



*Alocasia odora* observed in Taiwan by Yaling Lin via iNaturalist.org (CC BY-NC 4.0)

# Biodiversity data – GBIF

Anne-Sophie Archambeau | Node manager/Chair NSG

## Vision

---

A world in which the best possible biodiversity data underpins research, policy and decisions.



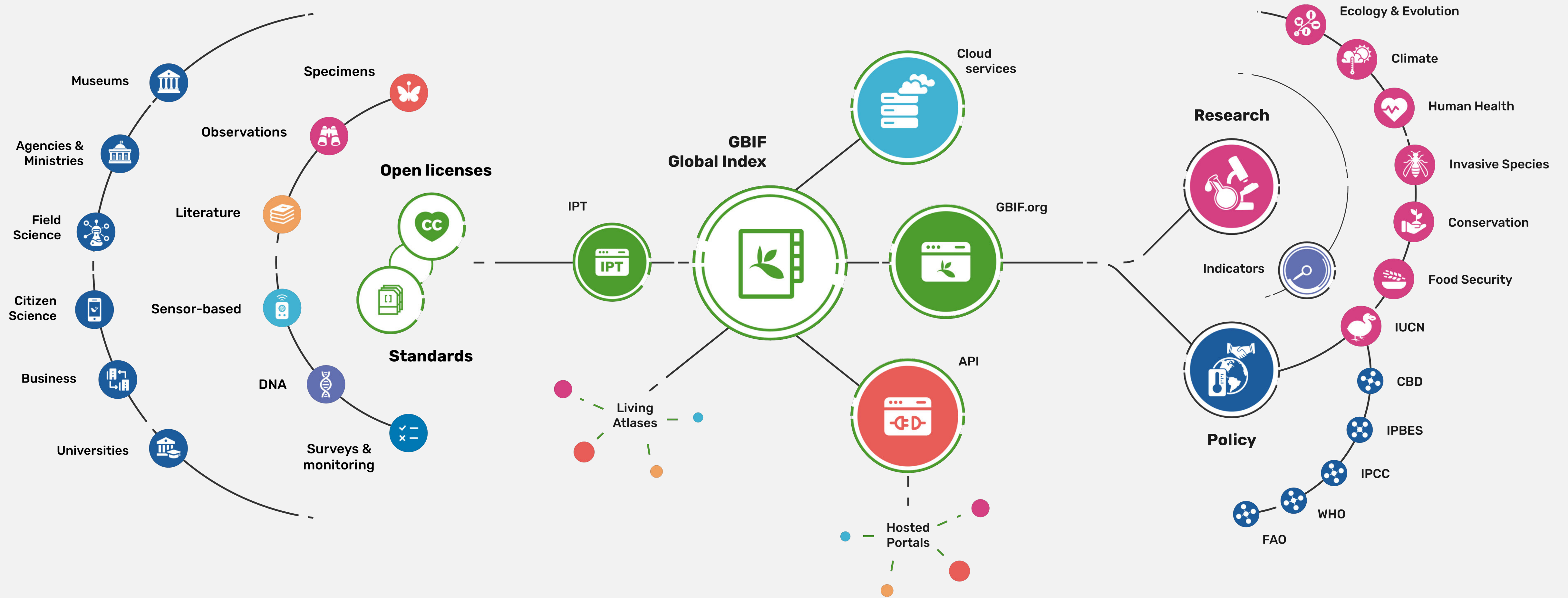
## Mission

---

To mobilize the data, skills and technologies needed to make comprehensive biodiversity information freely available for science and decisions addressing biodiversity loss and sustainable development



# Providing biodiversity evidence for research and policy



Datasets ●  
109,492

● Hosted portals  
23

Country  
Participants ●  
63

● Peer-review papers  
using data  
11,343

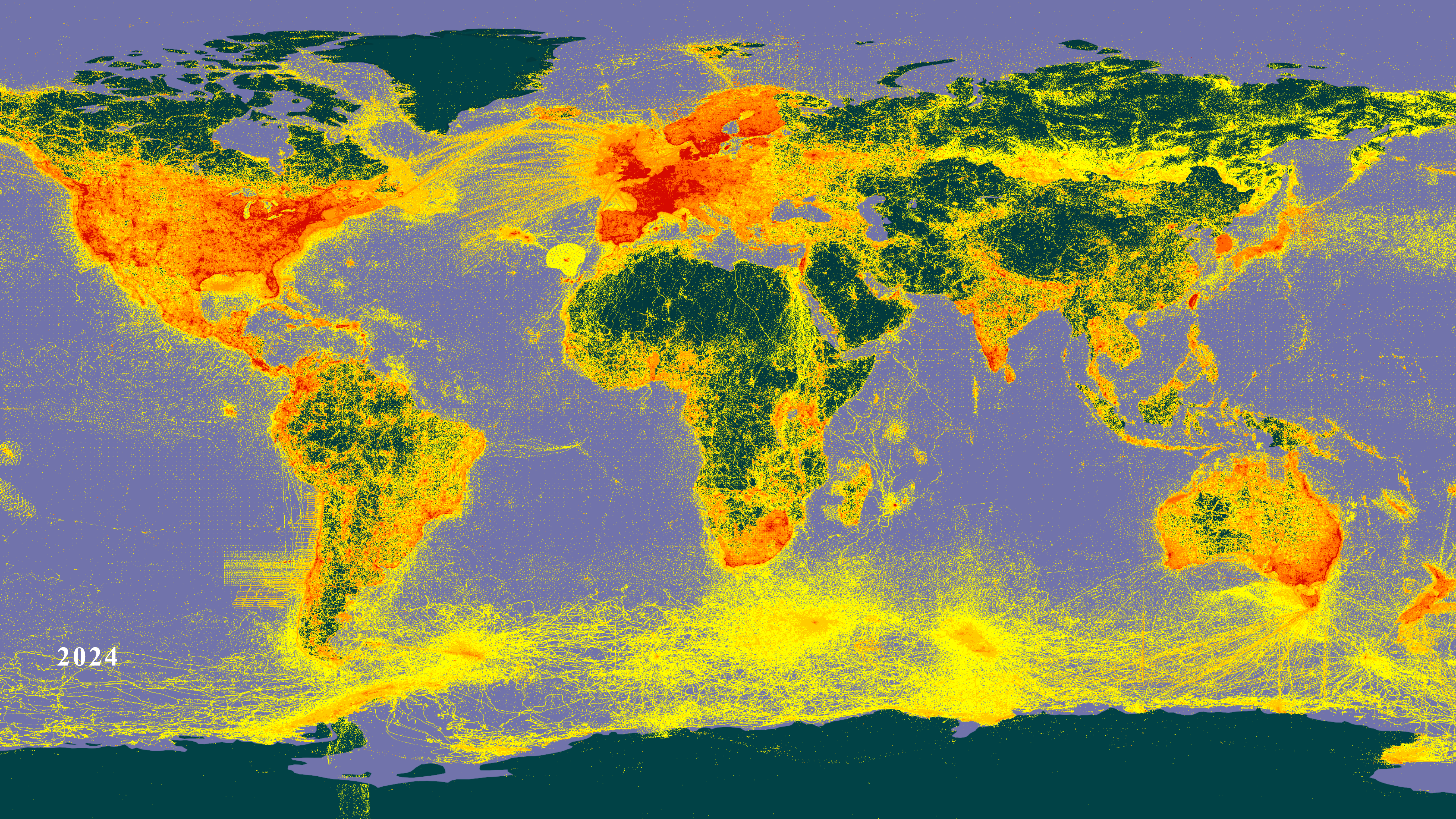
Organizational  
Participants ●  
43

● Average records  
downloaded per month (2024)  
201.5 billion

Publishers ●  
2,308

● Species  
occurrence records  
3,009,546,411





2024

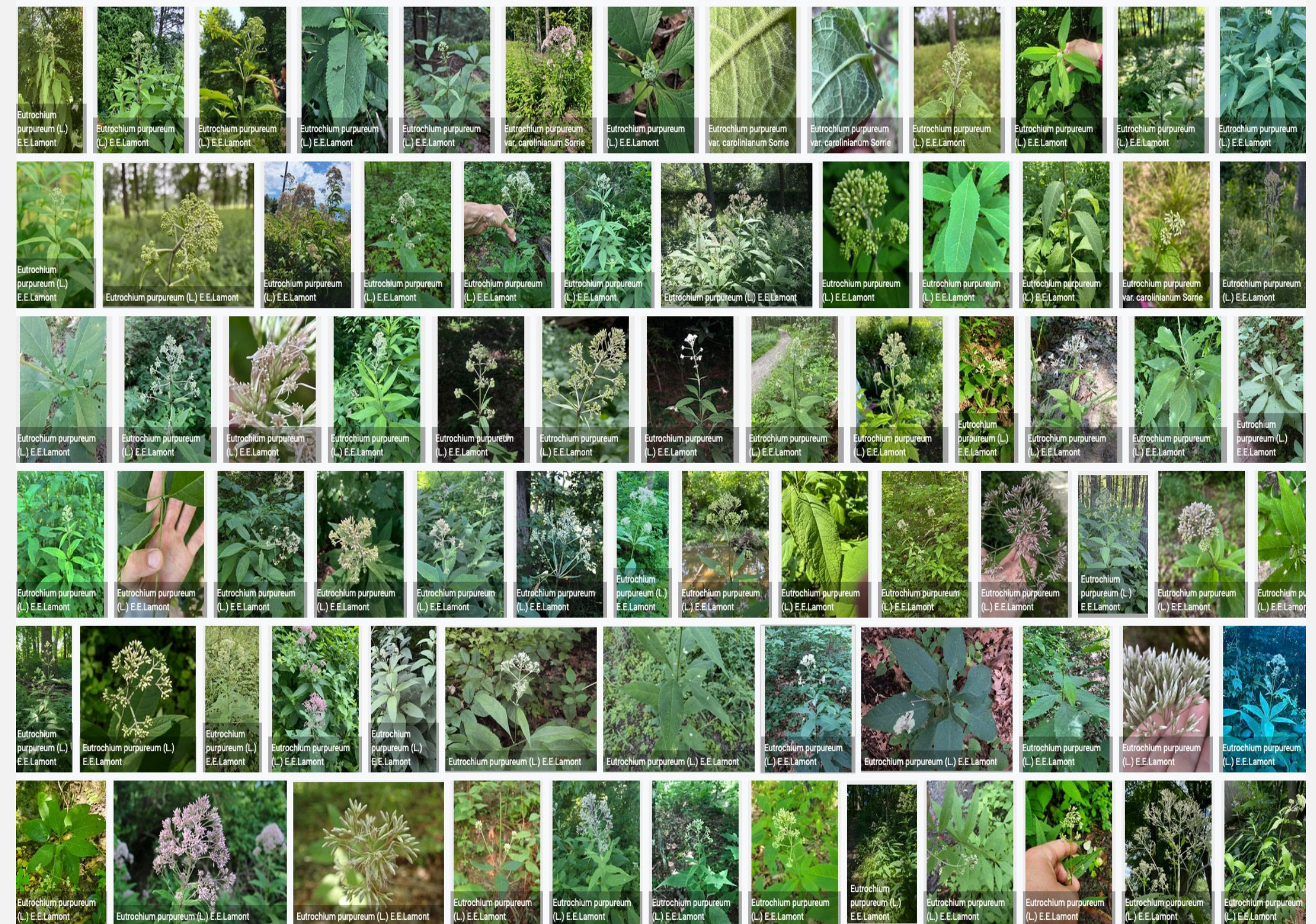
# Species occurrence records with multimedia evidence

**180.7 million records** with taxonomically identified images

- 120.6 million human observations
- 55.0 million specimens
- 2.8 million material samples
- 1.4 million fossil specimens

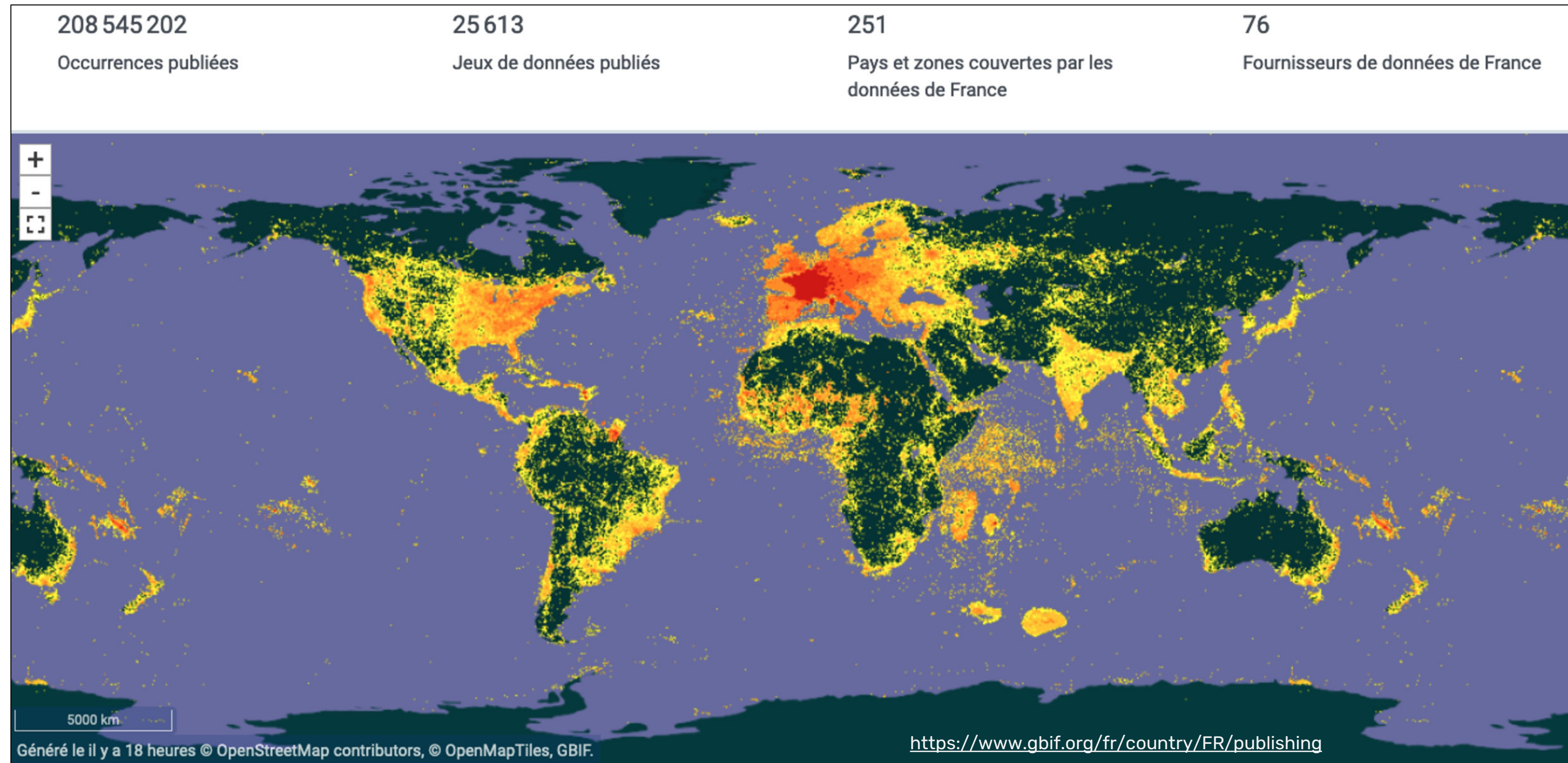
**1,420,991 audio files**

**10,418 videos**



# Data published by France

=> France is the 2<sup>nd</sup> country which provide data



Plus de 208,5 millions d'occurrences partagées en juin 2024 et 3 principaux éditeurs :

| Fournisseur de données                        | Compte    |
|---|-----------|
| ↻ UMS PatriNat (OFB-CNRS-MNHN), Paris         | 170352950 |
| ↻ Pl@ntNet                                    | 13856500  |
| ↻ MNHN - Museum national d'Histoire naturelle | 7744744   |

208,5 millions dont 13,89 Millions de données proviennent de la recherche  
- comprenant 11,79 millions de spécimens  
Et 20 millions des sciences participatives



# GBIF thematic priorities

2023 - 2027

## GBIF Strategic Framework

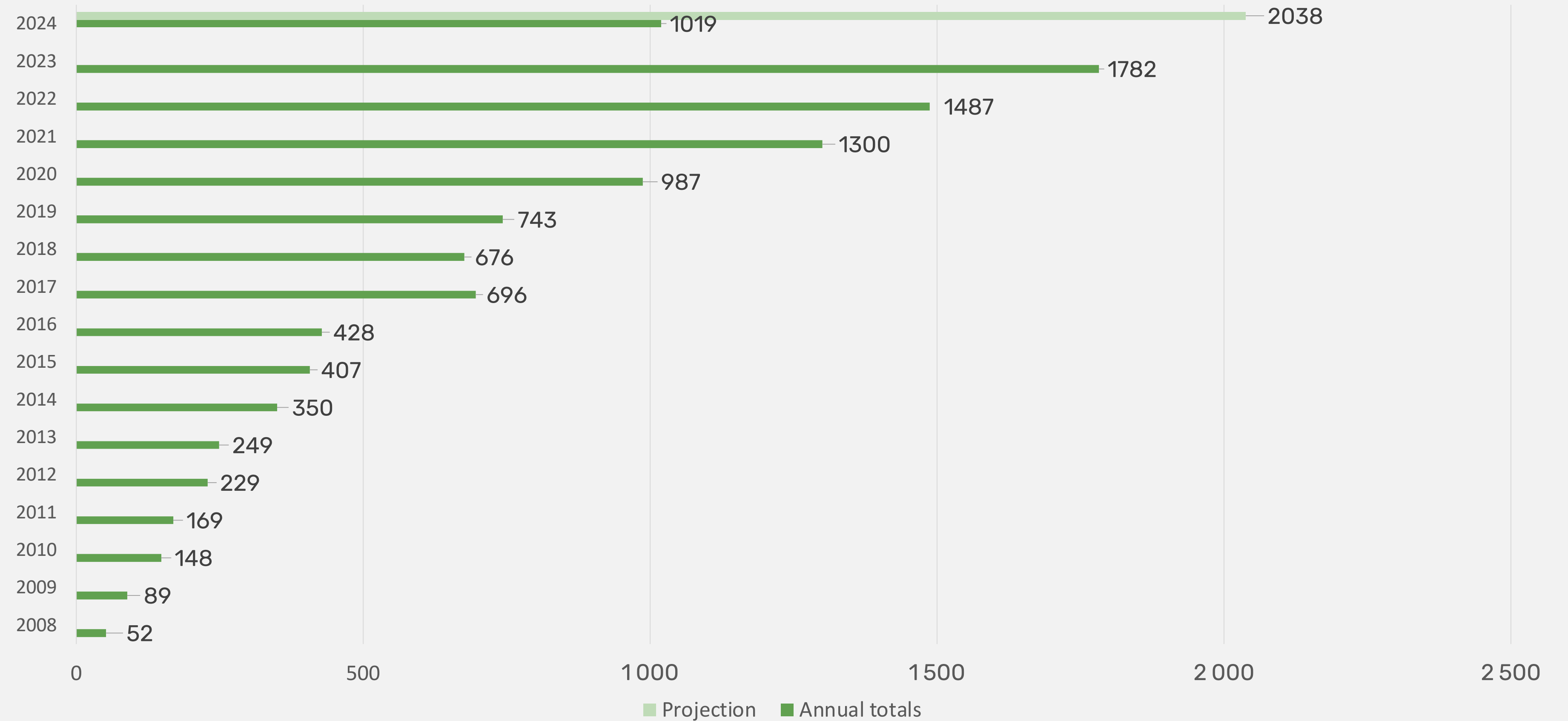




# Peer-reviewed publications using GBIF-mediated data



GBIF Secretariat (2024) GBIF Science Review No. 11.  
<https://doi.org/10.35035/d9pk-1162>

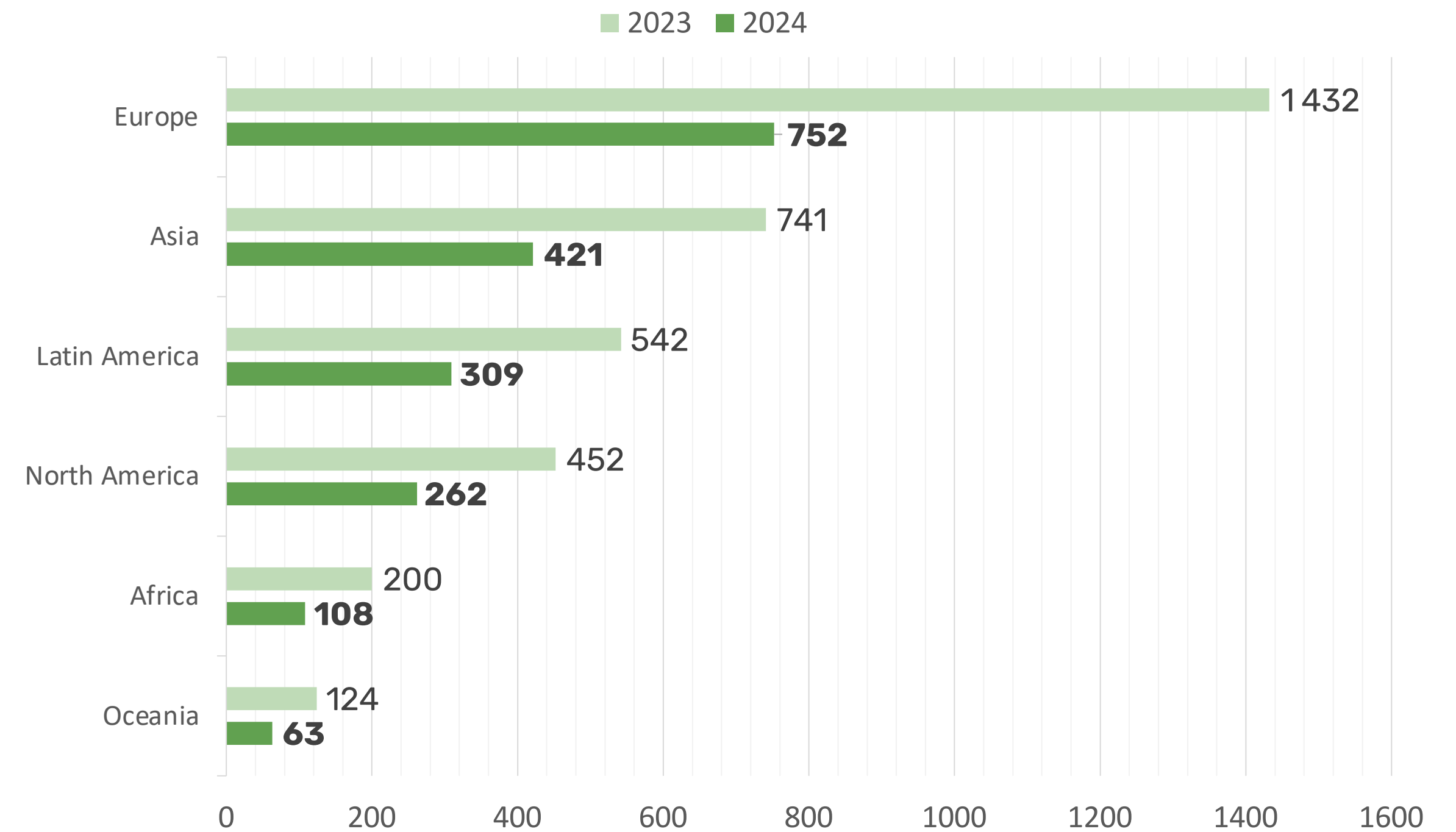


# Data use in peer-reviewed journals: 2024

Peer-reviewed uses by country

| End of Year |                       | 2024 total | 2023 total | 2023 rank |
|-------------|-----------------------|------------|------------|-----------|
| 1           | <b>China</b>          | <b>238</b> | 381        | 2         |
| 2           | <b>United States</b>  | <b>218</b> | 387        | 1         |
| 3           | <b>Brazil</b>         | <b>105</b> | 191        | 3         |
| 4           | <b>Germany</b>        | <b>95</b>  | 163        | 5<br>7    |
| 5           | <b>Mexico</b>         | <b>84</b>  | 129        | 7         |
| 6           | <b>United Kingdom</b> | <b>82</b>  | 182        | 4         |
| 7           | <b>Spain</b>          | <b>79</b>  | 130        | 6         |
| 8           | <b>Italy</b>          | <b>55</b>  | 79<br>108  | 10<br>8   |
| 9           | <b>Australia</b>      | <b>53</b>  | 94         | 9         |
| 10          | <b>France</b>         | <b>48</b>  | 108        | 8         |

Peer-reviewed uses by region





Resources 4

- France
- Country or area of coverage
- Literature type
- Journal article
- Relevance
- GBIF used
- Year of publication
- Topic
  - Agriculture 28
  - Biodiversity science 32
  - Biogeography 68
  - Citizen science 5
  - Climate change 123
  - Conservation 81
  - Data management 2
  - Data paper 16
  - Ecology 190
  - Ecosystem services 7
  - Evolution 95
  - Freshwater 16
  - Human health 16
  - Invasives 101
  - Marine 29
  - Phylogenetics 43
  - Species distributions 28

02

01

04

03

SEARCH RESOURCES | 636 RESULTS

ALL LITERATURE

DOWNLOAD AS TSV

Potential migration pathways of broadleaved trees across the receding boreal biome under future climate change Literature

Lima, J. Lenoir, J. Hylander, K. (2024) *Global Change Biology*  
 Climate change has triggered poleward expansions in the distributions of various taxonomic groups, including tree species. Given the ecological significance of trees as keystone species in forests and their socio-economic importance, projecting the potential future distributions of tree species is c...

broadleaved forest • climate adaptation • forestry • habitat suitability • range margins • range shift dynamics

Journal article Open access Peer-reviewed

Data referenced in work [DOI 10.15468/dl.27e5uv](#) [DOI 10.15468/dl.9xcvb7](#) [DOI 10.15468/dl.c88f6k](#)  
[DOI 10.15468/dl.g7g8qt](#) [DOI 10.15468/dl.qpzjwg](#) [DOI 10.15468/dl.shnu5h](#) [DOI 10.15468/dl.yh87rh](#)

Asymmetrical insect invasions between three world regions Literature

Isitt, R. Liebhold, A. Turner, R. Battisti, A. Bertelsmeier, C. Blake, R. ... - (2024) *NeoBiota*  
 AbstractThe geographical exchange of non-native species can be highly asymmetrical, with some world regions donating or receiving more species than others. Several hypotheses have been proposed to explain such asymmetries, including differences in propagule pressure, source species (invader) pools, ...

International trade • non-native insects • plants • propagule pressure • species pools

Journal article Open access Peer-reviewed

Accelerating and standardising IUCN Red List assessments with sRedList Literature

Cazalis, V. Di Marco, M. Zizka, A. Butchart, S. González-Suárez, M. Böhm, M. ... - (2024) *Biological Conservation*  
 The IUCN Red List of Threatened Species underpins much decision-making in conservation and plays a key role in monitoring the status and trends of biodiversity. However, the shortage of funds and assessor capacity slows the uptake of novel data and techniques, hamperi

Biodiversity conservation • Decision support • Extinction risk • GBIF • National Red List • Red List

Capture d'écran

[https://www.gbif.org/fr/resource/search?contentType=literature&literatureType=journal&relevance=GBIF\\_US&countriesOfResearcher=FR&peerReview=true](https://www.gbif.org/fr/resource/search?contentType=literature&literatureType=journal&relevance=GBIF_US&countriesOfResearcher=FR&peerReview=true)



# A new framework to combat the nature emergency



BIODIVERSITY CONVENTION CARTAGENA PROTOCOL NAGOYA PROTOCOL COUNTRIES PROGRAMMES

## KUNMING-MONTREAL GLOBAL BIODIVERSITY FRAMEWORK

### GBF HOME

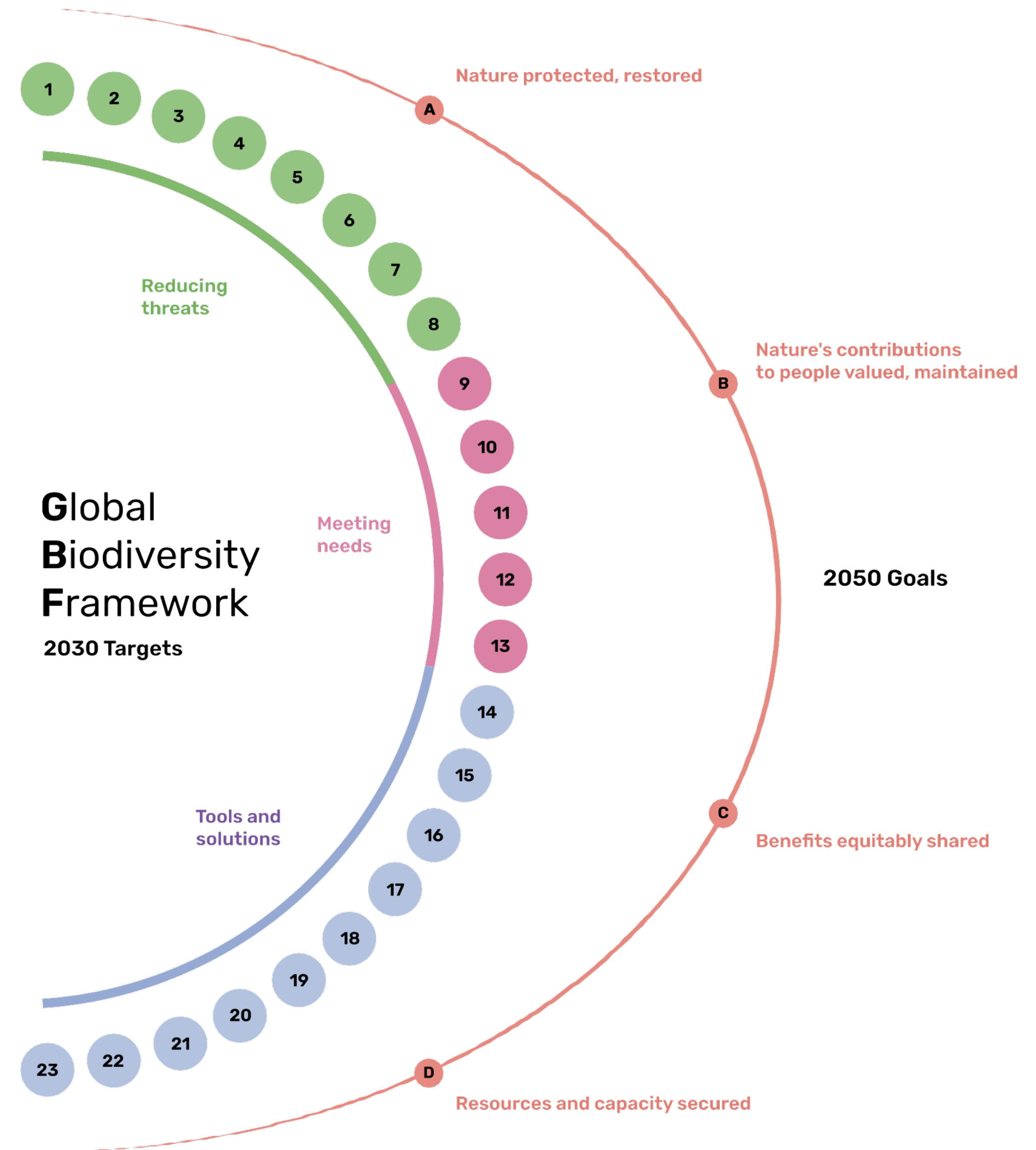
- INTRODUCTORY SECTIONS OF THE GBF
- 2050 VISION AND 2030 MISSION
- 2050 GOALS
- 2030 TARGETS (WITH GUIDANCE NOTES)
- IMPLEMENTATION AND SUPPORT MECHANISMS
- RESPONSIBILITY AND TRANSPARENCY
- COMMUNICATION, EDUCATION, AWARENESS AND UPTAKE
- RELATED DECISIONS

GBF HOME

WEDNESDAY // 9.6.2023

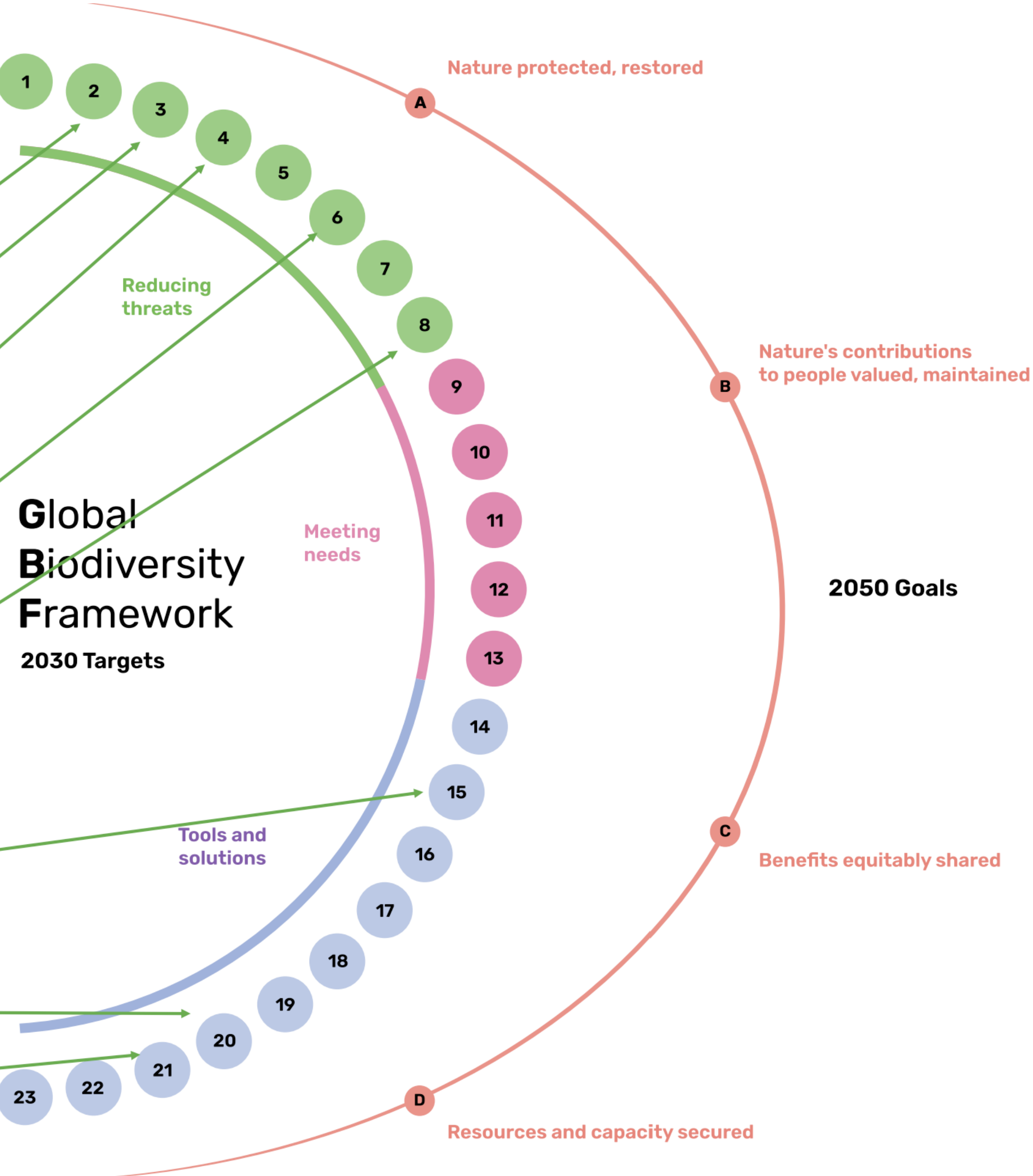


<https://www.cbd.int/gbf/>

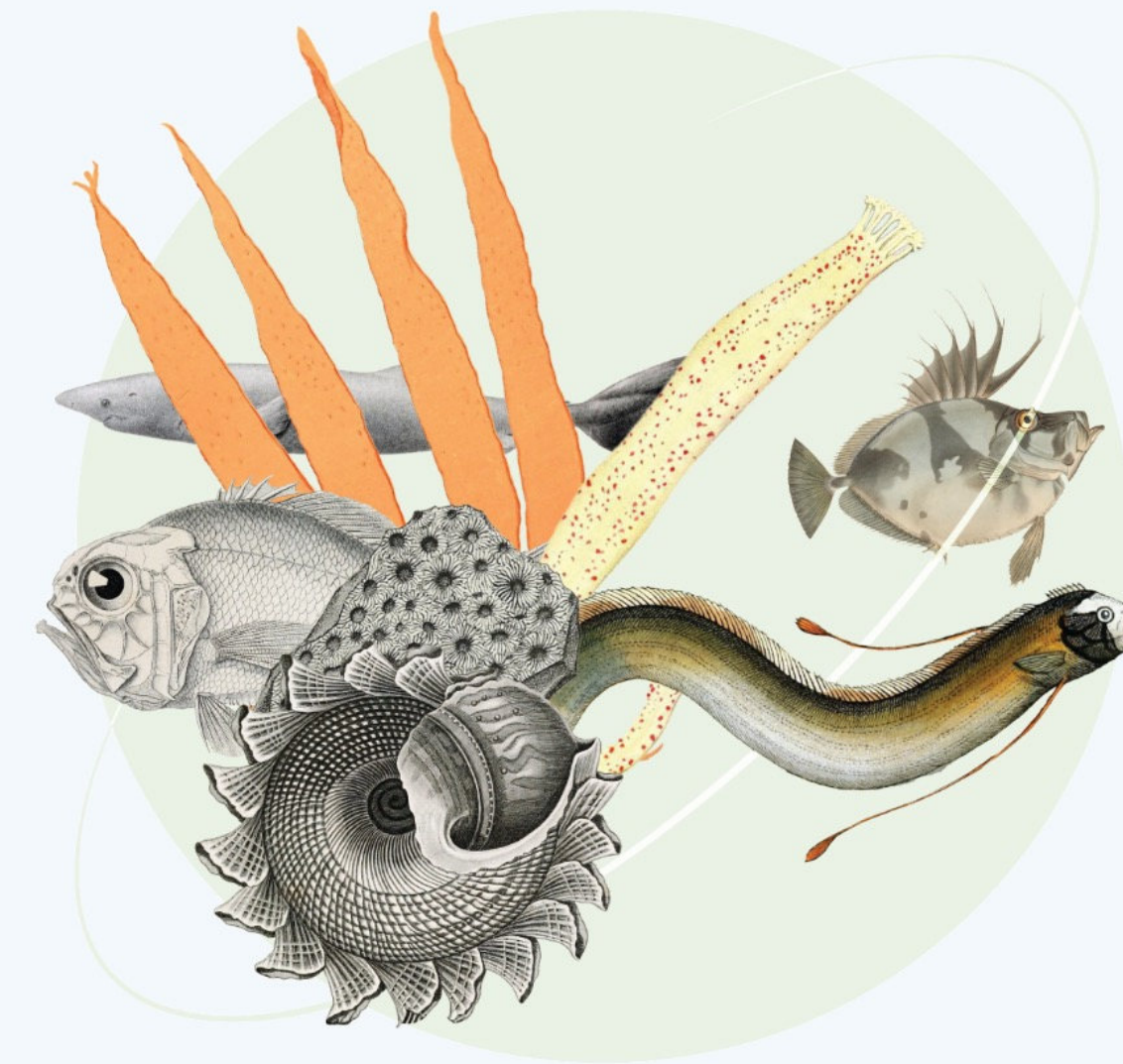
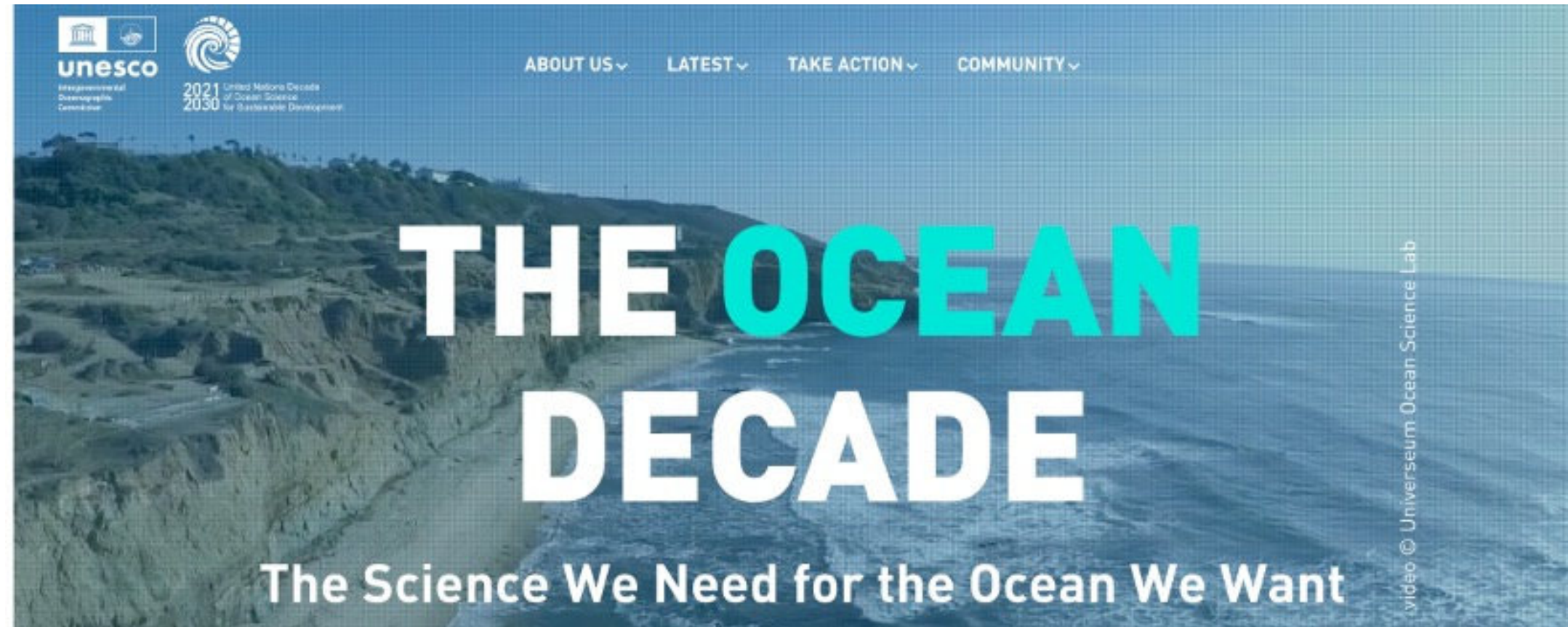


# A new framework to combat the nature emergency

| GBIF relevance                                    |
|---|
| Data to identify key biodiversity areas           |
| Data to monitor restoration                       |
| Data to locate, monitor protected areas           |
| Data for species conservation                     |
| Data on invasive species occurrence               |
| Data to model climate change impacts              |
| Platform for sharing EIA data                     |
| Capacity programmes for data mobilization and use |
| Core business of GBIF                             |



# GBIF and OBIS



## OBIS-GBIF Joint Strategy for Marine Biodiversity Data

OBIS • GBIF

Version febc05a, 2024-05-27 15:11:00 UTC

Webinars of 13 Juin 2024 available online



<https://www.gbif.org/fr/news/7w0xoodpNZRwnt1SjWfML0/obis-and-gbif-endorse-joint-strategy-and-action-plan-for-marine-biodiversity-data>





# ChecklistBank

## Taxonomic data infrastructure for everyone

The screenshot shows the ChecklistBank homepage. At the top left is the ChecklistBank logo and a navigation menu with 'Datasets', 'Tools', and 'COL24.7'. A 'Login' button is in the top right. The main header features a banner with the ChecklistBank logo and the tagline 'Index and repository for taxonomic data' over a background image of colorful flowers. Below the banner is a statistics table:

| Species in Catalogue of Life | Name Usages in ChecklistBank | Datasets in ChecklistBank | Latest COL Checklist |
|------------------------------|------------------------------|---------------------------|----------------------|
| 2,177,735                    | 350,244,666                  | 53,447                    | 2024-07-18           |

Below the table is a paragraph about the Catalogue of Life (COL) and GBIF's collaboration. To the right, a section titled 'Latest datasets added' lists three recent taxonomic updates. At the bottom, there is a footer with version information and logos for GBIF and the Global Core Biodata Resource.

Open data, open access taxonomic data publishing platform

### Tools for:

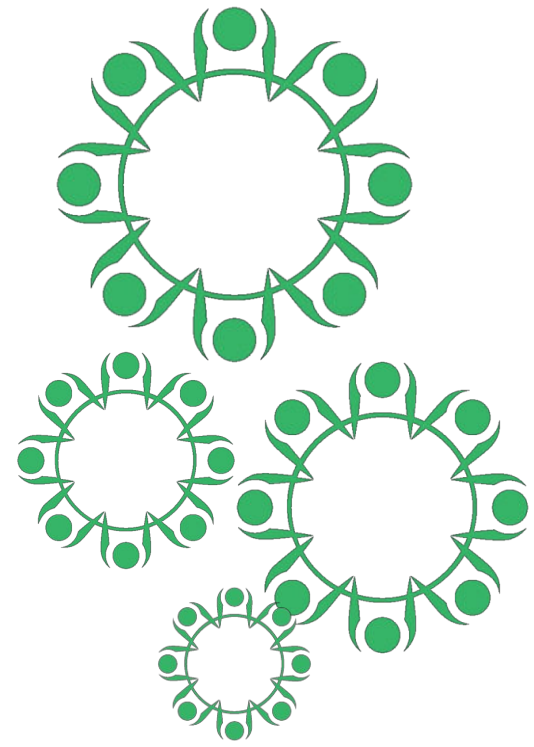
1. Importing lists
2. Comparing lists
3. Building lists

[checklistbank.org](https://checklistbank.org)

[api.checklistbank.org](https://api.checklistbank.org)



## Global checklists



*\*Management classification*

*\*Attach sectors*

*\*Editorial decisions*

*\*Persistent name ID's*



**Release**



**Catalogue of Life**



## Regional and management checklists



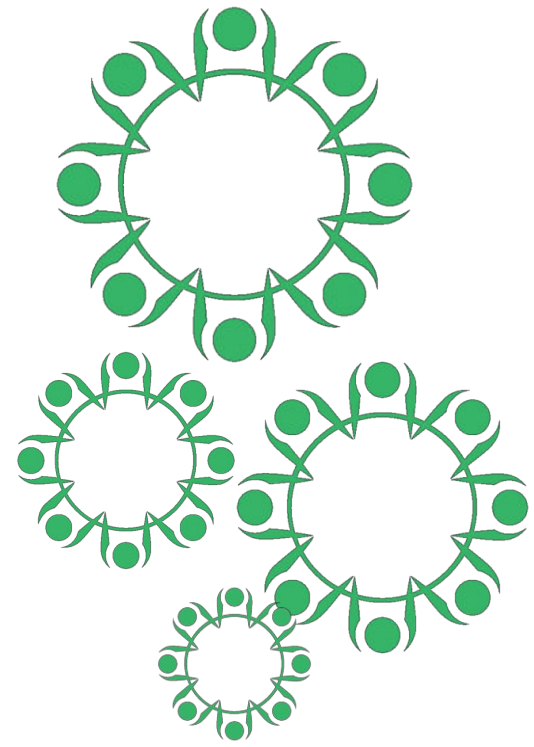
**ChecklistBank**

**A source** for  
the Backbone  
Taxonomy





## Global checklists



*\*Management classification*

*\*Attach sectors*

*\*Editorial decisions*

*\*Persistent name ID's*



**Release**

## Regional and management checklists



*\* Merge sectors*

*\* family and below*

*\* Homotypic grouping*

*\* Editorial decisions*

*\* Quality content check*



**Catalogue of Life**



**The only**  
source for the  
Backbone  
Taxonomy

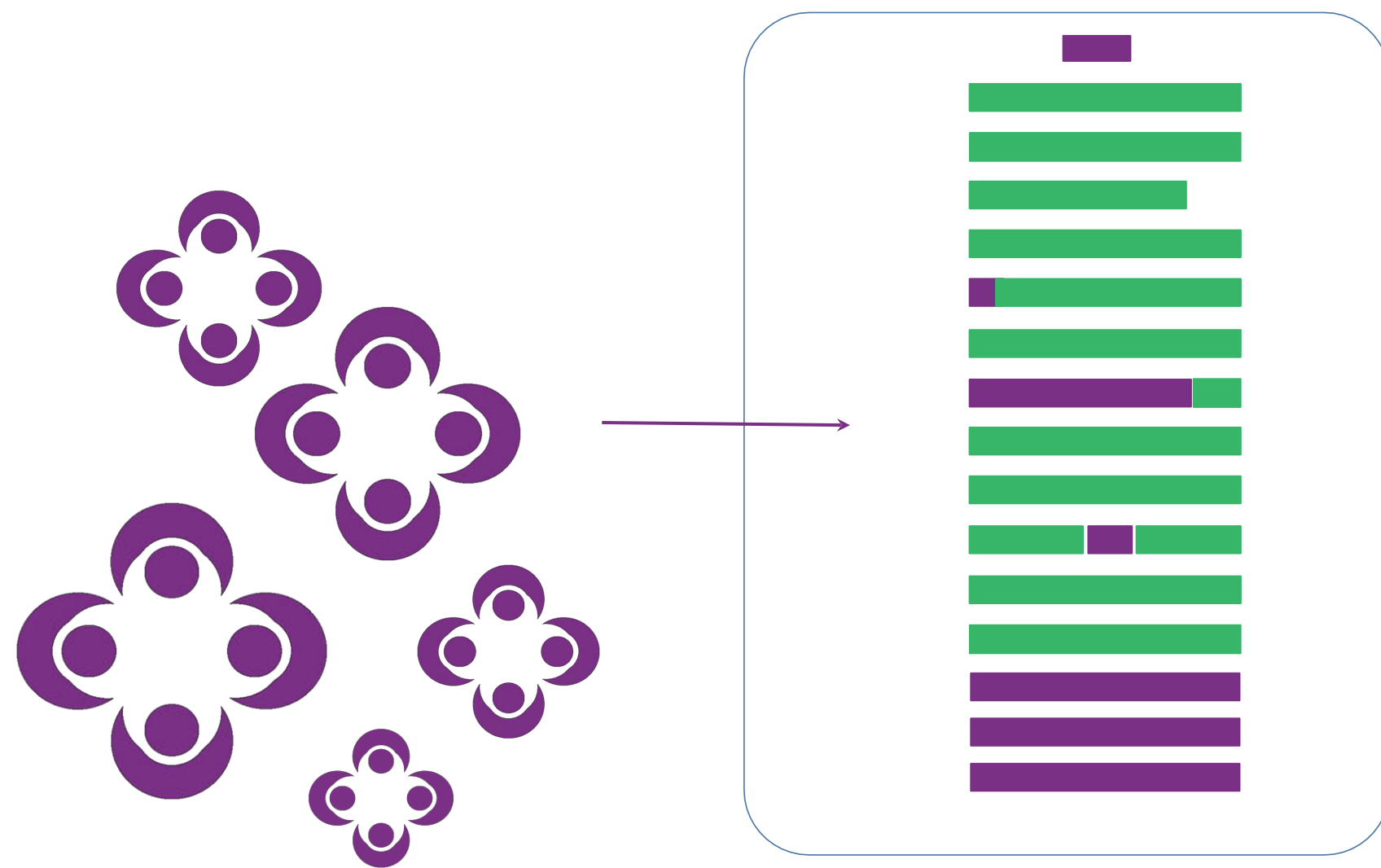


**ChecklistBank**



# You can contribute to a better COL

## 1. Publish or recommend a checklist



## 2. Issues report

Missing Chrysididae species #622

Open mdoering opened this issue last month · 3 comments

mdoering commented last month

There are quite a few species missing from the [Chrysididae](#) family in COL 24.1

The following species were existing in the SANBI species dictionary, but not found in COL:

- Acrotoma arnoldi
- Allocoelia bidens
- Allocoelia glabra
- Allocoelia latinota
- Allocoelia mocsaryi
- Brugmoia torrida
- Chrysidea minima
- Chrysidea pumila
- Chrysis alecto
- Chrysis alternans
- Chrysis angustula

<https://github.com/CatalogueOfLife/data/issues>

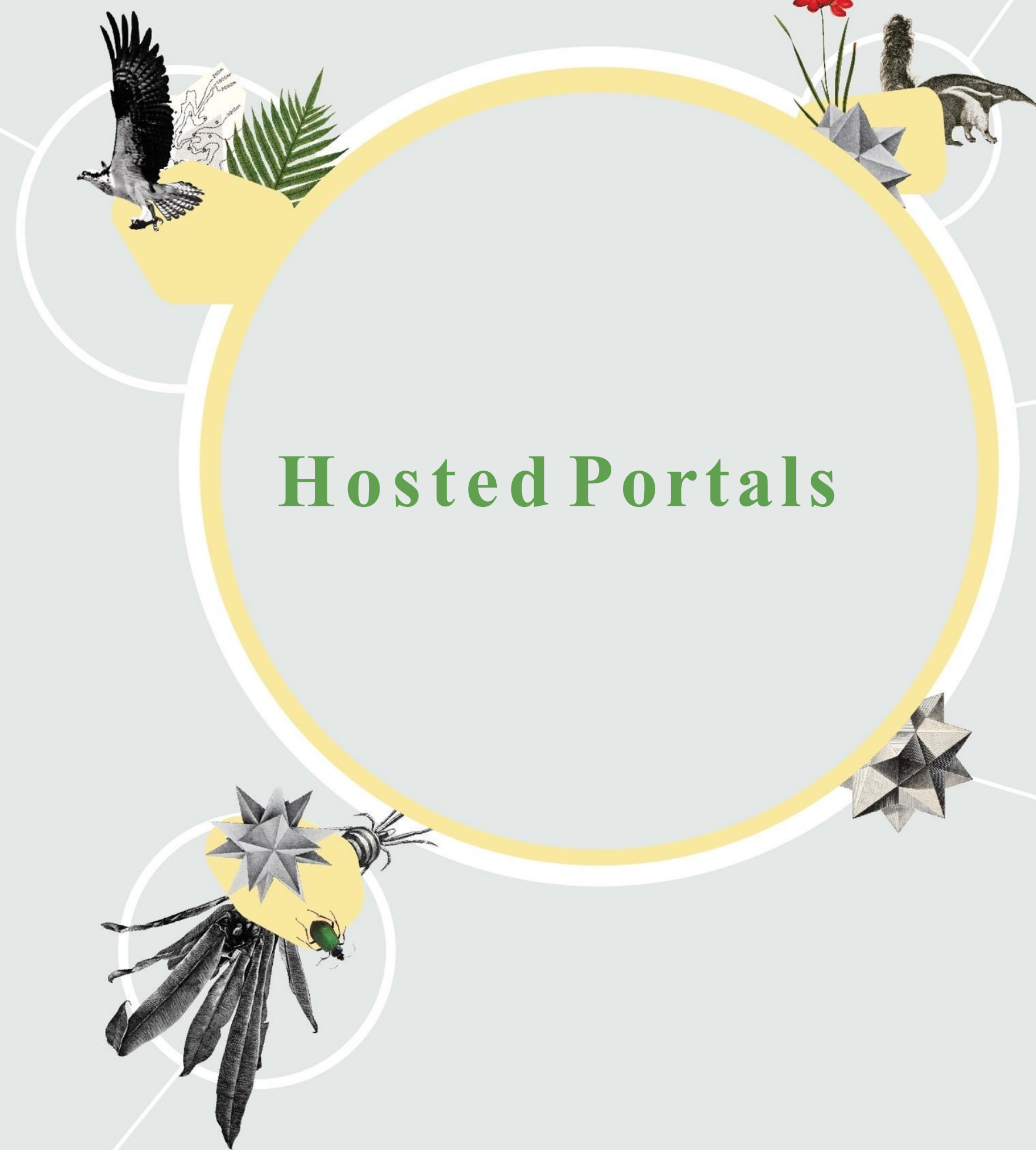
NEWS | 16 JANUARY 2024

## GBIF to serve as administrative host for Species 2000 Secretariat

*Agreement between GBIF Secretariat and the Catalogue of Life's legal body will support further collaboration on ChecklistBank and other joint activities*



# Hosted Portals



# Types of hosted portals



**4**  
Institutional



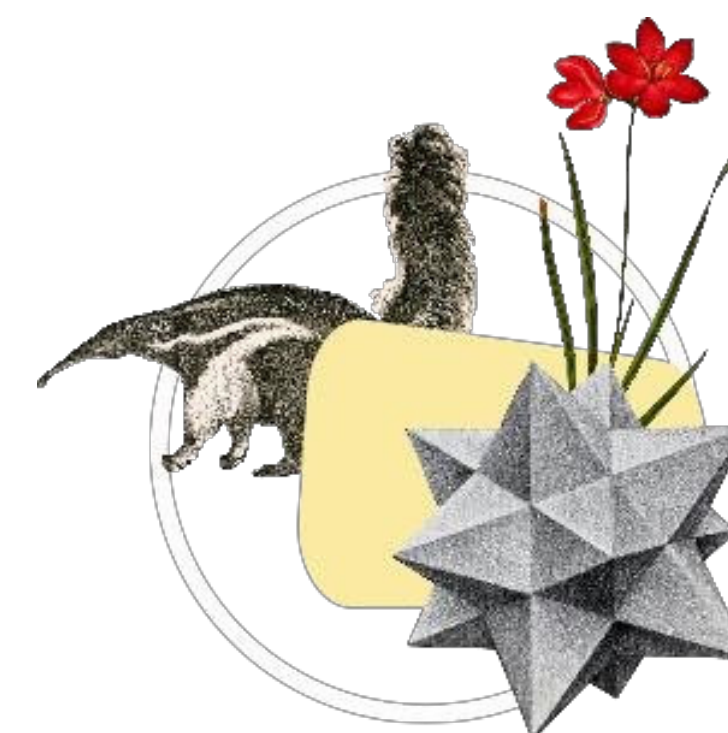
**12**  
GBIF Participant  
and GBIF Regional

**23**

Hosted portals in  
production



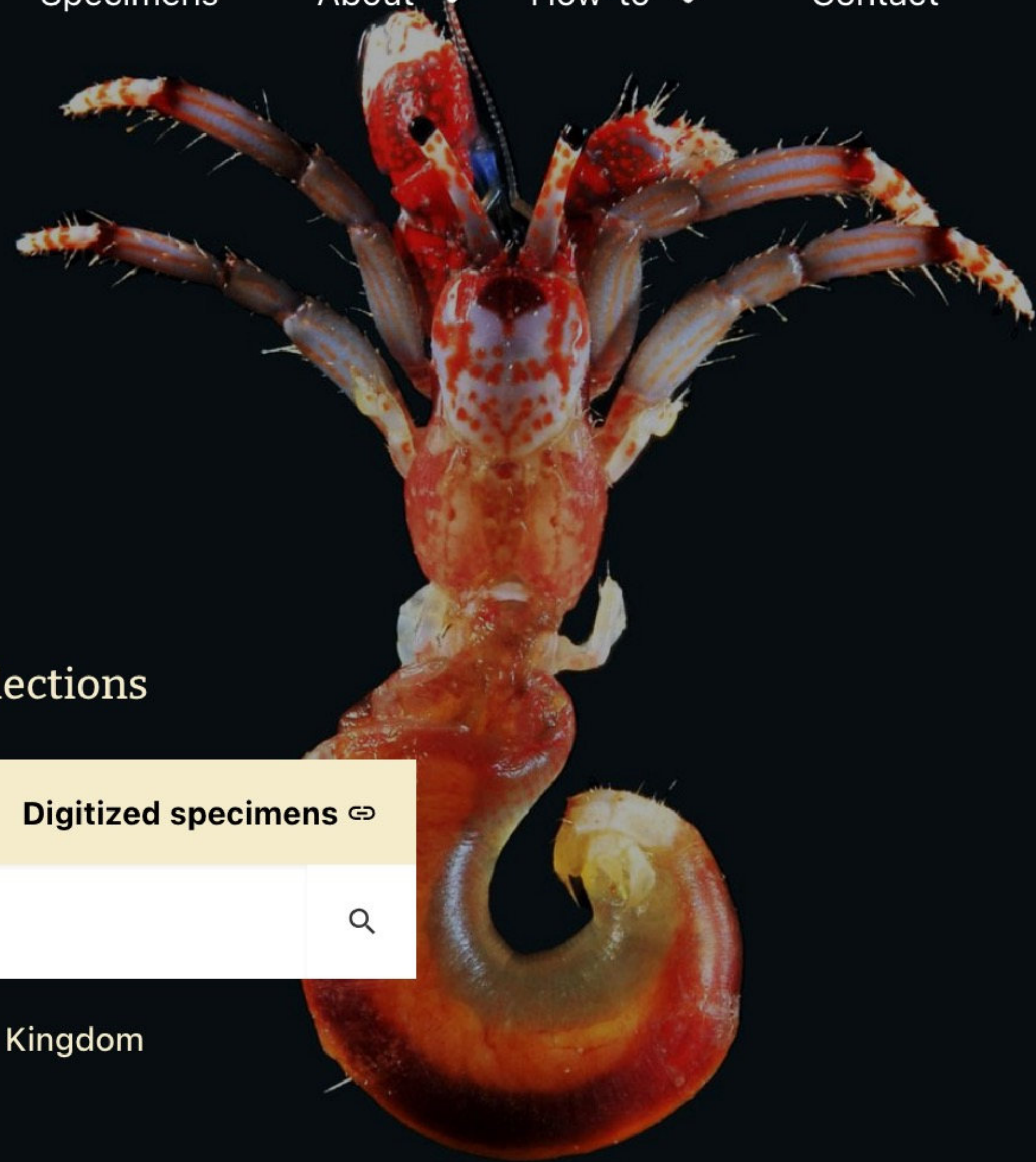
**29**  
In development



**7**  
Networks/ thematic  
networks

# Global Registry of Scientific Collections

A worldwide catalogue of scientific collections



Institution name

Digitized specimens ↻

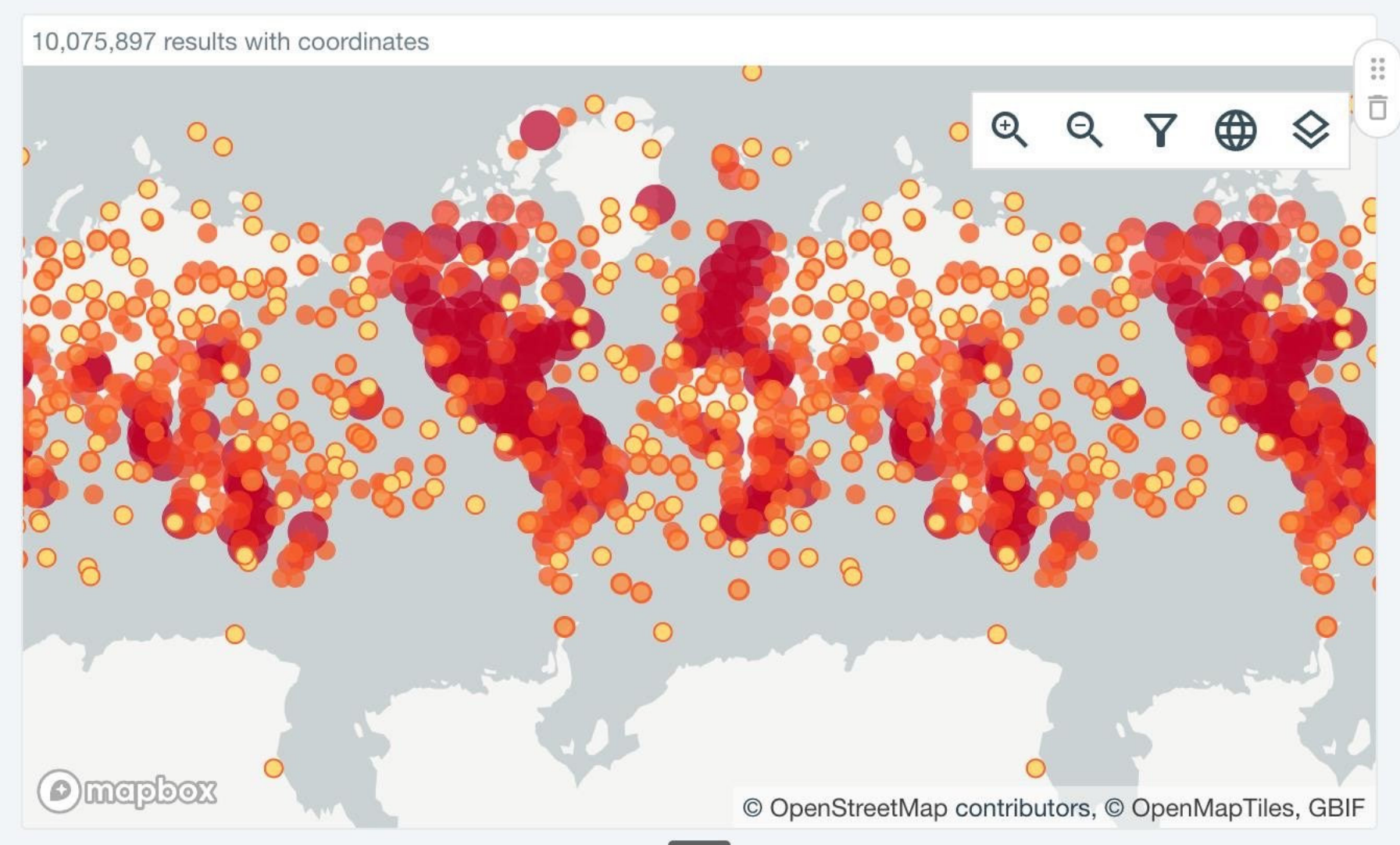
Search institutions




▶ Getting started?

📍 Institutions in United Kingdom


Scientific name : Diptera Verbatim scientific name Institution Collection Catalogue number Recorded by Identified by more




4,051,908 results with images




**BOLD:ACV8865 (*Sciaridae* sp.)**  
Feb 15, 2019 Costa Rica  
Sequenced




**BOLD:ABX1735 (*Anthomyza* sp.)**  
Jul 25, 2012 Germany  
Sequenced



**BOLD:ACZ8385 (*Chironomidae* sp.)**  
Apr 28, 2020 Costa Rica  
Sequenced



**BOLD:AAP5881 (cf. *Tanytarsus curvicristatus*)**  
Nov 22, 2017 Mexico  
Sequenced



***Richardia elegans* Wulp, 1899**  
Mexico  
Syntype

### Water body

Number of occurrences

1,269 results

|                  |        |
|------------------|--------|
| Ebro Basin       | 44,400 |
| Tugela           | 3,022  |
| Grytsjön         | 3,018  |
| Vaal River       | 2,791  |
| Great Berg River | 1,932  |
| Alsterån         | 1,872  |
| Buffalo River    | 1,426  |
| Limpopo river    | 1,327  |
| Skärsgölorna     | 1,325  |
| Stensjön         | 1,313  |

Previous Next

Non-interpreted values - same concept might appear with different names. 0.78% of all records have a value

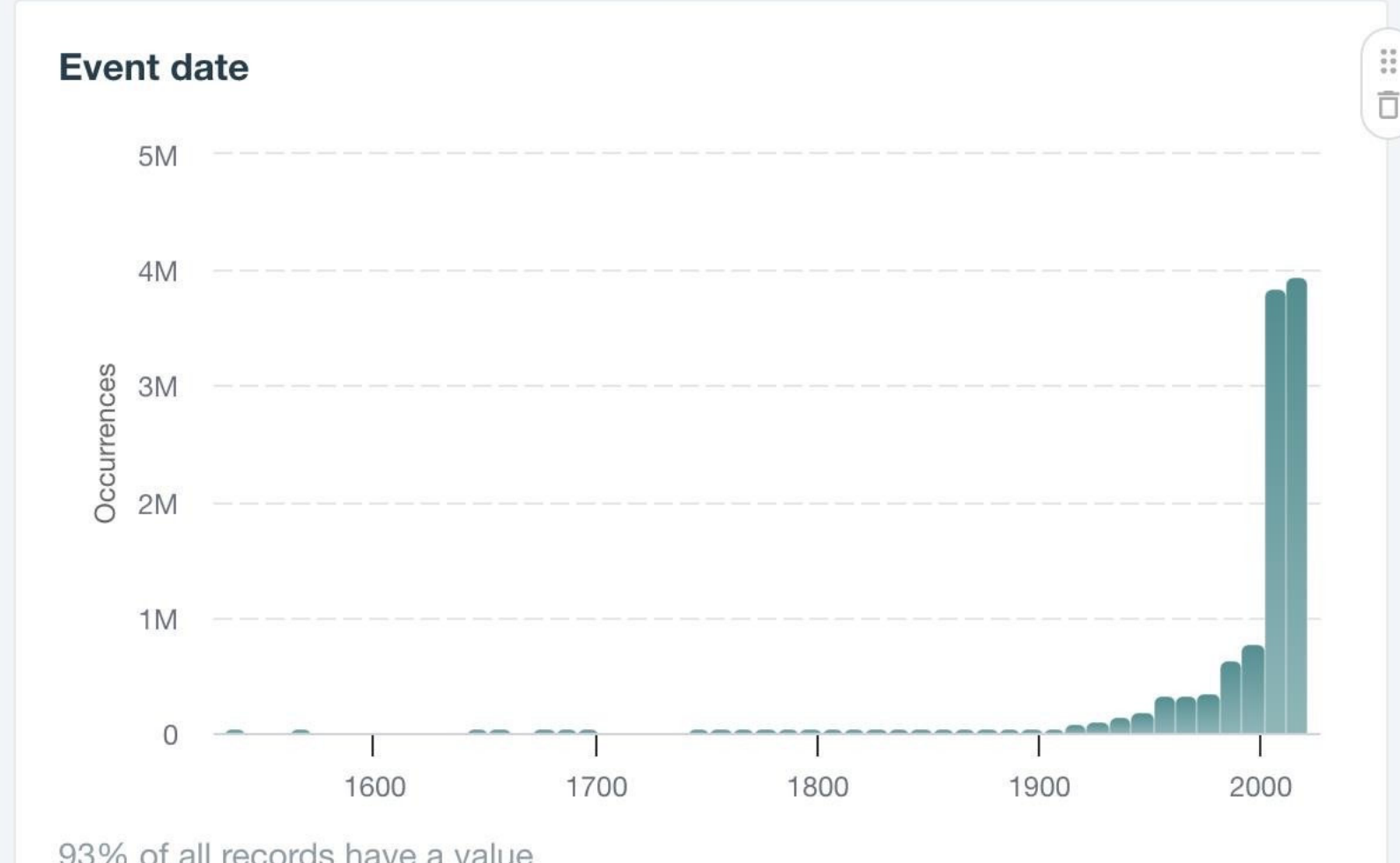
### Institution

Number of occurrences

567 results

|   |           |
|---|-----------|
| Centre for Biodiversity Genomics                            | 6,725,505 |
| National Biodiversity Institute, Costa Rica                 | 475,763   |
| Natural History Museum of Los Angeles County                | 463,831   |
| Natural History Museum, London                              | 380,852   |
| KwaZulu-Natal Museum  | 217,630   |
| Smithsonian Institution, National Museum of Natural History | 196,167   |
| Fundação Oswaldo Cruz                                       | 163,895   |
| Swedish Museum of Natural History (NRM)                     | 153,763   |
| Natural History Museum of Denmark                           | 153,394   |
| Naturalis Biodiversity Center                               | 144,396   |

Previous Next



### Recorded by

Number of occurrences

70,192 results

|                                       |           |
|---------------------------------------|-----------|
| D.Janzen, W.Hallwachs                 | 1,294,846 |
| D.Janzen, W.Hallwachs, J.A.Solano     | 323,818   |
| D.Janzen, W.Hallwachs, M.Zuniga Siles | 202,498   |
| BIOBus 2012                           | 198,449   |
| D.Janzen, W.Hallwachs, M.Obando       | 193,379   |
| D.Janzen, W.Hallwachs, R. Zuniga      | 175,316   |
| CBG Collections Staff                 | 141,451   |

Suggestions reviewed ●  
**1,933**

● Institutions  
**8,550**

Countries ●  
**116**

Global  
Registry  
of **Scientific  
Collections**

● Collections  
**8,112**

Country editors ●  
**57**








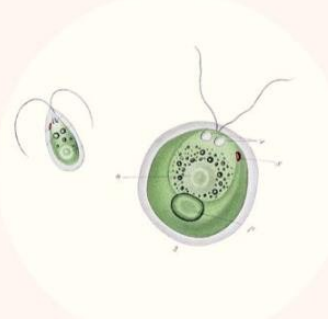









● Digitized specimens  
**228,680,952**



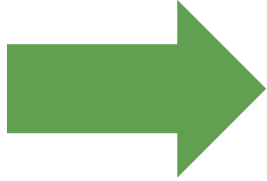
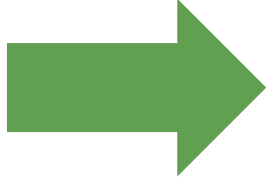


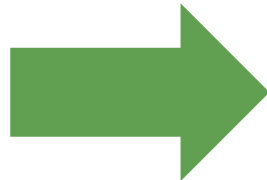
# New data model



# Case studies

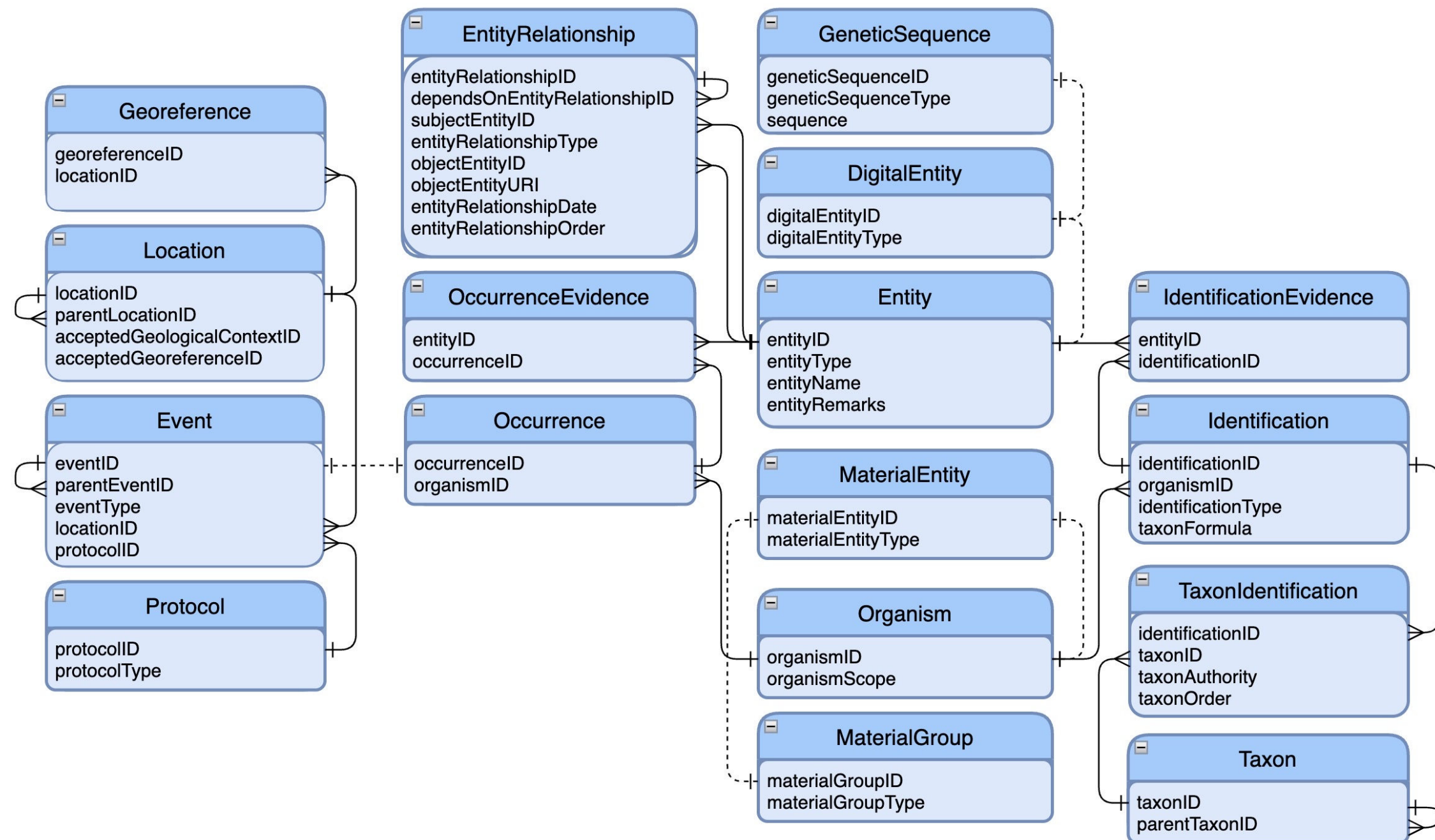
|   |  |  |   |   |
|---|--|--|---|---|
|  <p>Introduction</p>  |  <p>eDNA metabarcoding</p>    |  <p>Camera trap</p>                                     |  <p>Tissue samples</p>                           |  <p>Automatic moth trap</p>                |
|  <p>Global malaise programme</p>                            |  <p>iNaturalist</p>           |  <p>Specimens with media</p>                            |  <p>Environmental and community measurements</p> |  <p>Taxonomic treatments in literature</p> |
|  <p>Malaise trapping for reference barcode collection</p> |  <p>Taxonomic checklist</p> |  <p>Ecological survey data exchange specification</p> |  <p>Biotic interactions</p>                    |  <p>Recording absence data</p>           |

## Active work

|   |                               |   |
|---|-------------------------------|---|
|    | Collection Management Systems | A family of use cases related to collection management systems from the perspective of the material and associated digital objects.                           |
|   | Biotic Interactions           | Covers Events that provide evidence of interactions at the primary, organismal level (not at the derived taxonomic level).                                    |
|    | Camera Traps                  | Monitoring of an area, related sequences of images, machine generated annotations.  |
|   | Phylogenetic trees            | Covers links between specimens, sequences, material citations, and phylogenetic trees with the objective of enabling phylogenetic-based search and discovery. |
|  | Environmental DNA             | Covers sampling design and material subsampling; DNA based sequences to infer taxon occurrences.  |
|  | Ecological Inventories        | A family of use cases related to hierarchical inventory and monitoring- related events supporting observed presence, inferred absence and abundance.          |

# Working with 10 institutions

Instruction: "Map your data to this, please"



Specify ▼ dcc04c84-1ed3-11e3-bfac-90b11c41863e

University of Kansas, Biodiversity Institute and Natural History Museum > KUBI Ichthyology Collection

Catalog number 40560

## Psilorhynchus pseudecheneis

📍 Indrawati River at bridge crossing through Melamchi Town

🧬 Sequenced 🖼️ 14 images

---

### Catalogue item

|                  |                       |
|------------------|-----------------------|
| Preparations     | EtOH - 34; Tissue - 1 |
| Catalog number   | 40560                 |
| Recorded by      | Bentley, Andrew C     |
| Disposition      | In collection         |
| Institution code | KU                    |
| Collection code  | KUI                   |

### Collecting event

|               |            |
|---------------|------------|
| Field number  | RLM 08-46  |
| Event date    | 10/28/2008 |
| Event remarks | KWC 08-21  |

Display verbatim values

---

[See all 14](#)

### Occurrence

|                   |         |
|-------------------|---------|
| Organism quantity | 35      |
| Occurrence status | PRESENT |

Location

|                  |  |
|------------------|--|
| Higher geography | Asia, Nepal, Bagmati, Sindhupalchok                      |
| Continent        | Asia   |
| State province   | Bagmati  |
| County           | Sindhupalchok  |
| Locality         | Indrawati River at bridge crossing through Melamchi Town |

---

### Sequences

|                   |   |
|-------------------|---|
| Object entity iri | <a href="https://www.ncbi.nlm.nih.gov/nuccore/KM207611">https://www.ncbi.nlm.nih.gov/nuccore/KM207611</a> |
| Object entity iri | <a href="https://www.ncbi.nlm.nih.gov/nuccore/KM207658">https://www.ncbi.nlm.nih.gov/nuccore/KM207658</a> |
| Object entity iri | <a href="https://www.ncbi.nlm.nih.gov/nuccore/KM207657">https://www.ncbi.nlm.nih.gov/nuccore/KM207657</a> |
| Object entity iri | <a href="https://www.ncbi.nlm.nih.gov/nuccore/KM207634">https://www.ncbi.nlm.nih.gov/nuccore/KM207634</a> |
| Object entity iri | <a href="https://www.ncbi.nlm.nih.gov/nuccore/KM207612">https://www.ncbi.nlm.nih.gov/nuccore/KM207612</a> |
| Object entity iri | <a href="https://www.ncbi.nlm.nih.gov/nuccore/KM207635">https://www.ncbi.nlm.nih.gov/nuccore/KM207635</a> |

### Identification

|                         |   |
|-------------------------|---|
| Scientific name         | Psilorhynchus pseudecheneis   |
| Classification          |   |
| Scientific name (GBIF)  | Psilorhynchus pseudecheneis Menon & Datta, 1964   |
| Classification (GBIF)   | Animalia > Chordata > Cypriniformes > Psilorhynchidae > Psilorhynchus > Psilorhynchus pseudecheneis |
| Identified by           | Conway, Kevin W   |
| Remarks                 | RLM:3:N5,N10  |
| Verbatim identification | Psilorhynchus pseudecheneis   |
| Nature of ID            | unknown   |
| Date                    | December 18, 2008   |

---

### Roles

### Assertions

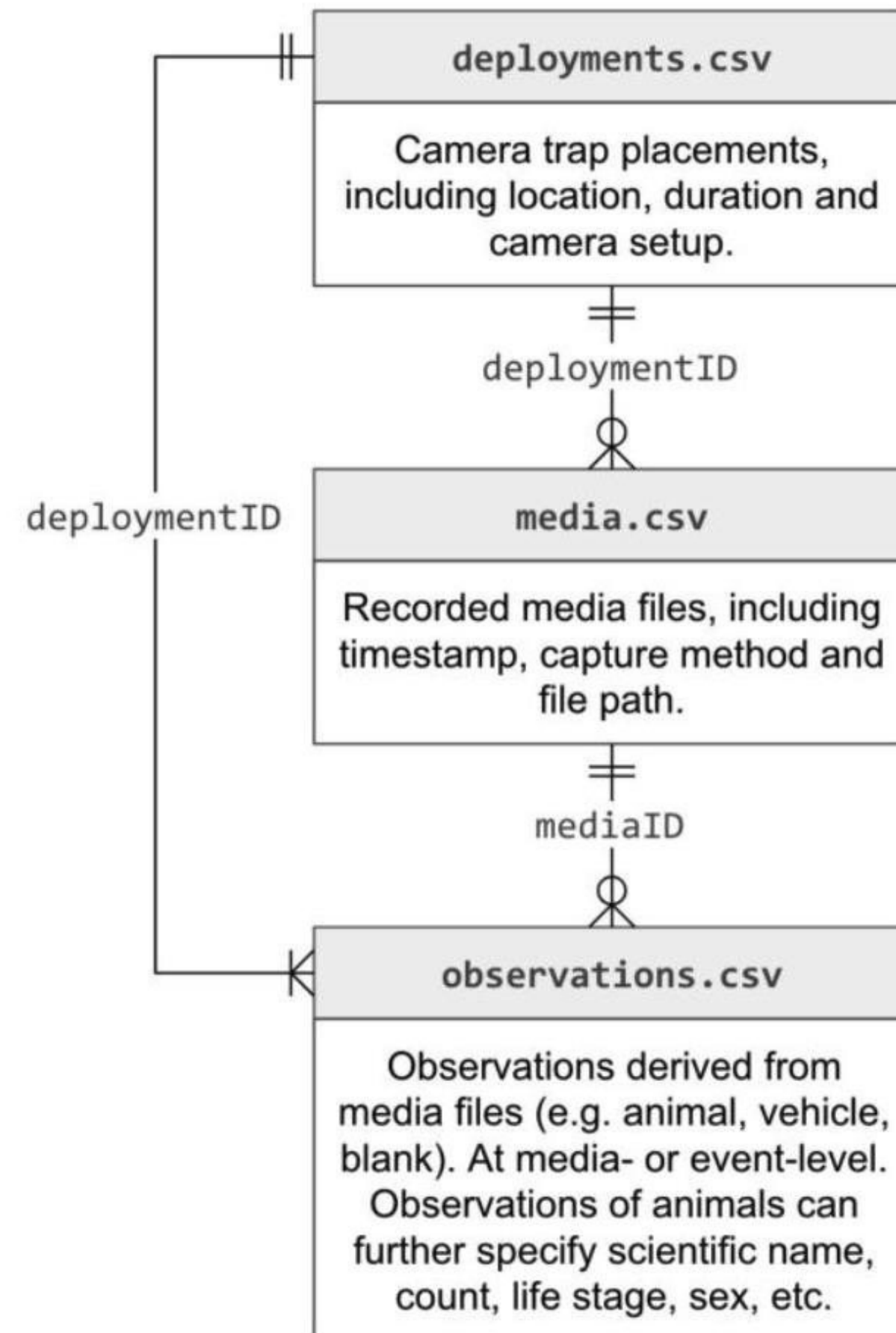
| Type   | Value          |
|--------|----------------|
| LENGTH | 16.3-88.4mm SL |

---

### Organism

|       |     |
|-------|-----|
| Scope | lot |
|-------|-----|

# Publishing model: CamtrapDP



Credit: Tom Rickman, CC BY-NC  
<https://www.inaturalist.org/observations/148735438>



NEWS | 14 FEBRUARY 2024

## GBIF releases IPT version 3

*Third major release of publishing software package adds support for new and future Frictionless Data Packages while retaining full support for current standards*



Common pheasant (*Phasianus colchicus* Linnaeus, 1758) from camera trap monitoring a waterway near Emlichheim, Germany. Photo (CC BY 4.0) via Research Institute for Nature and Forest (INBO).



# Announcement

## Launch of the Metabarcoding Data Programme

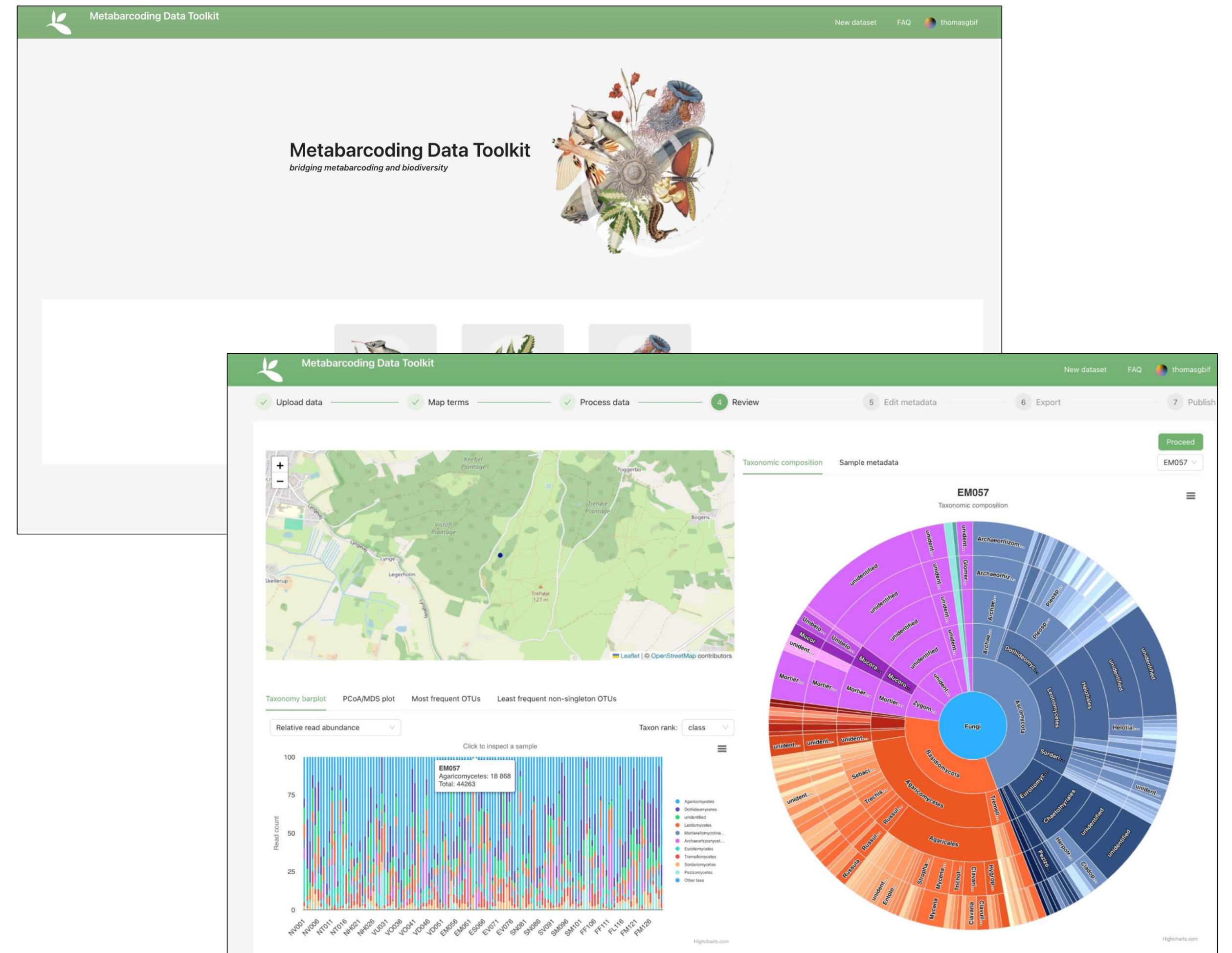


<https://gbif.org/dna>



# Metabarcoding Data Programme

- Aims to strengthen connection with the metabarcoding community
- Open to Participant Nodes
- Using the Metabarcoding Data Toolkit
- Pilot phase 2024 – 2026
  - Quarterly webinars to exchange feedback and guide the programme
  - Guide the tool development



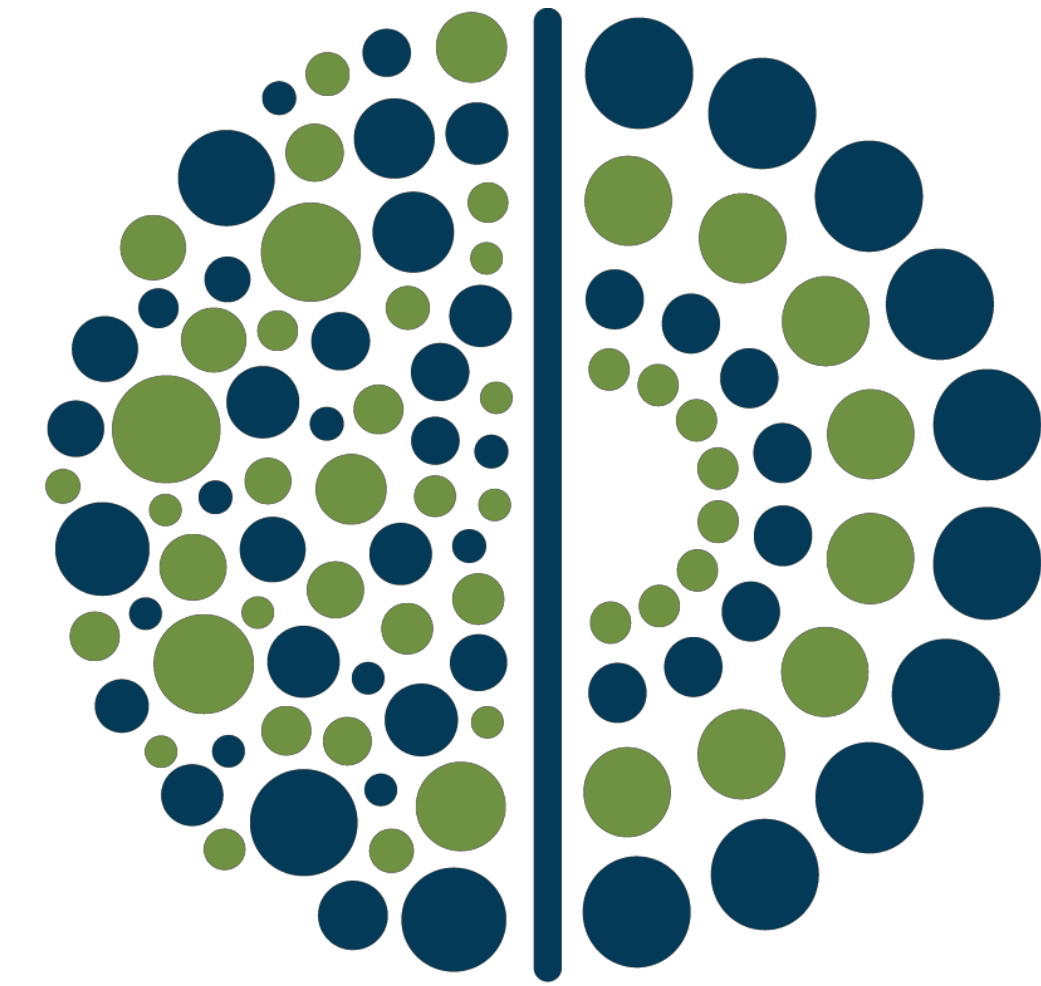
<https://mdt.gbif-uat.org>





## Humboldt Extension for Ecological Inventories

- A vocabulary to convey information about the context of sampling events, including e.g.
  - Target of survey
  - Effort spent
  - The metadata necessary to interpret the results
- Ratified by the TDWG standards group in 2024
- An extension suitable for use with Event-based Darwin Core data



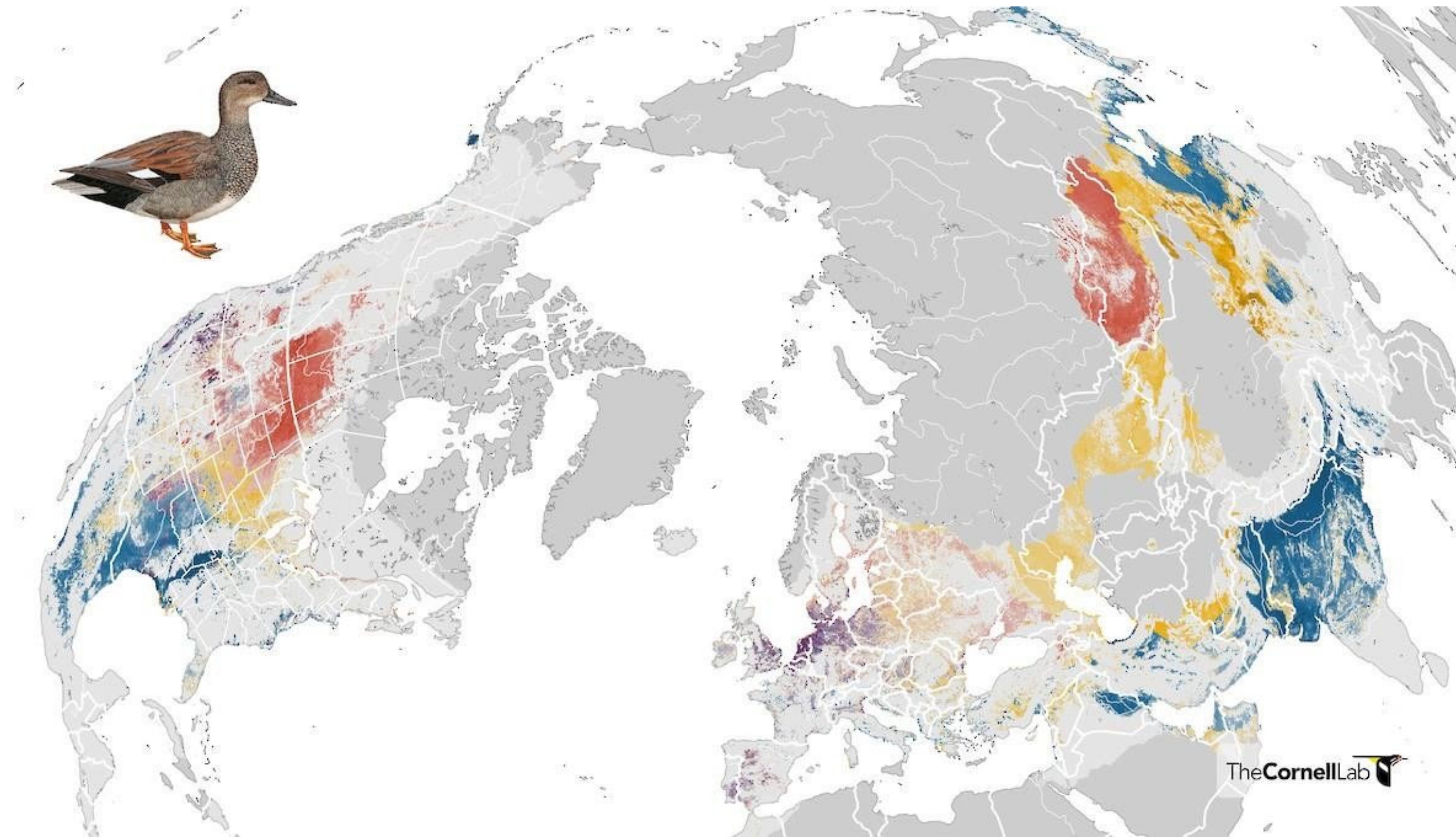
## Biodiversity Information Standards

**TDWG**



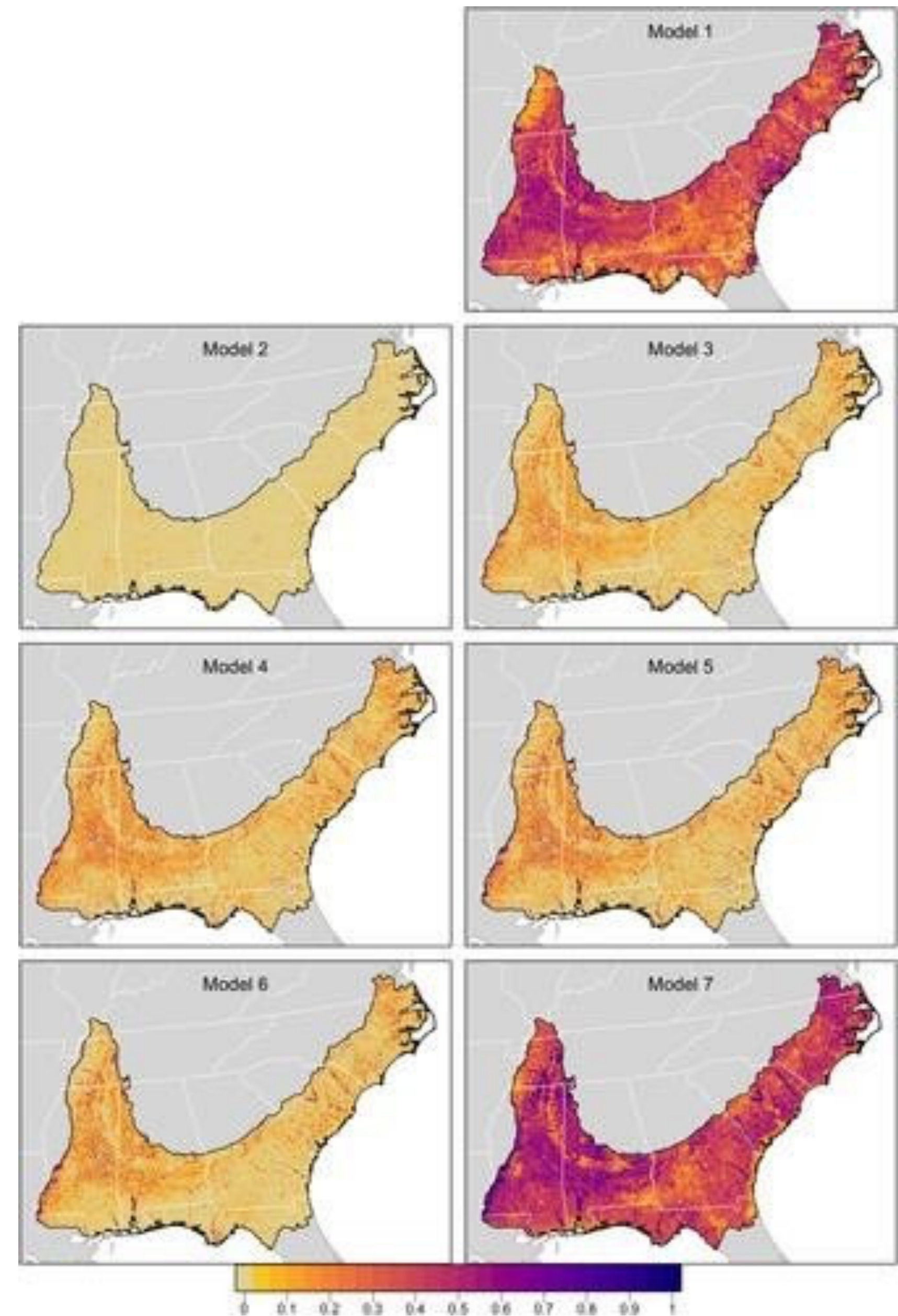
## Cornell Workshop

- Explored case studies that deal with hierarchical Events in datasets
  - Rapid inventory datasets (FieldMuseum)
  - Camera Trap data (INBO)
  - eBird
  - Long term monitoring (NEON)



## eBird Agreement

- Republish eBird as Event Data with Humboldt Ecological data extension
- Include effort (people, duration, distance) and completeness metadata on the events
- Supports explicit absence of detection and inferred absence of detection
- Improve ability to generate models from this dataset



## Connectez-vous Avec la Nature

Explorer et partager vos observations du monde naturel.


INSCRIVEZ-VOUS 

EXPLORER 

Get data How-to Tools Community About

PUBLISHER | SINCE FEBRUARY 9, 2012

## iNaturalist.org

ABOUT METRICS  HOME PAGE

100,734,988 OCCURRENCES 100,734,988 HOSTED OCCURRENCES 1 DATASET 5,442 CITATIONS

**Description:** iNaturalist.org is a website where anyone can record their observations from nature. Members record observations for numerous reasons, including participation in citizen science projects, class projects, and personal fulfillment.

**Endorsed by:** GBIF-US

**Installations:** [HTTP Installation](#)

**Administrative contact:** [Scott Loarie](#)

**Technical contact:** [Ken-ichi Ueda](#)

**Country or area:** [United States of America](#)

**Hosting:** 1 dataset ( 1 publisher • 1 country )

[Download activity report](#)

CONTACTS

|  |   |   |
|--|---|---|
| iNaturalist.org<br>United States of America<br><a href="http://www.inaturalist.org">http://www.inaturalist.org</a> | Ken-ichi Ueda<br>Technical point of contact<br><a href="mailto:kueda@inaturalist.org">kueda@inaturalist.org</a> | Scott Loarie<br>Administrative point of contact<br><a href="mailto:loarie@inaturalist.org">loarie@inaturalist.org</a> |
|--|---|---|

## Comment cela marche



1

Enregistrer vos observations



2

Partager avec d'autres naturalistes



3

Discuter de vos découvertes

## Contribuer à la science

Toute observation peut contribuer à l'étude de la biodiversité, du papillon le plus rare à la mauvaise herbe de jardin la plus courante. Nous partageons vos découvertes avec les bases de données scientifique comme le système mondial d'informations

Transfert des données depuis [www.inaturalist.org](http://www.inaturalist.org)...

Capture d'écran



Four organizations,  
one conference



# LIVING DATA 2025

Unified information  
driving transformation

October - 2025  
Bogota, Colombia





# Thank you!

Anne-Sophie Archambeau | [archambeau@gbif.fr](mailto:archambeau@gbif.fr)





Les entrepôts  
thématiques/institutionnels





## Entrepôts de confiance

Sur cette page, Recherche Data Gouv se fait le relai des entrepôts thématiques de confiance proposés par le Collège des Données de la recherche du Comité pour la science ouverte.

Le partage des données issues des activités de recherche implique de les mettre à disposition dans des entrepôts reconnus par les communautés scientifiques, thématiques, institutionnels ou généralistes.

Pour faciliter le travail des équipes de recherche, le collège des Données de la recherche propose une méthode d'identification des entrepôts thématiques de confiance, ainsi qu'une **première liste d'entrepôts découlant de la grille de critères d'analyse retenus à découvrir ci-dessous**.

**Cette liste est évolutive** et a vocation à être progressivement complétée par le collège des Données de la Recherche. **Elle n'est par ailleurs pas exhaustive** : ainsi, ne pas figurer dans la liste n'implique pas que l'entrepôt concerné n'est pas de confiance.

 N'oubliez pas, la dernière version à jour de la liste de référence est celle publiée sur le [site Ouvrir la Science](#).

# Ouvrir la science

The screenshot shows the website header with the logo 'Ouvrir la science!' and navigation links: 'LA SCIENCE OUVERTE', 'NOS ACTIONS', 'RESSOURCES', 'ACTUALITES', and 'QUI SOMMES NOUS?'. The breadcrumb trail reads: 'Accueil > Bibliothèque de la science ouverte > Sélectionner un entrepôt thématique de confiance pour le dépôt de données : méthodologie et analyse de l'offre existante'. A document cover is highlighted with a teal circle. The cover has a yellow background and contains the text: 'En pratique', 'Ouvrir la science!', 'Sélectionner un entrepôt thématique de confiance pour le dépôt de données : méthodologie et analyse de l'offre existante', and '2024'.

La note - produite par le collège Données de la recherche - propose une méthode d'identification des entrepôts thématiques en s'appuyant sur les critères d'exclusion et de description des entrepôts. Celle-ci est assortie d'une première liste découlant de la grille de critères d'analyse retenus.

# Ouvrir la science



## Sommaire

Contexte

Méthodologie

Axes de travail et livrables

Critères d'exclusion et de description des entrepôts

Résultats et analyse disciplinaire

Premiers éléments pour une mise à jour de la liste des entrepôts

Conclusion

Annexe : lettre de mission

## Sélectionner un entrepôt thématique de confiance pour le dépôt de données : méthodologie et analyse de l'offre existante

Collège Données de la recherche

*Mars 2024*

[Consulter la note sur HAL](#)

Le partage des données issues des activités de recherche nécessite de les mettre à disposition dans des entrepôts. Il est préconisé de privilégier les dépôts dans des entrepôts thématiques de confiance, comme les centres de référence thématiques de l'écosystème Recherche data gouv, ou à défaut sur des entrepôts génériques comme Recherche data gouv. Si aucun de ceux-là ne répond au besoin, le choix de l'entrepôt le mieux adapté (qu'il soit thématique, généraliste ou institutionnel) peut parfois être complexe.

Pour faciliter le travail des équipes de recherche, le collège des Données de la recherche propose une méthode d'identification des entrepôts thématiques de confiance, ainsi qu'une première liste découlant de la grille de critères d'analyse retenus.

Cette liste est évolutive et a vocation à être progressivement complétée.

[Télécharger la liste à jour](#)



# Ouvrir la science

ListesdesEntrepotsdeConfiance\_v1bis\_202407

Rechercher dans la fe

Partager

Accueil Insertion Dessin Mise en page Formules Données Révision Affichage

Calibri 11 A A Standard

Coller G I S Mise en forme conditionnelle Mettre sous forme de tableau Styles de cellule Insérer Supprimer Mise en forme Trier et filtrer Rechercher et sélectionner

E17 fx DOI

|    | A   | B   | C  | D  | E   | F  | G                             | H   | I  | J   | K   |
|----|---|---|--|--|---|--|-------------------------------|---|--|---|---|
|    | Nom   | URL   | Institution porteuse   | Modération   | Identifiant pérenne fourni  | Pérennité de l'infrastructure/Engagement sur la durée de préservation des données  | Discipline                    | Champs disciplinaires détaillés   | Données acceptées=mots-clés qui parlent à la communauté  | Embargo   | Limite de volume  |
| 14 | ELAR  | <a href="https://www.elararc.hive.org/">https://www.elararc.hive.org/</a>                   | Berlin-Brandenburg Academy of Sciences and Humanities                              | Oui, vérification des métadonnées et demande de corrections si nécessaire. La conformité du jeu de données à la politique d'Elar est également vérifiée lors du dépôt.   | Handle  | Pas d'engagement sur une durée précise. Si ELAR devait cesser ses activités, une procédure est prévue pour transférer l'ensemble des dépôts vers un autre entrepôt, par exemple DELAMAN <a href="https://www.delaman.org/">https://www.delaman.org/</a> , en fonction des caractéristiques et de la thématique des dépôts  | Sciences humaines et sociales | Linguistique  | Enregistrements sonores, métadonnées descriptives  | Possible, mais l'entrepôt encourage plutôt l'ouverture sans délai.  | Jusqu'à 100Go. Une procédure spécifique de dépôt est prévue pour les jeux volumineux. |
| 15 | European Nucleotide Archive                   | <a href="https://www.ebi.ac.uk/ena/browser/home">https://www.ebi.ac.uk/ena/browser/home</a> | European Molecular Biology Laboratory European Bioinformatics Institute (EMBL-EBI) | Les métadonnées sont modérées automatiquement via une checklist, à laquelle s'ajoute une vérification manuelle selon les cas. Prise de contact en direct avec les soumetteurs dans certains cas.   | Interne ( <a href="https://ena-docs.readthedocs.io/en/latest/submit/general-guide/accessions.html">https://ena-docs.readthedocs.io/en/latest/submit/general-guide/accessions.html</a> ) - Il est possible de choisir un identifiant externe (BioSample) | Existe depuis 1985.  | Biologie                      | Biologie moléculaire, génomique   | Séquences nucléotidiques. Données brutes et annotations  | Oui ( <a href="https://ena-docs.readthedocs.io/en/latest/faq/release/data-availability-policy.html">https://ena-docs.readthedocs.io/en/latest/faq/release/data-availability-policy.html</a> ) | Non   |
| 16 | GBIF Global Biodiversity Information Facility | <a href="https://www.gbif.org/">https://www.gbif.org/</a>                                   | GBIF   | Le GBIF fournit une modération à 2 niveaux :<br>1) au niveau de l'entité publiant les données (publisher), par exemple le laboratoire de recherche. Le GBIF s'assure que l'entité fournit des données dans le périmètre du GBIF. Ce contrôle est opéré une seule fois, lors de la création de l'entité déposante.<br>2) au moment du dépôt de données lui-même. Selon l'outil de dépôt, la modération diffère.<br>Avec l'utilisation de l'IPT (Integrated Publishing Toolkit, outil facultatif mais recommandé), une validation technique du contenu des données selon le standard Darwin core est automatiquement réalisée : vérification des identifiants, de l'absence de doublons et de la complétude des métadonnées obligatoires et basiques. Le site GBIF.org opère ensuite la réconciliation géographique/taxonomique et ajoute un "flag" des occurrences. | DOI   | En cas d'interruption d'accès aux données ou de rupture de service, le GBIF transférera les données auprès de ses partenaires afin d'assurer la préservation de long terme des données.<br>Les institutions qui partagent leurs données ("data publishers" ou éditeurs de données) sont seules responsables de la gestion, diffusion, mise à jour/correction et du possible retrait à tout moment de leurs données du GBIF. Les points nodaux GBIF (équipes des pays ou institutions participant du réseau GBIF) peuvent également installer des instances IPT pour le bénéfice d'institutions "data publisher" si celles-ci n'ont pas les moyens techniques | Sciences de l'environnement   | Sciences du vivant, Biodiversité, Biologie animale, Biologie végétale, Ecologie, Environnement; Ecosystèmes | Taxons, données d'occurrences, données d'échantillonnage, toutes standardisées selon les standards Darwin core ou ABCD.  | Oui   | Non   |
| 17 | Geovistory                                    | <a href="https://www.geovistory.org/">https://www.geovistory.org/</a>                       | LARHRA – KleioLab – Université de Berne  | Pas de modération des données mais échanges avec les porteurs de la plateforme sur les projets de recherche concernés.   | Identifiant interne   | Un projet en cours financé par swissuniversities vise à trouver des solutions de pérennisation du projet par la constitution d'une institution à but non lucratif. Les données seront prochainement déposées sur la grille de sauvegarde et  | Sciences humaines et sociales | Histoire, archéologie, histoire de l'art  | Données historiques reliant des sources, des textes et d'autres entités (personnes, objets physiques et conceptuels, lieux, événements, relations sociales, etc.), | Non   | Pas mentionné   |