Global biodiversity data as a source for vector-borne disease research and policy

Prof. Sylvie Manguin, IRD, UM

sylvie.manguin@ird.fr

BeOH « Ecosystems in the balance – January 23, 2025

GBIF—the Global Biodiversity Information Facility—is an **international network** and **data infrastructure** aimed at **providing anyone, anywhere, open access** to data about all types of **life on Earth**

Even mosquitoes …

GBIF

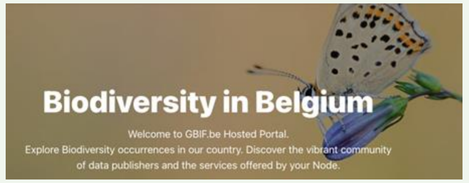
Vision

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**

GBIF



Datasets
110,345

Hosted portals
24

Country Participants
64

Peer-review papers using data
11,800

Organizational Participants
43

Average records downloaded per month (2023)
173.8 billion

Publishers
2,385

Species occurrence records
3,028,073,049

GBIF Memorandum of understanding (MoU)

Participants can be:

- a country
- an economy
- an intergovernmental or international organization
- an organization with an international scope

GBIF Governance

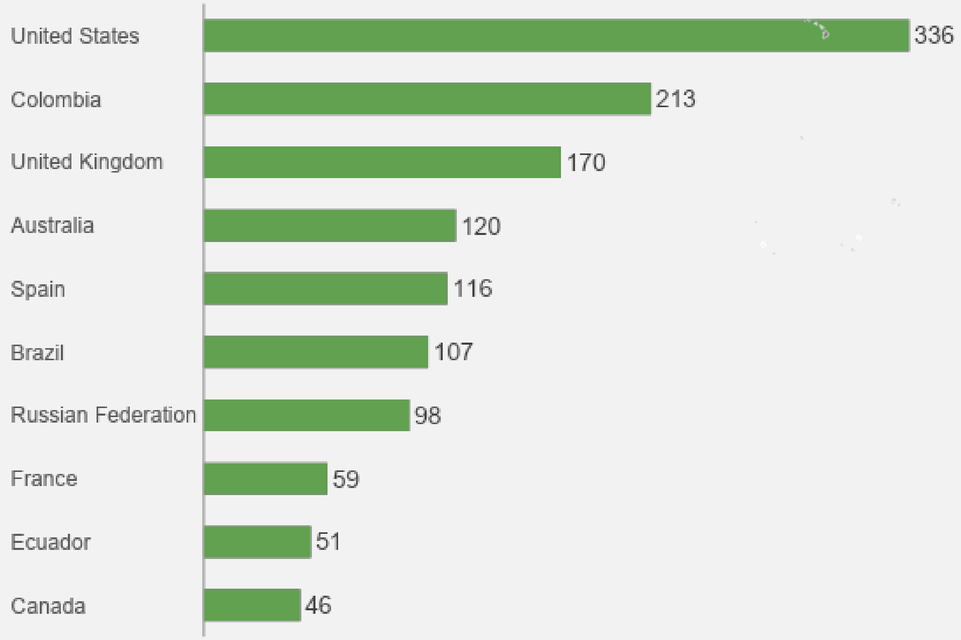
- GBIF Governing Board
  - The Executive Committee
  - Science Committee
  - Budget Committee
  - Participant Node Managers Committee (including the Nodes Steering Group)
    - Task groups

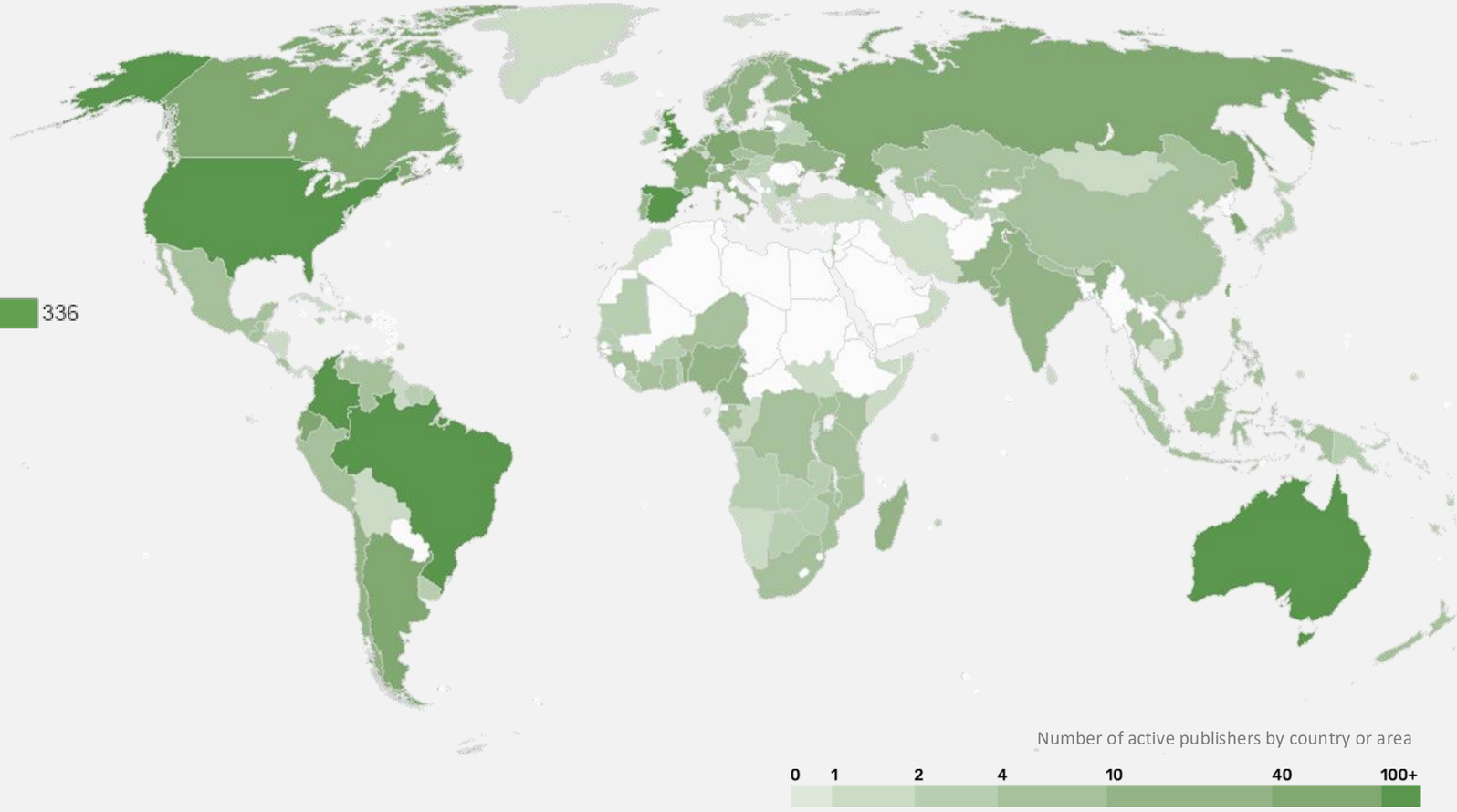
https://www.gbif.org/mou

1 January 2024

GBIF network of data publishing institutions

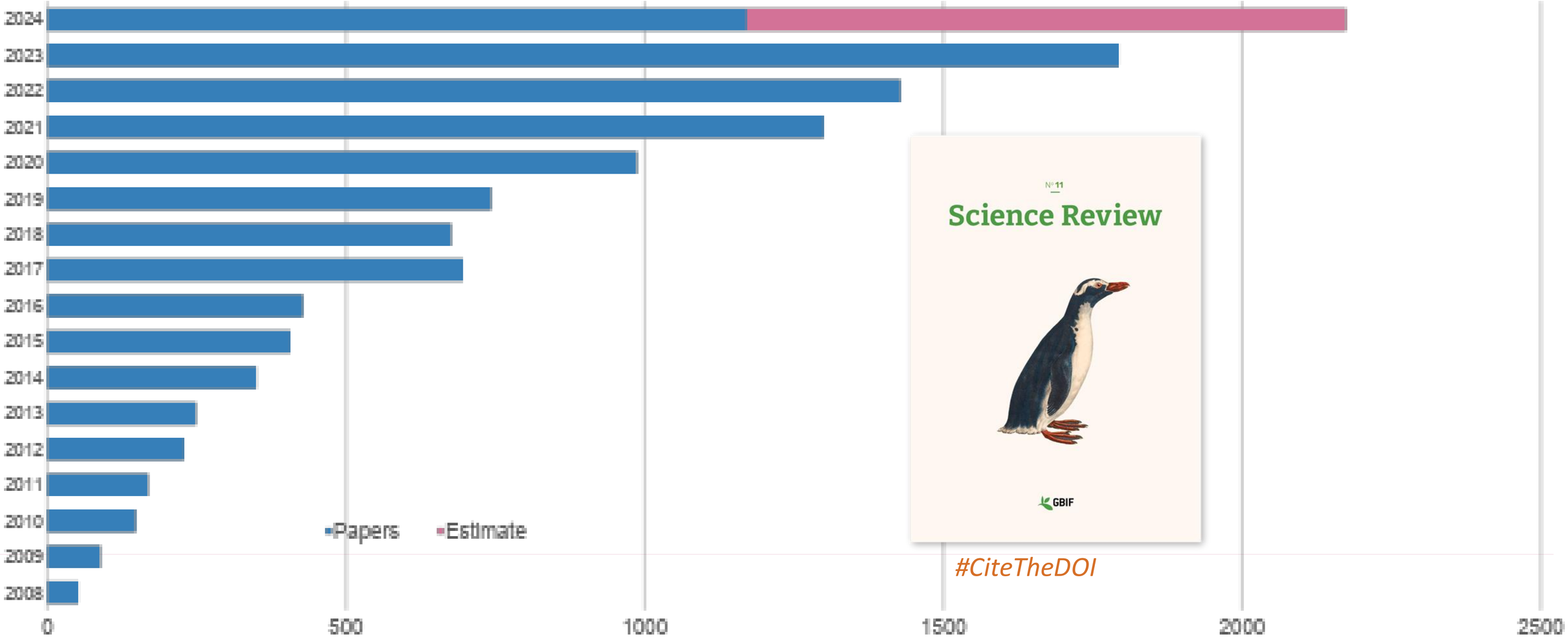
142 countries/areas with institutions sharing data through GBIF





https://www.gbif.org/publisher/search

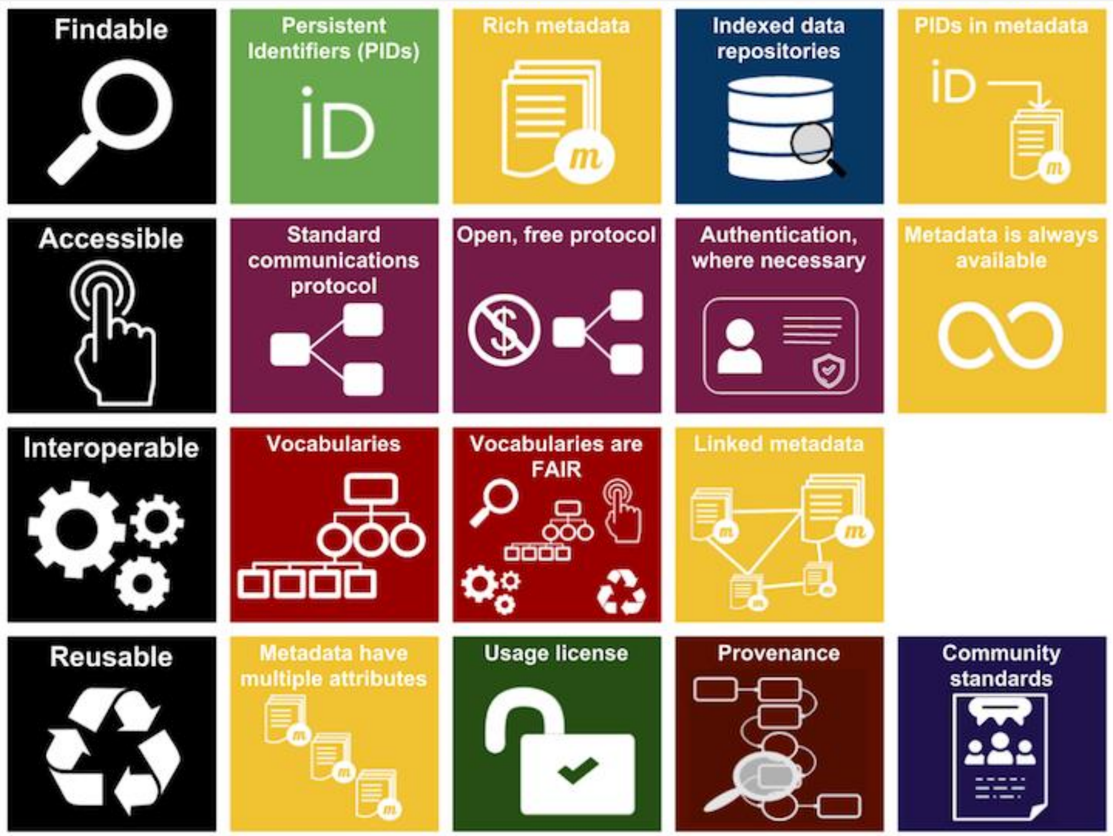
Peer-reviewed publications using GBIF-mediated data



#CiteTheDOI

https://www.gbif.org/resource/search?contentType=literature&literatureType=journal&relevance=GBIF_USED&peerReview=true

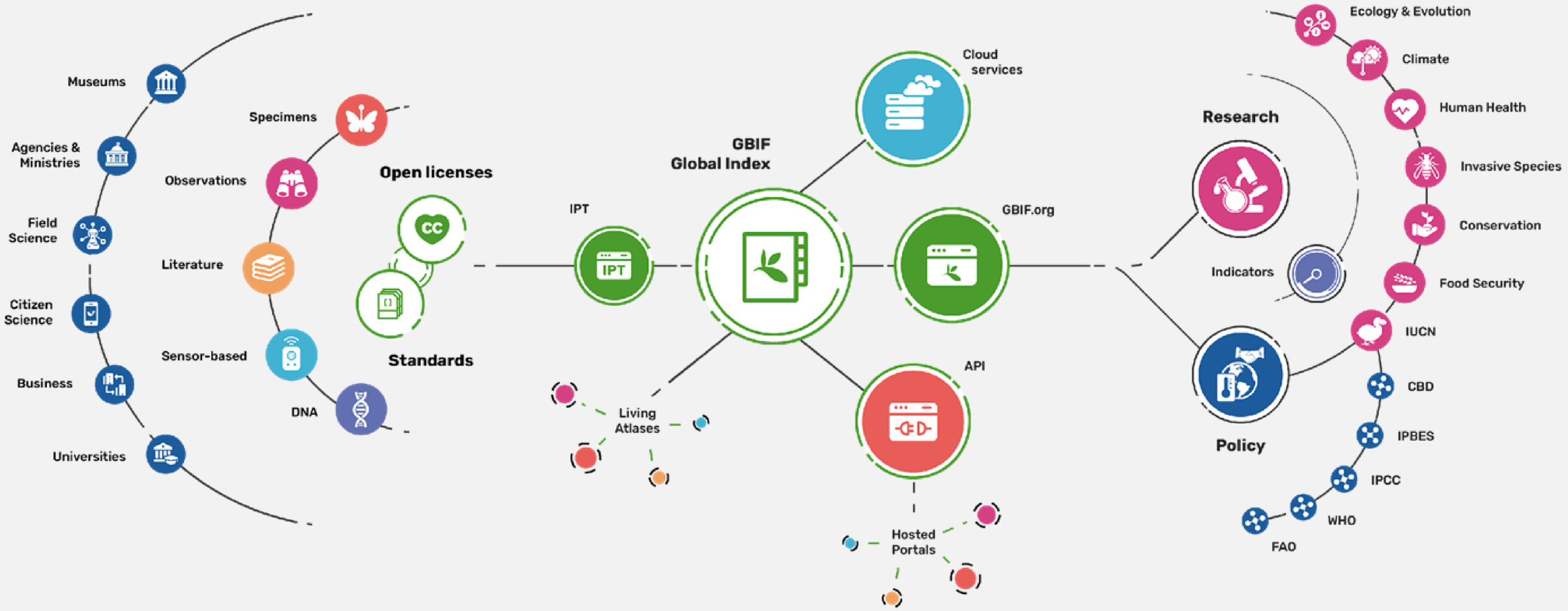
Why we need to go beyond open: FAIR



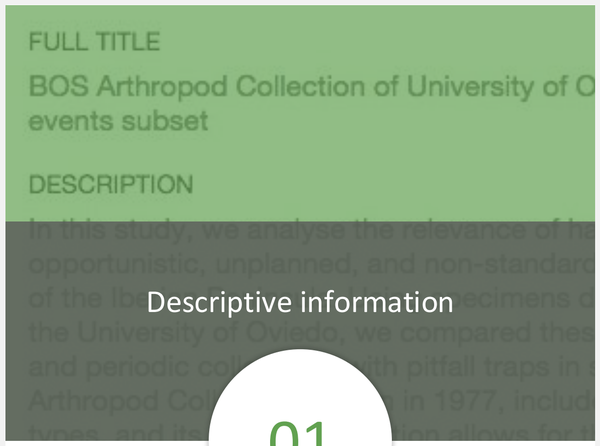
https://doi.org/10.1038/sdata.2016.18

https://www.gbif.org/resource/search?contentType=literature&year=2022,2022&literatureType=journal&topics=HUMAN_HEALTH&topics=DATA_PAPER&relevance=GBIF_PUBLISHED&peerReview=true&source=Gigabyte

Providing biodiversity evidence for research and policy



Data richness levels supported by GBIF



Descriptive information

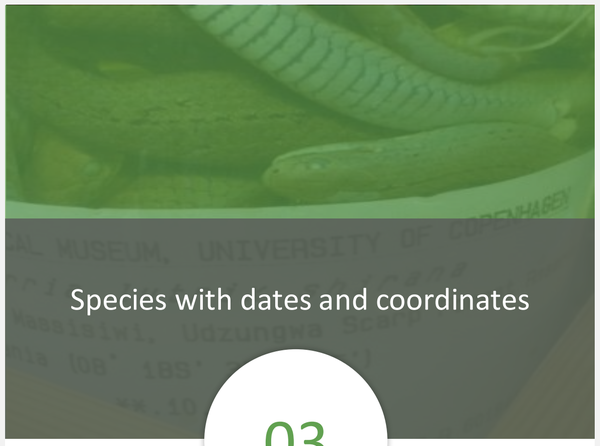
01

Dataset metadata

Species in countries and areas

02

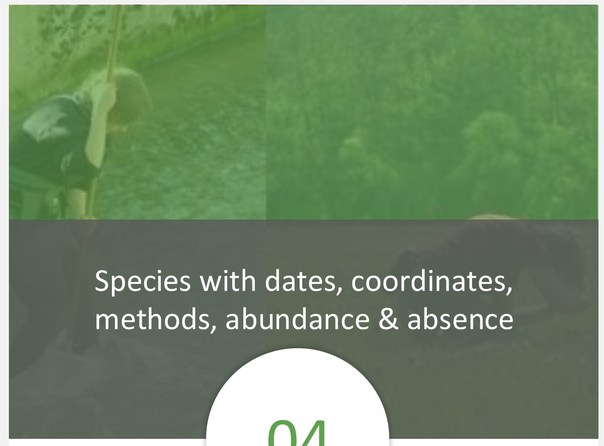
Species checklists



Species with dates and coordinates

03

Occurrence-only data

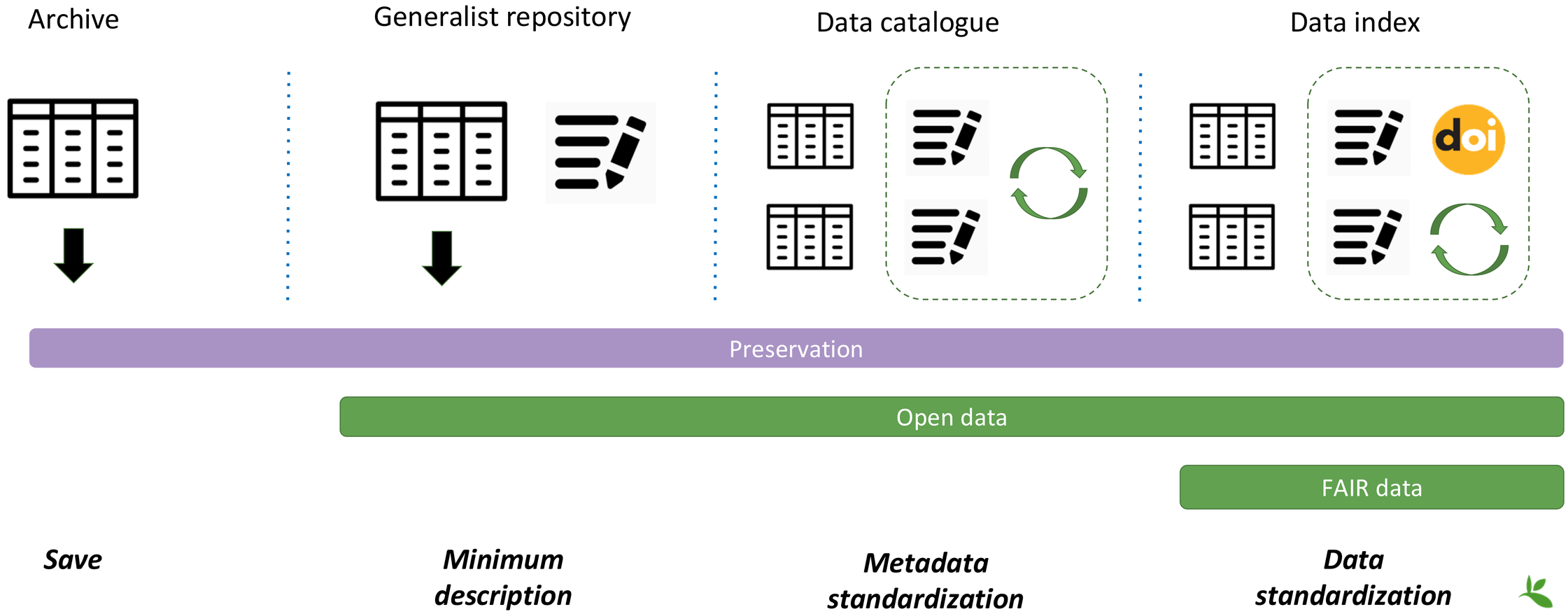


Species with dates, coordinates, methods, abundance & absence

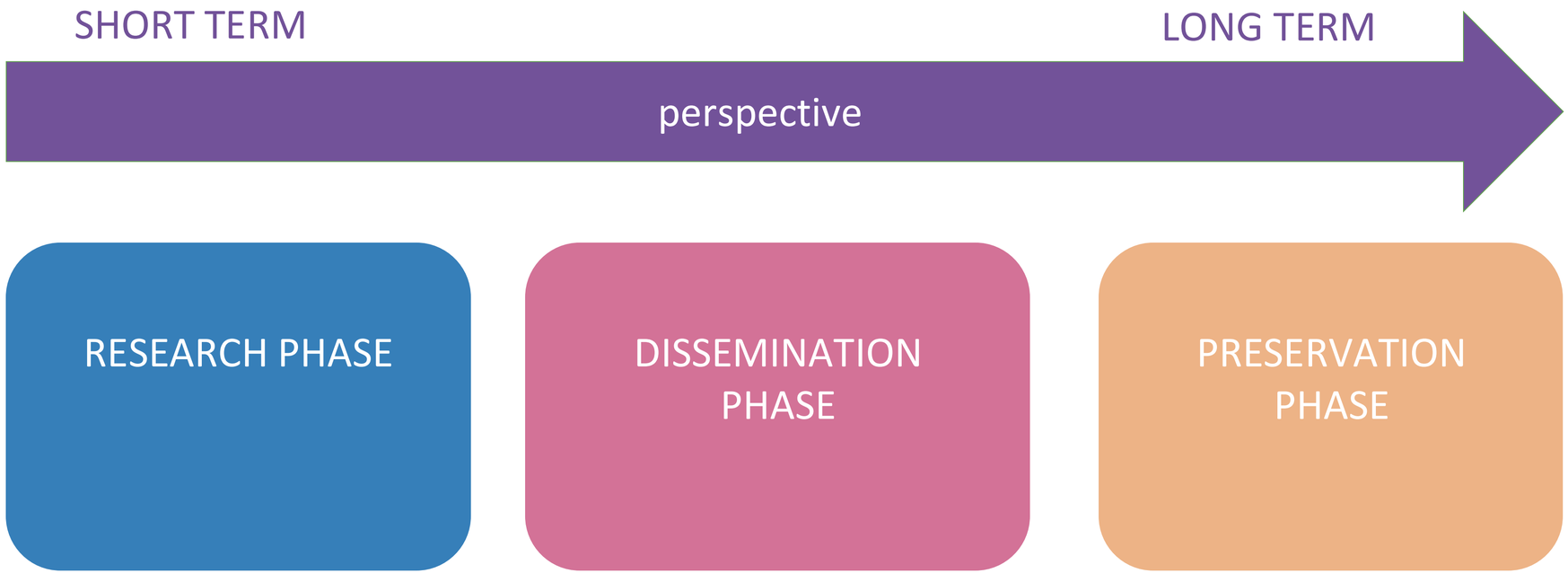
04

Sampling-event data

Data sharing in science: the WHERE choices



Data sharing in science: the WHEN choices

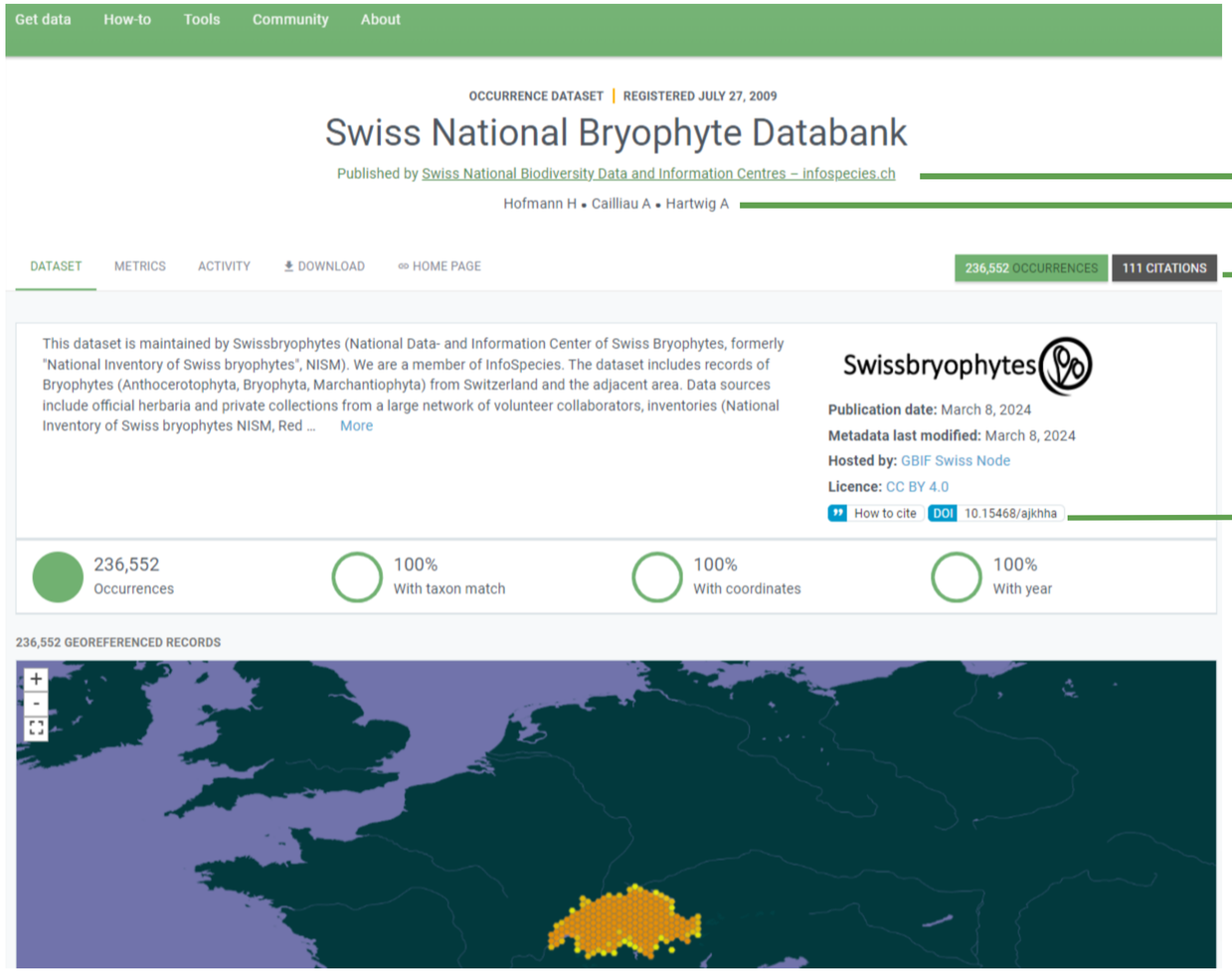


- file formats
- ownership
- storage
- backups

- share with whom
- embargo
- licensing
- metadata

- repository
- long-term manager

Data in 108,709 datasets: attribution, credit and affiliation

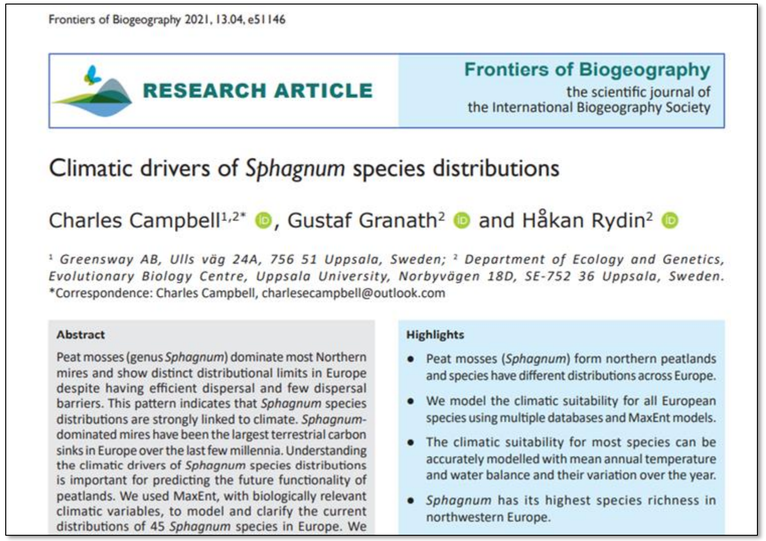


1 Affiliation

2 Authorship

3 Data citations

4 DOI


Frontiers of Biogeography 2021, 13.04, e51146

RESEARCH ARTICLE

Frontiers of Biogeography
the scientific journal of
the International Biogeography Society

Climatic drivers of *Sphagnum* species distributions

Charles Campbell[1,2*], Gustaf Granath[2] and Håkan Rydin[2]

[1] *Greensway AB, Ulls väg 24A, 756 51 Uppsala, Sweden;* [2] *Department of Ecology and Genetics, Evolutionary Biology Centre, Uppsala University, Norbyvägen 18D, SE-752 36 Uppsala, Sweden.*
*Correspondence: Charles Campbell, charlesecampbell@outlook.com

Abstract

Peat mosses (genus *Sphagnum*) dominate most Northern mires and show distinct distributional limits in Europe despite having efficient dispersal and few dispersal barriers. This pattern indicates that *Sphagnum* species distributions are strongly linked to climate. *Sphagnum*-dominated mires have been the largest terrestrial carbon sinks in Europe over the last few millennia. Understanding the climatic drivers of *Sphagnum* species distributions is important for predicting the future functionality of peatlands. We used MaxEnt, with biologically relevant climatic variables, to model and clarify the current distributions of 45 *Sphagnum* species in Europe. We

Highlights

- Peat mosses (*Sphagnum*) form northern peatlands and species have different distributions across Europe.
- We model the climatic suitability for all European species using multiple databases and MaxEnt models.
- The climatic suitability for most species can be accurately modelled with mean annual temperature and water balance and their variation over the year.
- *Sphagnum* has its highest species richness in northwestern Europe.

Hofmann H, Kiebacher T, Moser T, Meier M (2021). Swiss National Bryophyte Databank. Swiss National Biodiversity Data and Information Centres – infospecies.ch. Occurrence dataset https://doi.org/10.15468/ajkhha accessed via GBIF.org on 2022-04-28.

Task group to enhance GBIF-enabled research on species linked to human diseases

Experts to help guide mobilization and use of data on wild species that serve as hosts, vectors and reservoirs and cause disease in human communities

Tiger mosquito (*Aedes albopictus*), Côte d'Azur, France. Photo 2020 Martin Galli via iNaturalist Research-grade Observations, licensed under CC BY-NC 4.0.

2020-2022

NEWS | 9 FEBRUARY 2023

Renewed task group targets further advances for biodiversity data linked to human diseases

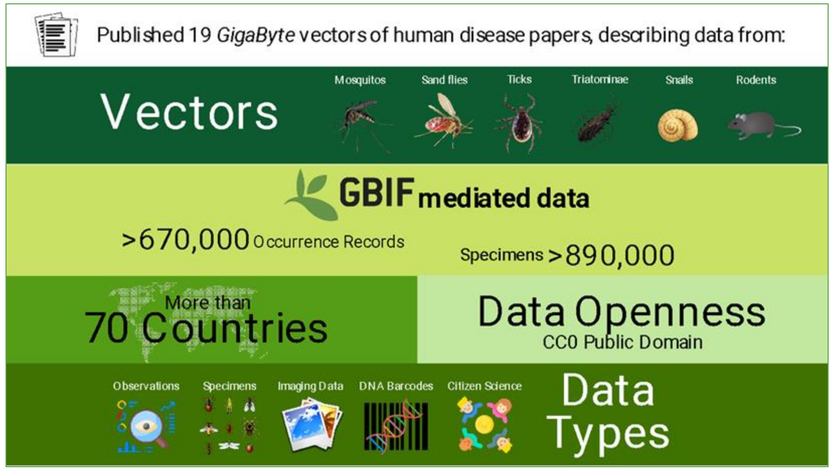
Final report from first term outlines improvements in discovery, access and use of data on wild hosts, vectors, and reservoirs of human diseases

Aedes scapularis, observed in Brazil. Photo 2022 César Favacho via iNaturalist Research-grade Observations, licensed under CC BY-NC 4.0.

2023-2025

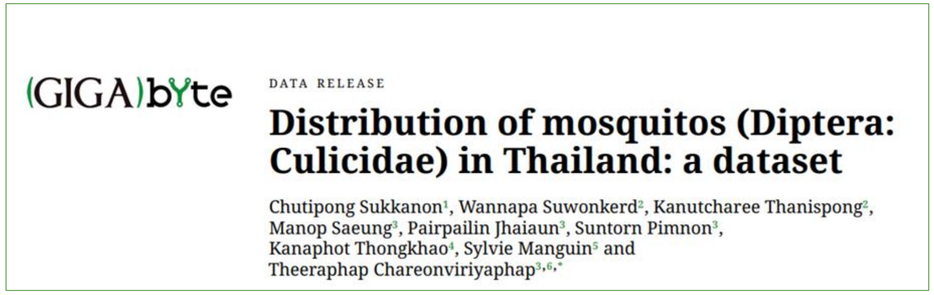
- The GBIF Secretariat has launched in 2020 a task group charged with helping to improve the discovery, access and use of biodiversity data linked to human diseases
- The task group on mobilization and use of biodiversity data for research and policy on human diseases is part of an ongoing effort to improve the completeness, relevance and fitness-for-use of biodiversity data shared through the GBIF network
- The Secretariat scoped the group in response to a growing range of studies that draw upon occurrence data for wild hosts, vectors, and reservoirs of human diseases

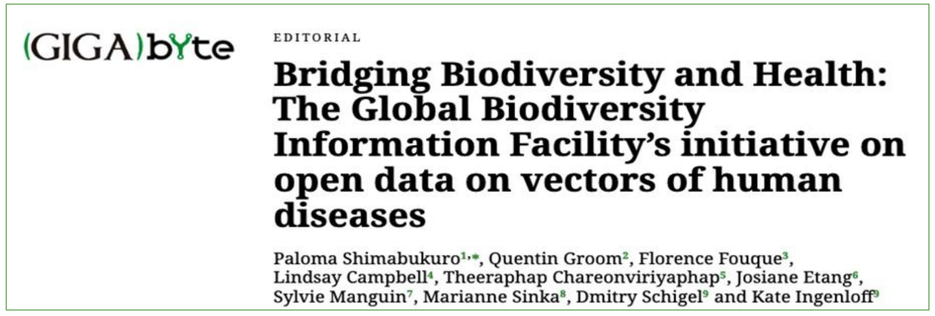

Get data
How-to
Tools
Community
About
GBIF and Health
Learn more about the connections that GBIF reveals between biodiversity and human health
Visual composition developed for the video GBIF & Health, 2022


Published 19 GigaByte vectors of human disease papers, describing data from:
Vectors
Mosquitos
Sand flies
Ticks
Triatominae
Snails
Rodents
GBIF mediated data
>670,000 Occurrence Records
Specimens >890,000
More than
70 Countries
Data Openness
CC0 Public Domain
Observations
Specimens
Imaging Data
DNA Barcodes
Citizen Science
Data Types


GBIF
Vectors of human disease series
Papers published : 20
Sort by : Published date (New)
Vector-borne diseases account for about one quarter of all infectious diseases. Although there has been significant progress for malaria, this progress is currently halting. Other diseases, such as those caused by arboviruses like dengue, chikungunya, yellow fever and Zika are expanding, with an increased number of cases and fatalities. There is a great need for data mobilization campaigns to improve data coverage to help research on these vector-borne diseases and human health. To address this need here we present collected series of Data Release papers with relevance for research on vectors of human vector-borne diseases. GigaScience Press has partnered with GBIF and supported by TDR, the Special Programme for Research and Training in Tropical Diseases, hosted at the World Health Organization to publish these papers. Data presented has all been shared to GBIF.org with high-quality data and metadata to improve data coverage to help research on vector-borne diseases and human health. With the publication of the second phase of papers sponsorship is now over, but the series is still open for submissions and please contact the GigaByte editors or GBIF helpdesk if you have any questions on how to submit.
With thanks to the GBIF expert task group on human diseases for oversight and support..
DOI 10.46471/GIGABYTE_SERIES_0002
TDR For research on diseases of poverty
UNICEF · UNDP · World Bank · WHO


One Health 16 (2023) 100484
Contents lists available at ScienceDirect
One Health
journal homepage: www.elsevier.com/locate/onehlt
ELSEVIER
Biodiversity data supports research on human infectious diseases: Global trends, challenges, and opportunities
Francisca Astorga, Quentin Groom, Paloma Helena Fernandes Shimabukuro, Sylvie Manguin, Daniel Noesgaard, Thomas Orrell, Marianne Sinka, Tim Hirsch, Dmitry Schigel


(GIGA)byte
DATA RELEASE
Distribution of mosquitos (Diptera: Culicidae) in Thailand: a dataset
Chutipong Sukkanon, Wannapa Suwonkerd, Kanutcharee Thanispong, Manop Saeung, Pairpailin Jhaiaun, Suntorn Pimnon, Kanaphot Thongkhao, Sylvie Manguin and Theeraphap Chareonviriyaphap


(GIGA)byte
EDITORIAL
Bridging Biodiversity and Health: The Global Biodiversity Information Facility's initiative on open data on vectors of human diseases
Paloma Shimabukuro, Quentin Groom, Florence Fouque, Lindsay Campbell, Theeraphap Chareonviriyaphap, Josiane Etang, Sylvie Manguin, Marianne Sinka, Dmitry Schigel and Kate Ingenloff


Asia Pacific Conference on Mosquito and Vector Control
Home
About
Program
Abstract submission
Sponsorship & Exhibition
Registration
Log in
PRE-CONFERENCE WORKSHOP
CALL FOR APPLICATIONS
World Health Organization
TDR For research on diseases of poverty
UNICEF · UNDP · World Bank · WHO
GBIF Global Biodiversity Information Facility
Selection of participants to a joint TDR-GBIF training workshop on data sharing on vectors to prevent and control vector-borne diseases within the context of multisectoral approaches
Deadline for submission: 30 July 2023, 5 pm (Thailand Time)

One Health 16 (2023) 100484

Contents lists available at ScienceDirect

One Health

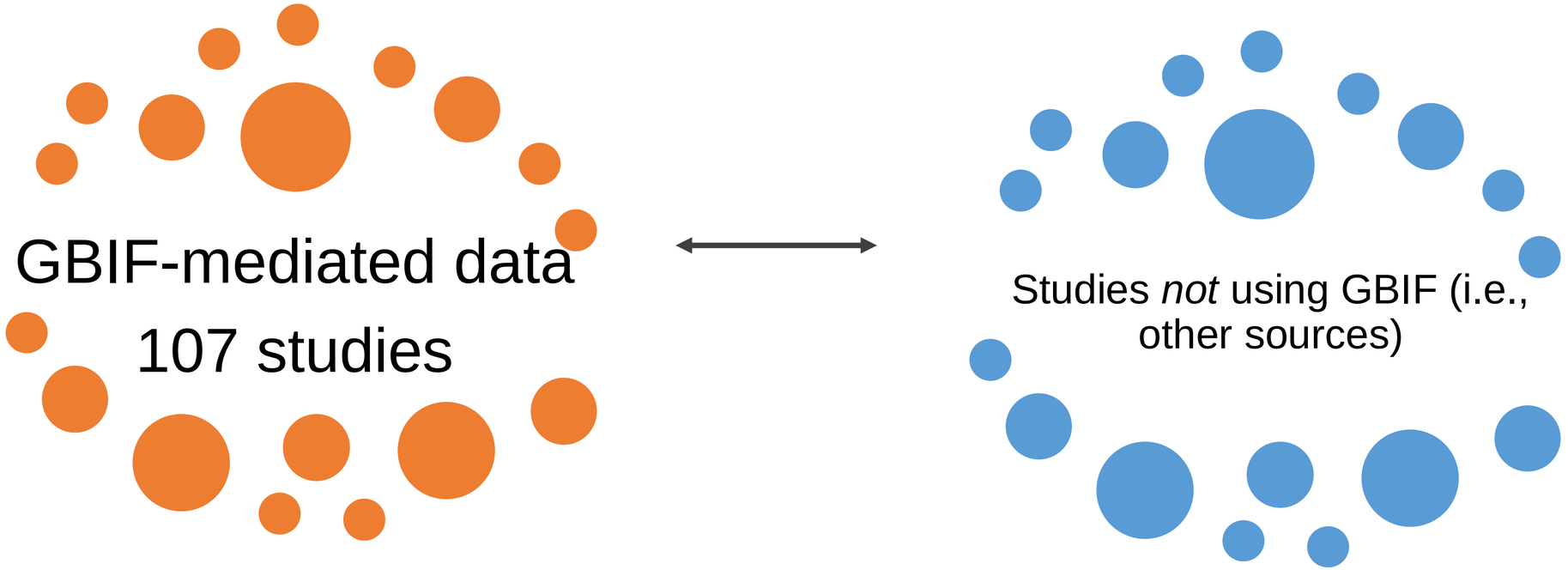
journal homepage: www.elsevier.com/locate/onehlt

Biodiversity data supports research on human infectious diseases: Global trends, challenges, and opportunities

Francisca Astorga [a,*], Quentin Groom [b], Paloma Helena Fernandes Shimabukuro [c], Sylvie Manguin [d], Daniel Noesgaard [e], Thomas Orrell [f], Marianne Sinka [g], Tim Hirsch [e], Dmitry Schigel [e]

- Landscaping study analyzing publications that used GBIF-mediated data and studies that have not
- Findings:
  - Most of the studies (particularly those using GBIF-mediated data) tended to analyze two or more species, mainly related to pathogens, highlighting an important need for increasing and improving available data on species of medical importance on GBIF
  - Very little data on pathogens were found in public data portals or repositories, again, opening a way to establish connections with the medical research community in order to collaborate with GBIF

- Literature tracking system maintained by GBIF
- Collection of studies related to infectious disease that *reuse species occurrence data*

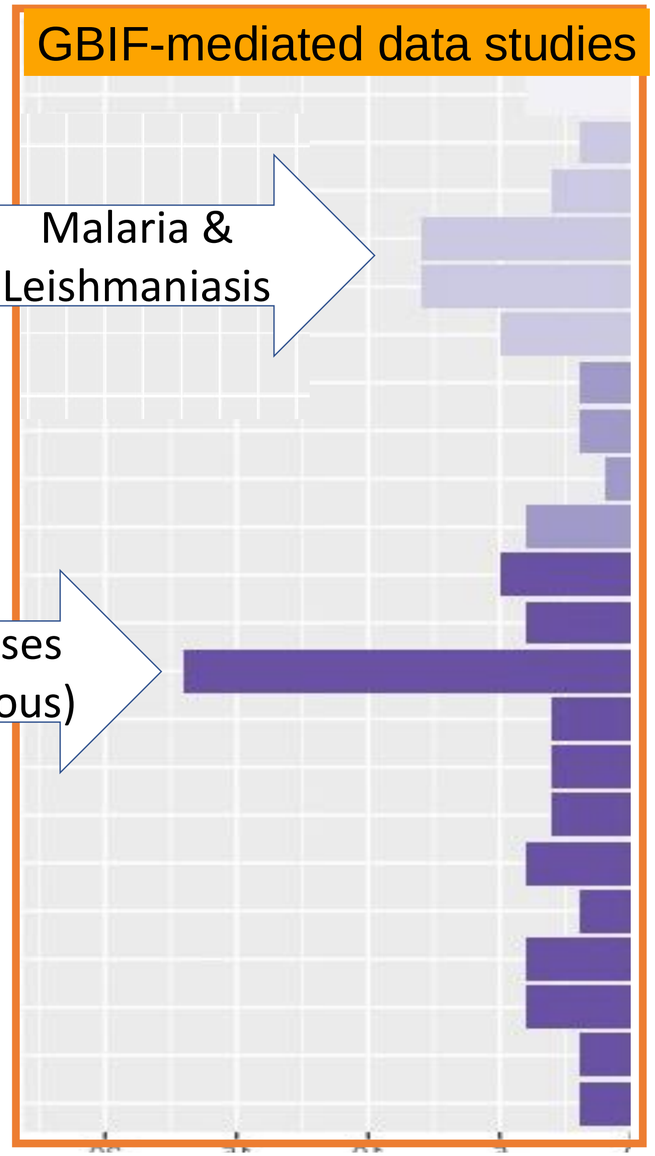


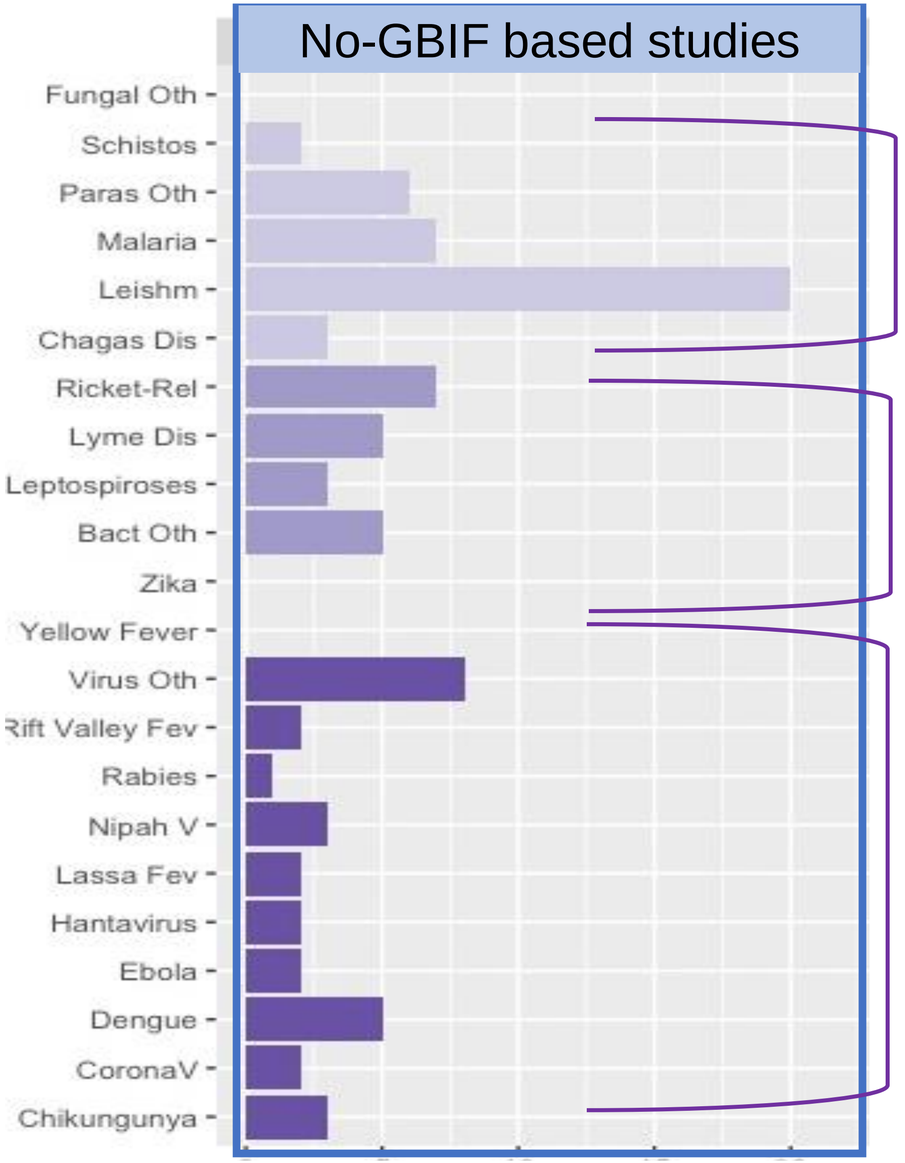
Results

What are the diseases covered by studies using data?

More biological and ecological sciences with host and vectors species

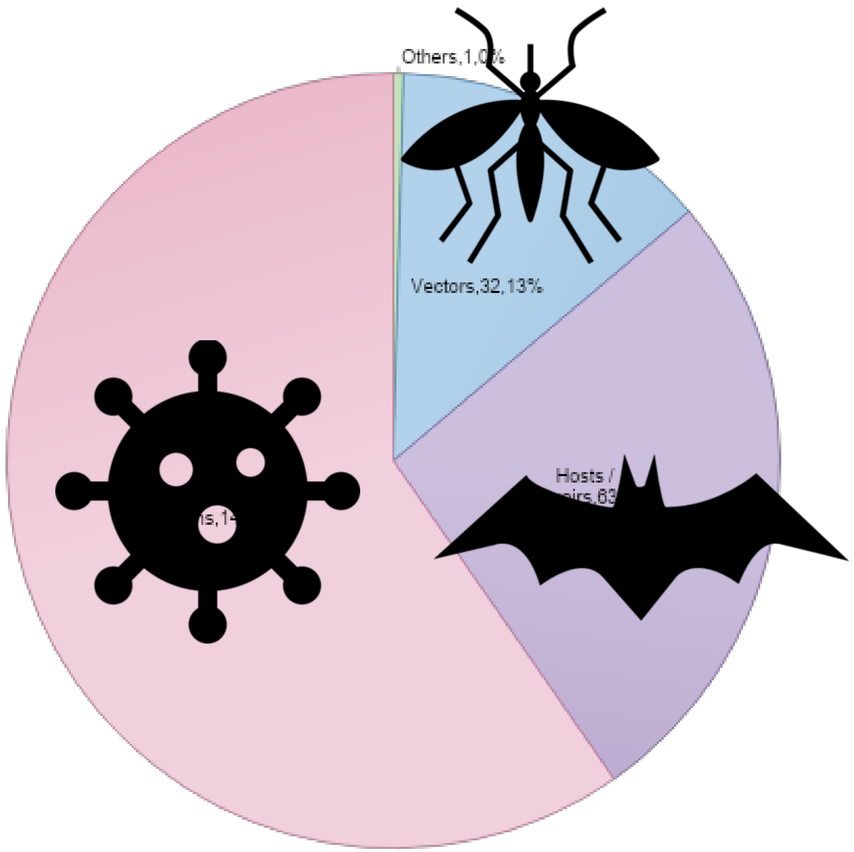
More medicine, public health, and veterinary sciences





Results:

For which epidemiological levels are data used?



Among the 214 studies there were 358 variables explored

- Most of them were pathogens (not serology)
- 1/3 host/reservoirs

(therefore, required corresponding occurrences data)

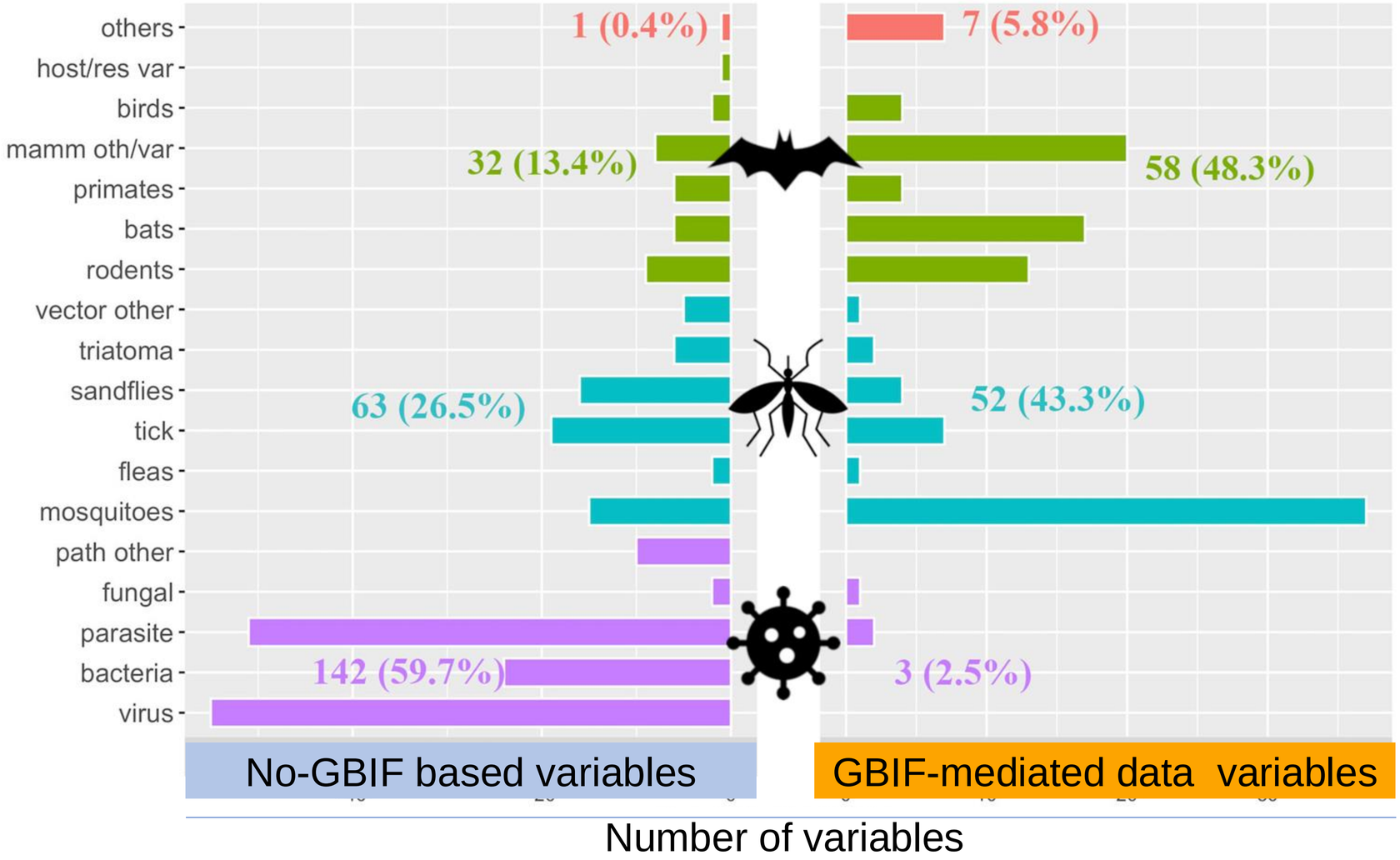
However, data sources (including GBIF) were unevenly used depending of the taxa and role of the species in disease dynamics

Results

When comparing with other sources, the biases is more evident.

Non-GBIF data sources are mainly used for pathogens

GBIF data sources are mainly broad scale ecological and health-related studies



GBIF Task Group

- The GBIF task group during more than 4 years was able to mobilize scientists around the world to use GBIF network for data openly shared under FAIR principles

- Training programs have been proposed to early-career scientists in vector data sharing during several international conferences (ICTMM in 2022, AMV in 2023 in Thailand)

- Large biodiversity datasets on hosts, vectors, and reservoirs of human diseases have been since published in Gigabyte Journal, sponsored by TDR/WHO

In conclusion, this task group has improved the visibility of GBIF in the field of vector-borne diseases through an ongoing effort to improve the completeness, relevance and fitness-for-use of biodiversity data on vectors and vector-borne diseases shared through the GBIF network

Conclusion

GBIF allows:

- **Recognition** for work carried out at the forefront, such as laboratory and field activities, with attribution and credit
- Increased awareness of the **importance of producing good quality data** by learning about the steps involved in producing and reusing data
- **Increased visibility** of institutions and compliance with regional, national and international standards/guidelines on open data
- **Contribution to global knowledge** of biodiversity
- Expanded **possibilities for collaboration** through exposure in an international repository
- **Tracking of data use** that can contribute to metrics and impact indicators of the work carried out.

Thank you for your attention!

Prof Sylvie Manguin

IRD, UM, Montpellier, France

sylvie.manguin@ird.fr