

## Informatics and data products

### Developments and plans

Tim Robertson, Andrea Hahn GBIF Secretariat



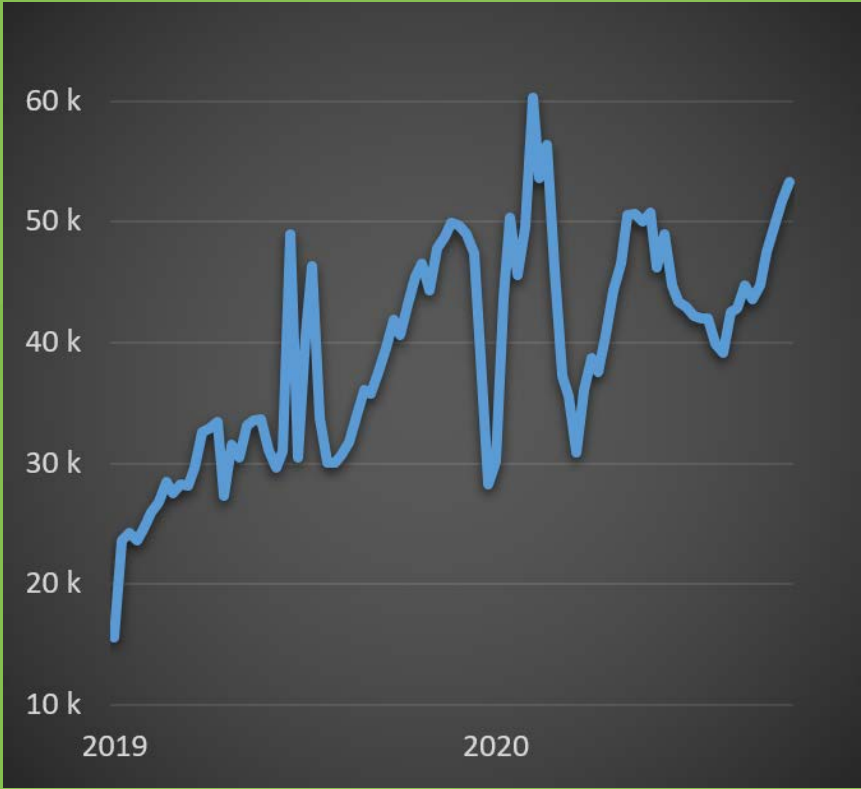




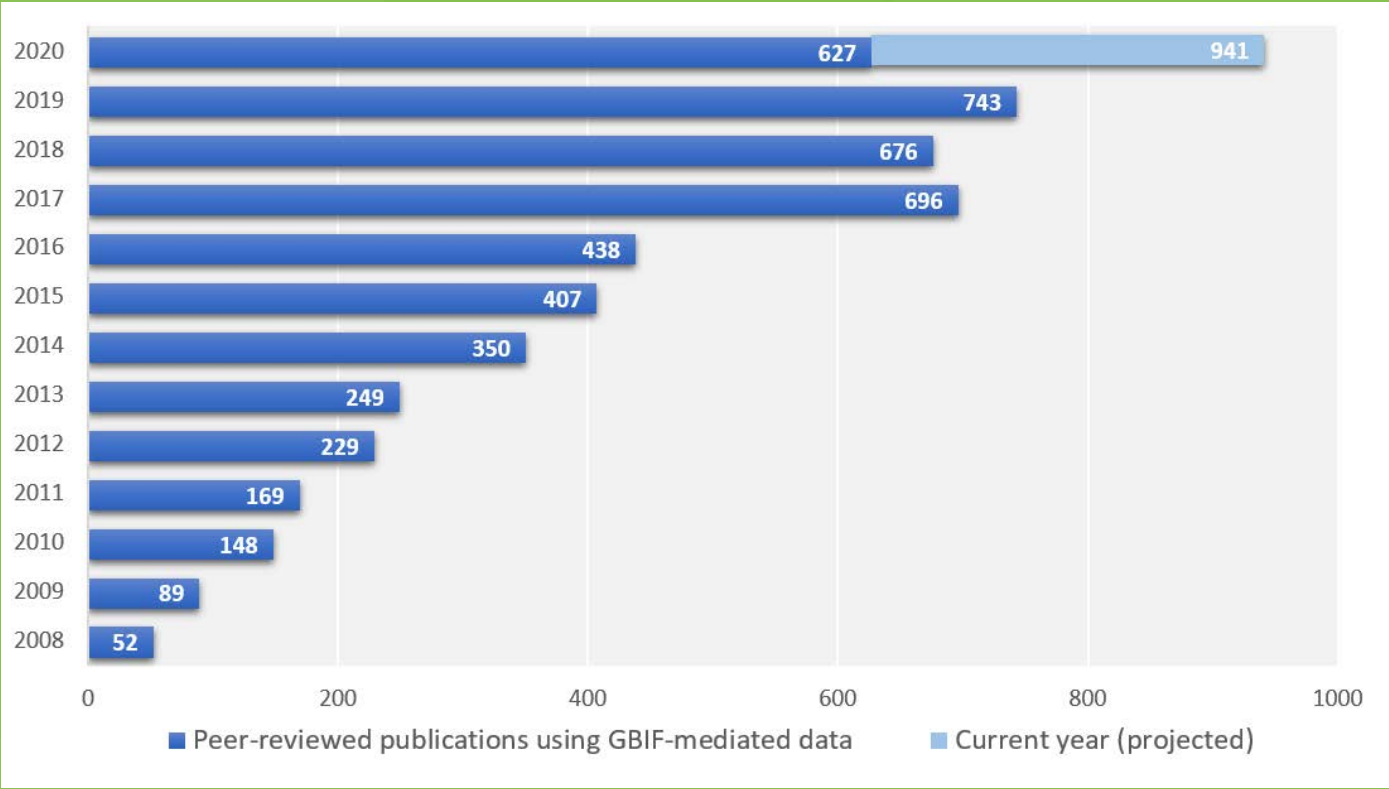
# GROWTH AND DATA PUBLISHING



# GROWTH IN USE



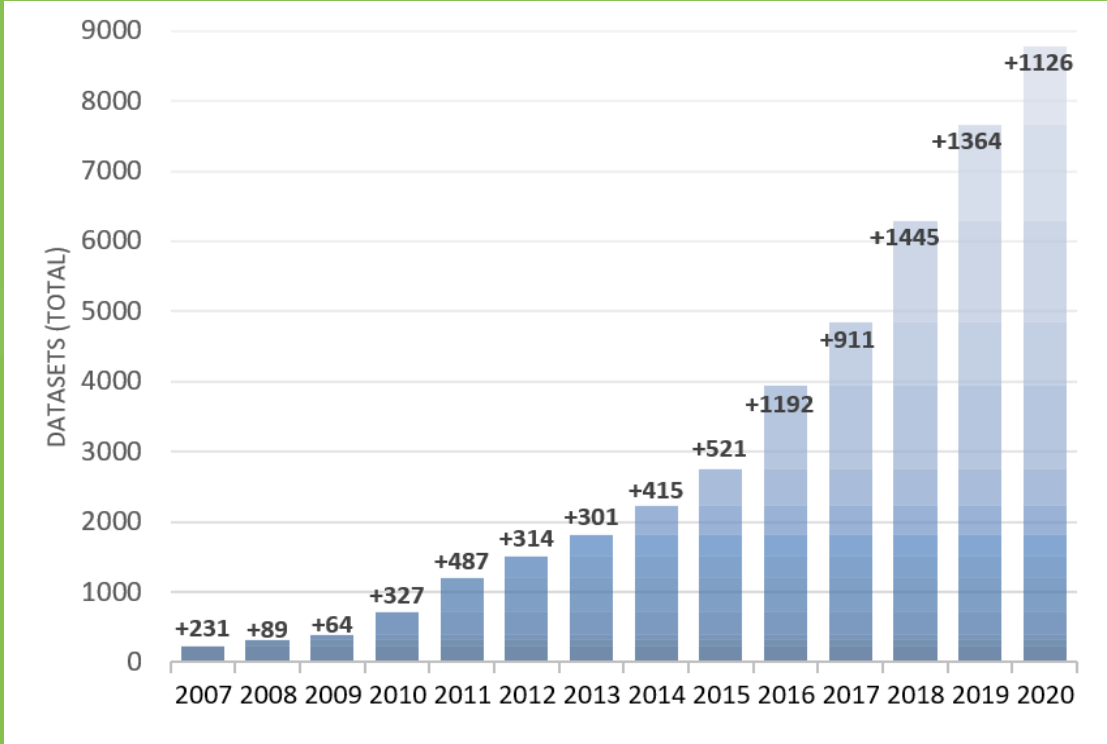
Users of GBIF.org per week  
~ 170k per month now (+20k over 2019)



status: Aug 2020

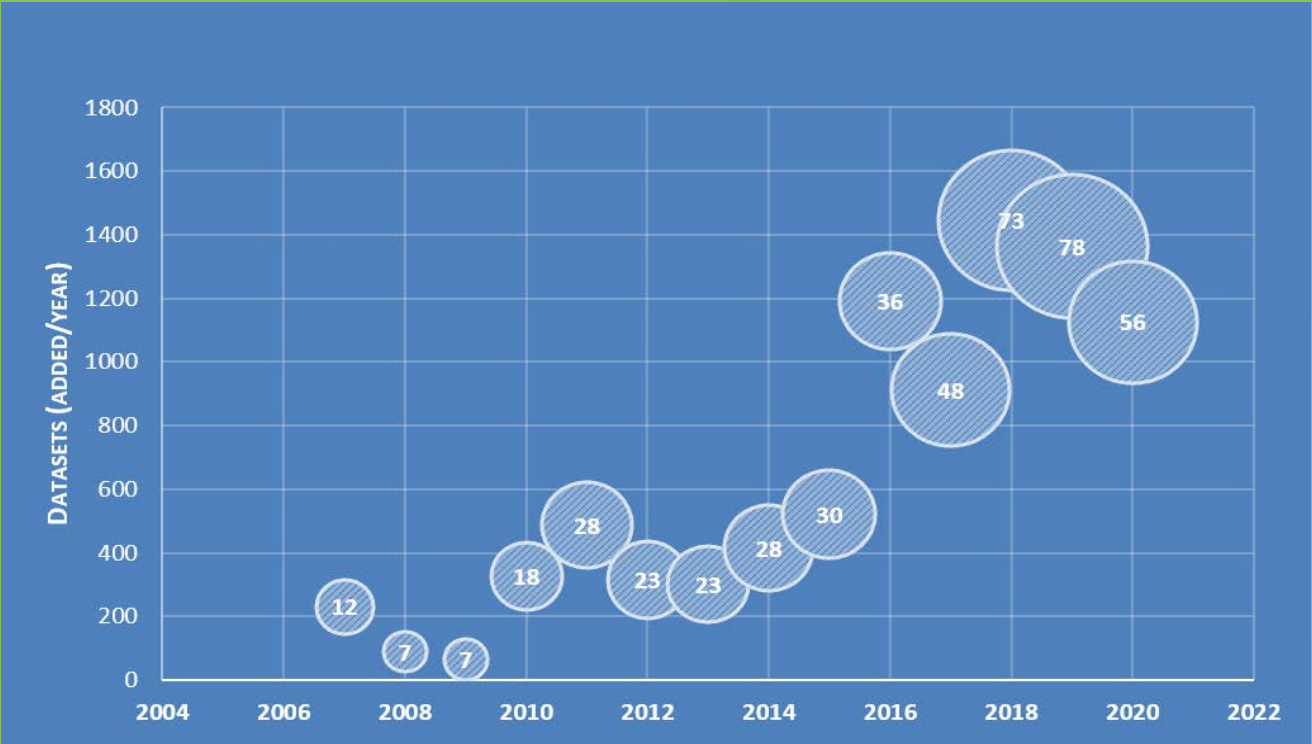
Publications citing use of GBIF-mediated data

# IPT (INTEGRATED PUBLISHING TOOLKIT)



not showing: 2385 datasets from UMS PatriNat (2020)

Datasets published through IPT installations,  
cumulative



status: Sept 2020

newly mobilized IPT datasets per year  
circles: number of countries involved





# NAMES INFRASTRUCTURE

In partnership with the Catalogue of Life



# MODERNIZING THE CATALOGUE OF LIFE

## Complete and completing (Nov 2020)

- New infrastructure deployed to manage data publication and assembly of the taxonomy
  - A new data exchange standard (COL Data Package)
  - An evolution of GBIF ChecklistBank -> COL ChecklistBank
  - A workbench for editors to review and assemble
  - A new suite of APIs and R-based library (rOpenSci)
  - Developed in collaboration, hosted by GBIF
- Deployment of public interfaces
  - New COL website (November)
  - Migration of historical annual checklists (2006-2019)

## Upcoming

- Expand the taxonomy with content necessary for GBIF
  - Automate reports on gaps
- Semi-automate assembly (with review)
- Integrate with GBIF.org
  - New backbone end 2020





## Communities

Prepare and curate  
checklists

## Data Publishing

COL Repository or other web  
repositories

## ChecklistBank

Storage and access to  
checklists

## Workbench

Construct and publish  
integrated checklist

## Users

Web and machine  
access

1

COLDP

DwC-A

ACEF

Spread-  
sheet

Text Tree

2a

Publish directly

Intermediate  
Repository

Primary datasets

2b

Register

Primary  
datasets

Standardised  
datasets

3

Secondary copies

4

- COL Checklist
- Sector source mappings
- Two views:
  - Reviewed (contributed sectors)
  - Comprehensive (incl. unreviewed)
- API for both views
- Links to data sources

5a

5b

Standardised access  
to any checklist



Catalogue of Life





# GBIF.ORG INFRASTRUCTURE





# NEW GBIF.ORG FILTERS

## Presence / Absence filter

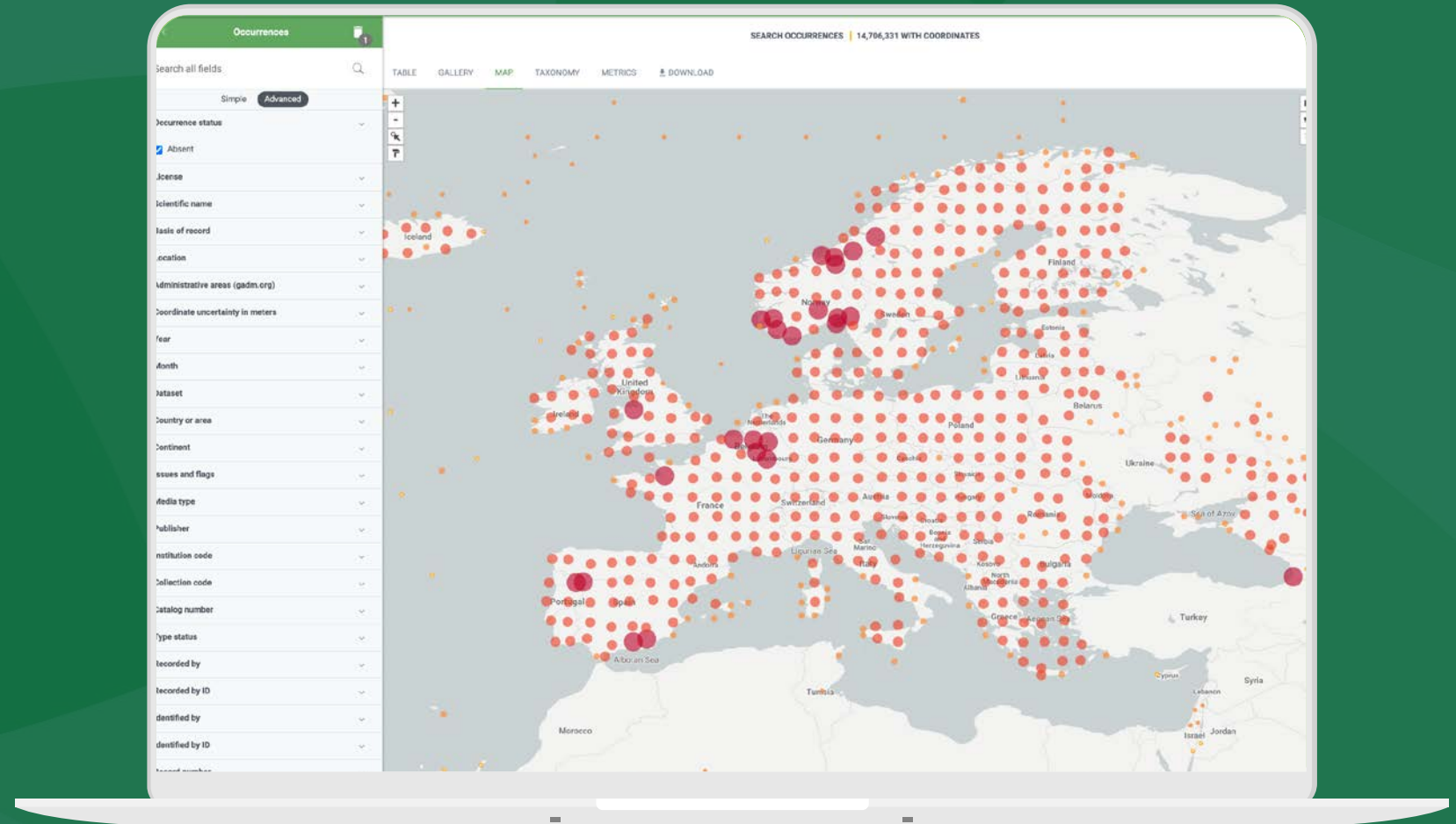
- All records declared as present or absent
- 14,7 million records of absence

## Search by person identifiers

- Recorded by or identified by
- Variety of identifier schemas (e.g. from ORCID, WikiData)
- Working for TDWG Attribution group

## Search by Administrative Areas

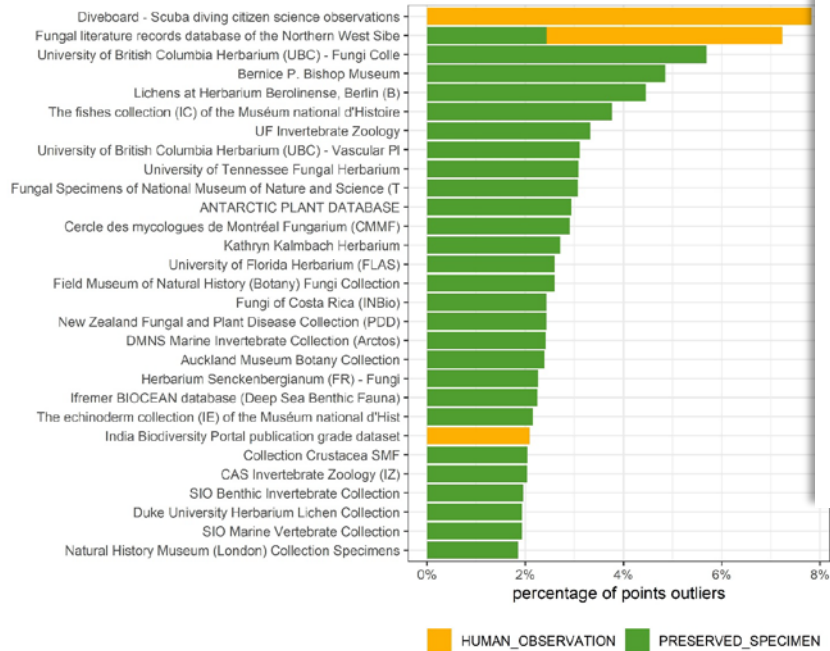
- Uses 3 levels from the GADM.org database e.g. country / state / county
- Using the API can report counts by e.g. county level



# DATA QUALITY - OUTLIER DETECTION

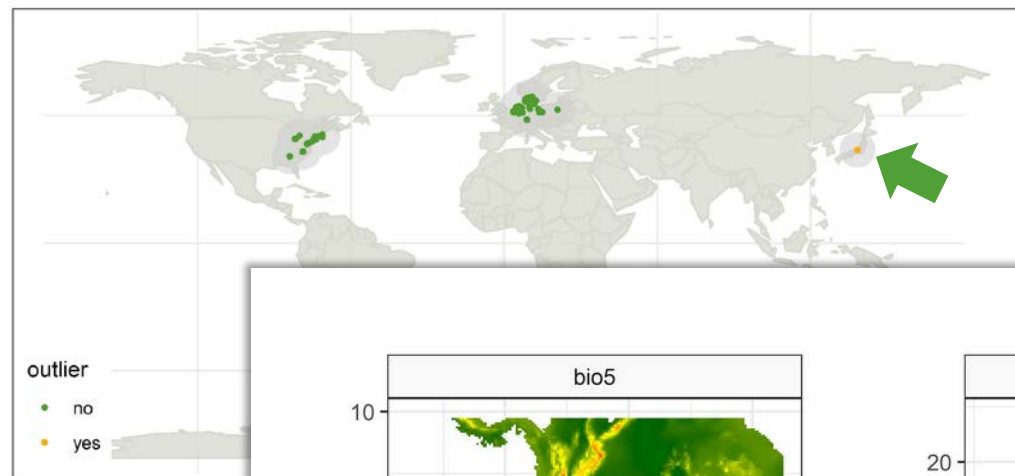
## Percentage of dataset outliers

percentage of unique points (datasetkey,specieskey,lat-lon) that are dbscan outliers

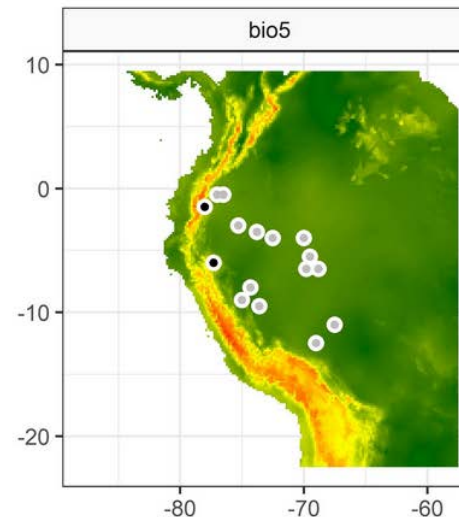


2020 datasets with more than 10K unique points  
DBSCAN settings: eps=1500km, minPts=3  
All outliers with establishmentmeans = NULL  
Only species with less than 20K unique lat-lon points run

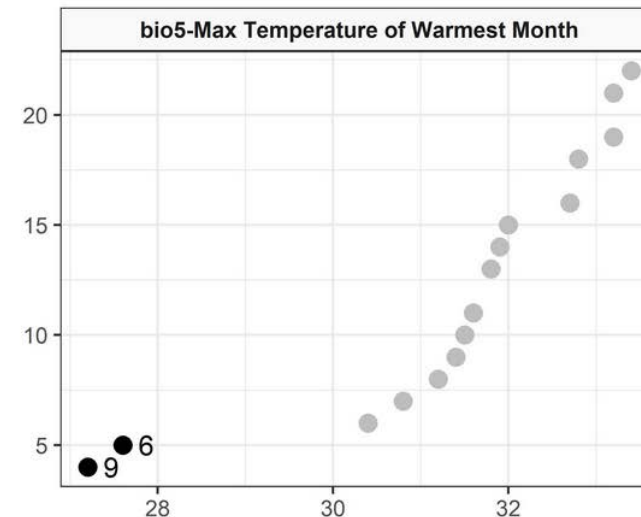
datasets – distance-based



species – distance-based



Outlier FALSE TRUE



species - environmental

Mammalia - Primates - Leontocebus fuscicollis  
specieskey=8493215  
outlier gbifids: 1920812594 2576278969 2244307771



ⓘ This is an experimental feature that highlights possible duplicate and/or related occurrences.

CURRENT

**Anochetus madagascarensis** Forel, 1887

Animalia > Arthropoda > Hymenoptera > Formicidae > Anochetus > Anochetus madagascarensis

Preparations: 100% EtOH

Dataset: AntWeb

Publisher: California Academy of Sciences

Basis of record: Preserved specimen

📍 12.8S, 45.1E 📅 February 15, 1999

1

RELATED OCCURRENCES

**Anochetus madagascarensis** Forel, 1887

Animalia > Arthropoda > Hymenoptera > Formicidae > Anochetus > Anochetus madagascarensis

Dataset: A revision of Malagasy species of Anochetus Mayr and Odontomachus Latreille (Hymenoptera: Formicidae).

Publisher: Plazi.org taxonomic treatments database

Basis of record: Preserved specimen

📍 12.8S, 45.1E 📅 February 15, 1999 📄 Treatment

Similar because Same accepted species Same date Same coordinates Same country Same recorder name

Details

2

**Anochetus madagascarensis** Forel, 1887

Animalia > Arthropoda > Hymenoptera > Formicidae > Anochetus > Anochetus madagascarensis

Dataset: International Barcode of Life project (iBOL)

Publisher: The International Barcode of Life Consortium

Basis of record: Material sample

📍 12.8S, 45.1E 📅 February 15, 1999 🖼️ 1 image 🧬 Sequenced

Similar because Same accepted species Same date Same coordinates Same country Identifiers overlap

Details

3



# Clustering

- Example shows a specimen record (1), a taxonomic revision (2), a DNA sequence (3)
- Grouped using location, identification, date, local identifiers and type status (so far) and fuzzy matching
- Stored as annotations against the records
- Interface allows easy comparison of the fields on the record (details)
- Purpose
  - Richer model of related information for researchers
  - Transferring information between institutional databases (images, sequences, new identifications) which may reduce effort spent
  - Aid in deduplicating data
  - Known to contribute towards data quality improvements
  - Contribution towards the Extended Digital Specimen / Open Digital specime concept (US and DiSSCo initiatives)

# USING GBIF IN CLOUD ENVIRONMENTS

## Growing interest in big data tools

- Google Big Query increasingly being used
- Google / Amazon / Microsoft etc clouds being used by several partners
- Opportunities with science computing clouds like the European Open Science Cloud
- Simplify research, ease integrations with other datasets

## New download formats

- Apache Avro format
- Bionomia-specific format to reduce computation on client side
- Services to easily move GBIF downloads onto cloud environments (underway)

## GBIF data as public datasets

- Strong interest expressed on the GBIF community forum
- Requires new citation service (underway)
- Exploring possibilities during 2021





# COLLECTIONS CATALOGUE





## COLLECTIONS CATALOGUE GLOBAL CONSULTATION

- Part of the EU Synthesys+ project and run as an initiative under the Alliance
- A virtual workshop held over 2 weeks
- Two two-hour preparatory webinars to introduce the Ideas Paper  
<https://doi.org/10.35035/p93g-te47>
  - Available in English, French, Spanish and simplified Chinese
- GBIF community forum <https://discourse.gbif.org/g/Collections-catalog>
- Daily summaries, in multiple languages
- Outcomes document to be drafted



# SOME OUTCOMES

- Many communities need information on collections
- Interest in a core set of common data elements
- Each community has specific needs; may document collections at different levels of granularity
- GBIF can:
  - bring together complementary information
  - promote a consistent approach to the common elements
  - help national nodes showcase their collections and the value they offer
  - Use the DOI-based citation and usage tracking to add value to nodes and to individual collection institutions.
- The catalogue may serve as a mechanism to attract funding; global partnerships to maintain biodiversity collections
- The GRSciColl Catalogue
  - Holds a basic core suitable to act as a stub for any collection
  - Can act as a first step towards better linking

# ENHANCING THE GRSCICOLL COLLECTIONS CATALOGUE

- Monthly synchronisation with Index Herbariorum
- Collaboration with iDigBio
  - Data imported from iDigBio catalogue
  - GRSciColl powering the iDigBio collections catalogue
  - Collaborative editing responsibility
- Early exploration of synchronisation with NCBI BioCollections\*
- Linking ~170M specimen-related occurrences to GRSciColl

2021

- Mature the processes around collaborative editing
- Deploy an enhanced collections catalogue portal
- Aim to broaden integrations (Australia)
- Explore a sample profile of the TDWG collections description standard

- NCBI <https://www.ncbi.nlm.nih.gov/biocollections>
- Visual concept of collection catalogue <https://labs.gbif.org/visual-concepts/>
- iDigBio collections powered by GBIF infrastructure <http://beta.idigbio.org/portal/collections> (beta version)

The screenshot displays the GRSciColl Collections Catalogue interface. At the top, there is a navigation bar with links: Natural History Collections, Institutions, Collections, Specimens, People, and About. Below this, a header section for 'Hacettepe University Biodiversity Advanced Research Center Herbarium' is visible, including a DOI (10.15468/5nille) and a list of major groups (Magnoliopsida, Liliopsida) and people (Jane Franklin, Viggo Mortensen). The main content area shows a list of specimens. The first specimen is 'HORNED PONDWEED' (*Lepilaena cylindrocarpa* Benth.), a holotype specimen from Taiwan, dated 13 July 2015, with 2 images and 16 citations. The second specimen is 'HORNED PONDWEED' (*Hyphaene guineensis* Schumach. & Thonn.), a paratype specimen from Estonia, dated 13 July 2015, with 2 images and 16 citations. The third specimen is also 'HORNED PONDWEED' (*Hyphaene guineensis* Schumach. & Thonn.), a paratype specimen from the British Indian Ocean Territory, dated 13 July 2015, with 2 images and 16 citations. Each specimen record includes a map of the location, a list of taxonomic ranks, and a 'More filters' button.





## HOSTED PORTALS

- Introduced at GB26 as part of work to lower technical threshold for participation in GBIF
- Raised interest from many GBIF Participants and partners
- Development work ongoing this year
  - Reworking our APIs (GraphQL-based)
  - Developing the first modules of user interfaces (ReactJS-based)
- Represents the next phase of infrastructure for GBIF
  - Will be consistent with GBIF
  - Will contribute to the envisaged collections catalogue

# THANK YOU!

[trobertson@gbif.org](mailto:trobertson@gbif.org)

[ahahn@gbif.org](mailto:ahahn@gbif.org)







## Capacity and Engagement during lockdown

Maheva Bagard Laursen, Laura Russell, Mélianie Raymond



**GBIF**

Global Biodiversity  
Information Facility

Image by Maheva Bagard Laursen



# UPDATE ON GBIF-LED PROGRAMMES

PROGRAMME

## Capacity Enhancement Support Programme

[ABOUT](#) [NEWS](#) [PROJECTS](#)



PROGRAMME

## BID: Biodiversity Information for Development

[ABOUT](#) [NEWS](#) [EVENTS](#) [PROJECTS](#)



PROGRAMME

## BIFA: Biodiversity Information Fund for Asia

[ABOUT](#) [NEWS](#) [EVENTS](#) [PROJECTS](#)



PROGRAMME

## Northern Eurasia

[ABOUT](#) [NEWS](#) [EVENTS](#) [PROJECTS](#)



# CAPACITY ENHANCEMENT SUPPORT PROGRAMME - CESP

Call for proposals for the 2020 Capacity Enhancement Support Programme

*Co-funding opportunity for collaborative projects targeting capacity enhancement and exchange to strengthen the GBIF network*



**19** Concept Notes from  
**17** GBIF participants

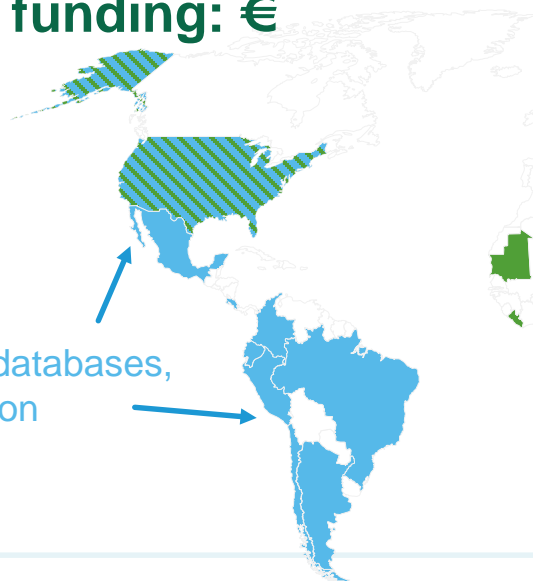
Requesting a total of  
€317,588

**10** proposals invited to  
submit a full proposals



**5 projects selected (total funding: €102,860):**

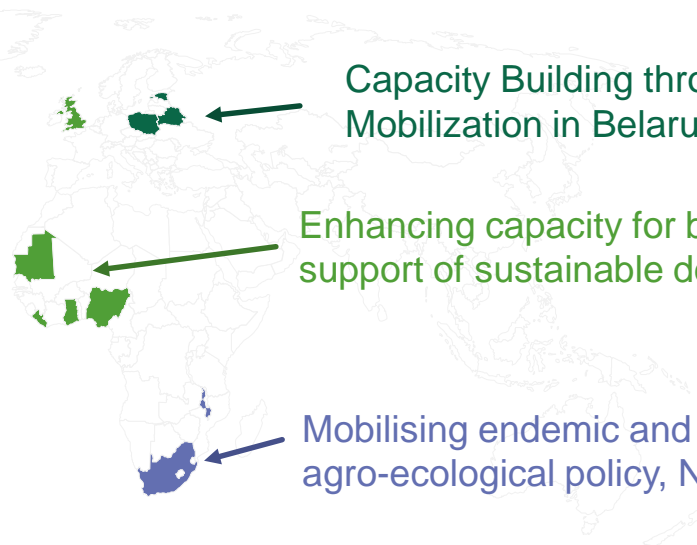
National Portals Workshop: taxonomic databases, species data information and visualization



Capacity Building through Regional Cooperation for Data Mobilization in Belarus, Latvia and Lithuania

Enhancing capacity for biodiversity data mobilization and use in support of sustainable development in West Africa

Mobilising endemic and economically useful species data for agro-ecological policy, National Red List and KBA assessments



# NORTHERN EURASIA

# BioDATA



UiO Natural History Museum



Diku

Direktoratet for  
innsamlingsforskning  
og kvalitetsutvikling  
Høyere utdanning



## Coordinated by:

- Natural History Museum of Oslo
- GBIF Norway

## Support from:

- GBIF Secretariat

## Objective:

Biodiversity data management skills for students through:

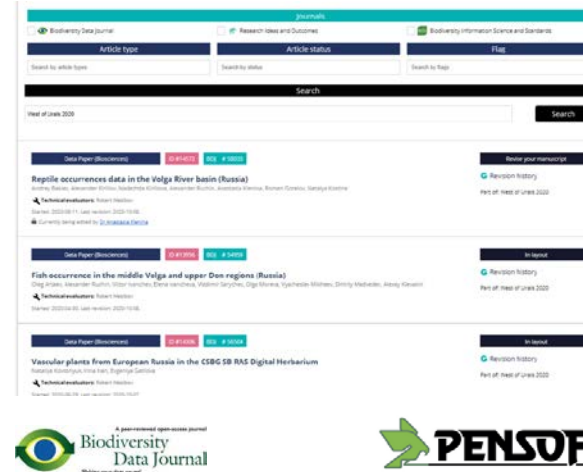
- Training courses
- Translation
- Capacity building

## Support for capacity development & Call for data papers from European Russia:

Funded by: €30,000 **FinBIF**  
INNOVATING PRACTICAL SOLUTIONS  
FOR FAIR BIODIVERSITY INFORMATICS

## Key activities:

- Training events
- Call for data papers: an incentive for data mobilization





# BIODIVERSITY INFORMATION FOR DEVELOPMENT (BID)

(closed) BID call for proposals: Sub-Saharan Africa 2020

*The Biodiversity Information for Development programme aims to enhance capacity for effective data mobilization and use for sustainable development. DEADLINE: 27 AUGUST 2020*



BID call for proposals: Caribbean 2020

*The Biodiversity Information for Development programme aims to enhance capacity for effective data mobilization and use for sustainable development. DEADLINE: 26 NOVEMBER 2020*



## 3 calls for proposals:

- Sub-Saharan Africa (closed)
- Caribbean (Open)
- Pacific (Open)

BID call for proposals: Pacific 2020

*The Biodiversity Information for Development programme aims to enhance capacity for effective data mobilization and use for sustainable development. DEADLINE: 26 NOVEMBER 2020*



## 4 grant types:

- Institution-level
- National
- Regional
- Data use

## Early results from the BID call for proposal in Sub-Saharan Africa:

- 215 concept notes
- From 39 countries in the region



*Concept notes received under the BID call in Sub-Saharan Africa*

# TIMELINE

**June 2020:**

Africa Call

**Sept. 2020:**

Caribbean Call  
Pacific Call

**April 2021 – March 2023**

Projects from Africa

**July 2021 – June 2023:**

Projects from the  
Caribbean & Pacific

**Second quarter 2021**

BID workshop in  
Africa

**Third quarter 2021**

BID workshop in  
Caribbean & Pacific

**July 2023 – August 2023**

Impact studies and  
metrics

**September 2023**

Report on BID  
impact

**September 2023**

End event

BID calls

Project implementation

Capacity enhancement

Impact analysis

End event

NEWS | 7 SEPTEMBER 2020  
**BID call for Development proposals: Caribbean 2020**  
Biodiversity Information for Development programme aims to enhance capacity for effective data mobilisation for sustainable development. DEADLINE: 26 NOVEMBER 2020



# BIODIVERSITY INFORMATION FUND FOR ASIA (BIFA)

Call for proposals promoting the mobilization and use of biodiversity data in Asia

*Biodiversity Information Fund for Asia (BIFA) programme invites new round of proposals from the region*



**32** Concept Notes from  
**13** countries in Asia

Requesting a total of  
€340,214

**13** proposals invited to  
submit a full proposals



Cleaning and digitizing plant specimen records  
from Helongjiang Province

Digitization of mycological collection in Nepal

Enriching the Bhutan Biodiversity Portal through digitization  
and mobilization of specimen collections

Digitizing and Databasing of bee specimens in Thailand

Supporting conservation through mobilizing ecological data from  
Kalimantan, Indonesia

**9 projects selected (total funding: €117,829):**

Mobilizing invertebrate monitoring data in South-East Asia

Using citizen science to study whales and dolphins in  
Vietnamese waters

Digitization of specimens at the University of the Philippines  
Los Baños Museum of Natural History

Mobilizing Indonesian butterfly collections at the  
Museum Zoologi Bogor (II)

<https://www.gbif.org/programme/82629/bifa-biodiversity-information-fund-for-asia>





CESP

09 December 2019

Two stage selection process:

- Concept note
- Full proposal

15 June 2020

Notification sent to projects selected to be funded

June 2020

Start of the project

October 2020

Webinar on Project implementation

31 December 2021

End of the project

- Final report

Call for proposals

Contract procedure

Project implementation (12 to 18 months)

BIFA

25 November 2019

Two stage selection process:

- Concept note
- Full proposal

14 May 2020

Notification sent to projects selected to be funded

July 2020

Start of the project

July 2020

(virtual) Data mobilization workshop

October 2020

Webinar on Progress reporting

31 December 2021

End of the project

- Final report

Call for proposals

Contract procedure

Project implementation (12 to 18 months)

BID

10 June 2020

Sub-Saharan Africa

Two stage selection process:

- Concept note
- Full proposal

07 September 2020

Pacific &amp; Caribbean

Two stage selection process:

- Concept note
- Full proposal

March 2021

Selected projects from Africa

June 2021

Selected projects from the Caribbean and Pacific

April 2021 – March 2023

Projects from Africa

July 2021 – June 2023:

Projects from the Caribbean &amp; Pacific

Second quarter 2021

BID workshop in Africa

Third quarter 2021

BID workshop in Caribbean &amp; Pacific

Call for proposals

Contract procedure

Project implementation (24 months)

Activities related to GBIF led-programme in 2020 - 2021

# STREAMLINING PROCESSES TO INCREASE EFFICIENCY



GBIF Grants Portal

Login Now:

Sign in

[Reset or create password](#)

**New to the GBIF Grants Portal?**

**Create and register your organizational profile**

Grants available via the GBIF Grants Portal are only available and awarded to organizations. In order to be considered for funding, the first step is to introduce yourself and tell us about your organization. Click on the "Create an account now" button below to start.

Note that you can not edit your profile after submission, so please ensure you enter accurate information when completing the registration form.

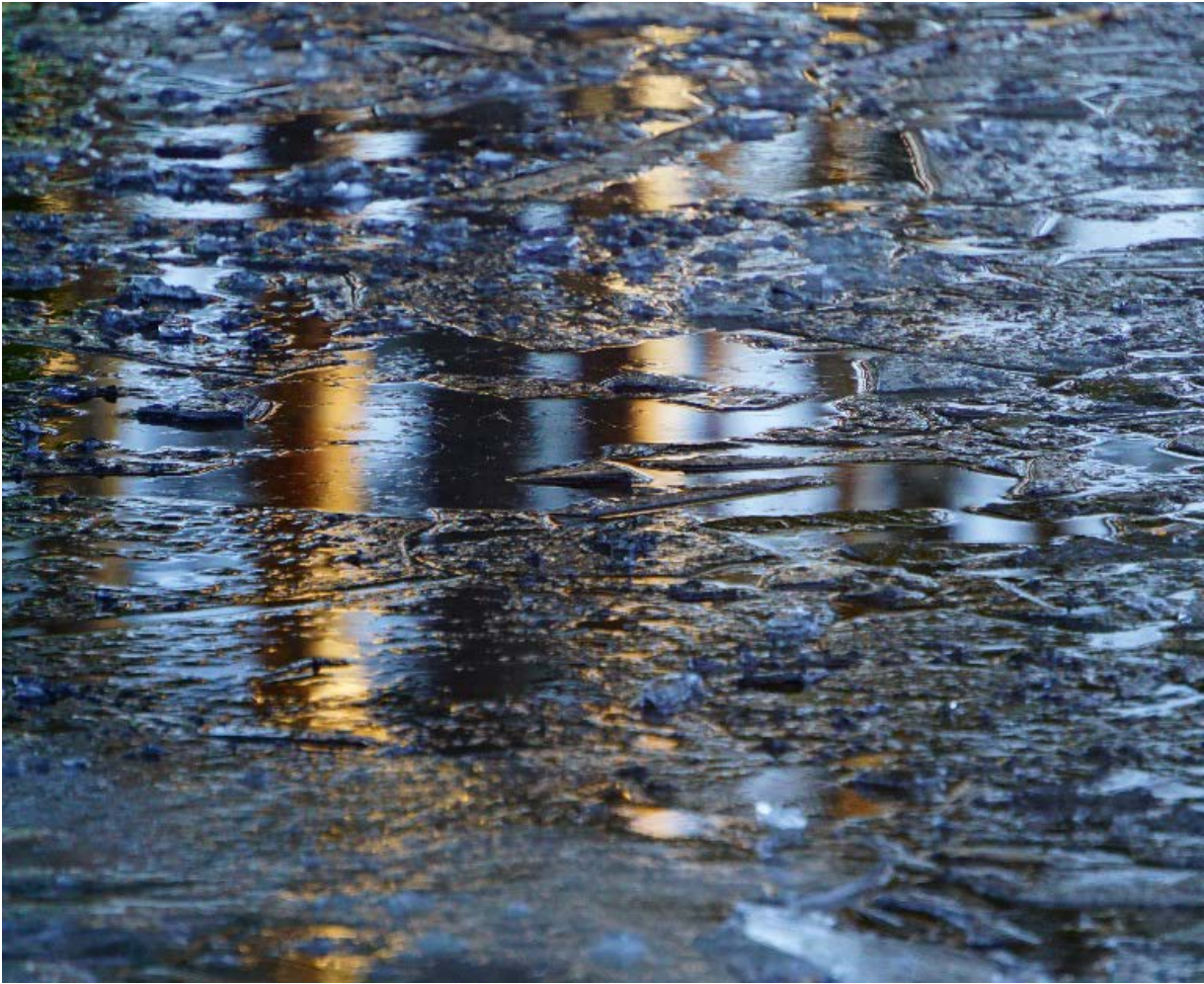
Once submitted, you will receive an email notification from GBIF within 1-2 working days with your login information. Do not forget to check your spam folder for the email.

This login will allow you to access the portal and apply to open calls. If your application is successful, the portal will also provide access to reporting templates.

Create an account now



[Privacy Policy](#) [Accessibility](#)







## IMPORTANCE OF THE GBIF COMMUNITY

- Grant review process
- Selection panel
- Project mentor
- Capacity enhancement workshop (Mentors & Trainers)
- Translations



# UNEXPECTED CIRCUMSTANCES



Image by Maheva Bagard Laursen



# DOCUMENTATION

- OpenRefine guide:  
<https://doi.org/10.15468/doc-gzjg-af18>
- Sensitive species guide:  
<https://doi.org/10.15468/doc-5jp4-5g10>
- Georeferencing best practices guide:  
<https://doi.org/10.15468/doc-gg7h-s853>
- Georeferencing quick reference guide:  
<https://doi.org/10.35035/e09p-h128>
- Georeferencing calculator manual:  
<https://doi.org/10.35035/gdwq-3v93>
- Publishing DNA-derived data guide:  
<https://doi.org/10.35035/doc-vf1a-nr22>

Get data   How-to   Tools   Community   About       [Login](#)

NEWS | 17 SEPTEMBER 2020

## Georeferencing documents released for GBIF community review

*Deadlines for input from community review: 1 November 2020*



Commerson's dolphin (*Cephalorhynchus commersonii*), off the coast of Chubut, Argentina. Photo 2018 Gabriel Laufer via iNaturalist research-grade observations, in the public domain under CC0.

GBIF has released three new technical documents for community review and welcomes input and feedback on each of the following publications:

- [Georeferencing Best Practices](#), by Arthur Chapman & John Wieczorek
- [Georeferencing Quick Reference Guide](#), by Paula Zermoglio, Arthur Chapman, John Wieczorek, Maria Celeste Luna & David Bloom
- [Georeferencing Calculator Manual](#), by David Bloom, John Wieczorek & Paula Zermoglio

Taken together, these three georeferencing documents provide essential tools for improving the practice of spatially interpreting locations, particularly (but not exclusively) for those working with biological occurrence data from museum-based scientific collections. The Best Practices guide provides theoretical background and methods for georeferencing descriptive localities along with common terms

# Grants and funding

Informational videos and tutorials for helping you successfully apply for funding through a GBIF-managed grant programme

Share



Project reporting

GBIF



Financial reporting

GBIF



Objective and risk assessments

GBIF



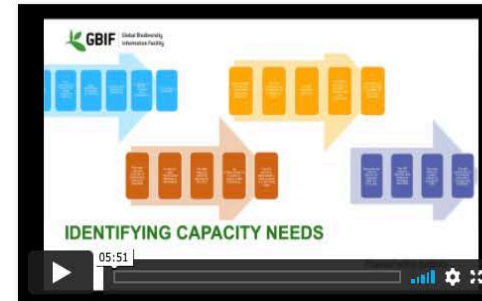
Project deliverables

GBIF



Budget details

GBIF



Identifying Capacity Needs

GBIF



## COURSE VIDEOS

# Data Mobilization Course

These videos are part of a series of presentations used in the GBIF Biodiversity Data Mobilization course. The biodiversity data mobilization curriculum was developed as part of the Biodiversity Information Development (BID) Programme funded by the European Union. The curriculum was developed and expanded by GBIF community Trainers, Mentors and Students. You can subscribe to the full course via our e-Learning platform hosted by GBIF Spain. <https://elearning.gbif.es/>

1 - 12 of 16

Share



### Introduction to GBIF

GBIF



### GBIF: La biblioteca de la vida

GBIF

¿Te imaginas cuánto conocimiento podría ser generado, si las personas pudieran encontrar y usar los datos sobre biodiversida...



### Foundations - Definitions

GBIF



### Foundations - Software

GBIF



### Foundations - Structures

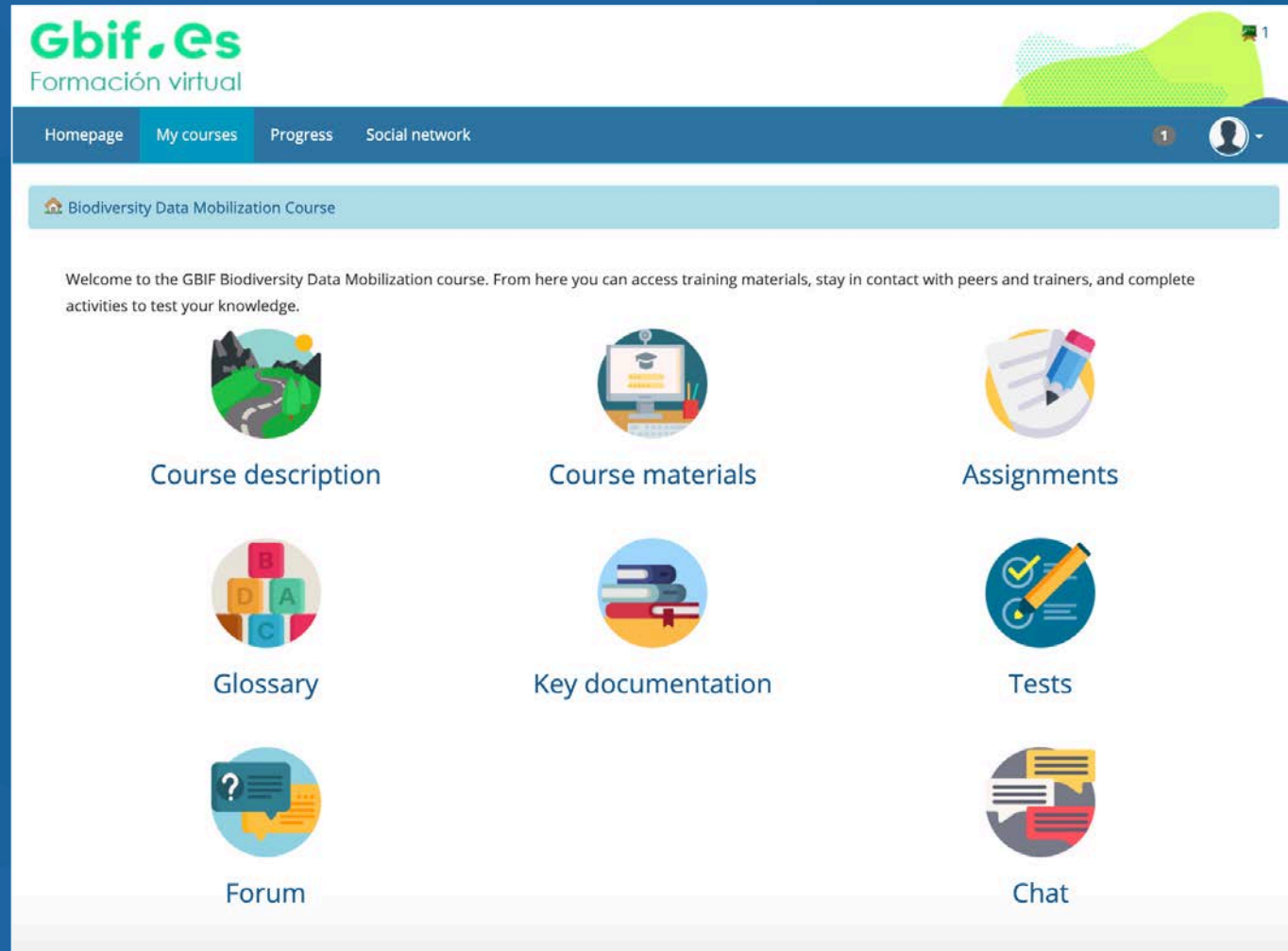
GBIF



### Foundations - Data Quality

GBIF

# E-LEARNING PLATFORM



The screenshot displays the Gbif.es virtual training platform interface. At the top left is the logo "Gbif.es" with the tagline "Formación virtual". To the right of the logo is a green abstract graphic. Below the logo is a navigation bar with links: "Homepage", "My courses", "Progress", and "Social network". On the right side of the navigation bar, there is a notification icon with the number "1" and a user profile icon. Below the navigation bar, a light blue banner reads "Biodiversity Data Mobilization Course". A welcome message follows: "Welcome to the GBIF Biodiversity Data Mobilization course. From here you can access training materials, stay in contact with peers and trainers, and complete activities to test your knowledge." Below the welcome message are nine circular icons arranged in a 3x3 grid, each with a corresponding label: "Course description" (landscape with a path), "Course materials" (computer monitor with a graduation cap), "Assignments" (notepad and pencil), "Glossary" (four colored blocks labeled A, B, C, D), "Key documentation" (stack of books), "Tests" (pencil and checkmark), "Forum" (speech bubble with a question mark), and "Chat" (speech bubbles). The "Tests" icon is missing a label.

**Gbif.es**  
Formación virtual

Homepage My courses Progress Social network

Biodiversity Data Mobilization Course

Welcome to the GBIF Biodiversity Data Mobilization course. From here you can access training materials, stay in contact with peers and trainers, and complete activities to test your knowledge.

Course description Course materials Assignments

Glossary Key documentation Tests

Forum Chat

# VIRTUAL WORKSHOP

Data Mobilization  
for Asia 2020

6 July – 31 Aug





# COURSE AGENDA

6-21 July	22 July	23 July	24 July - 31 August
Self-paced	Self-paced	Self-paced	Self-paced
Section I 6-17 July	Section II: Planning	Section II: Data management	Section II: Data publishing
03:00-05:00 UTC, 21 JULY Zoom Session - 1	03:00-05:00 UTC, 22 JULY Zoom Session - 3	03:00-05:00 UTC, 23 JULY Zoom Session - 5	03:00-05:00 UTC, 24 JULY Zoom Session - 7
Lead: Laura Welcome and introductions	Leads: Sharon and Dave Planning activity, discussion and Q&A	Leads: Sharon and Dave Data management activity	Leads: Sharon and Dave Data publishing practice and Q&A
Self-paced	Self-paced	Self-paced	Self-paced
Review: Section I Foundations	Section II: Data capture	Section II: OpenRefine	Review Use Case II and Use Case III
07:00-09:00 UTC, 21 JULY Zoom Session - 2	07:00-09:00 UTC, 22 JULY Zoom Session - 4	07:00-09:00 UTC, 23 JULY Zoom Session - 6	07:00-09:00 UTC, 24 JULY Zoom Session - 8
Leads: Laura and Sophie Foundations discussion and Q&A	Leads: Laura and Sophie Data capture discussion and Q&A	Leads: Laura and Sophie Data management discussion and Q&A	Leads: Laura and Sophie Final review and conclusion
Self-paced	Self-paced	Self-paced	Self-paced
Section II: Planning	Section II: Data management	Section II: Data publishing	Section III 27 July - 31 Aug

# COMMENTS FROM COURSE PARTICIPANTS

Thank you very much for all our mentors' efforts in making the classes engaging and worthwhile. It's my first time to have a full-blown virtual training workshop and I definitely learned a lot, at my own pace. The technology utilized is good for me as I was able to catch up and understand the coursework better.

There were some minor issues that I faced with the e-learning platform - doesn't load and have to re-sign in a couple of time. - Vimeo could be reconsidered as a platform for video upload as it cannot be accessed in Indonesia (not sure about any other countries with the same issue)

Indeed, it went well. I could be more better if we would have discussed more among the BIFA-Grantees and Mentors.

The interenet is sometimes not effective

The online platform used was efficient and effective. However, I still believe that physical "hands on" training will be the best way to indulge in the subject and learn with more passion.

Thank you very much for all your efforts in making this training workshop worth our time and effort. There really is no replacement for the physical meetings and interactions that make learning better. However, with the current health concerns, your efforts to make this course possible is really commendable! Again, thank you very much to all our mentors.

# LESSONS LEARNED

- Ability to teach more participants
- Planning across disparate time zones is difficult
- Continue to design online courses with options for virtual or in-person activities.
- Find ways to virtually replicate networking between individuals





# WORKSHOP MENTORS AND TRAINERS



1. Ahmad Taufiq, Oliver Griffin, Dieta Sudarma, Abdulla Ghazal
2. Melissa Liu, Saunak Pal, Arman Pili, Wei Song Hwang
3. Sharon Grant, Sophie Pamerlon, David Bloom

## REGIONAL MEETINGS IN 2020

Support regional coordination

Facilitate regional  
collaboration

Provide feedback to the  
Secretariat and NSG

Develop stronger regional  
engagement



GBIF Community Webinar:  
Framing the 2020 virtual regional  
meetings



# VIRTUAL REGIONAL NODES MEETINGS

## North America

5-6 May 2020

39 attendees

## Europe & Central Asia

11-12 May 2020

39 attendees

## Latin America & Caribbean

18 & 20 May 2020

38 attendees

## Africa

10 & 12 June

42 attendees

## Asia

17 July 2020

14 attendees





# BID REGIONAL ENGAGEMENT WEBINARS



## Latin America & Caribbean

19 May 2020

101 attendees

## Africa

11 June 2020

94 attendees

## Pacific

7 September 2020

29 attendees



# GBIF ASIA VIRTUAL SUMMIT

## Asia

20 July 2020

144 attendees



Panel discussion to discuss the issues raised by the 20 year review, including the recommendation to **build Eastern participation** (R5)



# CHALLENGES OF VIRTUAL MEETINGS FOR OUR COMMUNITY ENGAGEMENT



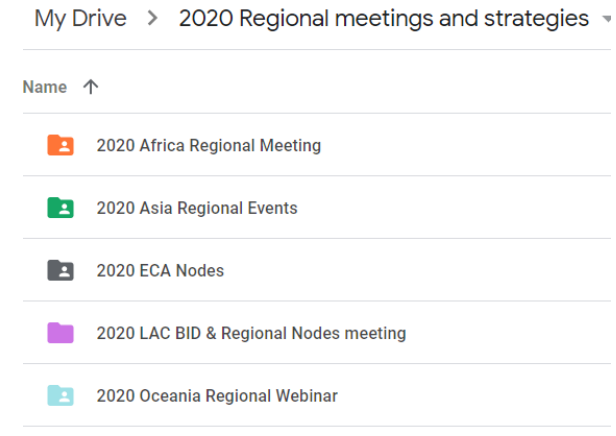
- More difficult to manage large group discussions and consensus forming



- Not easy to allow for informal discussions, coffee break conversations, personal connections

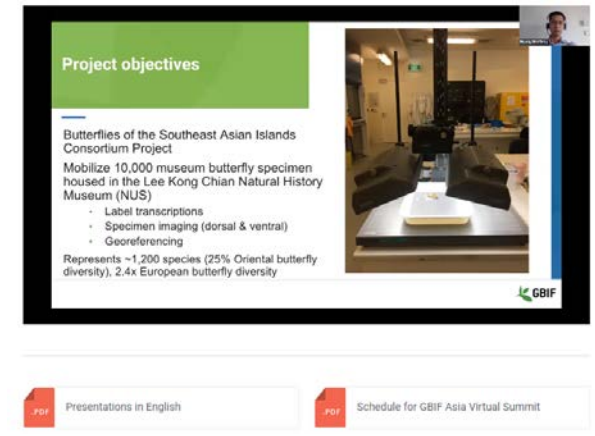
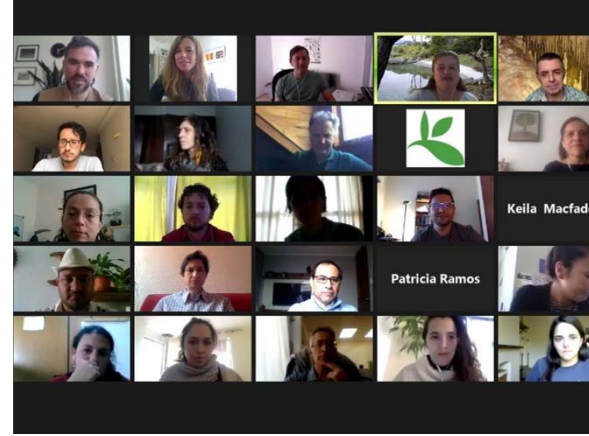


- Complications of working across time zones



- At least as much time (or more time) spent on preparation as for in person meetings

# ADVANTAGES OF VIRTUAL MEETINGS FOR OUR COMMUNITY ENGAGEMENT



- Ability to invite more people to our events
- Easy to record meetings for those that can't attend at the specified time
- Saving of travel costs and travel time (emissions and jet lag)
- Some community building effect, especially in small groups
- Good for sharing information and updates

## STRONG COMMUNITY

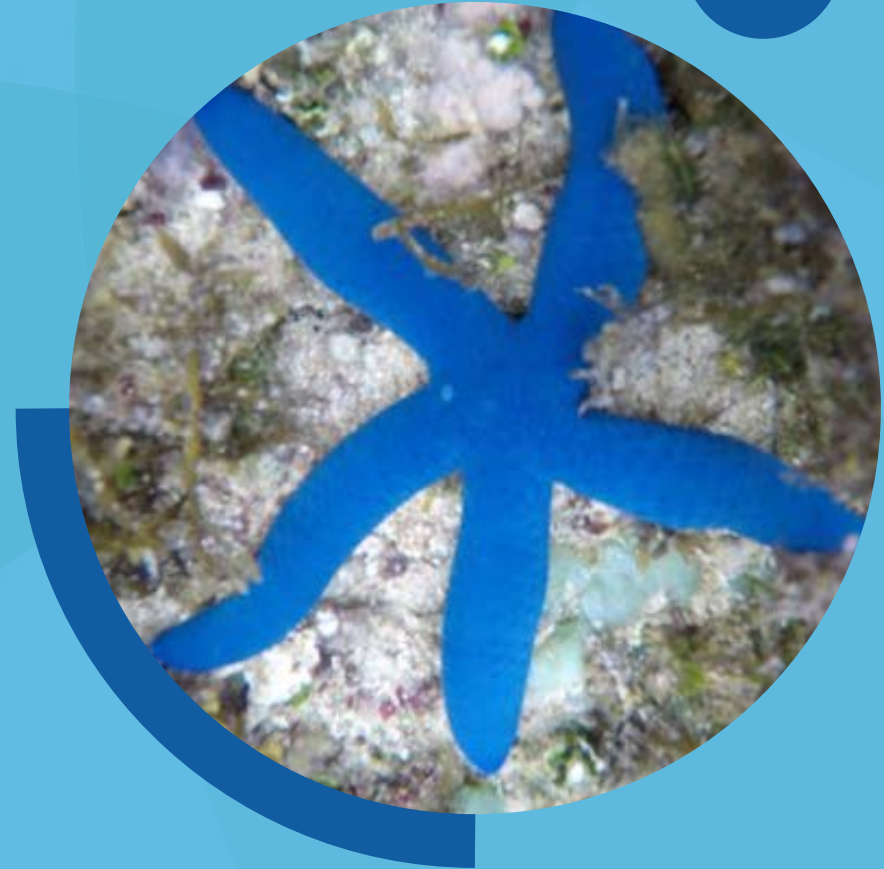
Thanks to the  
community for the  
flexibility and  
dedication to  
supporting the capacity  
and engagement work  
this year





# THANK YOU

mblaursen@gbif.org  
larussell@gbif.org  
mraymond@gbif.org



*Linckia laevigata* – Mélianie Raymond – CCBY - <https://www.inaturalist.org/photos/8403155>



## Private Sector Engagement

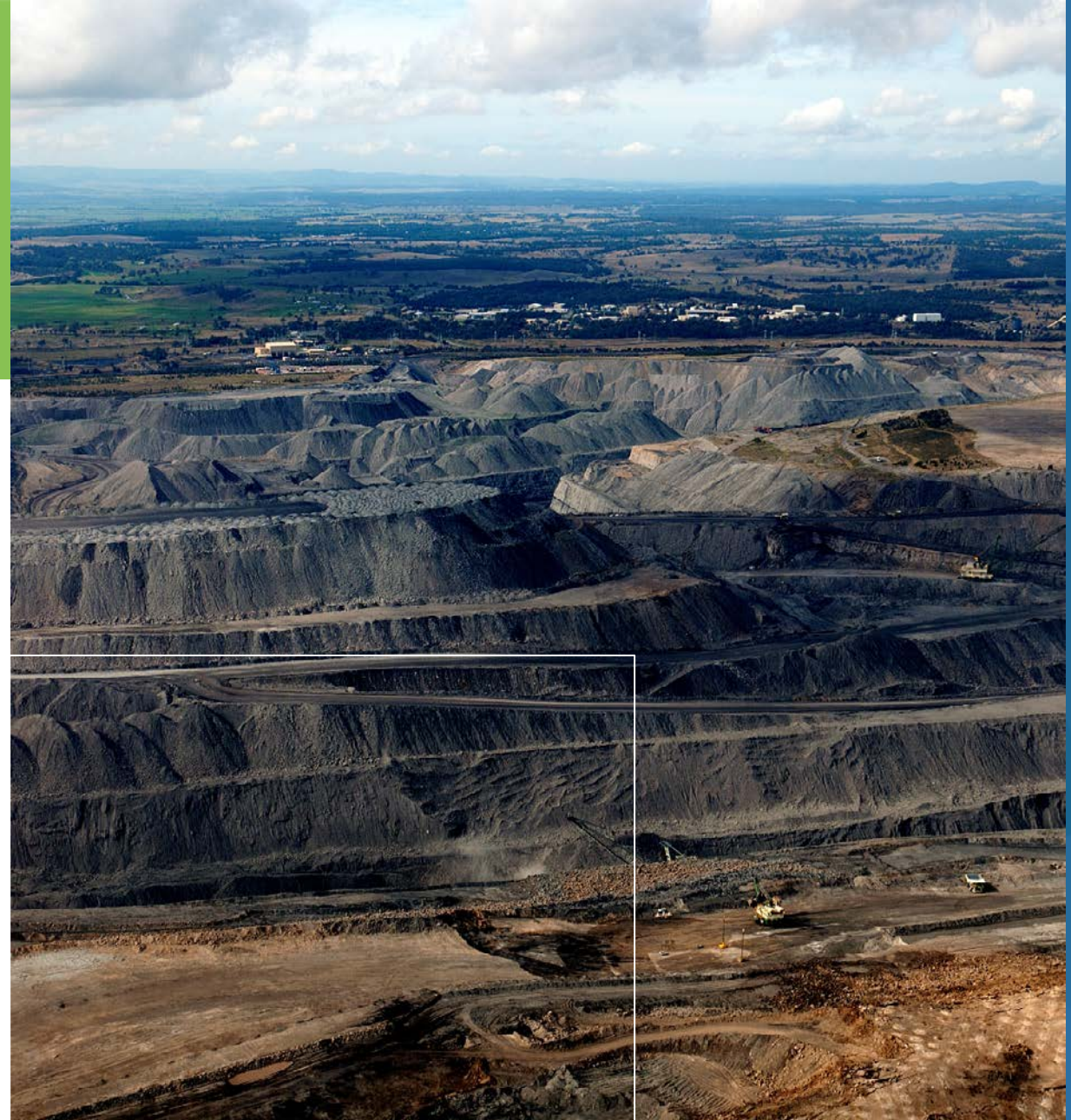
Andrew Rodrigues | Programme Officer





# BUSINESS AND BIODIVERSITY DATA

- Business uses biodiversity data to mitigate the impacts of their operations through environmental and social impact assessments
- Companies want to build their social licence to operate and demonstrate their commitment to international development goals
- Data sharing represents a low cost opportunity with high impact potential





# DATA FROM BUSINESS AND GBIF

- 27 private sector companies are publishing data through GBIF have published 1,419,240 records in GBIF
- Private companies are willing to share data when there is pro-active engagement
- Other engagement activities in past
  - Eye on Earth
  - International Association for Impact Assessments

Company	Activity sector	Country	Datasets	Occurrence records
<a href="#">Aigos SAS</a>	Consulting	Colombia	3	2,404
<a href="#">Anadarko Colombia Company</a>	Energy	Colombia	3	282
<a href="#">Arctic Research and Consulting DA</a>	Consulting	Norway	1	8,914
<a href="#">Asplan Viak AS</a>	Engineering	Norway	14	3,779
<a href="#">Biofokus</a>	Consulting	Norway	1	508,993
<a href="#">Biolog J.B. Jordal AS</a>	Consulting	Norway	1	177,814
<a href="#">Celsia</a>	Energy	Colombia	1	35
<a href="#">Chevron Australia</a>	Energy	Australia	1	2,048
<a href="#">Ecofact</a>	Consulting	Norway	1	11,273
<a href="#">EDP - Energias de Portugal</a>	Energy	Portugal	73	601,636
<a href="#">Faun Naturforvaltning AS</a>	Consulting	Norway	1	3,788
<a href="#">Federación Nacional de Cafeteros</a>	Agriculture	Colombia	6	26,840
<a href="#">Hatovial SAS</a>	Engineering	Colombia	1	1,898
<a href="#">INERCO</a>	Consulting	Colombia	1	1,090
<a href="#">Isagen</a>	Energy	Colombia	8	19,352
<a href="#">LafargeHolcim</a>	Mining	Spain	2	35
<a href="#">Moam SAS</a>	Consulting	Colombia	1	1,781
<a href="#">Multiconsult</a>	Consulting	Norway	1	308
<a href="#">Naturrestaurering AS</a>	Consulting	Norway	2	515
<a href="#">NNI</a>	Consulting	Norway	2	3,116
<a href="#">Oleoducto Bicentenario</a>	Energy	Colombia	3	2,074
<a href="#">Rådgivende Biologer</a>	Consulting	Norway	5	15,214
<a href="#">Stratos</a>	Consulting	Colombia	1	849
<a href="#">SUEZ</a>	Energy	Spain	1	56,847
<a href="#">SWECO Norge AS</a>	Engineering	Norway	1	1,139
<a href="#">Terrasos</a>	Consulting	Colombia	4	9,781
<a href="#">Total</a>	Energy	France	1	324

# SHIFT TOWARDS PUBLICATION THROUGH GBIF

Move towards making biodiversity data openly and freely available

- New legislation e.g. Western Australia

*"I welcome agreement to progress digital environmental impact assessment. Ensuring WA and the country has efficient and effective environmental regulation systems is vital to protect our environment for future generations and providing certainty to industry."*

- Hon Stephen Dawson MLC, Minister for Environment; Disability Services; Electoral Affairs – Government of Western Australia

- Open data organisational policies

The image is a composite graphic. At the top, a banner with a blue and green background features the text "DIGITALLY TRANSFORMING ENVIRONMENTAL ASSESSMENT" in white and orange boxes, with the subtitle "Leveraging Digital Information to Streamline Environmental Assessment and Approvals". Below this, on the left, is a small image of a mouse. To the right is a screenshot of a GBIF dataset page titled "Marine Mammal Observation - Total E&P Angola 2006-2011". The page shows metadata, a list of occurrences (324), and a map of the study area. At the bottom, logos for "The Western Australian Biodiversity Science Institute" and "Western Marine Institute" are displayed.

# SHIFT TOWARDS PUBLICATION THROUGH GBIF

Financial institutions are also demanding transparency and open data in the projects they finance

## - Equator Principles

*“The EPFI will encourage the client to share commercially non-sensitive Project-specific biodiversity data with the Global Biodiversity Information Facility (GBIF) and relevant national and global data repositories, using formats and conditions to enable such data to be accessed and re-used in future decisions and research applications.”*



Regions	No. of projects with environmental impact assessments	% of all GBIF-mediated data generated by AFD financing	% of non-bird, GBIF-mediated data generated by AFD financing
Southern Africa	8	0,07	2
East Africa	25	1,17	23,8
North Africa	21	5,04	157,2
West and Central Africa	74	12,15	23,4
Total for Africa	128	0,85	15,5
Central America (only Mexico)	2	0,03	0,3
South America	15	0,07	1,4
The Antilles	2	1,17	21,5
Total for Latin America and the Caribbean	19	0,07	1,16
Central Asia (only Uzbekistan)	1	4,87	60
East Asia (only China)	13	1,91	60,7
West Asia	19	2,26	127,9
South Asia	18	0,12	20,4
South-east Asia	12	1,39	36
Total for Asia	63	0,37	39,4

## • Development Banks – Data4Nature



# NEXT STEPS

- Consolidating existing guidance and expertise
  - [OpenPSD Project](#)
  - IAIA-GBIF
  - Equator Principles
- Capacity building for data publishing
  - Training materials
  - Helpdesk services
- Continued promotion of biodiversity data sharing by business
  - Supporting campaigns such as Data4Nature
  - Engaging in sustainable business fora
  - Developing the GBIF business case
  - Engaging with sectoral associations and bodies

# THANK YOU

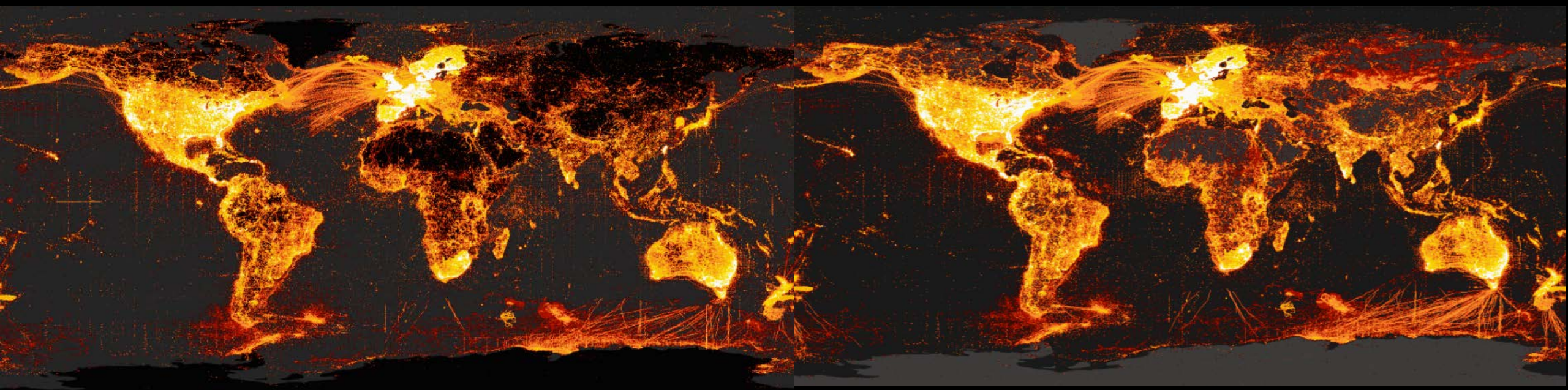
Andrew Rodrigues  
Programme Office for Participation and Engagement  
[arodrigues@gbif.org](mailto:arodrigues@gbif.org)



2018



2020



981,464,491



1,602,089,347





# DNA-DERIVED DATA IN GBIF GUIDANCE AND TRAINING

Dmitry Schigel | Scientific officer



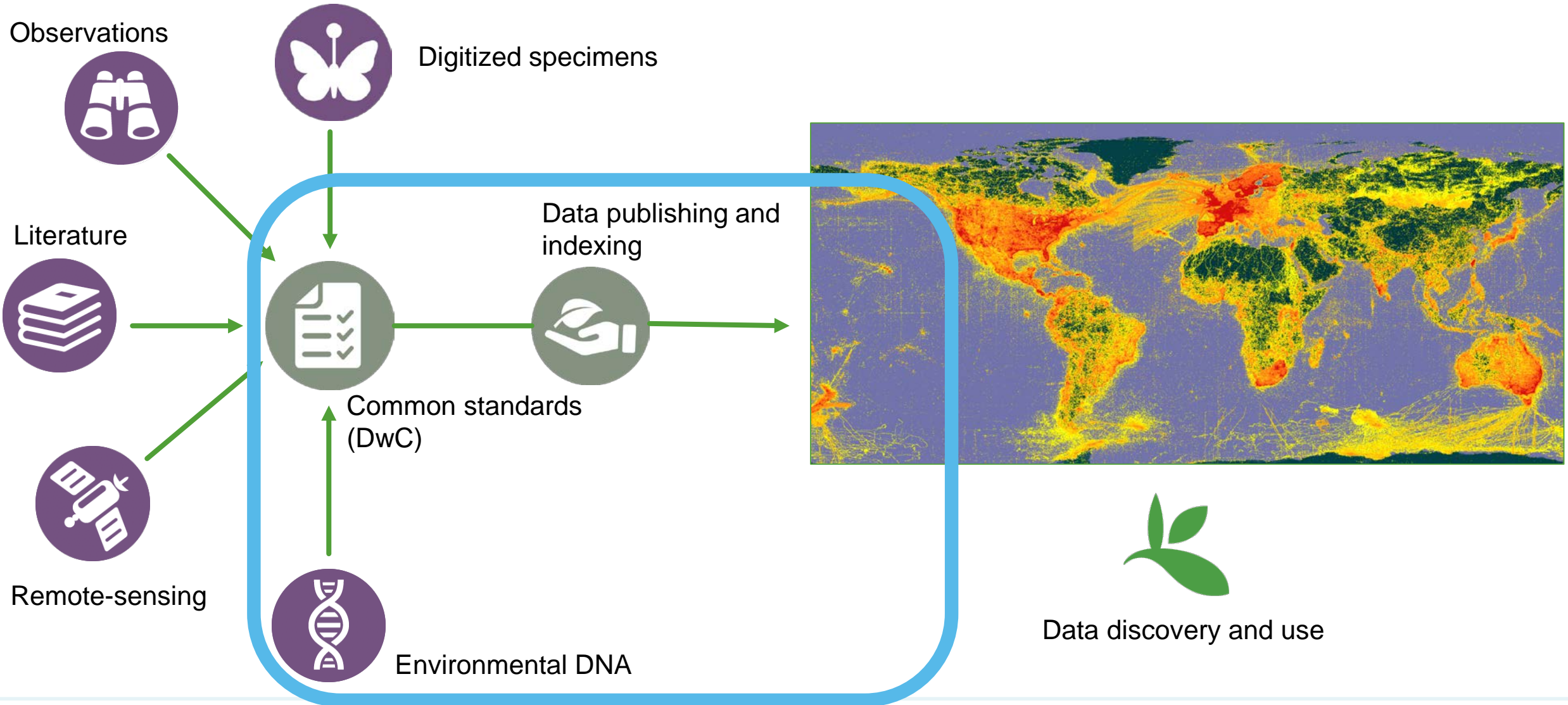
**GBIF**

Global Biodiversity  
Information Facility

October 2020

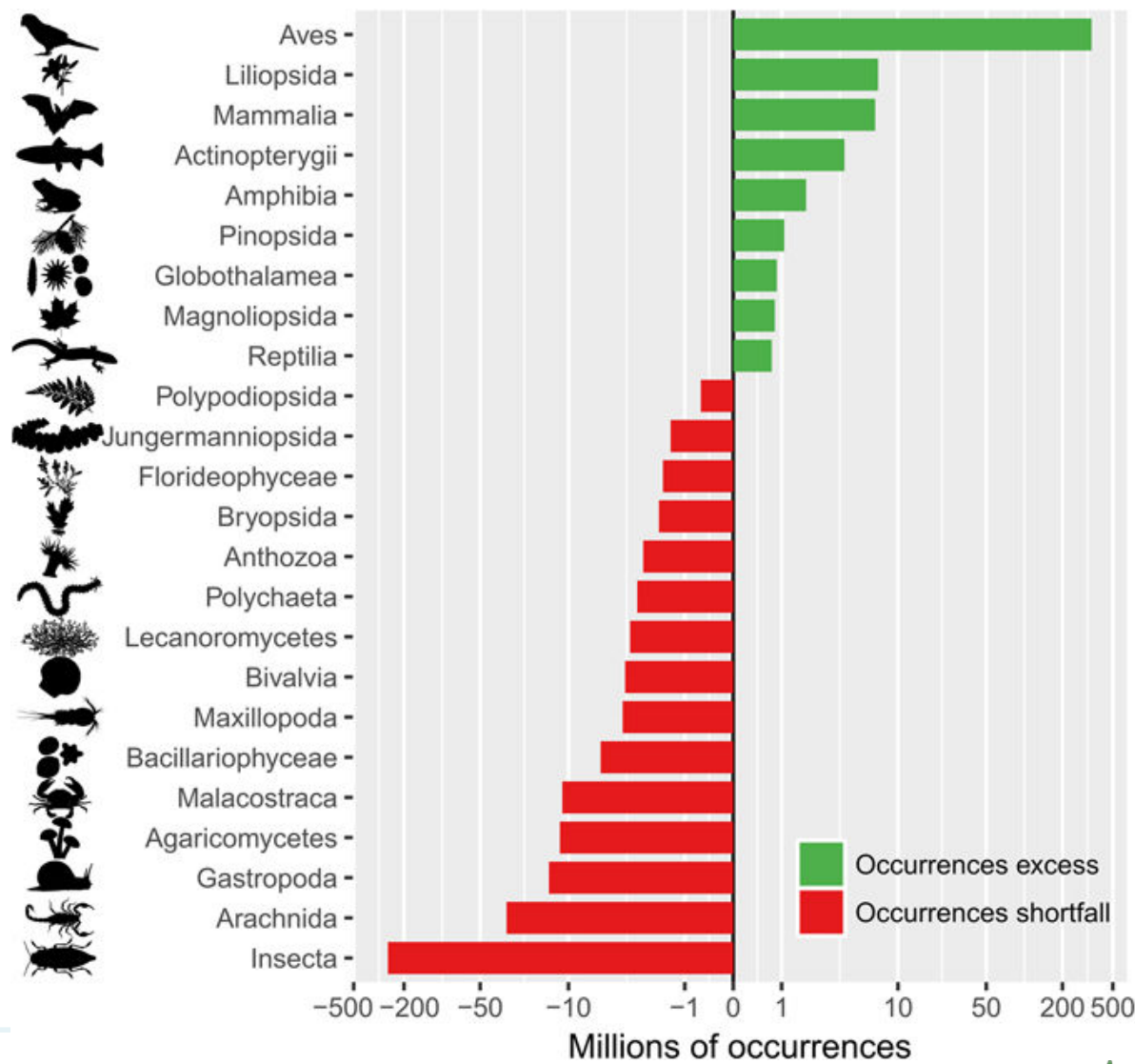
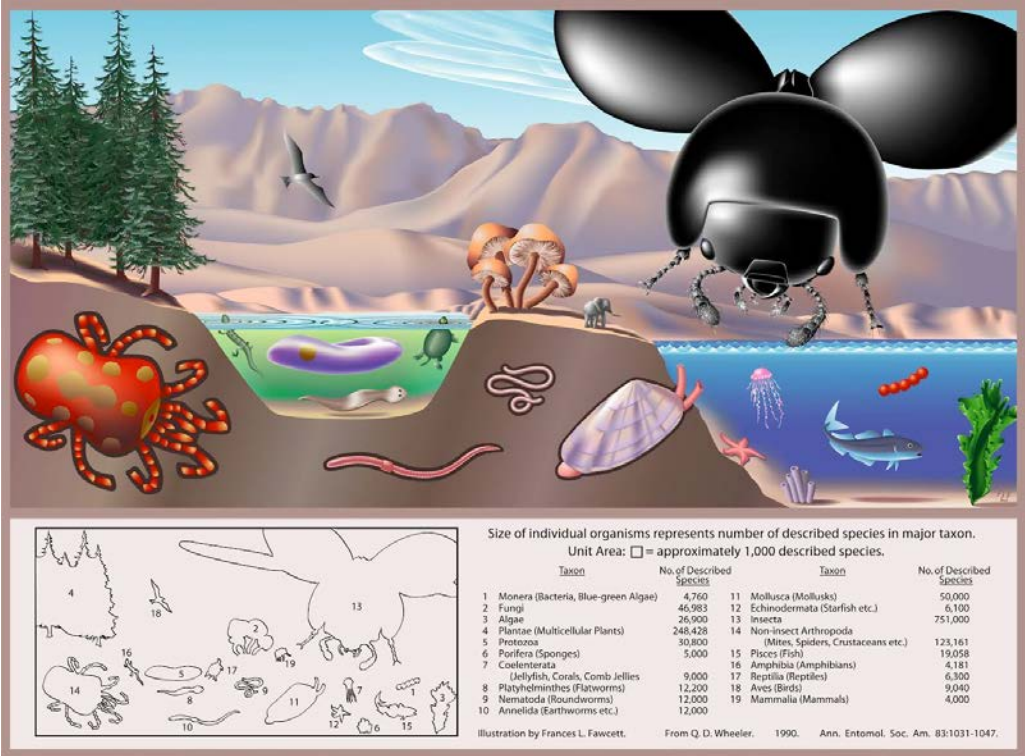


# A WINDOW ON EVIDENCE ABOUT WHERE SPECIES HAVE LIVED, AND WHEN



# GLOBAL BIODIVERSITY VS. DIGITALLY AVAILABLE DATA

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



Trout et al. Nature Scientific Reports 2017



DATA STREAMS IN GBIF

DNA-based evidence

natural history collections

human observations etc.

Sequence names

“taxon hypotheses”, OTUs

UNITE: SH

BOLD: BINs

SILVA

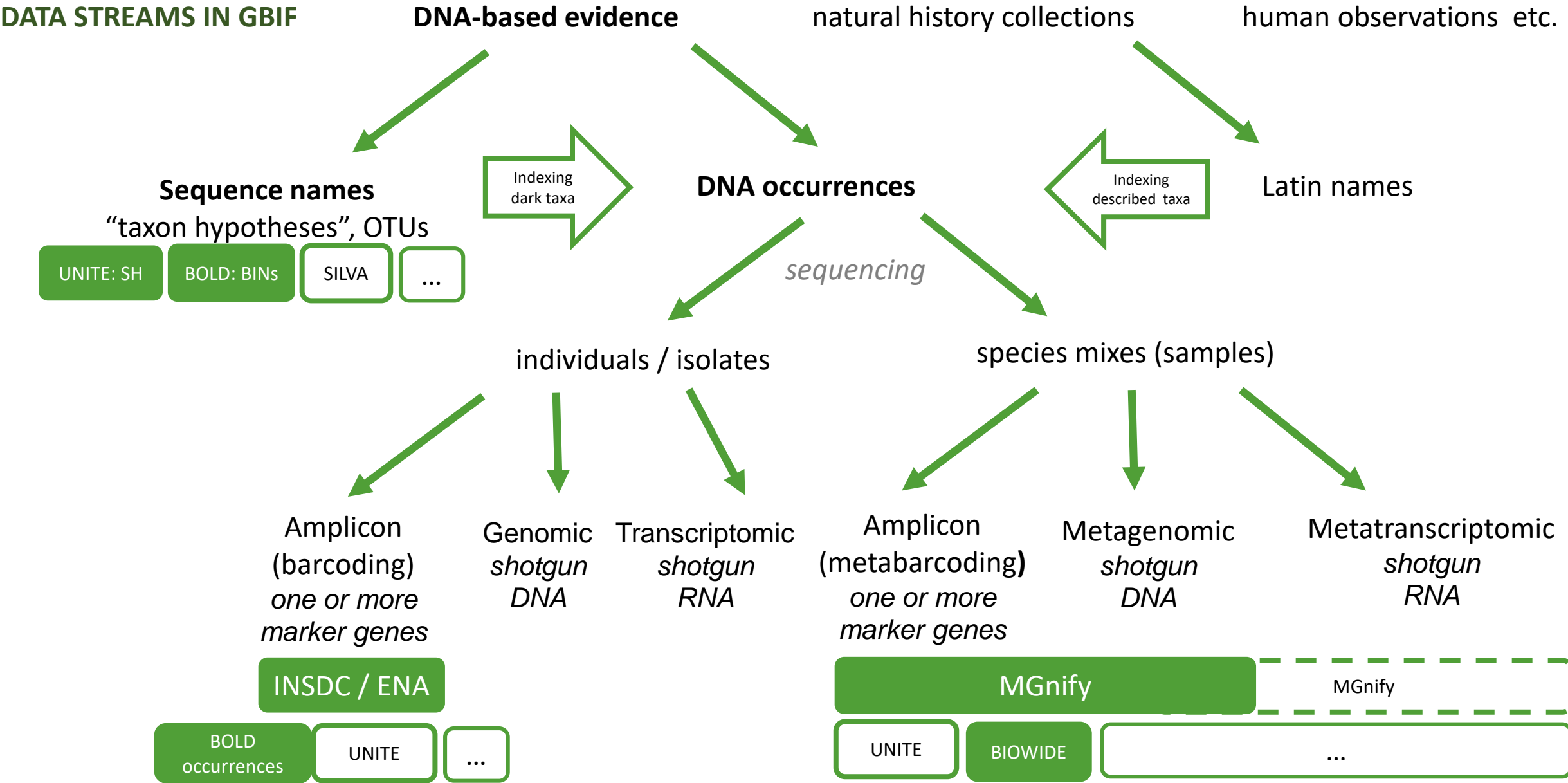
...



DNA occurrences



Latin names



DATA STREAMS IN GBIF

DNA-based evidence

natural history collections

human observations etc.

Sequence names

“taxon hypotheses”, OTUs

UNITE: SH

BOLD: BINs

SILVA

...

Indexing  
dark taxa

DNA occurrences

Indexing  
described taxa

Latin names

*sequencing*

individuals / isolates

species mixes (samples)

Targeted detection  
qPCR

Norway

?

Amplicon  
(barcoding)  
*one or more  
marker genes*

INSDC / ENA

BOLD  
occurrences

UNITE

...

Genomic  
*shotgun  
DNA*

Transcriptomic  
*shotgun  
RNA*

Amplicon  
(metabarcoding)  
*one or more  
marker genes*

Metagenomic  
*shotgun  
DNA*

Metatranscriptomic  
*shotgun  
RNA*

MGnify

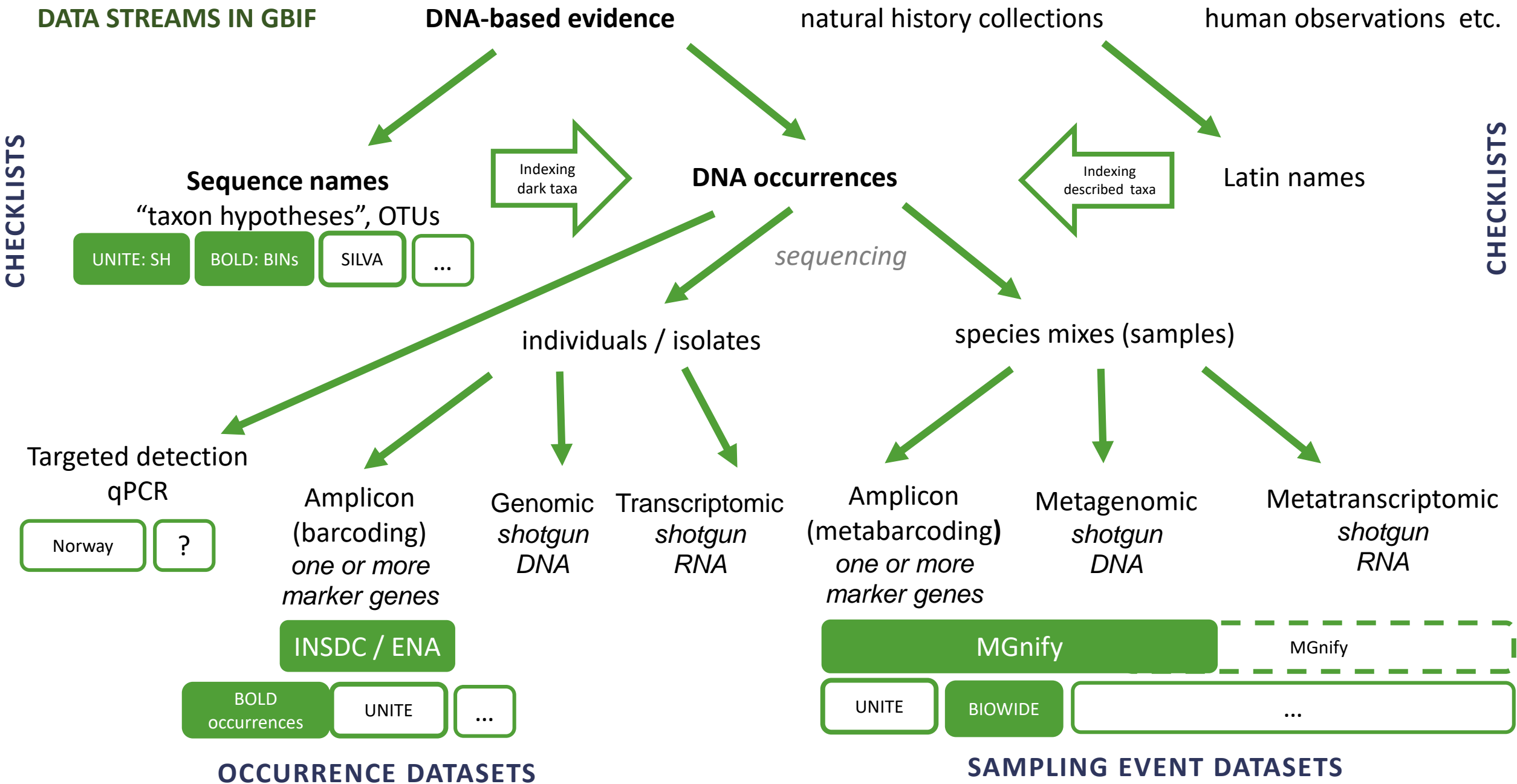
UNITE

BIOWIDE

...

MGnify



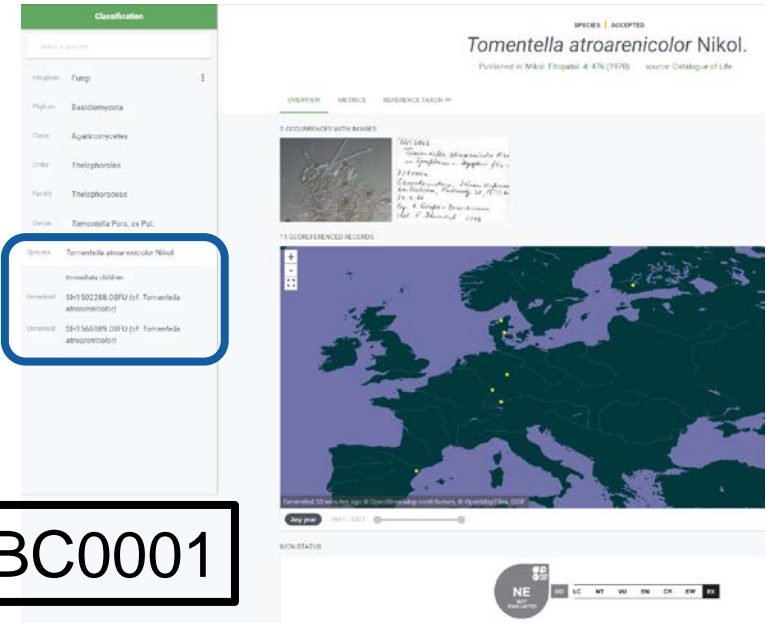


OPERATIONAL  
TAXONOMIC  
UNITS

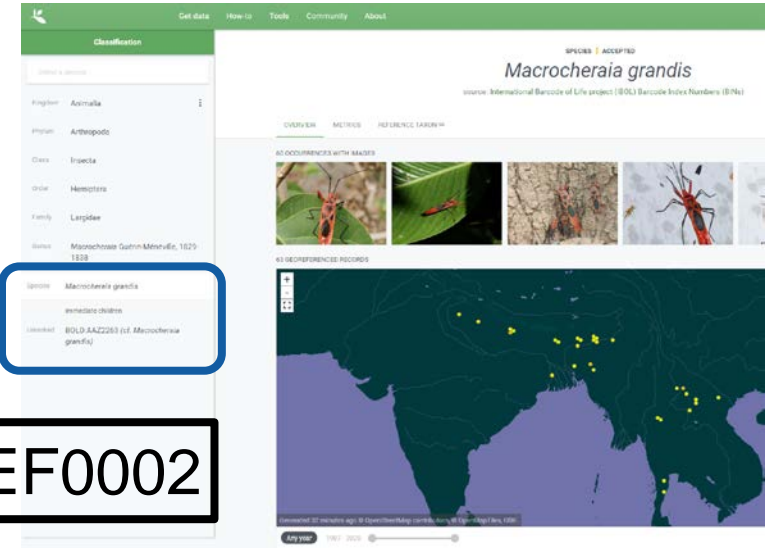


GBIF  
backbone taxonomy

SH ABC0001



OTU = **SH**,  
Species hypothesis



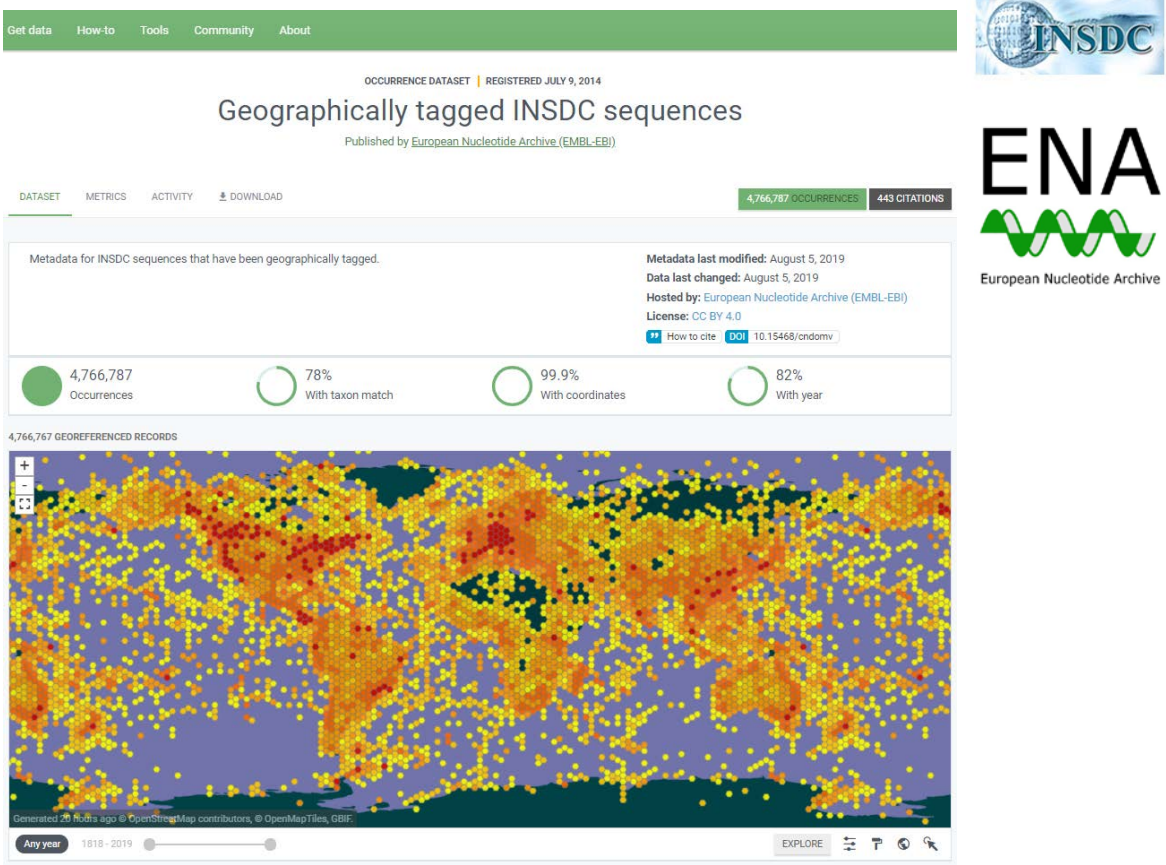
OTU = **BIN**,  
Barcode  
identification number



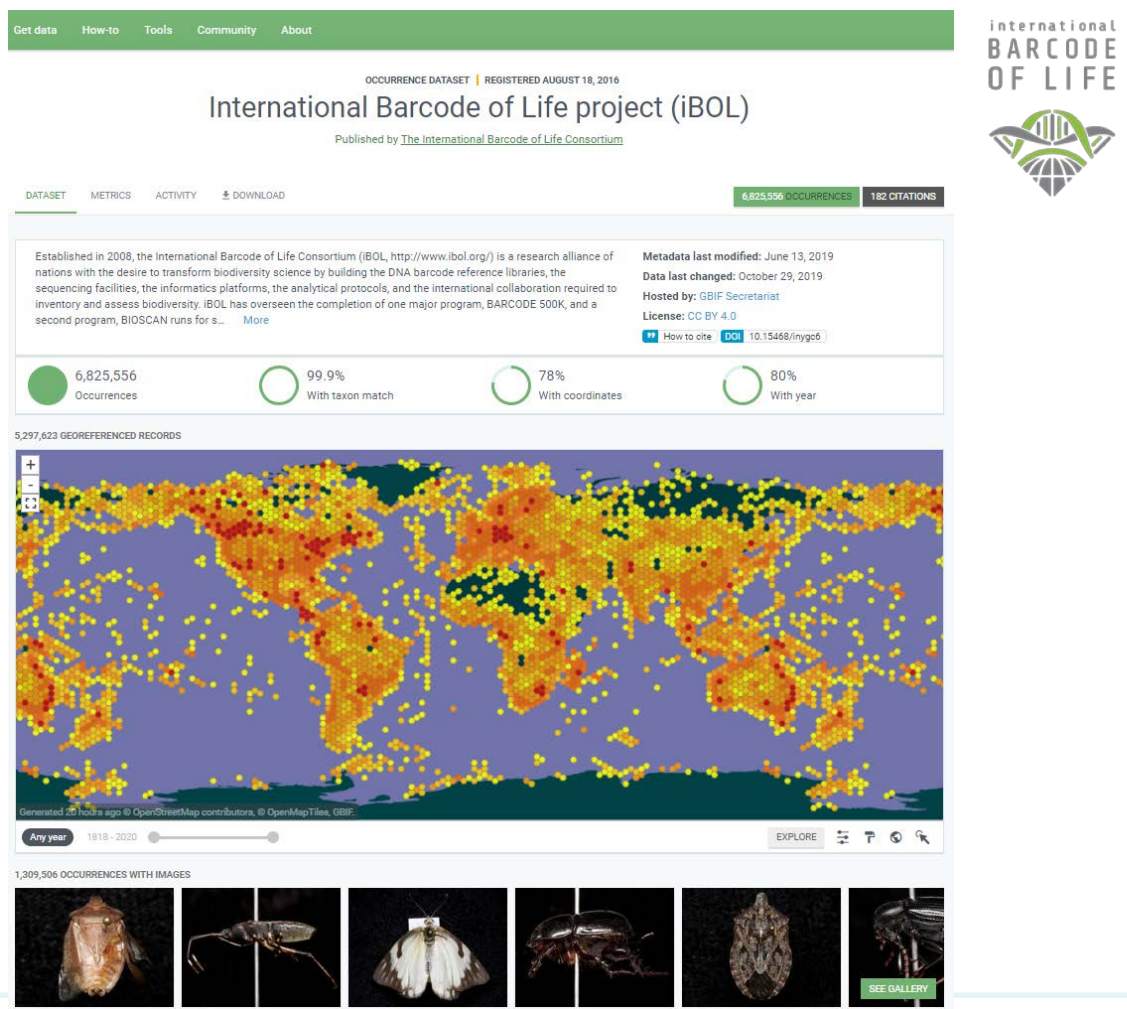
BIN DEF0002

# INDIVIUDUAL SEQUENCES WITH COORDINATES

## European Nucleotide Archive: 4.8M records



## International Barcode of Life: 6.8M records





# METABARCODING: MGNIFY AND MORE

PUBLISHER | SINCE MARCH 13, 2019

MGNify

ABOUT METRICS HOME PAGE

20,107,151 OCCURRENCES

920 DATASETS

41 CITATIONS

**Description:** MGNify is a free to use analysis platform and database of analysed microbiome sequencing projects, hosted by the European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI). Microbiome research typically involves the study of all genomes present within a specific environment,

such as soil, seawater or a human body site. Underpinned by dramatically mainstream. The approach can provide unique insights into the complex p surroundings, each other, and, in some cases, their host. MGNify offers an diversity and functional/metabolic potential of environmental samples. Us held within the database. In addition, users can request analysis of any ap following publication: EBI Metagenomics in 2017: enriching the analysis o Mitchell, Maxim Scheremetjew, Hubert Denise, Simon Potter, Aleksandra T Petra ten Hoopen, Blaise Alako, Clara Amid, Darren J. Wilkinson, Thomas F

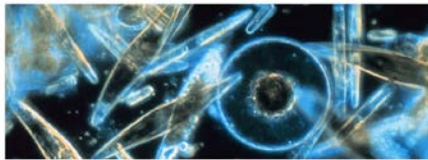
**Endorsed by:** National Biodiversity Network

**Administrative contact:** Robert Finn

**Technical contact:** MGNify Helpdesk

**Country or area:** United Kingdom of Great Britain and Northern Ireland

Biodiversity infrastructures to crosslink metagenomics and species occurrence data



ne the taxonomic ed public datasets , please refer to the i (2017) Alex L. Fiona M. I. Hunter,



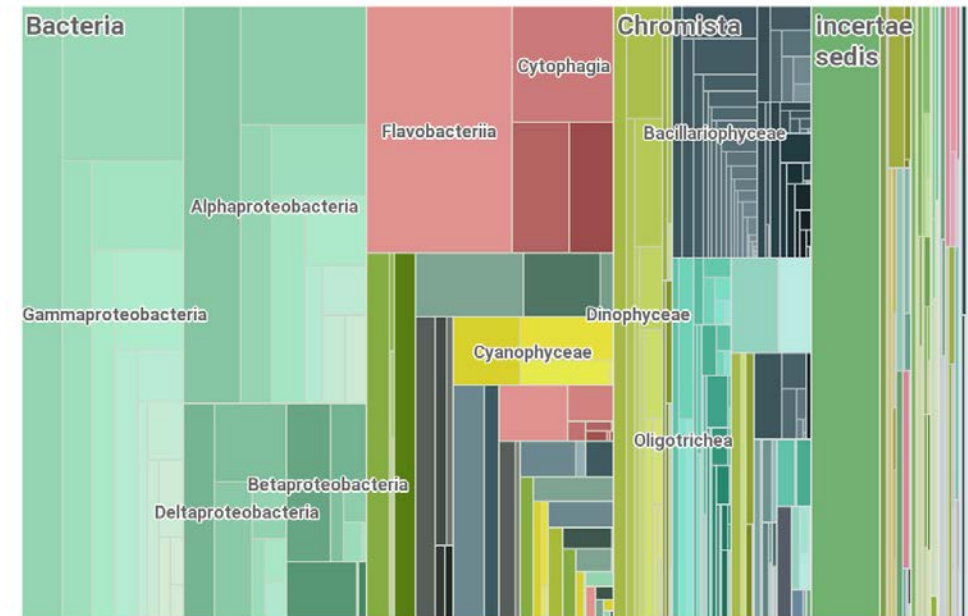
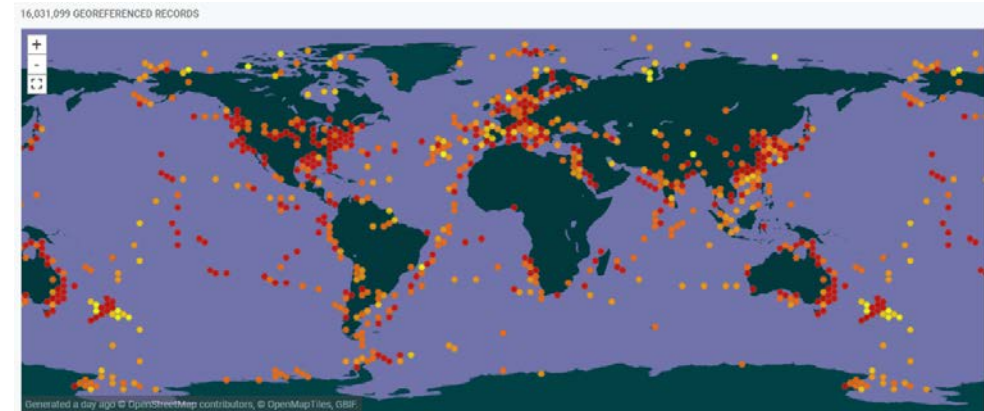
## CONTACTS

MGNify  
European Molecular Biology Laboratory, European  
Bioinformatics Institute (EMBL-EBI), Wellcome  
Genome Campus  
Hinxton CB10 1SD  
Cambridge

MGNify Helpdesk  
Technical point of contact  
metagenomics-help@ebi.ac.uk

Robert Finn  
Administrative point of contact • Point of contact  
rdf@ebi.ac.uk  
+441223492679

920 datasets: 20.1M new records  
14.8M Bacteria  
2.3M Chromista



<https://www.gbif.org/publisher/ab733144-7043-4e88-bd4f-fca7bf858880>

# NEW GUIDE: PUBLISHING DNA-DERIVED DATA THROUGH BIODIVERSITY DISCOVERY PLATFORMS

Mapping and data publishing

Cross-platform

About 40 pages long "cookbook"

- ❖ **Introduction**
- ❖ **Categorization**
- ❖ **Mapping\***
- ❖ **Visuals**
- ❖ **Future prospects**
- ❖ **Resources**

*\*Darwin Core and MIxS based*



## Publishing DNA-derived data through biodiversity data platforms [Community review draft]

Anders F. Andersson · Andrew Bissett · Anders G. Finstad · Frode Fosøy · Marie Grosjean · Michael Hope · Thomas S. Jeppesen · Urmas Kõljalg · Daniel Lundin · R. Henrik Nilsson · Maria Prager · Cecilie Svenningsen · Dmitry Schigel – Version 90868d9, 2020-10-15 12:04:43 UTC

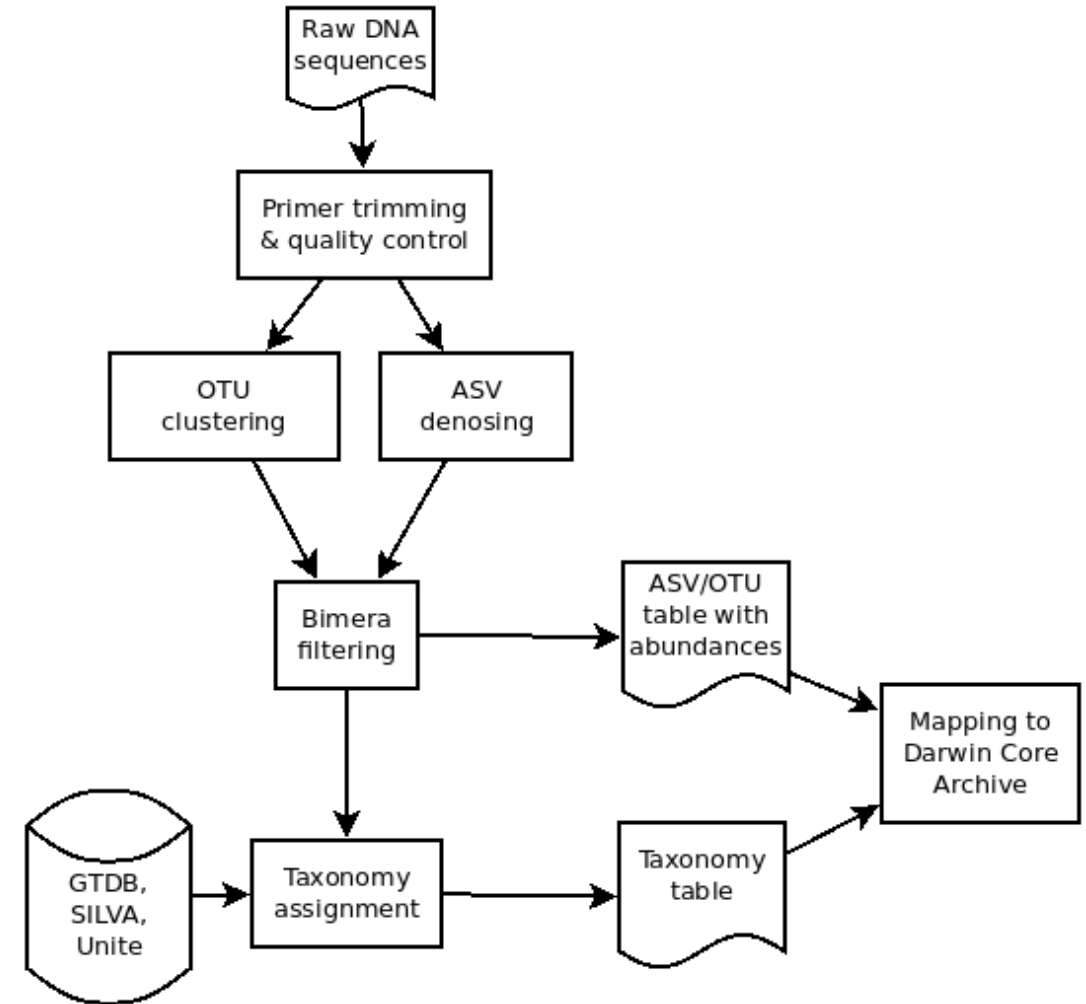
This document is also available in [PDF format](#).



# PUBLISHING SEQUENCE-DERIVED DATA: THE “LEARN” SECTION

## Introduction

- ❖ Rationale
- ❖ Audiences
- ❖ DNA derived occurrence data
- ❖ Biodiversity data
- ❖ Processing workflows
- ❖ Taxonomy of sequences





# PUBLISHING SEQUENCE-DERIVED DATA: THE “DO” SECTION

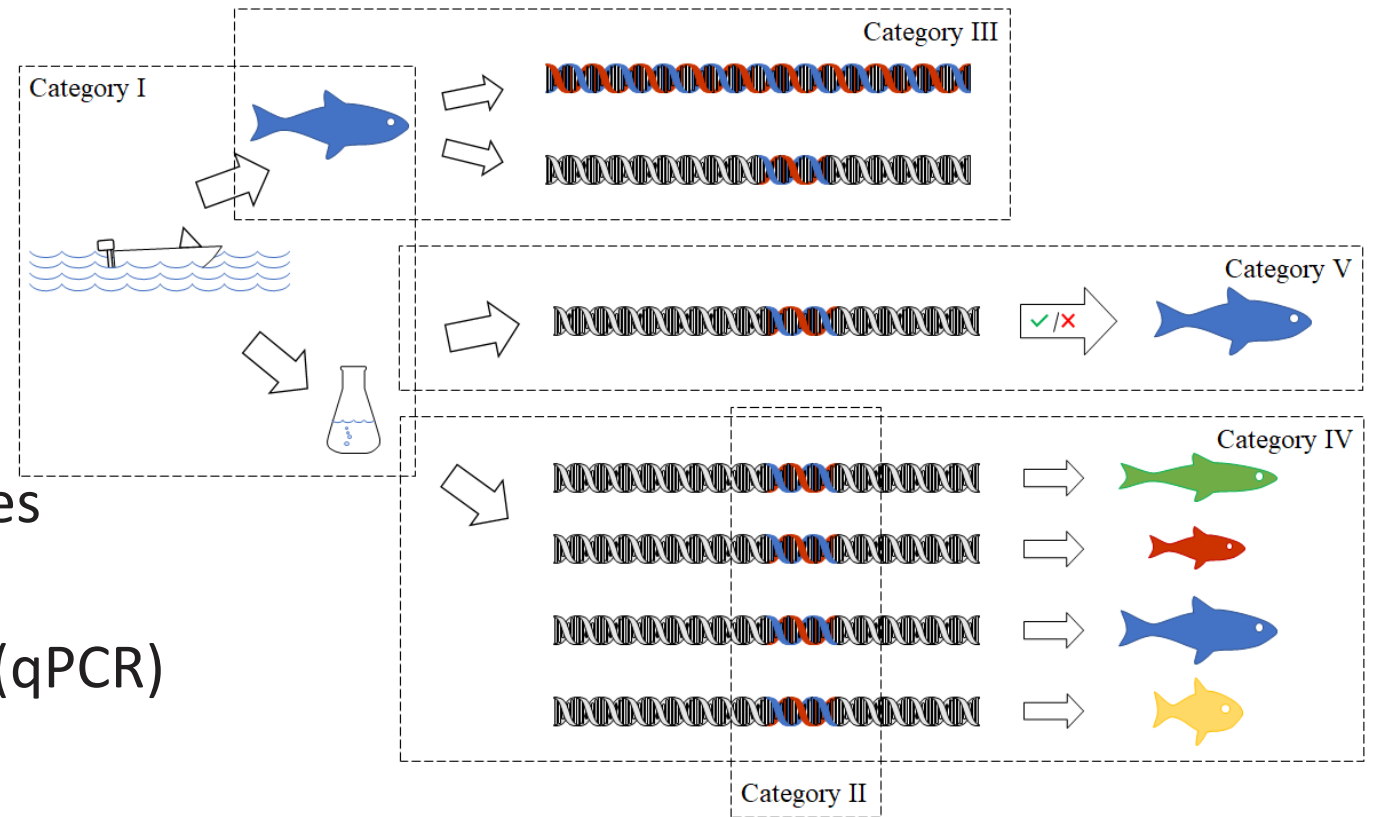
## Data packaging and mapping

### Categorisation of data

- I Sequence-based occurrences
- II Enriched occurrences
- III Targeted species detection (qPCR)
- IV Name references
- V Metadata-only

### Mapping

### Examples



# NEW TRAINING COURSE ON GBIF AND BOLD SKILLS – WITH BIODATA /GBIF NORWAY

Home UiO > Natural History Museum

For employees Norwegian website Search

UiO Natural History Museum

Home Visiting Research and collections About the museum

Research and collections

Research projects

BioDATA


Activities

People

## Data publication course in Tbilisi

Planned BioDATA data publishing course in Tbilisi to focus on publishing molecular data in GBIF. Notice that the course is postponed because of the COVID-19 travel recommendations - the dates will be updated.

Time and place: Jan 1, 2021, Tbilisi, Georgia  
Add to calendar



Red Cabbage Bug (*Eurydema ornata*) CC-BY-ds401

**DATES**

GBIF / BOLD course on data management skills was originally scheduled Tuesday 7 July - Friday 10 July (four full days from 9:00 till 17:00). This was immediately after CaBOL meeting: Tuesday, June 30 - Friday, July 3 and BioBlitz: Friday 3 July (evening) - Monday 6 July. Notice that the course is postponed because of the COVID-19 travel recommendations - the dates will be updated.

**SCOPE**

Data management skills for publishing data through BOLD and GBIF data platforms. This is an observation/specimen -> published record course that does not include wet lab steps. Examples of the past, to be modified:

2018 [https://www.fortbio.uio.no/events/courses/2018/Data\\_mobilization\\_Baikal.html](https://www.fortbio.uio.no/events/courses/2018/Data_mobilization_Baikal.html)  
2019 <https://sisu.ut.ee/publish-sequence-datasets>

**Organizer**

BioDATA, GBIF and FortBio

Accelerating biodiversity research through DNA barcodes, collection and observation data

## Work in progress

*Why publish in GBIF?*

*Barcoding, data and biodiversity*

*Biodiversity databases in Georgia*

*Open data as a first-class research citizen*

*Databases as a research tool*

*Why do we need to identify species*

*Concept of DNA barcoding*

*Data in and data out: recognize and understand your data*

*Principles of data organization and personal data management*

*Data structure: standards*

*Data citation*

*Data exposure: why and when*

*Data papers*

*Barcode reference repositories*

*Quality control*

*Use case Bombus and legumes: uncovering pollination mysteries*

## KEY FACTS

- New GBIF course
- Tbilisi, Georgia
- 2021, 4 days
- Onsite or virtual
- GBIF and BOLD
- publish and use
- DNA-derived FAIR data
- Learn and practice

# THANK YOU

Dmitry Schigel

[dschigel@gbif.org](mailto:dschigel@gbif.org)

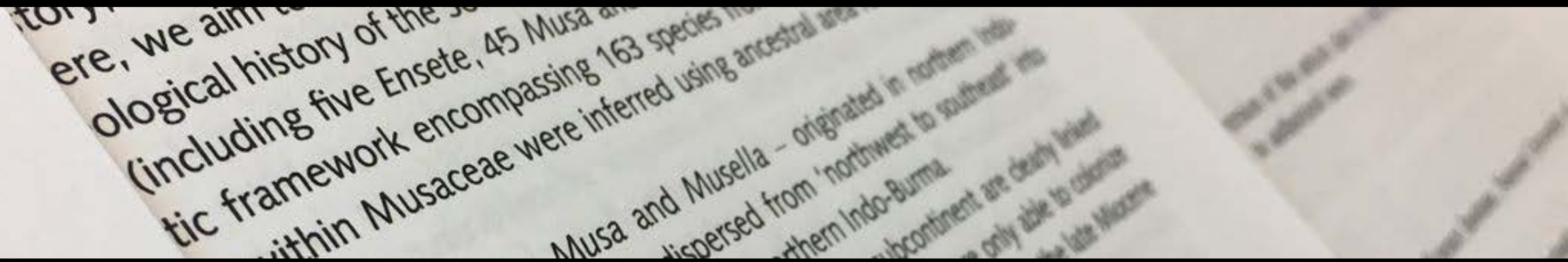






# Data integration enables biodiversity synthesis

*A literature review of GBIF-enabled studies*



---

*Mason Heberling*<sup>1</sup>, Joe Miller<sup>2</sup>, Daniel Noesgaard<sup>2</sup>, Scott Weingart<sup>3</sup>, Dmitry Schigel<sup>2</sup>

<sup>1</sup>Carnegie Museum of Natural History Pittsburgh, PA USA; heberlingm@carnegiemnh.org

<sup>2</sup>Global Biodiversity Information Facility Secretariat, Copenhagen, Denmark

<sup>3</sup>Carnegie Mellon University, Pittsburgh, PA USA

*GB27 Meeting, October 2020*

# Free and open access to biodiversity data

gbif.org

OCCURRENCES

SPECIES

DATASETS

PUBLISHERS

RESOURCES

Search



WHAT IS GBIF?

ABOUT GBIF UNITED STATES OF AMERICA

*Precis archesia subsp. ugentensis* observed in Adi, DR Congo by Thomas Desjardes. Photo via GBIF

Occurrence records

1,605,663,540

Datasets

54,622

Publishing institutions

1,655

Peer-reviewed papers using data

5,059



News

GBIF nodes collaborations confirmed for funding through capacity programme



News

Award winner addresses data bias while assessing trends in boreal butterfly diversity



News

Award winner applies machine learning to model host-pathogen relationships & interactions



News

Georeferencing documents released for GBIF community review



GBIF | Global Biodiversity Information Facility

# Free and open access to biodiversity data

gbif.org

OCCURRENCES

SPECIES

DATASETS

PUBLISHERS

RESOURCES

Search



WHAT IS GBIF?

ABOUT GBIF UNITED STATES OF AMERICA

*Precis archesia subsp. ugentensis* observed in Adi, DR Congo by Thomas Desjardes. Photo via GBIF

Occurrence records

1,605,663,540

Datasets

54,622

Publishing institutions

1,655

Peer-reviewed papers using data

5,059



News

GBIF nodes collaborations confirmed for funding through capacity programme



News

Award winner addresses data bias while assessing trends in boreal butterfly diversity



News

Award winner applies machine learning to model host-pathogen relationships & interactions



News

Georeferencing documents released for GBIF community review



[Get data](#)[How-to](#)[Tools](#)[Community](#)[About](#)[Resources](#)[SEARCH RESOURCES](#) | 5,059 RESULTS

Search



ALL

LITERATURE

Country or area of researcher



Country or area of coverage



Literature type

☒ Journal article

Relevance

☒ GBIF used

Year



Topic



Dataset



Publisher



Peer-reviewed

☒ Yes

Open access



Download key



Journal



### Assessing spatial and temporal biases and gaps in the publicly available distributional information of Iberian mosses

Literature

Ronquillo, C. Alves-Martins, F. Mazimpaka, V. Sobral-Souza, T. Vilela-Silva, B. G. Medina, N. ... - (2020) Biodiversity Data Journal

One of the most valuable initiatives on massive availability of biodiversity data is the Global Biodiversity Information Facility, which is creating new opportunities to develop and test macroecological knowledge. However, the potential uses of these data are limited by the gaps and biases associated...

Biodiversity data • Bryophyta • Global Biodiversity Information Facility • IberBryo • Iberian Peninsula • Inventory completeness

[Journal article](#) [Open access](#) [Peer-reviewed](#)Data referenced in study [DOI 10.15468/dl.eujakg](#) [DOI 10.15468/dl.ogvrsc](#)

### Dioecy is associated with high genetic diversity and adaptation rates in the plant genus *Silene*

Literature

Muyle, A. Martin, H. Zemp, N. Mollion, M. Gallina, S. Tavares, R. ... - (2020) Molecular Biology and Evolution

About 15,000 angiosperm species (~6%) have separate sexes, a phenomenon known as dioecy. Why dioecious taxa are so rare is still an open question. Early work reported lower species richness in dioecious compared to non-dioecious sister clades, raising the hypothesis that dioecy may be an evolutionary...

[Journal article](#) [Open access](#) [Peer-reviewed](#)Data referenced in study [DOI 10.15468/dl.bwdu9z](#) [DOI 10.15468/dl.fim8f2](#) [DOI 10.15468/dl.fxocf](#)[DOI 10.15468/dl.mfyds](#) [DOI 10.15468/dl.poqxc1](#) [DOI 10.15468/dl.vizuiv](#) [DOI 10.15468/dl.vmvebr](#)[DOI 10.15468/dl.wizsb6](#) [DOI 10.15468/dl.yrvenq](#) [DOI 10.15468/dl.yywbkr](#)

### Richness and distribution patterns of elasmobranchs in the San Andres, Providencia and Santa Catalina Archipelago: is this area a hotspot of these spe...

Literature

[Get data](#)[How-to](#)[Tools](#)[Community](#)[About](#)[Resources](#)[SEARCH RESOURCES](#) | 5,059 RESULTS

Search



ALL

LITERATURE

Country or area of researcher



Country or area of coverage



Literature type

☒ Journal article

Relevance

☒ GBIF used

Year



Topic



Dataset



Publisher



Peer-reviewed

☒ Yes

Open access



Download key



Journal



### Assessing spatial and temporal biases and gaps in the publicly available distributional information of Iberian mosses

Literature

Ronquillo, C. Alves-Martins, F. Mazimpaka, V. Sobral-Souza, T. Vilela-Silva, B. G. Medina, N. ... - (2020) Biodiversity Data Journal

One of the most valuable initiatives on massive availability of biodiversity data is the Global Biodiversity Information Facility, which is creating new opportunities to develop and test macroecological knowledge. However, the potential uses of these data are limited by the gaps and biases associate...

Biodiversity data • Bryophyta • Global Biodiversity Information Facility • IberBryo • Iberian Peninsula • Inventory completeness

[Journal article](#) [Open access](#) [Peer-reviewed](#)Data referenced in study [DOI 10.15468/dl.eujakg](#) [DOI 10.15468/dl.ogvrsc](#)

### Dioecy is associated with high genetic diversity and adaptation rates in the plant genus *Silene*

Literature

Muyle, A. Martin, H. Zemp, N. Mollion, M. Gallina, S. Tavares, R. ... - (2020) Molecular Biology and Evolution

About 15,000 angiosperm species (~6%) have separate sexes, a phenomenon known as dioecy. Why dioecious taxa are so rare is still an open question. Early work reported lower species richness in dioecious compared to non-dioecious sister clades, raising the hypothesis that dioecy may be an evolutionar...

[Journal article](#) [Open access](#) [Peer-reviewed](#)Data referenced in study [DOI 10.15468/dl.bwdu9z](#) [DOI 10.15468/dl.fim8f2](#) [DOI 10.15468/dl.fxocf](#)[DOI 10.15468/dl.mfyds](#) [DOI 10.15468/dl.poqxc1](#) [DOI 10.15468/dl.vizuiv](#) [DOI 10.15468/dl.vmvebr](#)[DOI 10.15468/dl.wizsb6](#) [DOI 10.15468/dl.yrvnq](#) [DOI 10.15468/dl.yywbk](#)

### Richness and distribution patterns of elasmobranchs in the San Andres, Providencia and Santa Catalina Archipelago: is this area a hotspot of these spe...

Literature



# The Big Questions

1. How are GBIF-mediated data used? (*the what*)
2. Have uses changed through time? (*the when*)
3. Where are GBIF-mediated studies used? (*the where and who*)
4. How does GBIF-mediated papers compare to potential use? (*the future*)



# Growth of GBIF data

Rapid growth in data.

+1,150% since 2007!

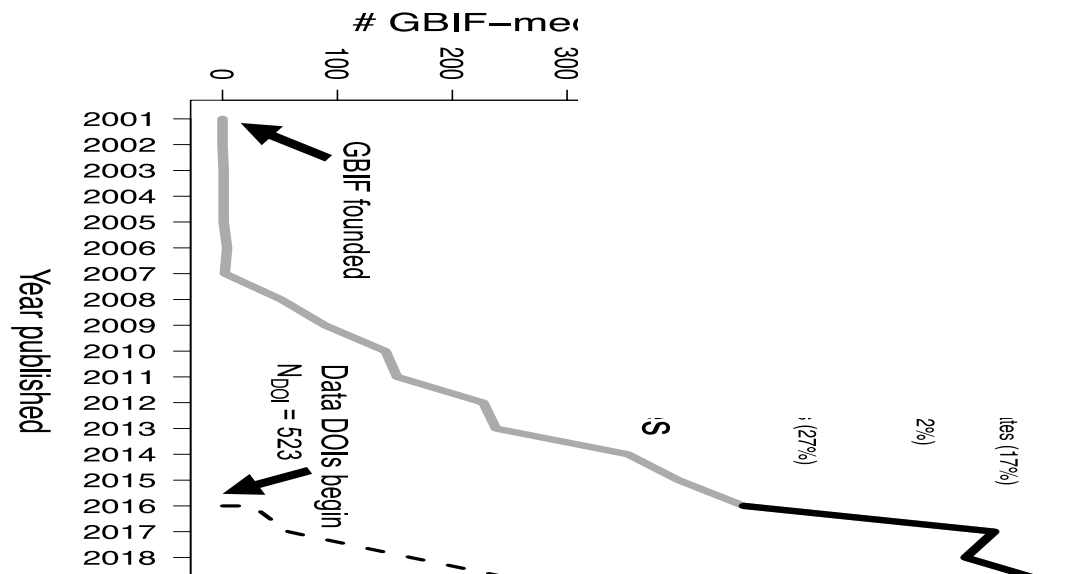
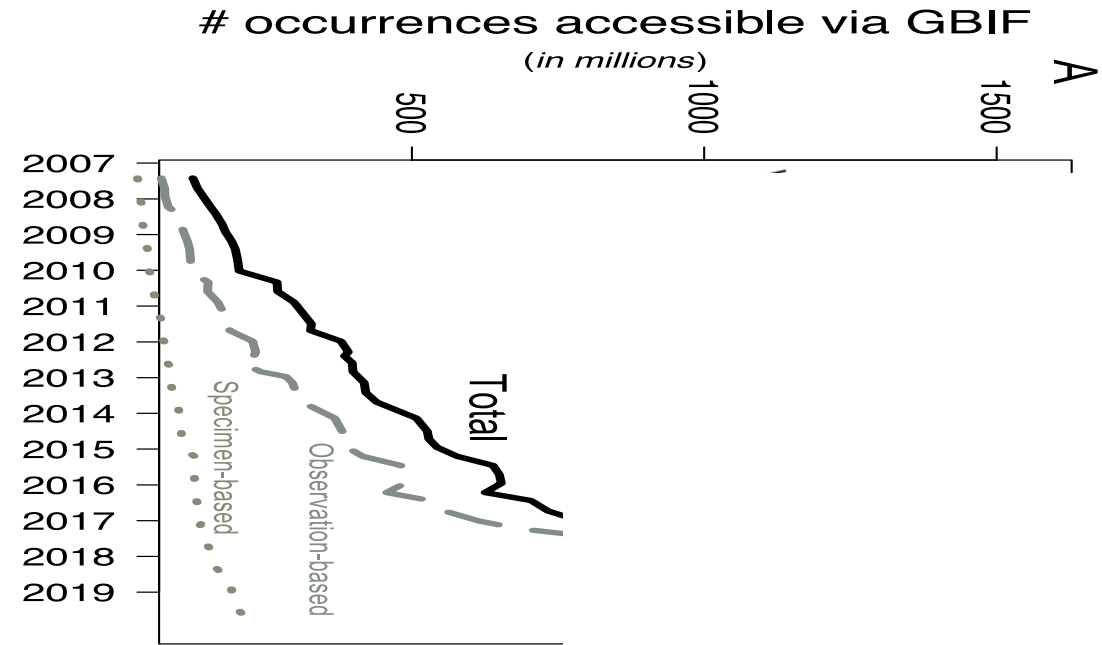
Rapid growth in data use.

2019 alone = 723 studies

Datasets are being used.

26,046 GBIF-mediated datasets have >1 citation!

\*median citations per dataset = 11



# Growth of GBIF data

Rapid growth in data.

+1,150% since 2007!

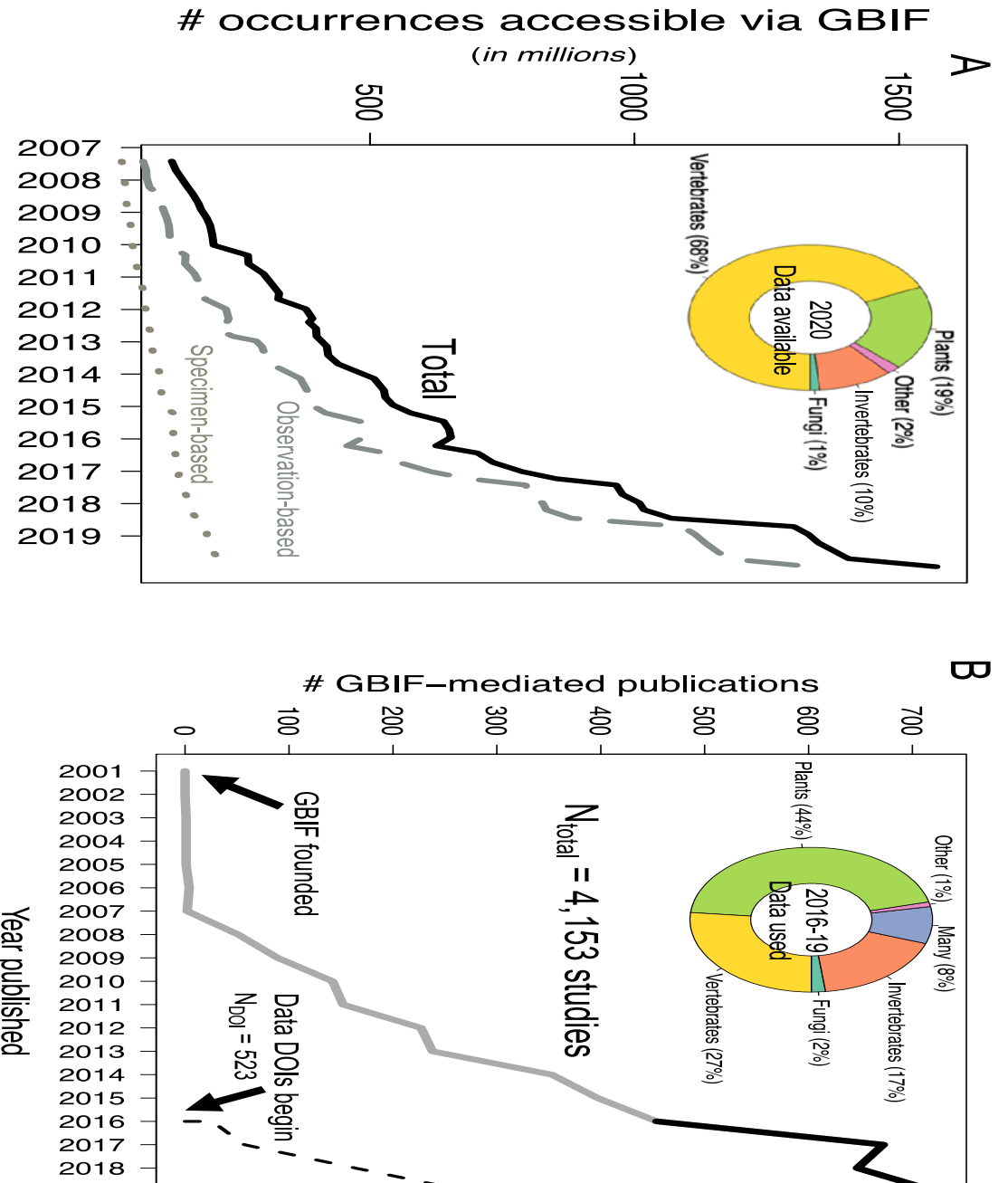
Rapid growth in data use.

2019 alone = 723 studies

Datasets are being used.

26,046 GBIF-mediated datasets have >1 citation!

\*median citations per dataset = 11



# The what

*How are GBIF-mediated data used in scientific papers?*

## Approach:

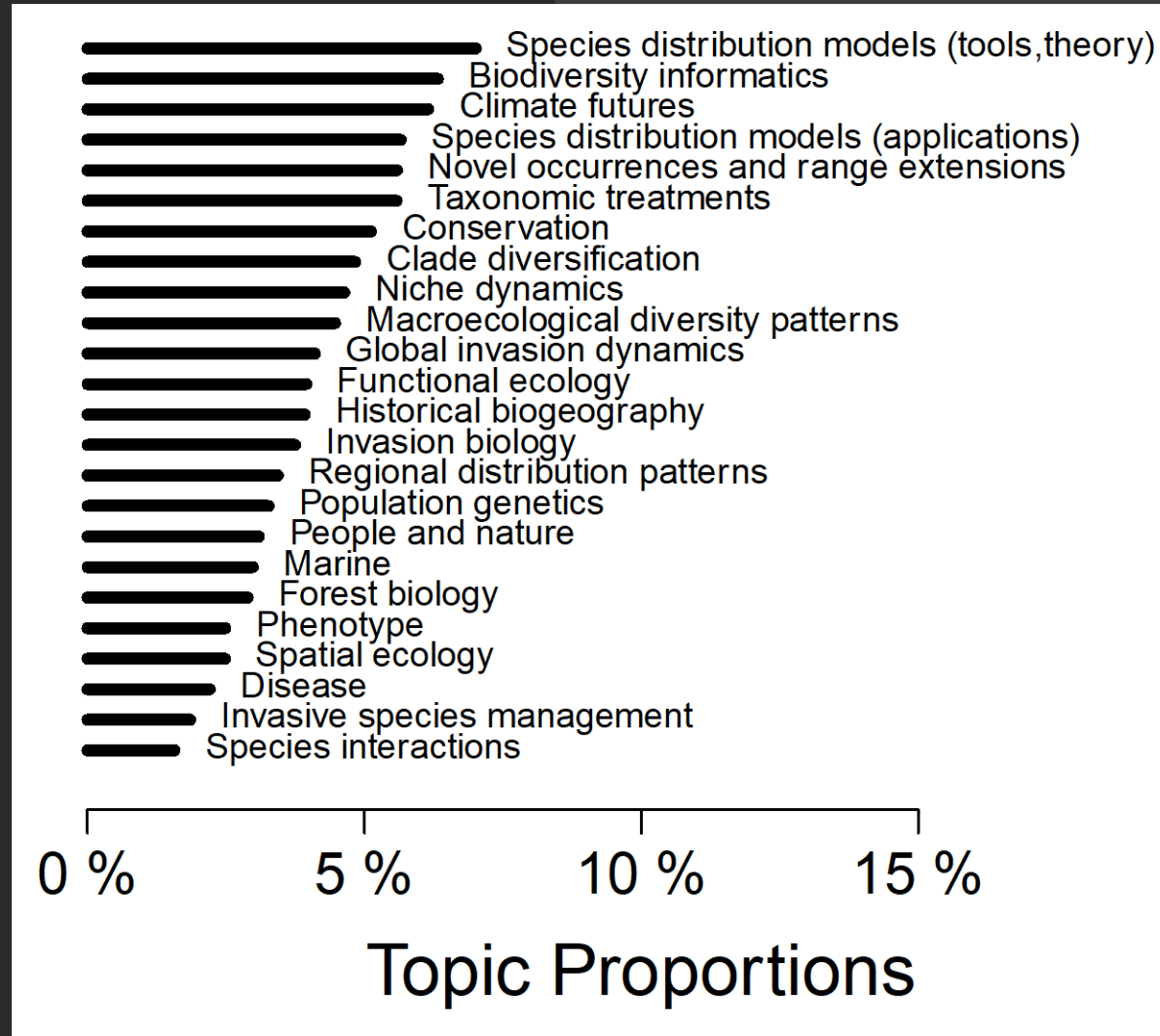
Computational text analysis of 4,035 studies using GBIF.

Topics emerge from computation of word associations.

## Results:

24 topics across a range of content areas

Species distribution modeling, a clear major use.





# The when

*Have major uses of GBIF data changed through time?*

## Approach:

Compare **pre-2016** to **recent (2016-19)**

Color = **growth (red)** or **decline (blue)**.

## Results:

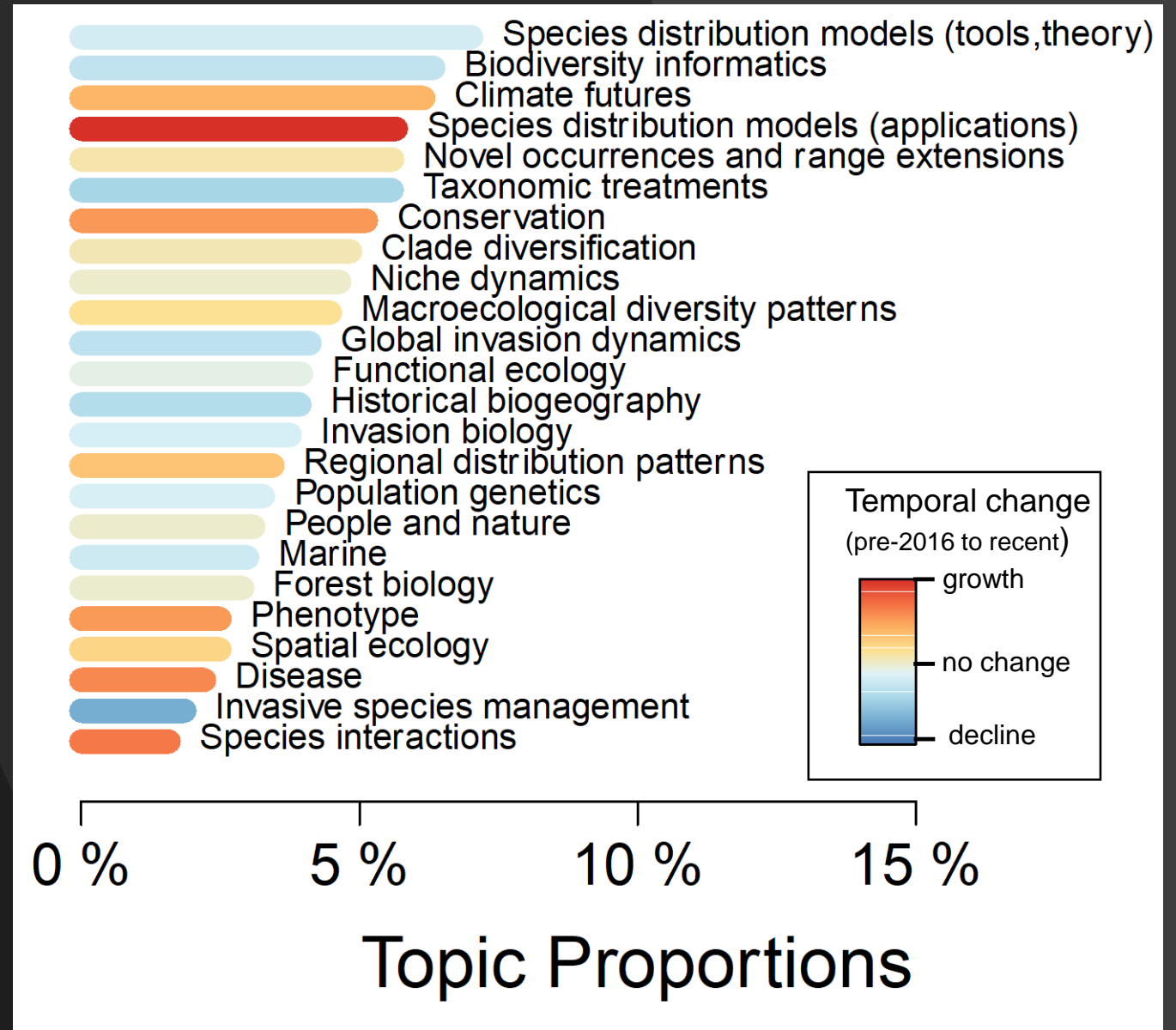
### Increasing topics +:

Species distribution models (applications), conservation, disease, species interactions, climate futures

### Decreasing topics - :

Species distribution models (tools), Biodiversity informatics, invasion management, taxonomic treatments

**\*\*NOTE:** relative change between time periods.  
E.g., 'decreasing' topics can remain common!

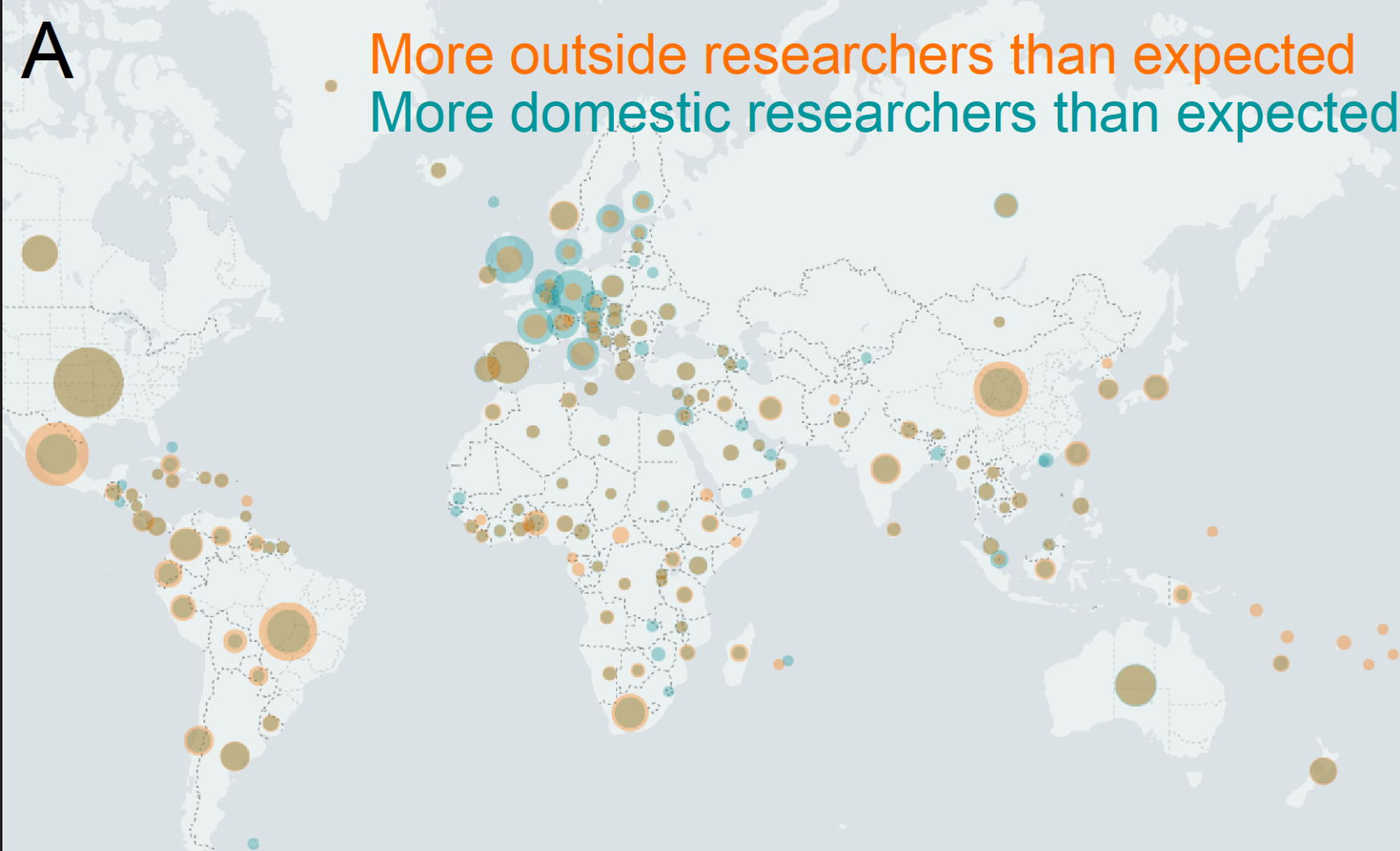


# The who and where

*GBIF data use includes a global authorship, though not evenly distributed.*

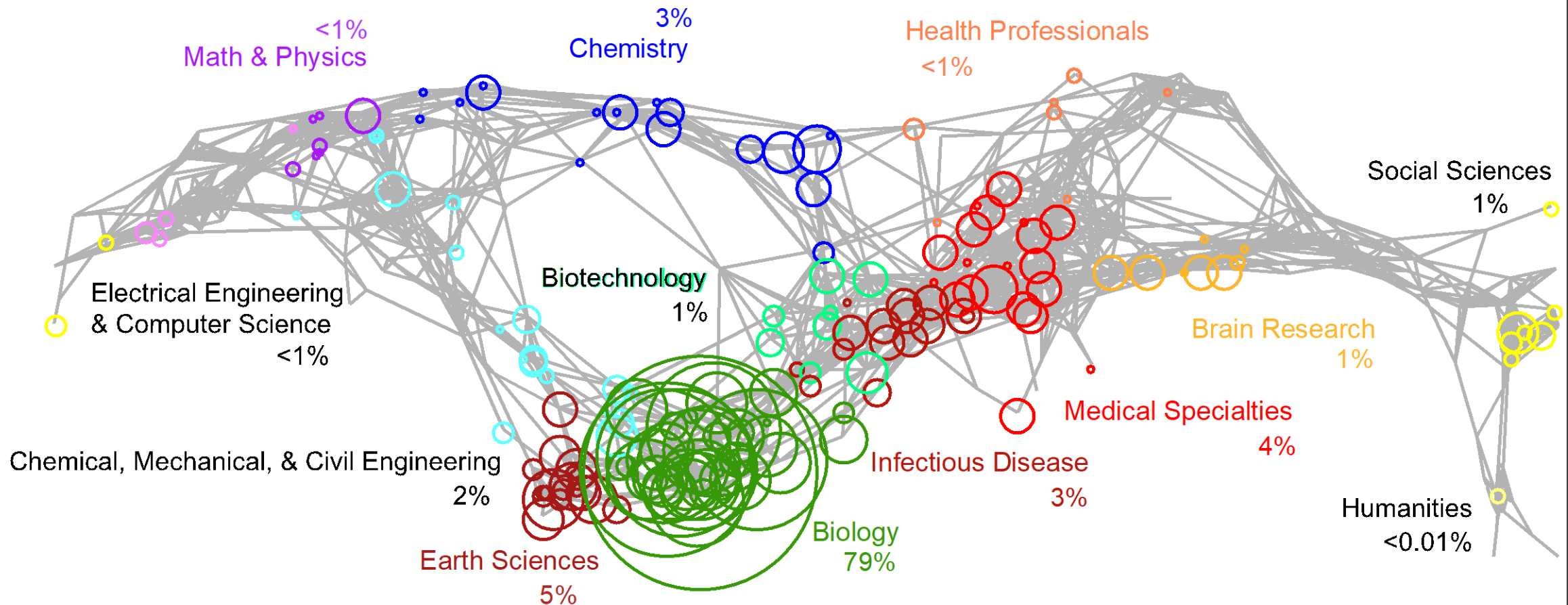
A

More outside researchers than expected  
More domestic researchers than expected



# The future: The GBIF Map of Science

## *GBIF research use spans many disciplines*



Map generated via Sci2 Tool





## GBIF-mediated data enables:

- Connection of disparate data types and sources
- Diverse topic areas and cross-disciplinary research use
- Increasingly global network of research



## Evidence for:

- Areas to prioritize and foster
- Continued data digitization and publishing
- Creating, connecting and improving biodiversity data

# Thanks!

Mason Heberling  
heberlingm@carnegiemnh.org



## Data integration enables biodiversity synthesis

*A literature review of GBIF-enabled studies*

---

Mason Heberling<sup>1</sup>, Joe Miller<sup>2</sup>, Daniel Noesgaard<sup>2</sup>, Scott Weingart<sup>3</sup>, Dmitry Schigel<sup>2</sup>

<sup>1</sup>Carnegie Museum of Natural History Pittsburgh, PA USA

<sup>2</sup>Global Biodiversity Information Facility Secretariat, Copenhagen, Denmark

<sup>3</sup>Carnegie Mellon University, Pittsburgh, PA USA

*GB27 Meeting, October 2020*



# Citations and use of GBIF-mediated data

Daniel Noesgaard | Science Comms Officer



**GBIF**

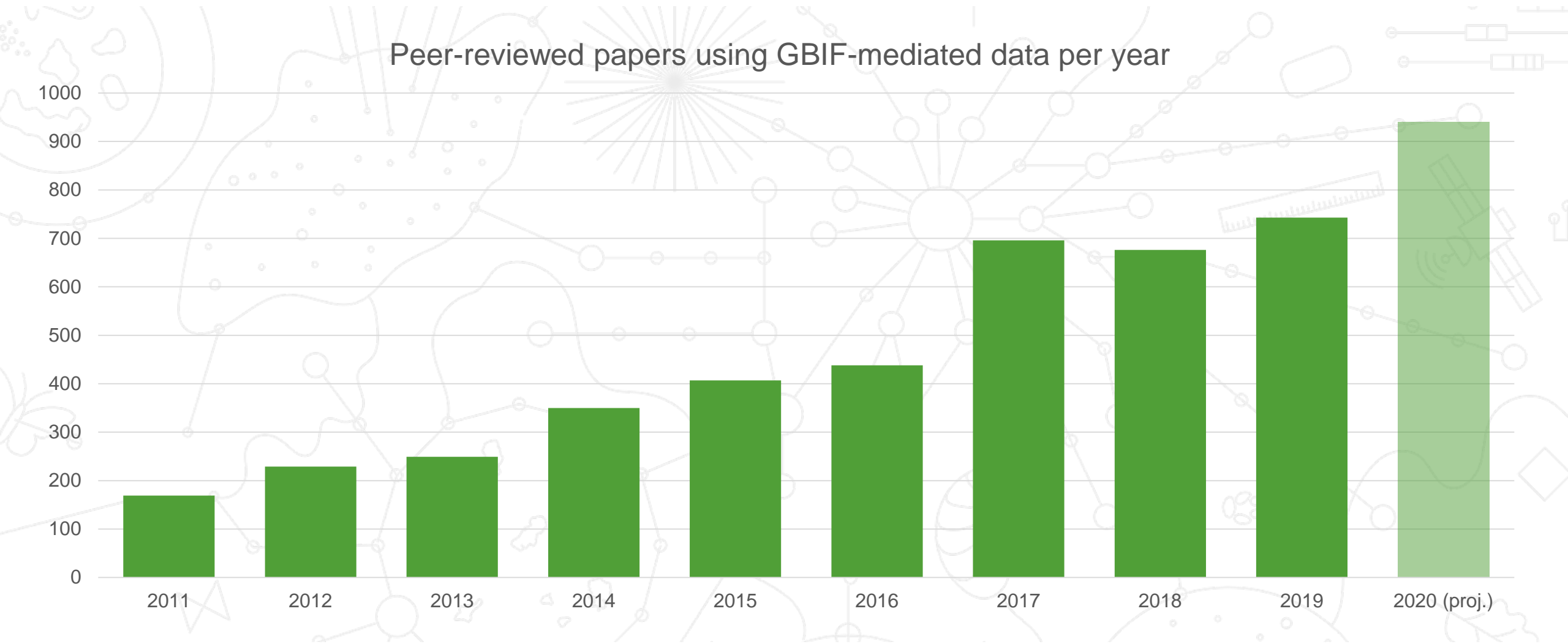
Global Biodiversity  
Information Facility





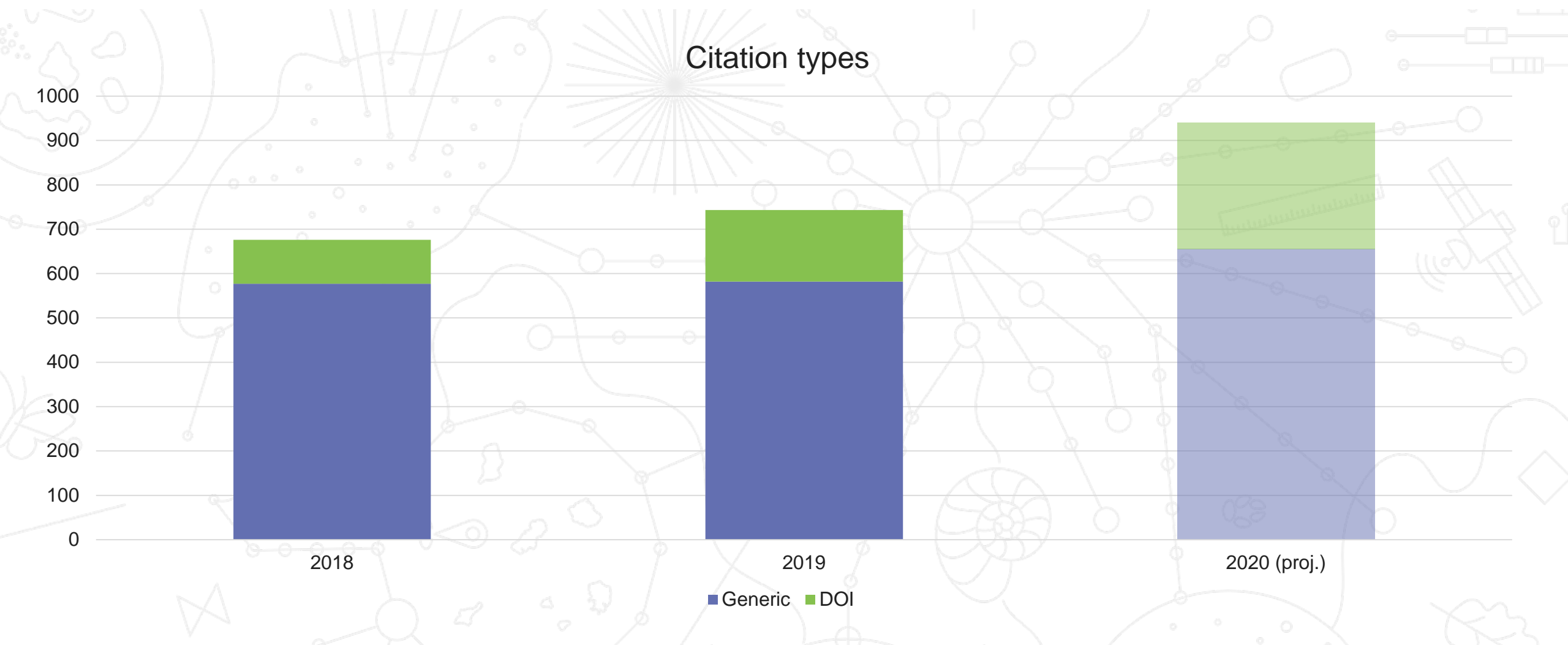
# TEN YEARS OF GROWTH

Peer-reviewed papers using GBIF-mediated data per year



30 per cent annual growth on average—almost 1,000 citations in 2020 predicted

# USE OF DOIS IN CITATIONS



DOIs are now used in nearly a third of papers using GBIF-mediated data

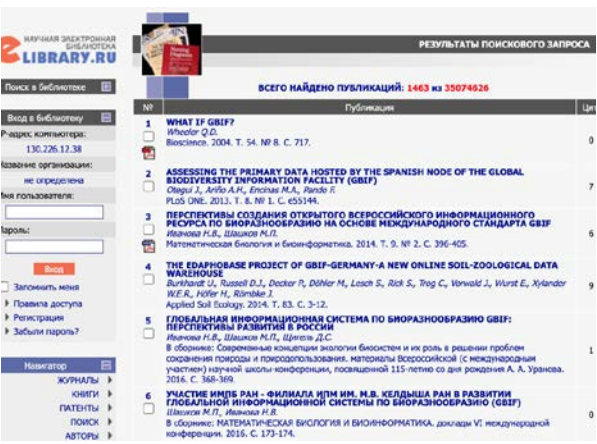
# NEW SOURCES/LANGUAGES OF LITERATURE



Chinese



Japanese



Russian



Others?

400 papers added

Thanks to Dr Luo Maofang (IB-CAS): Bulk import of ~400 papers in Chinese via CNKI (China National Knowledge Infrastructure)





# NEW CITATION SERVICE

Cloud environment  
exports



Search API derived  
datasets (rgbif, etc.)



## Simple service

- DOI of original “download” (e.g. a monthly snapshot)
- A list of dataset identifiers/DOI preferably with record counts
- A URL for the filtered dataset

## Return

- A new DOI presenting the cleaned data
- Linked to the specific contributing datasets with counts (attribution)
- Provides access to the filtered dataset (reproducibility)

DERIVED DATASET | 12 OCT 2020  
**235,112,532 occurrences**

DOI 10.21373/dd.vt2x99

DOWNLOAD  1 CITATION

FILTER APPLIED 13 MAY 2020

RERUN QUERY

**Citation:** GBIF.org (13 May 2020) GBIF Derived dataset <https://doi.org/10.21373/dd.vt2x99>

**License:** [CC BY-NC 4.0](#)

**File:** 4 KB Simple

**Involved datasets:** [11](#)

Make sure to read the [data user agreement](#) and [citation guidelines](#).

# #CITETHEDO

Daniel Noesgaard  
GBIF Secretariat







# Positioning GBIF for the the post-2020 Global Biodiversity Framework

Tim Hirsch | Deputy Director

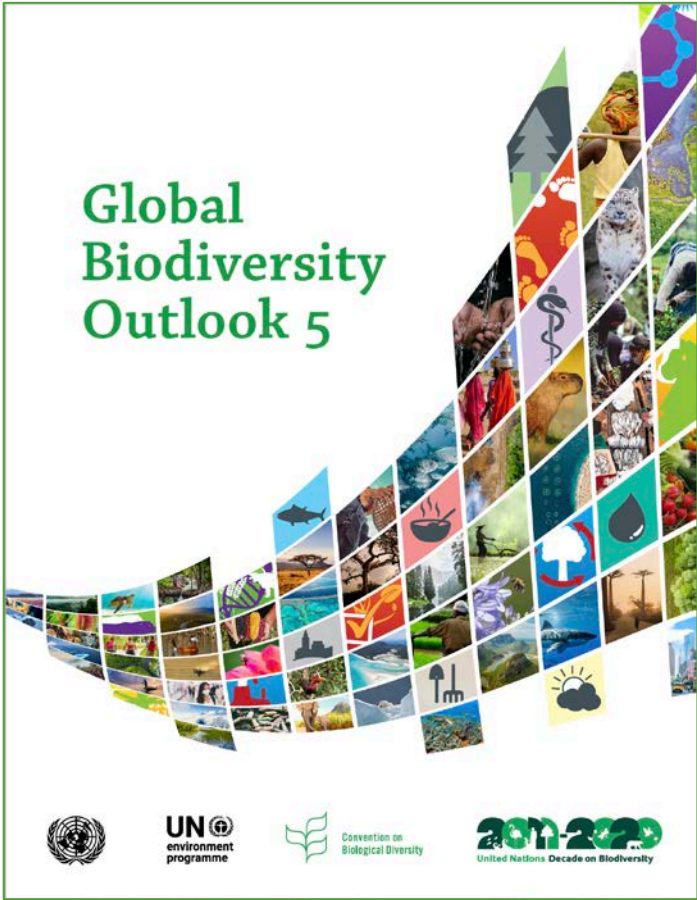


**GBIF**

Global Biodiversity  
Information Facility



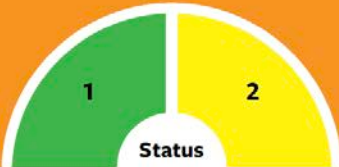
# 5<sup>TH</sup> GLOBAL BIODIVERSITY OUTLOOK



## SHARING INFORMATION AND KNOWLEDGE

### TARGET ELEMENTS

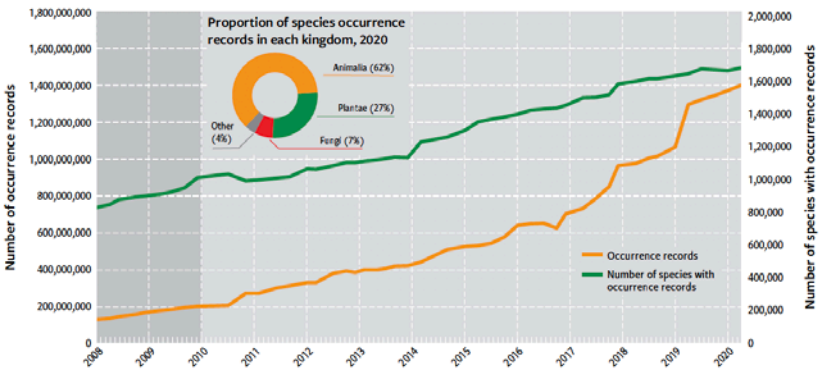
1. Biodiversity knowledge, science and technologies improved
2. Biodiversity knowledge, science and technologies shared



### Summary of target achievement

Significant progress has been made since 2010 in the generation, sharing and assessment of knowledge and data on biodiversity, with big-data aggregation, advances in modelling and artificial intelligence opening up new opportunities for improved understanding of the biosphere. However, major imbalances remain in the location and taxonomic focus of studies and monitoring. Information gaps remain in the consequences of biodiversity loss for people, and the application of biodiversity knowledge in decision making is limited. **The target has been partially achieved** (*medium confidence*).<sup>1</sup>

Figure 19.2. Growth in GBIF-mediated species occurrence records<sup>12</sup>



Mobilization of open-access data through the Global Biodiversity Information Facility (GBIF). The lines show the number of species occurrence records over time, and the number of species having occurrence records.

# GBIF SUPPORTING BIODIVERSITY INDICATORS



- Trend in invasive alien species introductions (through Global Register of Introduced and Invasive Species)



- Species Protection Index
- Protected Area Representativeness Index

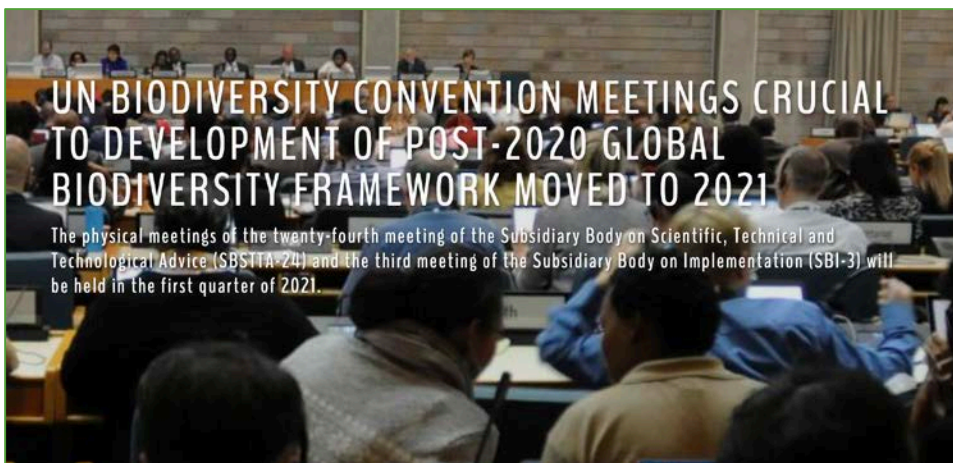


- Comprehensiveness of conservation of socioeconomically/culturally valuable species
- Agrobiodiversity Index
- Crop Wild Relative Index





- Growth in species occurrence records accessible through GBIF
- Species Status Information Index

# DELAYED TIMETABLE FOR NEW GLOBAL BIODIVERSITY FRAMEWORK



- Subsidiary body meetings (SBSTTA 24, SBI3) - Canada | March 2021
- Final meeting of Open-ended Working Group - Cali, Colombia | ? 2021
- COP 15 (UN Biodiversity Conference) - Kunming, China | ? 2021

**UN**  
environment  
programme

**Convention on  
Biological Diversity**

**CBD**

Distr.  
GENERAL

CBD/POST2020/PREP/2/1  
17 August 2020

ORIGINAL: ENGLISH

---

PREPARATIONS FOR THE POST-2020  
BIODIVERSITY FRAMEWORK

UPDATE OF THE ZERO DRAFT OF THE POST-2020 GLOBAL BIODIVERSITY FRAMEWORK

*Note by the Co-Chairs*

**I. BACKGROUND**

1. In decision [14/34](#), the Conference of the Parties set out the process for developing a post-2020 global biodiversity framework, established the Open-ended Working Group on the Post-2020 Global Biodiversity Framework to support this process and designated its Co-Chairs. Subsequently, the Working Group at its first meeting requested the Co-Chairs and the Executive Secretary, with the oversight of the Bureau, to continue the preparatory process in accordance with decisions [14/34](#), [CP-9/7](#) and [NP-3/15](#), and to prepare documentation, including a zero draft text of the post-2020 global biodiversity framework<sup>1</sup> for consideration by the Working Group at its second meeting. Pursuant to these requests, a zero draft of the post-2020 global biodiversity framework was issued for consideration by the Working Group at its second meeting (CBD/WG2020/2/3).

## OPPORTUNITIES FOR GBIF

- Strong role for data and monitoring within new targets (draft new Target 19)
- Implementation support mechanisms (Knowledge generation, management and sharing; capacity development)
- Critical role of primary data for monitoring and reporting




# ANALYSING DATA NEEDS FOR GLOBAL BIODIVERSITY FRAMEWORK

NEWS | 25 FEBRUARY 2020

## Call for proposals: Analysis of biodiversity data needs in the post-2020 framework

*Research contractor to carry out a nine-month fixed-term assignment analysing primary biodiversity data needs for the post-2020 global biodiversity framework*  
**DEADLINE 30 MARCH 2020**



Pallas' sailer (*Neptis sappho*), observed in Kunming, China, 23 October 2019. Photo by benanna via iNaturalist Research-grade Observations, licensed under CC BY-NC 4.0

The GBIF Secretariat seeks applications to carry out an analysis of the anticipated contributions of primary biodiversity data to the post-2020 Global Biodiversity Framework.

- Competitive call for analysis of post-2020 data needs
- Entries evaluated by panel including external experts
- Contract awarded to VertNet consortium (Rob Guralnick, David Bloom, Paula Zermoglio)
- Work timeline adjusted due to changes in CBD timetable
- Project under way from September 2020, publication 2nd quarter 2021

# PROJECT PLAN

1. Historic inputs of primary biodiversity data (PBD) into CBD processes
2. Projects and products that use PBD to support the post-2020 Framework
3. PBD sources used in EBVs, Indicators, and other products
4. Dependencies on PBD within the post-2020 Framework
5. Conclusions and next steps



Research



Analysis



Report

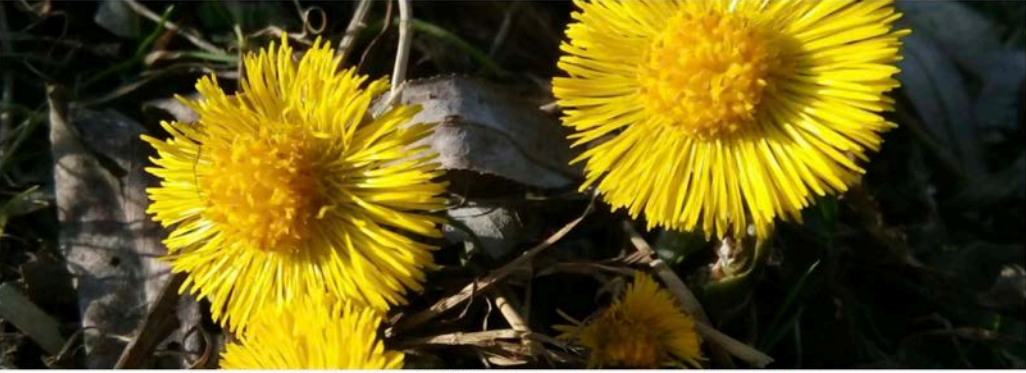
# HARMONIZING NATIONAL PLATFORMS

PROJECT | IN PROGRESS

## Using the CBD Clearing-House Mechanism to strengthen biodiversity data acquisition and data sharing

1 Jun 2019 - 31 Dec 2020 € 25,000

ABOUT NEWS & EVENTS



*Colts-foot (Tussilago farfara), Perwez, Belgium. Photo 2019 by Mathieu Gilbert, via Naturalist research-grade observations, licensed under CC-BY-NC 4.0.*

During the last two years, the Parties to the Convention on Biological Diversity (CBD) and its Secretariat has been developing a new Content Management System for the Clearing House Mechanism (CHM) of the Convention. The tool, named "Bioland", is interoperable with others biodiversity databases.

The principle goals of this project is to reinforce links between CHM and GBIF communities, and to develop concrete synergies, for example a specific GBIF-compatible plugin for Bioland.

National CHM teams in this project include: DR Congo, Cameroon, Mauritania, Madagascar, Bénin, Togo, Guinea, Belgium and France.

**Project Progress**

As a part of the first phase of the project, the Bioland tool became active and available for partners to use. A workshop was held in the start of October 2019 to give the partners involved in the project access to Bioland as well as full proficiency. The workshop also included training in the GBIF process, data acquisition and methods, so in the end of the workshop partners were ready to start the implementation of the Bioland tool in their national CHM.

- Project funded under Capacity Enhancement Support Programme (CESP)
- Developing plugin for incorporating GBIF-published data in national Clearing-House Mechanism (CHM) sites via Bioland tool
- Encouraging data-sharing from national CBD platforms via GBIF Integrated Publishing Toolkit (IPT)
- Starting with 8 francophone African countries, aim for wider adoption



# RELEVANT 2021 WORK PROGRAMME ITEMS

## Activity 1e (Expand national participation):

- Actively seek new Voting Participant countries, and return of former Voting Participants, to broaden the funding base ahead of the 2023-2027 financial period; in particular highlighting to governments the value of GBIF in meeting data and capacity needs for national implementation of the post-2020 global biodiversity framework under the Convention on Biological Diversity (CBD)

## Activity 5c (Support biodiversity assessment):

- Publish outcomes of the analysis of primary data needs for the post-2020 Global Biodiversity Framework, promote at relevant fora including the COP15 meeting of the CBD in Kunming, China

# THANK YOU

[thirsch@gbif.org](mailto:thirsch@gbif.org)



[www.gbif.org](http://www.gbif.org)

