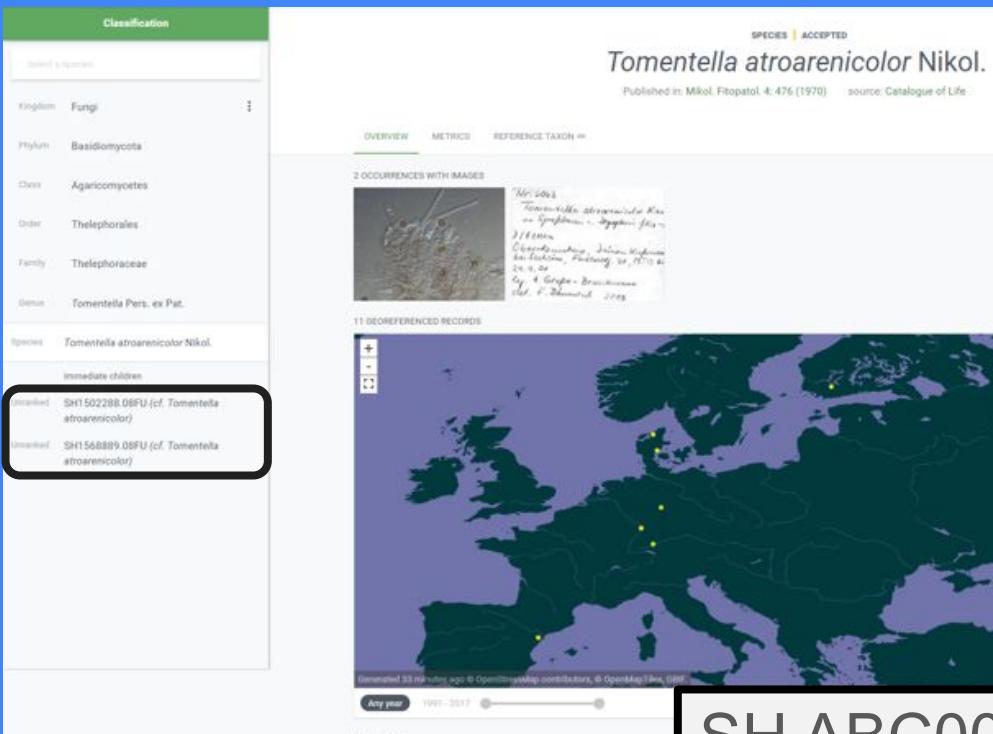
GBIF translates traditional nomenclature into **Operational Taxonomic Units (OTUs)**



# OTUs and the GBIF Backbone



Species Hypothesis (SH) numbers [DOI]

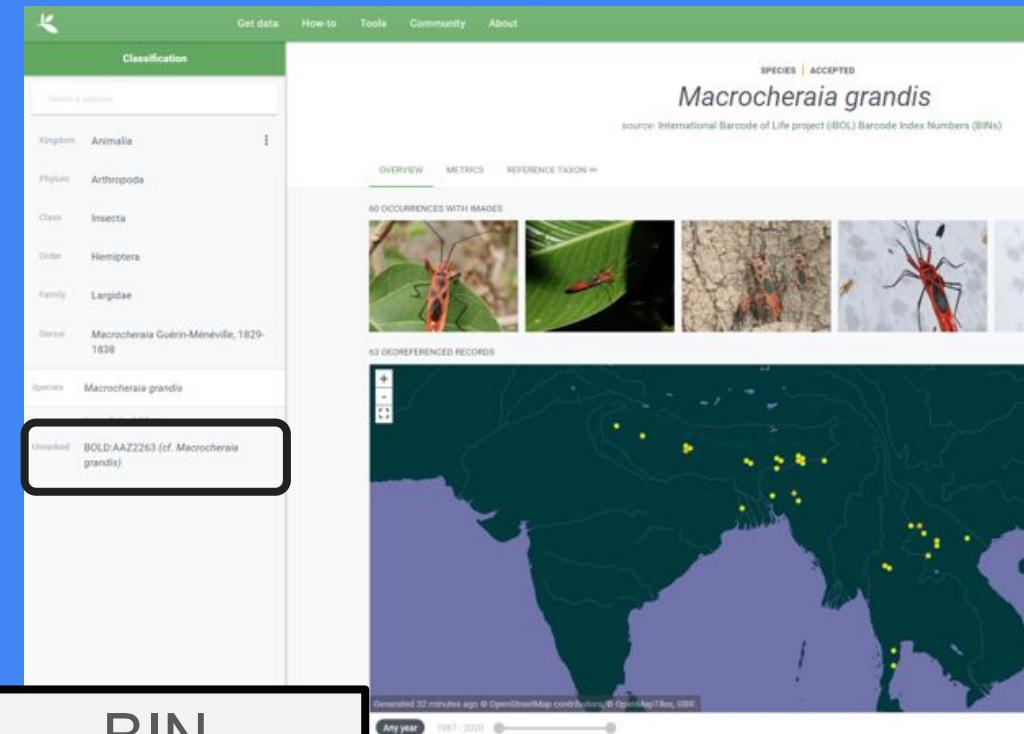


SH ABC0001

GBIF  
backbone  
taxonomy



BIN  
DEF0002



Barcode Identification Number (BIN)

international  
BARCODE  
OF LIFE

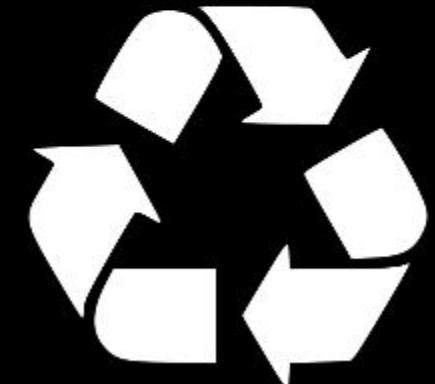
# Machine-Readability Requires Persistent Identifiers

*The purpose of identifiers is  
... to name things  
... making it possible to refer to them*

- To uniquely identify something it needs a **persistent identifier**, a PID.
- A Persistent Identifier is **globally unique, persistent, and resolvable**.
- A PID is resolvable when it allows both **human and machine** users to access an object or its representation, and its *Kernel Information*.
- **Kernel Information** is a structured record that contains information (metadata) about the referred object, such as a pointer to the location where the data for the object can be found.



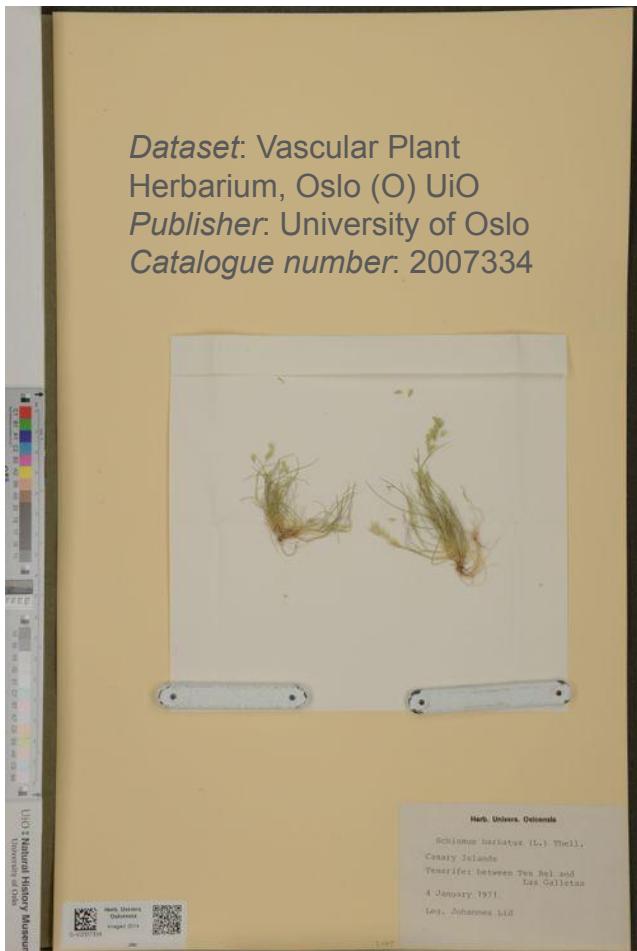
F indable A ccessible I nteroperable R eusable



FAIR data is about ***machine-readable*** data

*researchers & museums need to do more than simply post their data on the web for it to be **re-usable**.*

# GBIF & Identifiers



<https://www.gbif.org/occurrence/1095052193>

Get data How-to Tools Community About

Occurrence

Catalogue number	2007334
Occurrence ID	<a href="urn:catalog:O:V:2007334">urn:catalog:O:V:2007334</a>
Occurrence status	PRESENT
Other catalogue numbers	0574816d-3d99-41b8-b3b8-c6035de0e929
Recorded by	Johannes Lid
Recorded by ID	<a href="https://www.wikidata.org/wiki/Q94522">https://www.wikidata.org/wiki/Q94522</a>

Identification

Date identified	1971-01-04T00:00:00
Identified by	Johannes Lid
Identified by ID	<a href="https://www.wikidata.org/wiki/Q94522">https://www.wikidata.org/wiki/Q94522</a>

Location

Country or area	Spain
Country code	ES
Locality	Tenerife: Tenerife: between Ten Bel and Las Galletas.

Other

Identifier	<a href="urn:catalog:O:V:2007334">urn:catalog:O:V:2007334</a>
Record licence	<a href="http://creativecommons.org/licenses/by/4.0/legalcode">http://creativecommons.org/licenses/by/4.0/legalcode</a>
Modified	2015-05-04T00:00:00.000+00:00

Record

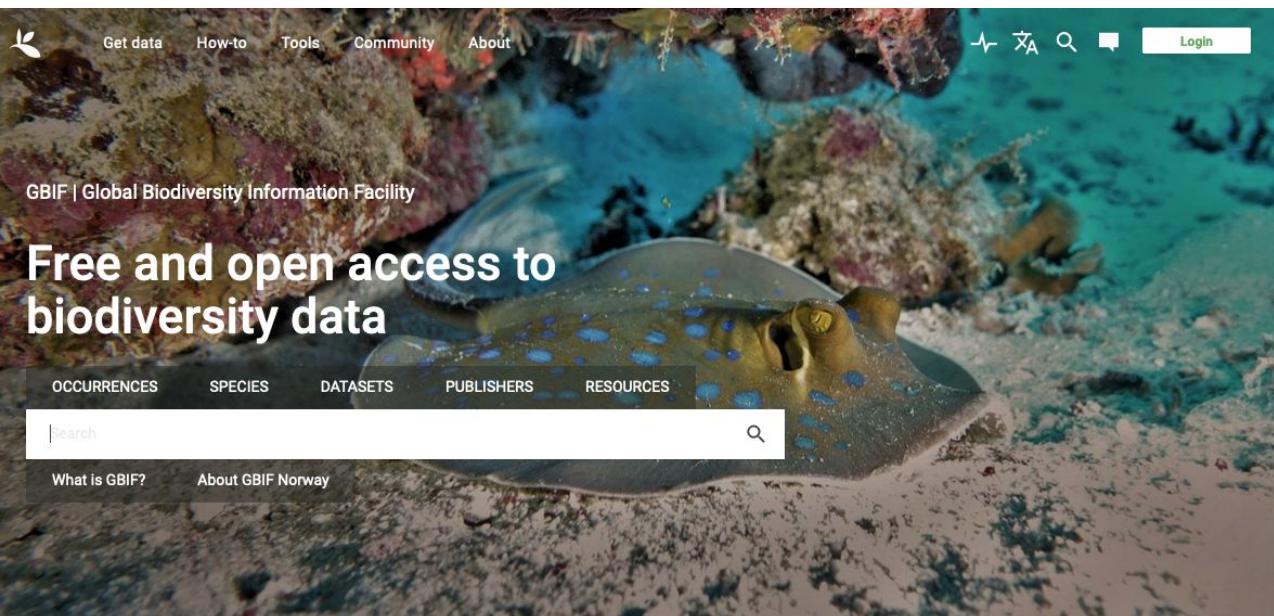
Basis of record	Preserved specimen
Collection code	V
Dataset name	Macaronesia
Institution code	O University of Oslo
Institution ID	<a href="https://ror.org/01xthb56">https://ror.org/01xthb56</a> University of Oslo

Event

Day	4
Month	1
Year	1971
Event date	1971-01-04T00:00:00

Taxon

Kingdom	Plantae
Phylum	Tracheophyta
Class	Liliopsida
Order	Poales
Family	Poaceae
Genus	Schismus
Specific epithet	barbatus
Infraspecific epithet	
Scientific name	Schismus barbatus (L.) Thell.
Rank	Species
Taxonomic status	Accepted



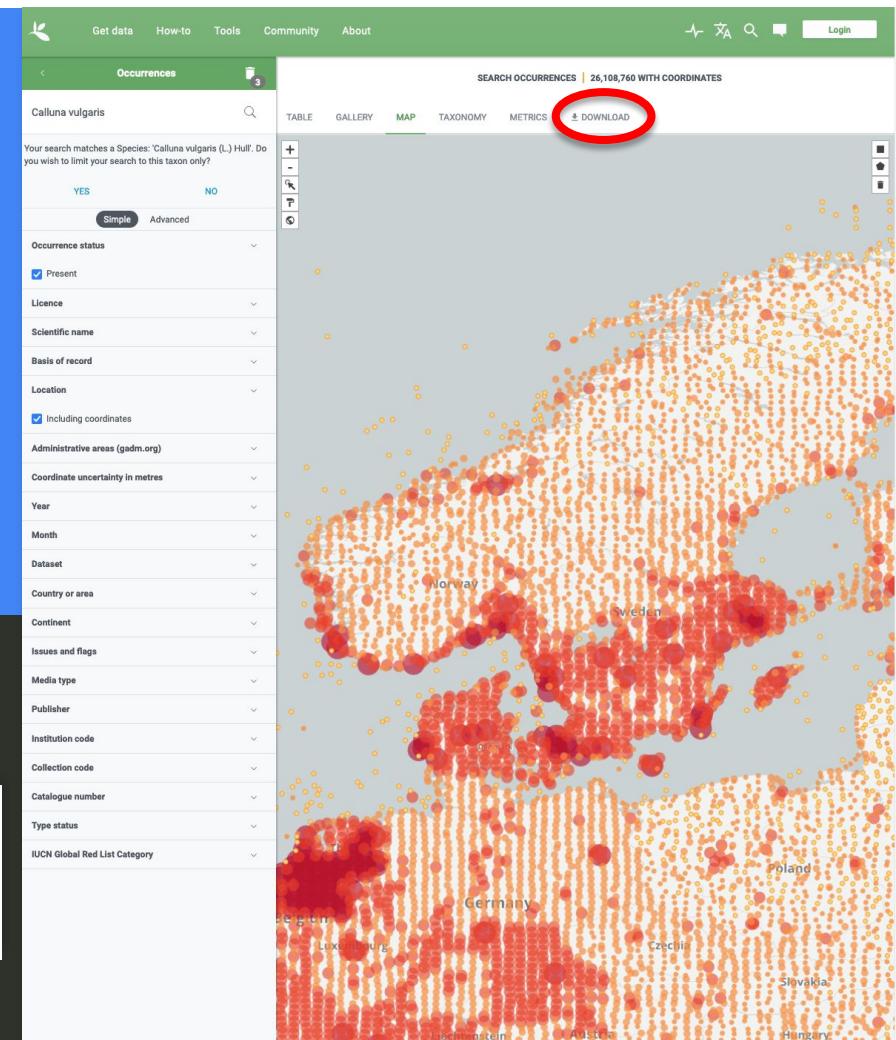
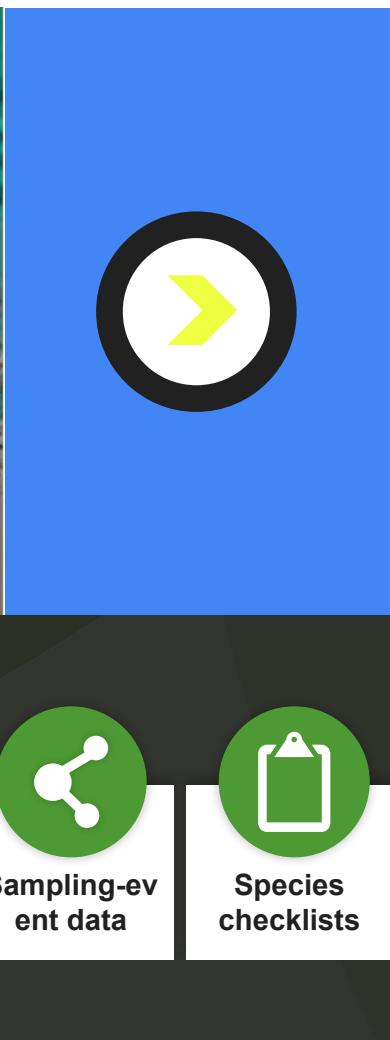
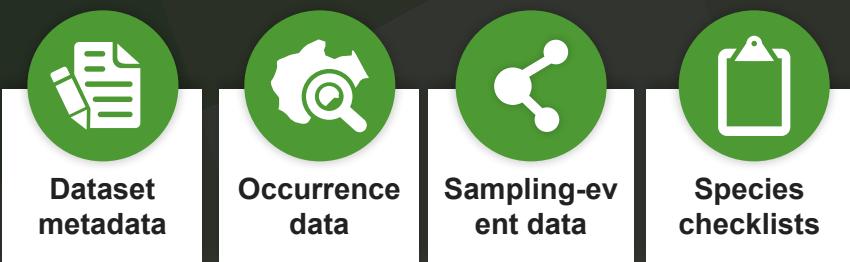
Free and open access to biodiversity data

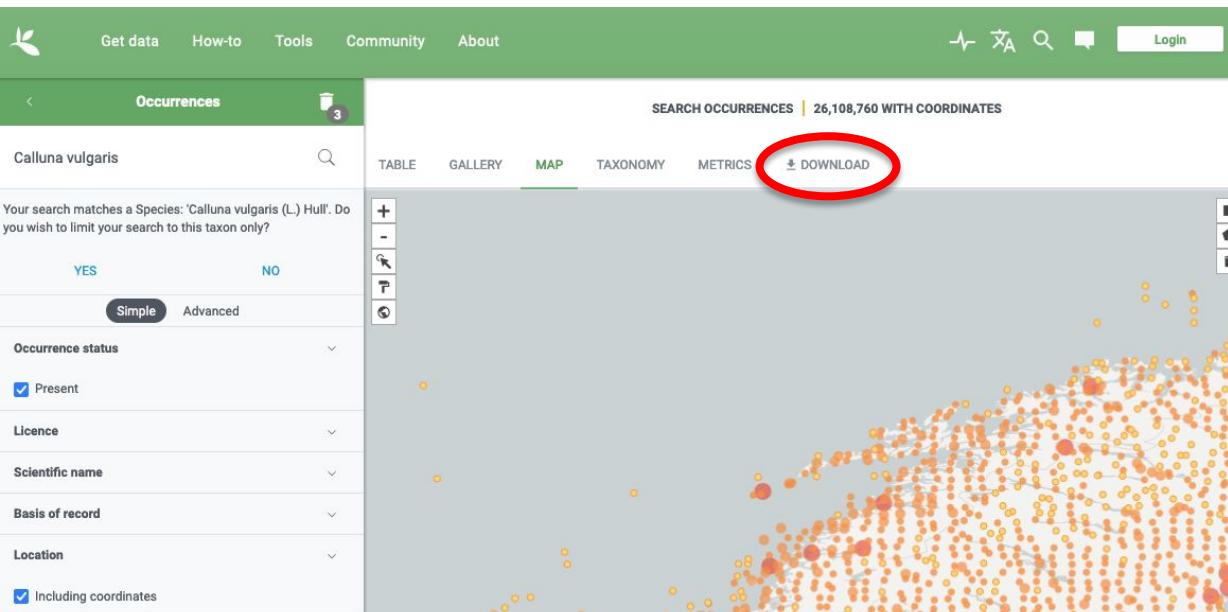
OCCURRENCES SPECIES DATASETS PUBLISHERS RESOURCES

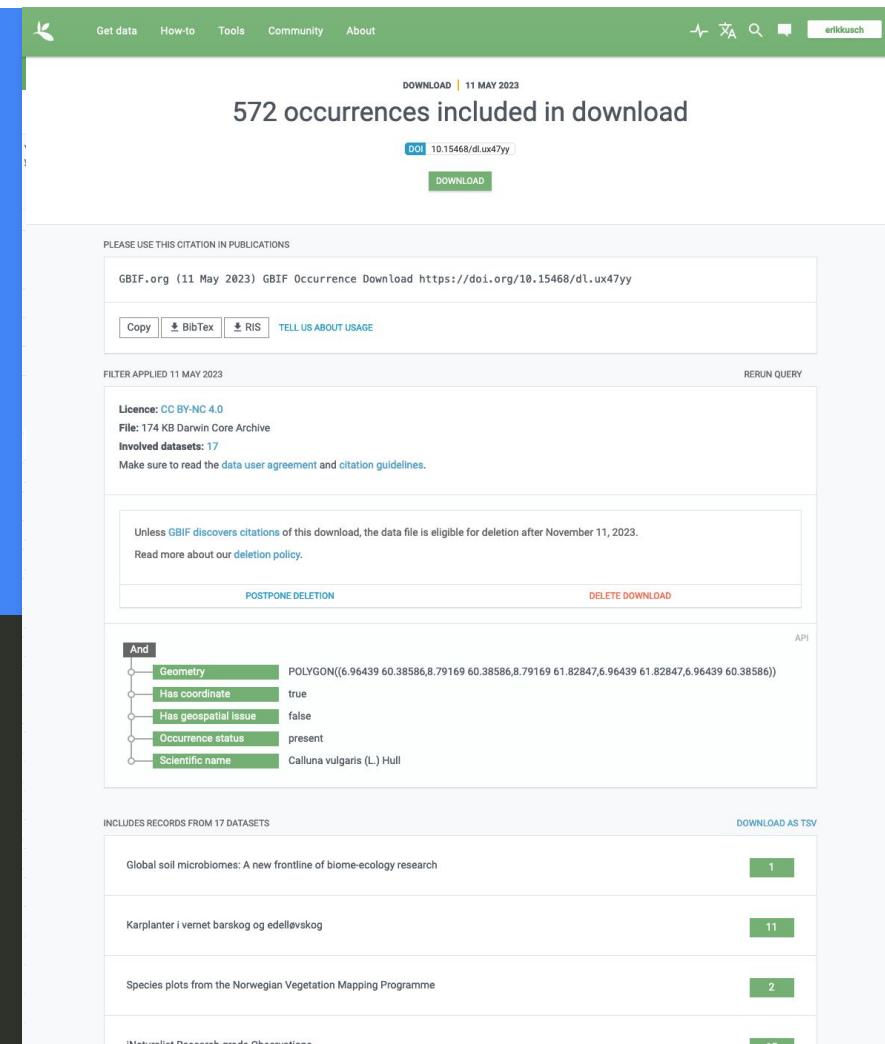
Search

What is GBIF? About GBIF Norway

- GBIF trawling for data via GBIF Data Portal
- Search can be refined via filters







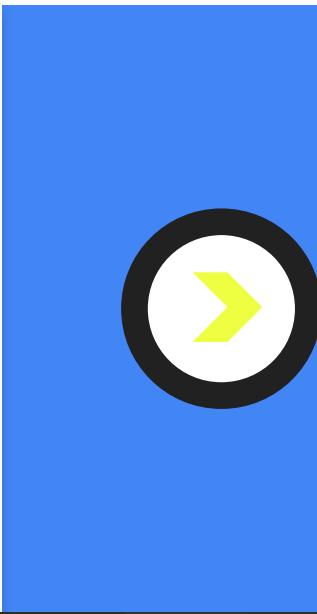
**Download through the GBIF Data Portal is a three-step process:**

1. Select desired data
2. Stage download & wait for GBIF to finish processing
3. Download final product

```

occ_initial      <- occ_search(
  taxonKey = sp_key,
  country = "NO",
  year = "2000,2022",
  basisOfRecord = "HUMAN_OBSERVATION",
  occurrenceStatus = "PRESENT"
)

```



```

res <- occ_download(
  pred("taxonKey", sp_key),
  pred("basisOfRecord", "HUMAN_OBSERVATION"),
  pred("country", "NO"),
  pred("hasCoordinate", TRUE),
  pred_gte("year", 2000),
  pred_lte("year", 2022)
)

```

```
res_data <- occ_download_import(res_get)
```

```

res_get <- occ_download_get(res)

## List of 12
## $ key           : chr "0256615-23022409556074"
## $ doi          : chr "10.15468/dl.awgk9s"
## $ license       : chr "http://creativecommons.org/licenses/by-nc/4.0/legalcode"
## $ request       :List of 5
## $ created       : chr "2023-05-22T19:49:39.661+00:00"
## $ modified      : chr "2023-05-22T19:50:53.445+00:00"
## $ eraseAfter    : chr "2023-11-22T19:49:39.577+00:00"
## $ status        : chr "SUCCEEDED"
## $ downloadLink  : chr "https://api.gbif.org/v1/occurrence/download/request/0256615-23022409556074.zip"
## $ size          : int 6513729
## $ totalRecords  : int 24904
## $ numberDatasets: int 47
## - attr(*, "class")= chr "occ_download_meta"
## - attr(*, "format")= chr "DWCA"

```

## Discovery & Download takes four R function calls:

1. Initial Data Search: `occ_search(...)`
2. Asynchronous download:
  1. `occ_download(...)`
  2. `occ_download_get(...)`
  3. `occ_download_import(...)`

# Long-Term Download Accessibility

We store all download files as long as possible.

The download metadata page will always resolve, but the file itself might be removed in the future.

We strive to store all downloads, but prioritize downloads that have been cited.





Global Biodiversity  
Information Facility

## Accrediting GBIF Data

# BY THE NUMBERS | 10<sup>th</sup> November 2025

Species occurrence records

**3 567 323 913**



Country Participants

**68**

Organizational Participants

**42**

Datasets

**118 721**



Records downloaded per month in 2022

**>119 billion**



Publishers

**2 645**



Peer-review papers using data

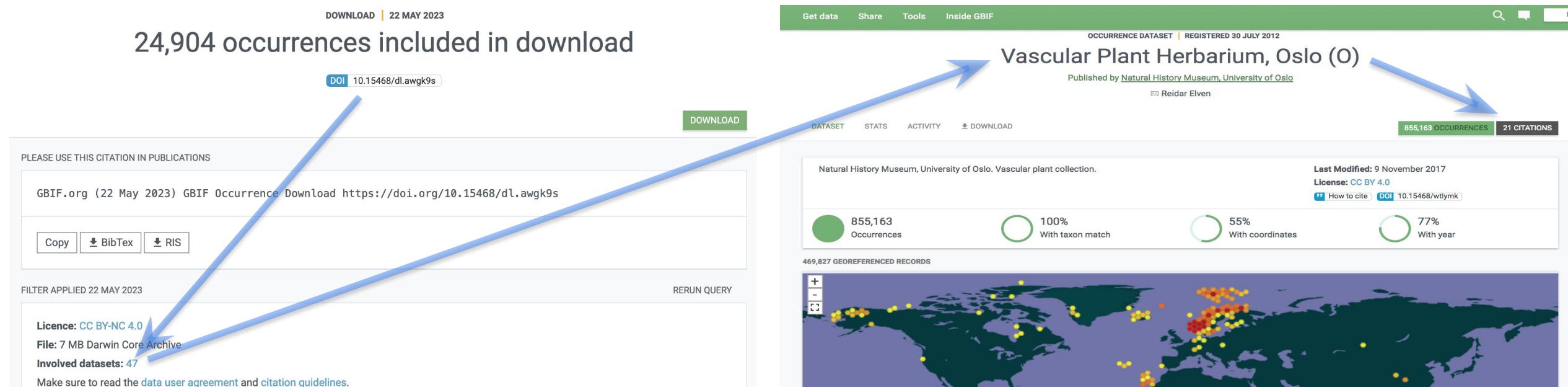
**13 763**

# How to Cite Data Mediated by GBIF

- 1. Download data from GBIF.org**
- 2. and receive recommended citation with a **download DOI****
- 3. Cite the DOI in published research or other work**

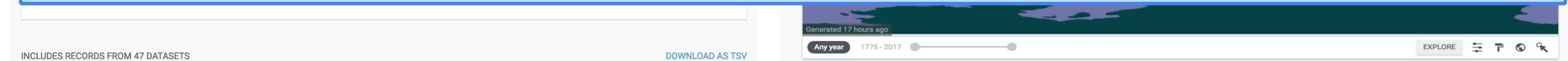
*Example:* GBIF.org (9 November 2021) GBIF Occurrence Download <https://doi.org/10.15468/dl.xxxxxxx>

# Downloads and Datasets are Assigned DOIs



Citing the data **download DOI** will resolve to the **dataset DOIs** assigned for each dataset contributing data records to the download set.

This way, **all data publishers** contributing data records **will be accredited!**





## Dataset DOIs

Source dataset #1

iD

ROR

UUID



Source dataset #2

iD

ROR

UUID



Source dataset #3

iD

ROR

UUID



## Download DOI

GBIF download

doi

WIKIDATA

UUID

iD

Filter & download

Process & archive

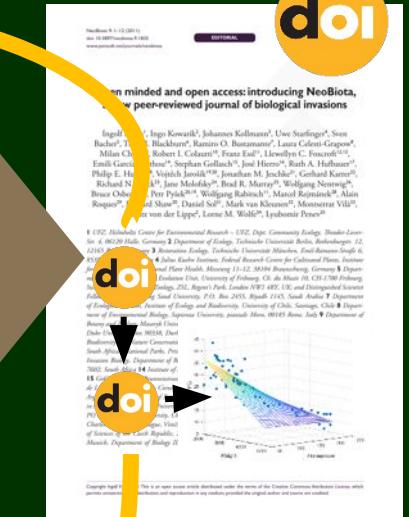
Final state of data

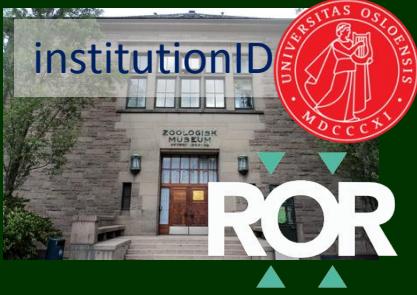
iD

UUID

zenodo doi

## Bibliographic DOI





## Dataset DOIs

Source dataset #1

doi

UUID

UUID

Source dataset #2

doi

UUID

Source dataset #3

doi

UUID

## Download DOI

GBIF download

doi

UUID

UUID

UUID

UUID

Process & archive

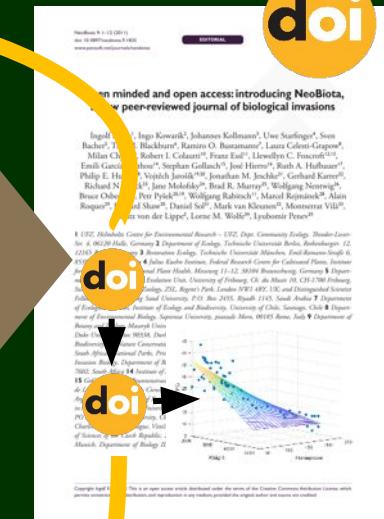
Final state of data

iD

UUID

zenodo doi

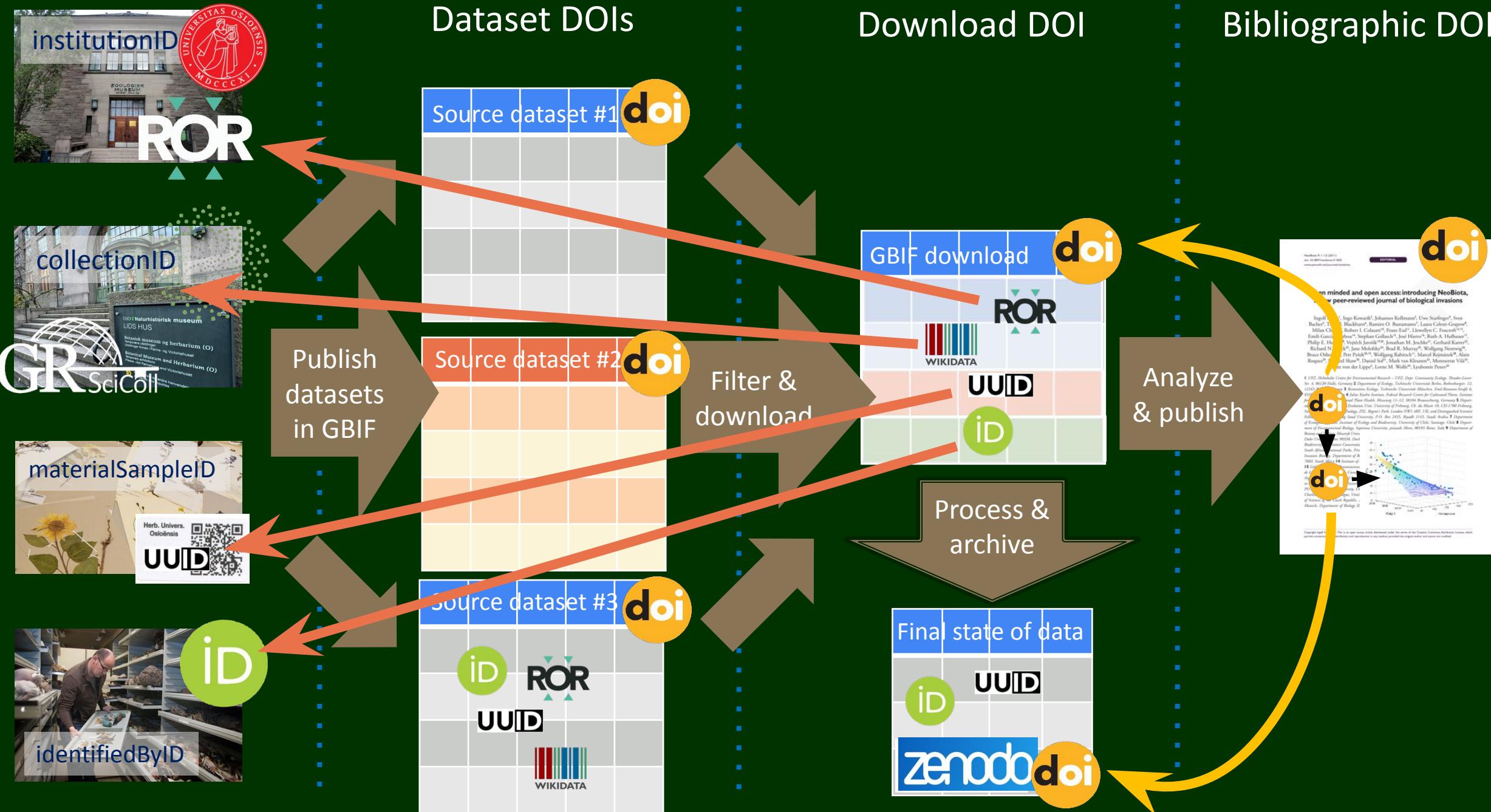
## Bibliographic DOI



## Dataset DOIs

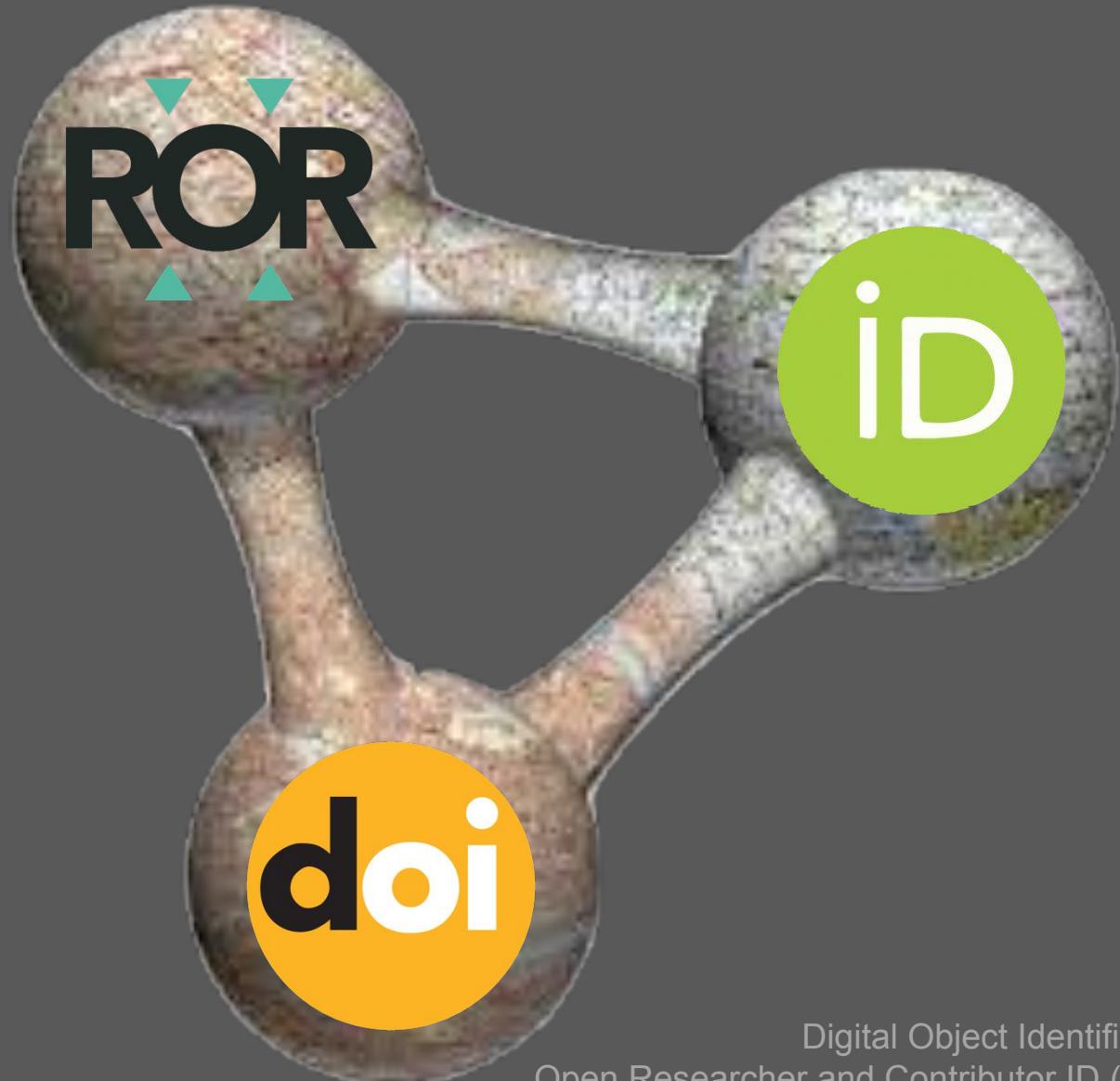
## Download DOI

## Bibliographic DOI



ROR for institutions  
ORCID for curators  
*DOI for datasets*  
(GRSciColl UUID for collections)

*will enable the linking of museum collection specimens to scientific litterature and scientific actors (authors, curators, etc)*



Digital Object Identifier (DOI)  
Open Researcher and Contributor ID (ORCID)  
Research Organisation Registry (ROR)

# Example workflow in R

```
library(rgbif)

Sys.setenv(
  GBIF_USER  = "your_username",
  GBIF_PWD   = "your_password",
  GBIF_EMAIL = "you@example.org"
)

# Submit download
taxon_key <- name_backbone("Amanita muscaria")$usageKey
req <- occ_download(pred("taxonKey", taxon_key))

# Wait
repeat {
  s <- occ_download_meta(req$key)$status
  message("Status: ", s)
  if (s == "SUCCEEDED") break
  if (s %in% c("FAILED", "KILLED")) stop("GBIF download failed: ", s)
  Sys.sleep(10)
}

# Download & unzip
zip_path <- occ_download_get(req$key, overwrite = TRUE)
unzip(zip_path, exdir = req$key)
```

# New GBIF CUBE SQL

- Docs <https://techdocs.gbif.org/en/data-use/data-cubes>
- <https://www.gbif.org/occurrence/download/sql#create>
- Available columns

[https://techdocs.gbif.org/en/data-use/api-sql\\_download#sql-columns](https://techdocs.gbif.org/en/data-use/api-sql_download#sql-columns)

```
library(rgbif)
occ_download_sql("SELECT datasetKey, countryCode, COUNT(*) FROM occurrence
WHERE continent = 'EUROPE' GROUP BY datasetKey, countryCode")
```

## API SQL Downloads

### NOTE

This is an **experimental feature**, and the implementation may change throughout 2025.

The screenshot shows a web-based interface for generating an SQL query. At the top, there's a navigation bar with links for 'Get data', 'How-to', 'Tools', 'Community', and 'About'. Below the navigation is a green header bar with the text 'CREATE' and 'ABOUT' (ABOUT is underlined). In the center, there's a large text area containing an SQL query:

```
SELECT
  countrycode,
  specieskey,
  species,
  COUNT(*) occ_count
FROM
  occurrence
WHERE
  occurrence.continent = 'EUROPE'
  AND CAST(occurrence.classkey AS INTEGER) = 131
  AND occurrence.specieskey IS NOT NULL
  AND occurrence.species IS NOT NULL
GROUP BY
  occurrence.countrycode,
  occurrence.specieskey,
  occurrence.species
```

Below the SQL editor is a green 'DOWNLOAD' button. At the bottom of the page, there's a note: 'The easiest way to download and explore data is via the occurrence search user interface. But for complex queries and aggregations, the SQL editor provides more freedom.' There are also buttons for 'Occurrence search' and 'SQL editor'.

# Exercise #1

Get all human observations of Rangifer tarandus (reindeer) in Norway from 2016–2025, only records with coordinates and no known geospatial issues.

```
library(rgbif)

reindeer_key <- name_backbone(name = "Rangifer tarandus")$usageKey

req1 <- occ_download(
  pred("taxonKey", reindeer_key),
  pred("country", "NO"),
  pred_gte("year", 2016),
  pred_lte("year", 2025),
  pred("basisOfRecord", "HUMAN_OBSERVATION"),
  pred("hasCoordinate", TRUE),
  pred("hasGeospatialIssue", FALSE),
  type = "and"
)
req1
```

# Exercise #2

Download occurrences for *Apis mellifera* OR *Bombus terrestris* in Norway, Sweden, or Finland during June–August, with images.

```
library(rgbif)

apis_key    <- name_backbone("Apis mellifera")$usageKey
bombus_key <- name_backbone("Bombus terrestris")$usageKey

req2 <- occ_download(
  pred_or(
    pred("taxonKey", apis_key),
    pred("taxonKey", bombus_key)
  ),
  pred_in("country", c("NO", "SE", "FI")),
  pred_in("month", c(6, 7, 8)),
  pred("mediaType", "StillImage"),
  pred("hasCoordinate", TRUE),
  type = "and"
)
req2
```

# Exercise #3

Within a polygon approximating part of Jotunheimen, get PRESERVED\_SPECIMEN records for vascular plants.

```
library(rgbif)

# Vascular plants (Tracheophyta) key:
tracheophyta_key <- name_backbone("Tracheophyta", rank = "phylum")$usageKey

# Rough WKT polygon (lon/lat) around a section of Jotunheimen
poly_wkt <- "POLYGON((7.2 61.2, 9.0 61.2, 9.0 61.9, 7.2 61.9, 7.2 61.2))"

req3 <- occ_download(
  pred("taxonKey", tracheophyta_key),
  pred_within(poly_wkt),
  pred("basisOfRecord", "PRESERVED_SPECIMEN"),
  pred("hasCoordinate", TRUE),
  pred("hasGeospatialIssue", FALSE),
  type = "and"
)
req3
```

# Exercise #4

Marine mammals anywhere in the world, depth  $\leq$  200 m, 2006–2025, coordinates only.

```
library(rgbif)

mammalia_key <- name_backbone("Mammalia", rank = "class")$usageKey
# Filter to families commonly marine via higher taxonomy is tricky;
# we keep Mammalia + depth filter as a simple marine proxy.
req4 <- occ_download(
  pred("taxonKey", mammalia_key),
  pred_gte("year", 2006),
  pred_lte("year", 2025),
  pred_lte("depth", 200),
  pred("hasCoordinate", TRUE),
  type = "and"
)
req4
```

# Exercise #5

Introduced occurrences of *Pinus radiata* in Spain or Portugal, 2015–2025, licensed CC0 or CC-BY only.

```
library(rgbif)

pinus_radiata_key <- name_backbone("Pinus radiata")$usageKey

req5 <- occ_download(
  pred("taxonKey", pinus_radiata_key),
  pred_in("country", c("ES", "PT")),
  pred_gte("year", 2015),
  pred_lte("year", 2025),
  pred_in("license", c("CC0_1_0", "CC_BY_4_0")),
  pred("establishmentMeans", "INTRODUCED"),
  pred("hasCoordinate", TRUE),
  type = "and"
)
req5
```