

# Global Biodiversity Information Facility (GBIF)

*...for science*

Dmitry Schigel | Scientific officer



# 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



**1** Building the evidence to advance scientific research and understanding of global biodiversity

1



**2** Supporting policy responses and knowledge transfer that address urgent societal challenges around planetary change

2



2023 - 2027

# GBIF Strategic Framework

**4** Driving innovation to advance biodiversity-related knowledge

4



**3** Enabling the network to meet future needs and challenges

3



# GBIF strategic framework 2023 - 2027

## Building the evidence to advance scientific research and understanding of global biodiversity

### *Objectives*



- Reduce knowledge gaps by helping the network to set targets for consolidating data coverage across thematic, taxonomic, phylogenetic, spatial and temporal dimensions.
- Enable, expand and diversify the uptake and application of GBIF-mediated data.
- Support the evolution of fundamental data-driven biodiversity research and its application across methods, scales and disciplines in life and environmental sciences.
- Ensure meaningful data exchange and reuse of data through citations and measurements, promoting and consistently applying accepted standards that adhere to best practices and sustain the highest aspirations in open and data-intensive science.
- Increase GBIF's relevance to research, promote recognition, expand involvement and improve best practices for data in academia and higher education.



Datasets ●  
121,395

● Hosted portals  
69

● Peer-review papers  
using data (1 Oct 2025)  
14,415

● Average records  
downloaded per month (2025)  
297 billion

● Species  
occurrence records  
3,660,126,436

Country  
Participants ●  
69



Organizational  
Participants ●  
42

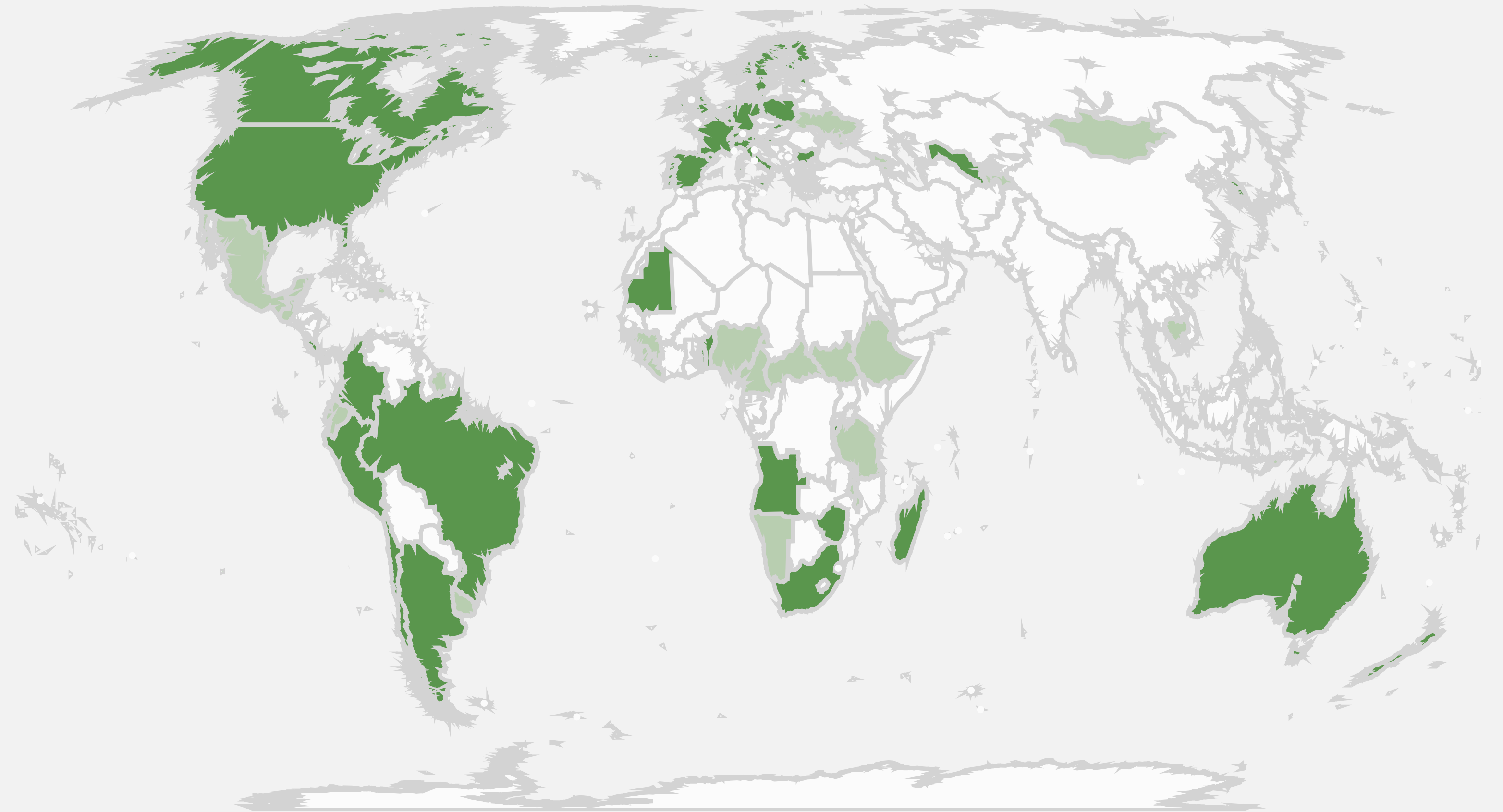
Publishers ●  
2,669





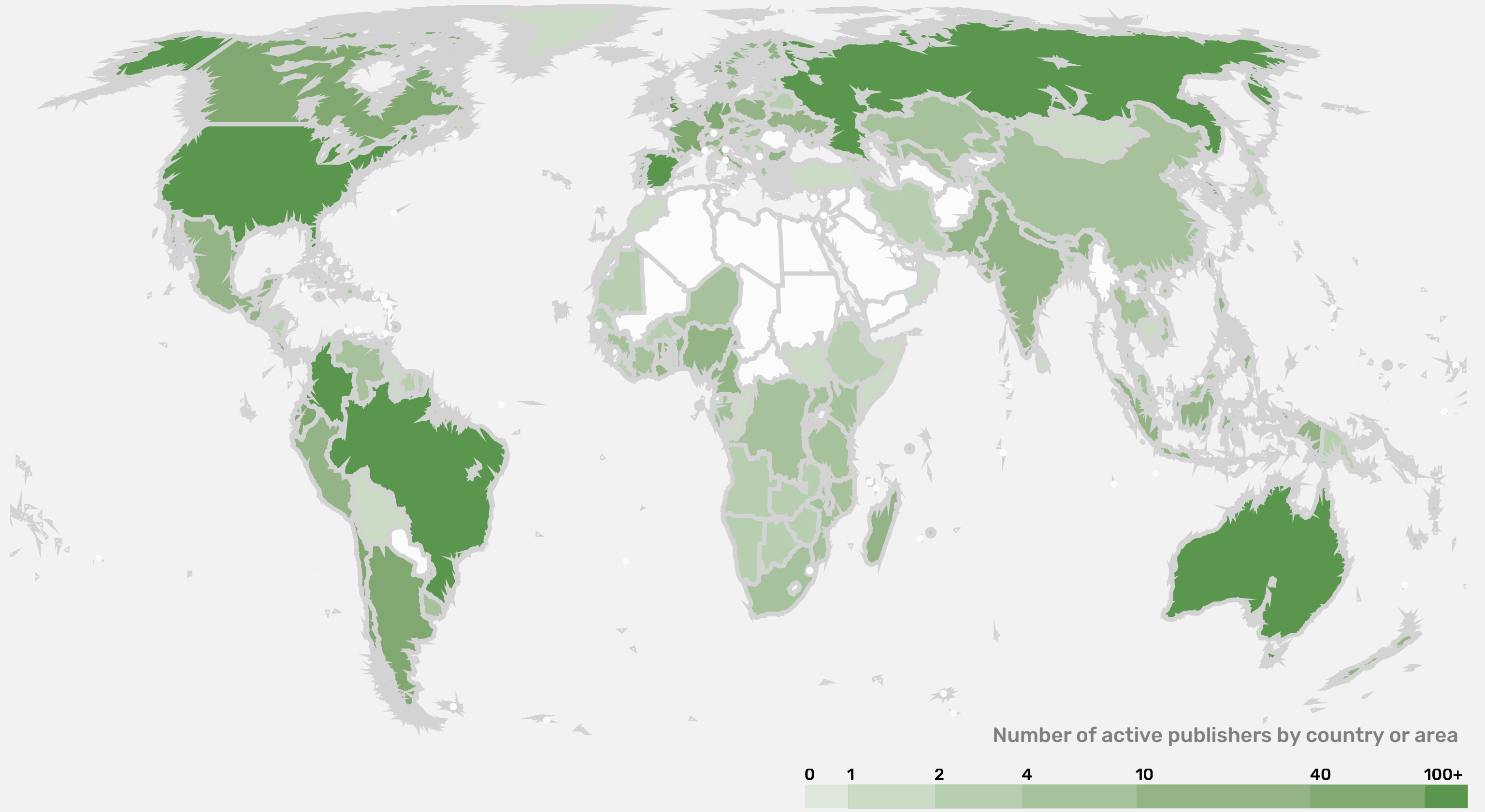
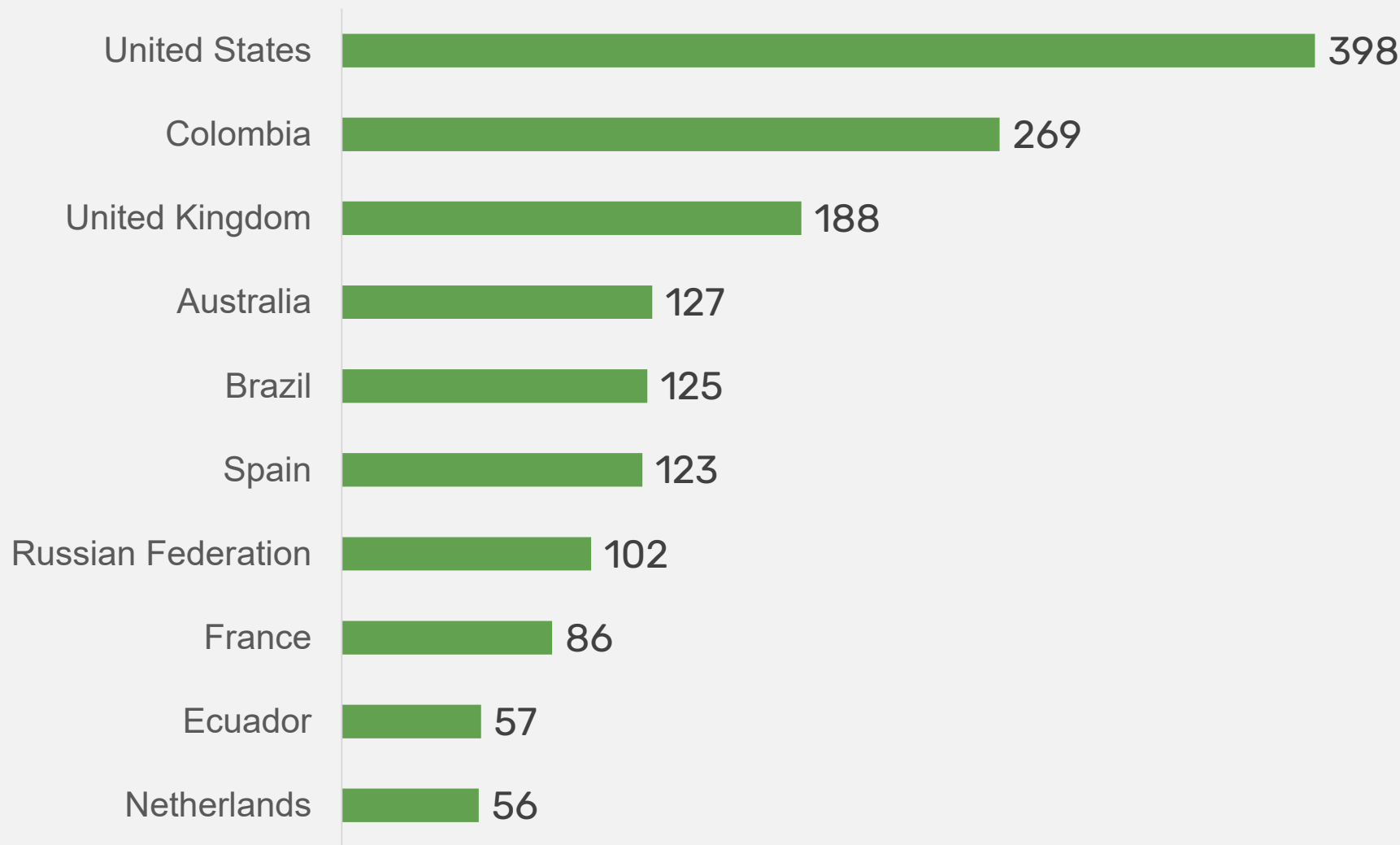
# GBIF Participant countries

-  Voting Participants
-  Associate Participants



# GBIF network of data publishing institutions

**148** countries/areas with institutions sharing data through GBIF



# GBIF.org traffic by country: 2025

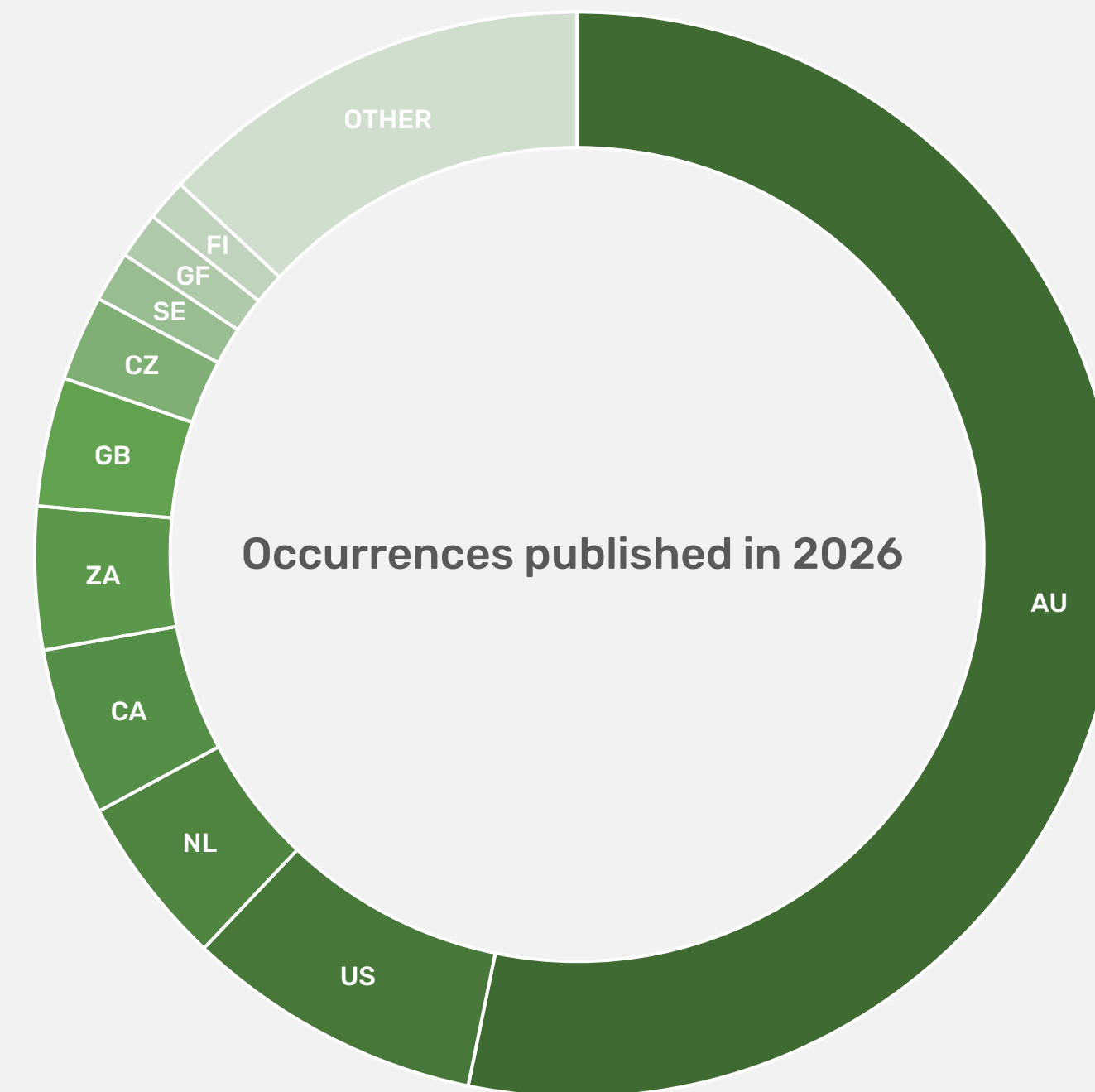
Rank	Country/Territory	Visitors	Visits	Effective visits*	% Effective visits	2025 rank	Views / visit
1	United States	609,750	735,409	117,665	16 %	1	2.46
2	China	313,153	658,326	157,998	24 %	4	2.37
3	Russian Federation	188,313	221,239	26,549	12 %	10	2.5
4	France	102,980	123,635	61,818	50 %	9	6.19
5	United Kingdom	72,723	95,179	34,264	36 %	22	5.49
6	Germany	68,413	82,941	44,788	54 %	17	8.78
7	Colombia	66,470	81,212	43,854	54 %	11	6.37
8	Vietnam	62,883	64,236	3,854	6 %	12	1.59
9	Spain	61,423	76,784	44,535	58 %	13	7.94
10	Brazil	61,353	72,191	36,096	50 %	2	6.28
<b>TOTAL</b>		<b>1,607,458</b>	<b>3,108,830</b>	<b>1,025,914</b>	<b>33 %</b>		<b>4.39</b>

SOURCE: Plausible.io. \* Effective visits = total visits - (total visits x bounce rate)



# Occurrences published by country: 2026

Rank	Country/Territory	2026	2025 total	2025 rank
1	Australia	52,256,917	53,466,496	2
2	United States	8,694,274	203,021,107	1
3	Netherlands	4,996,292	21,639,084	8
4	Canada	4,920,733	32,338,096	5
5	South Africa	4,183,639	3,205,585	24
6	United Kingdom	3,766,760	41,629,288	3
7	Czechia	2,496,906	862,103	43
8	Sweden	1,474,542	15,314,346	10
9	French Guiana	1,368,501	47,406	116
10	Finland	1,218,436	4,211,838	20
<b>OTHER COUNTRIES + AREAS</b>		-12,817,312	237,985,556	
<b>TOTAL</b>		<b>72,559,688</b>	<b>613,827,594</b>	

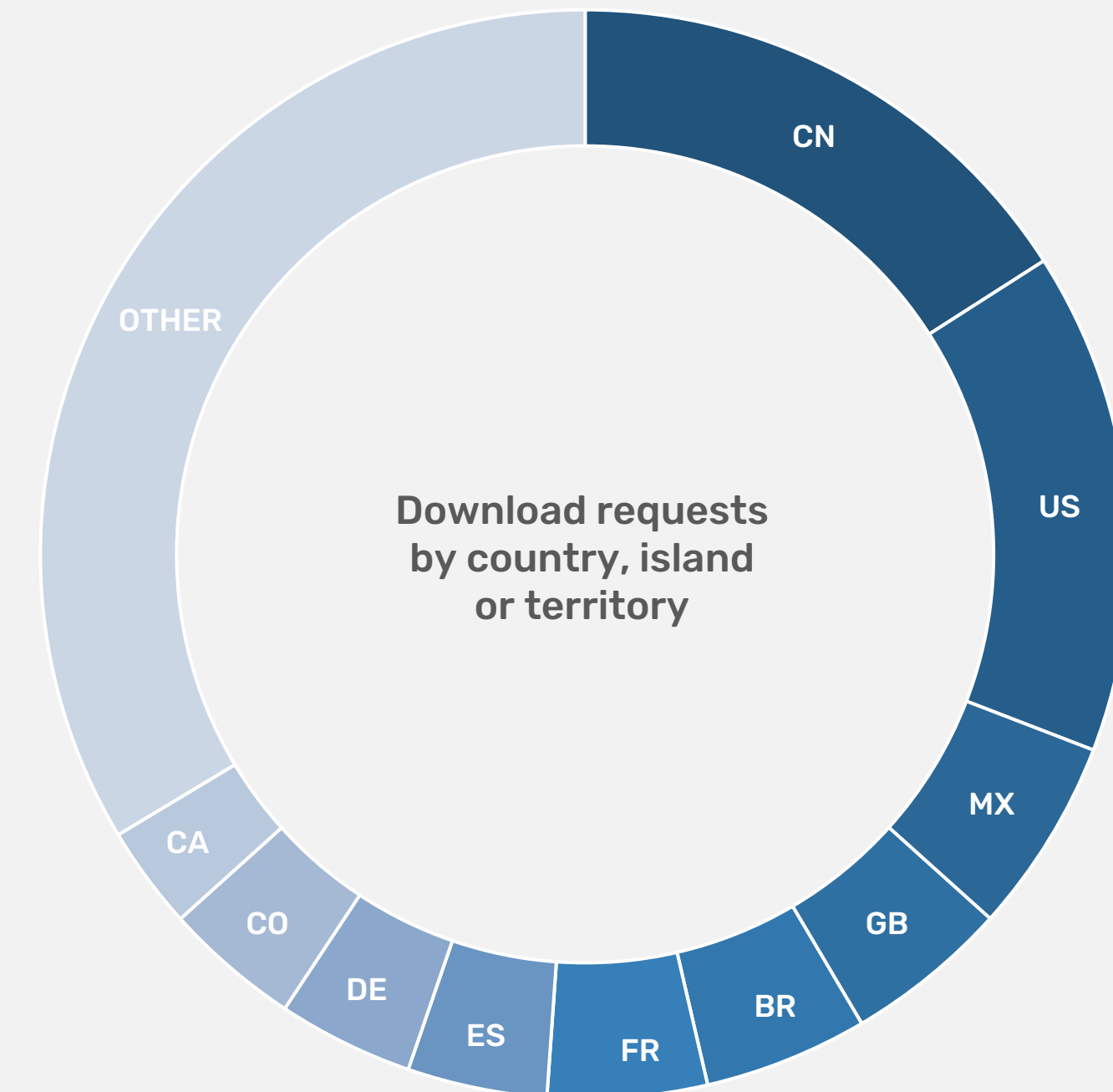


# Data download requests by country: 2025

Rank	Country	2026	2025 total	2025 rank
1	China	16,966	68,869	1
2	United States	15,876	46,554	2
3	Mexico	6,182	27,758	3
4	United Kingdom	5,983	15,124	6
5	Brazil	5,179	19,629	5
6	France	5,039	10,125	8
7	Spain	4,412	15,030	7
8	Germany	4,294	9,628	10
9	Colombia	4,293	22,565	4
10	Canada	3,258	9,678	9
<b>OTHER COUNTRIES + AREAS</b>		35,778	131,765	
<b>TOTAL</b>		<b>107,260</b>	<b>376,725</b>	

174

countries, islands & territories where users have ≥5 requested downloads in 2026

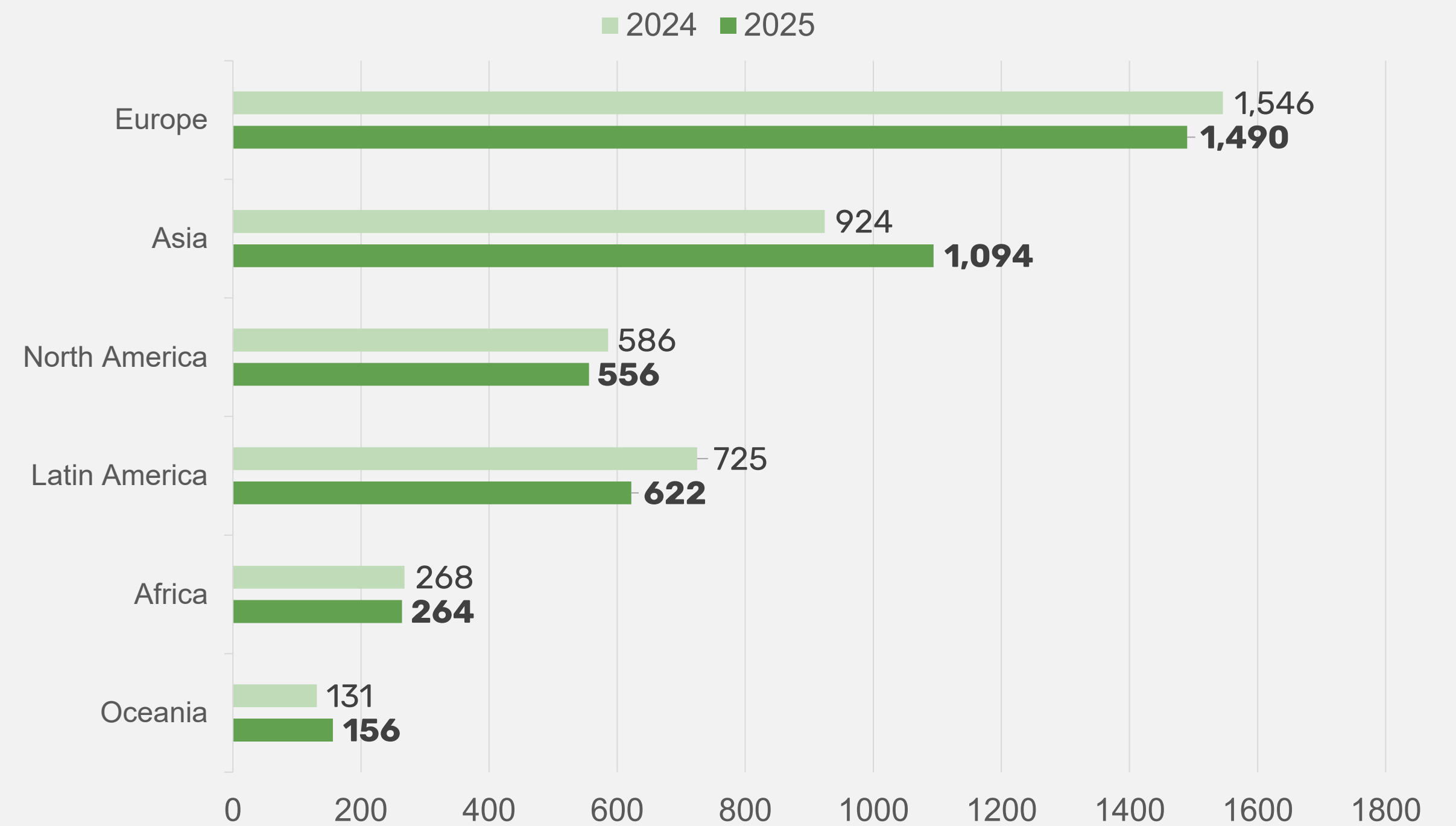


# Data use in peer-reviewed journals: 2025

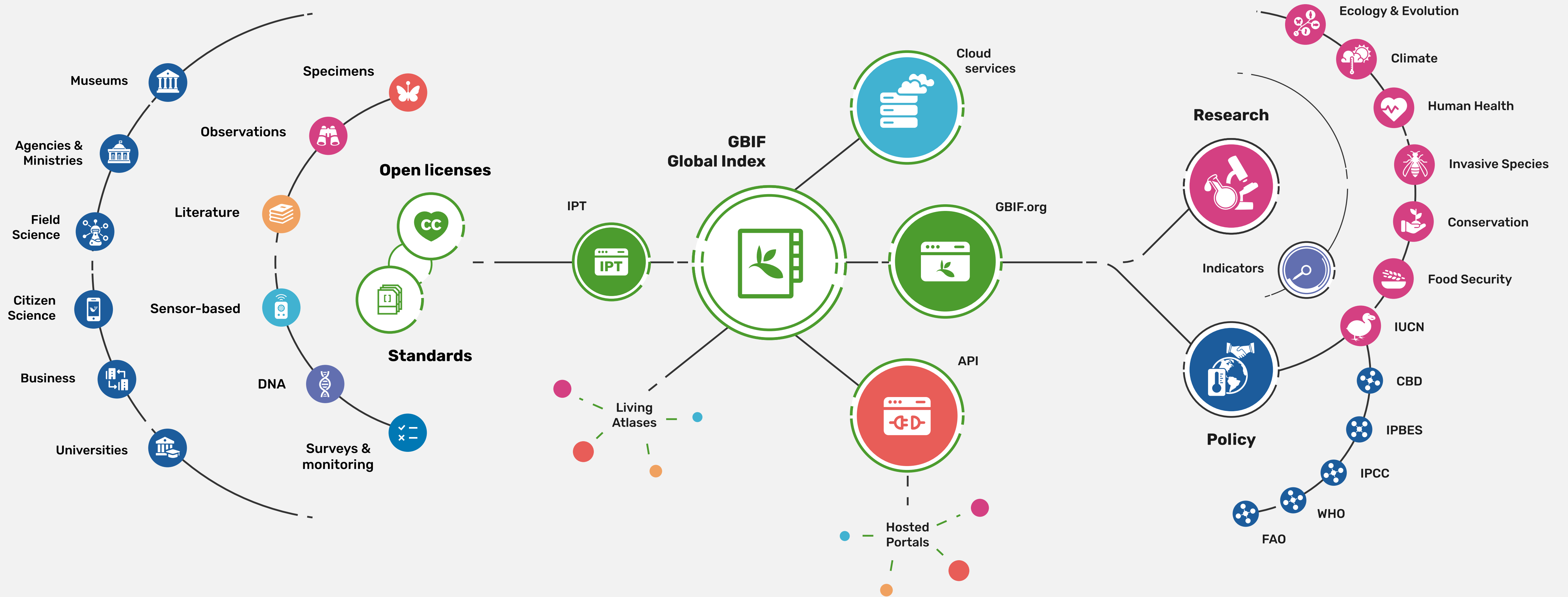
Peer-reviewed uses by country

End of Year		2025 total	2024 total	2024 rank
1	China	656	529	1
2	United States	446	486	2
3	Brazil	226	238	3
4	United Kingdom	206	186	5
5	Germany	181	186	6
6	Spain	141	171	7
7	Australia	121	104	10
8	Mexico	115	186	4
9	Canada	106	100	11
10	India	93	78	12

Peer-reviewed uses by region



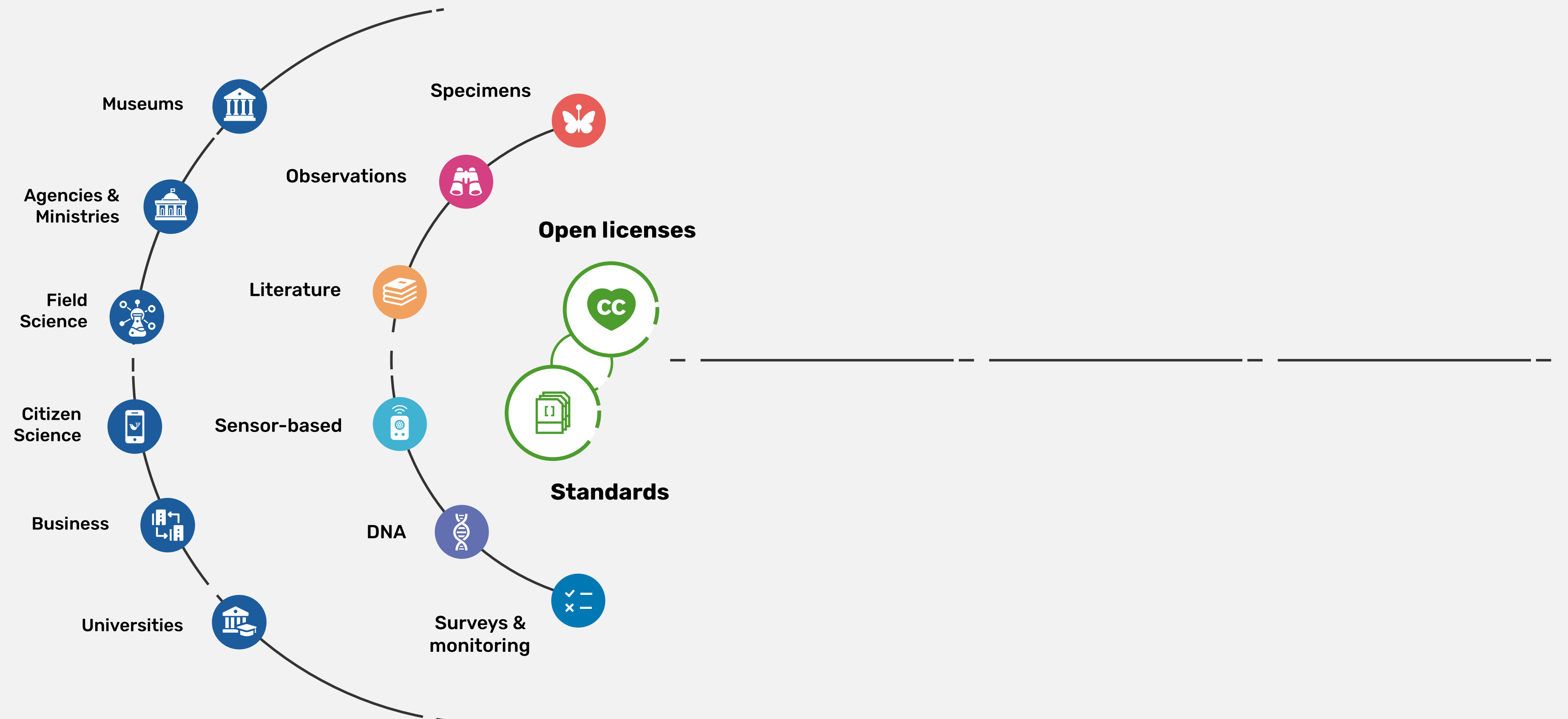
# Providing biodiversity evidence for research and policy



# Sources of biodiversity evidence

**Create**

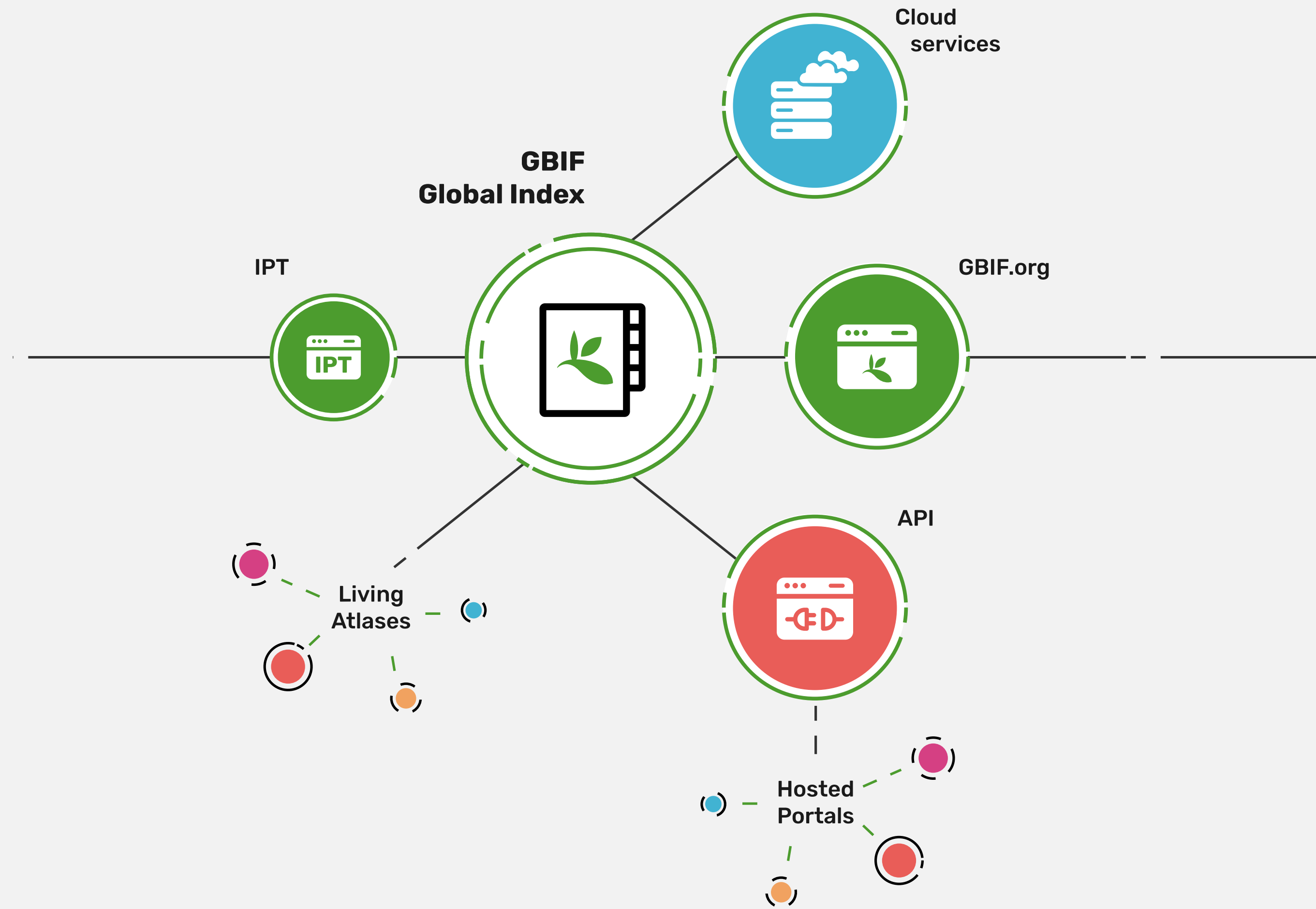
*Combine sources of evidence*



# Access to biodiversity evidence

Share

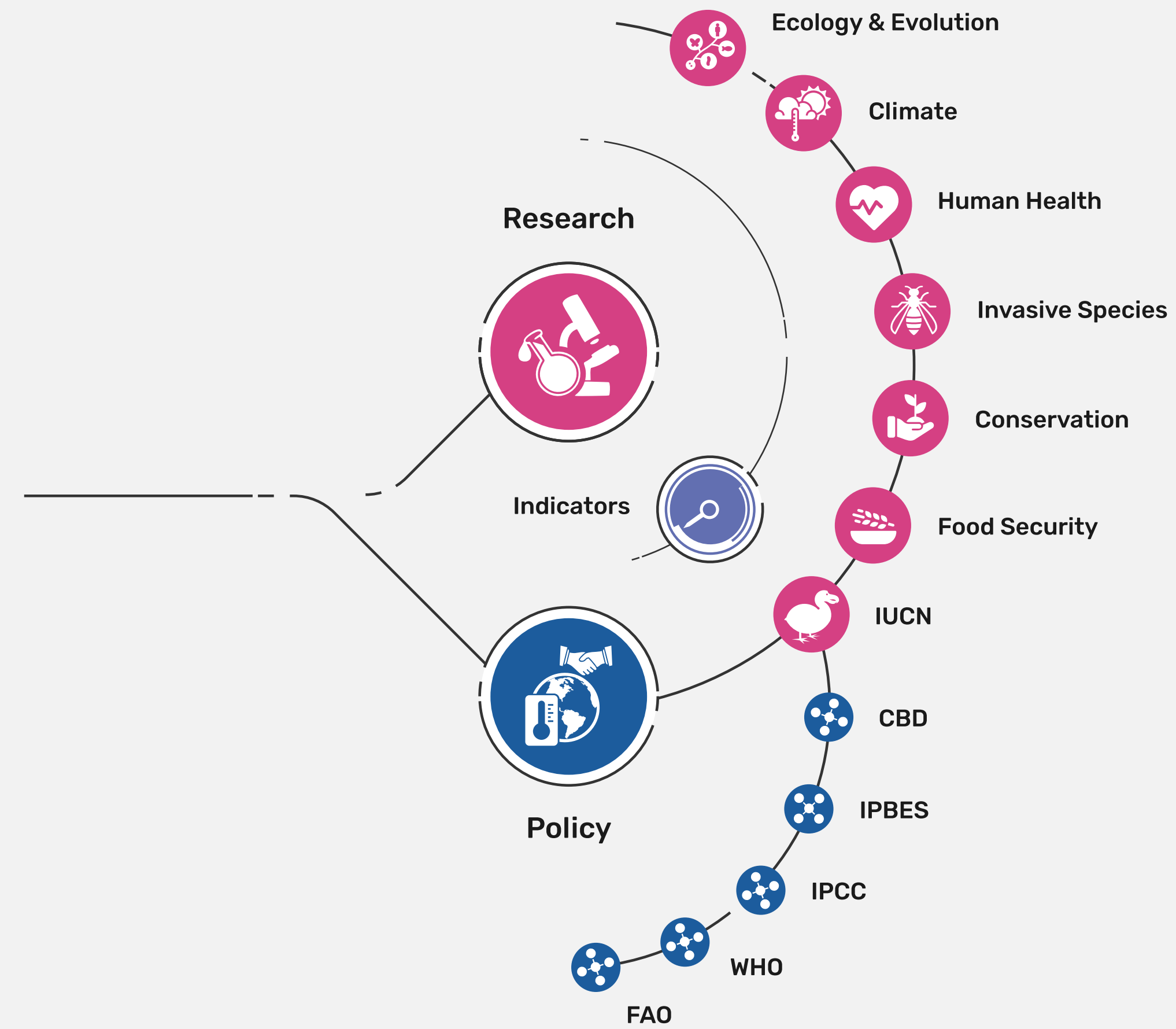
*FAIR and open access*



# Uses of biodiversity evidence

## Transform

*Apply and use data*



# Science and research



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

Get data | How-to | Tools | Community | About

OCCURRENCE DATASET | REGISTERED JULY 27, 2009

## Swiss National Bryophyte Databank

Published by [Swiss National Biodiversity Data and Information Centres – infospecies.ch](https://www.infospecies.ch)

Hofmann H • Cailliau A • Hartwig A

DATASET | METRICS | ACTIVITY | DOWNLOAD | HOME PAGE

236,552 OCCURRENCES | 111 CITATIONS

This dataset is maintained by Swissbryophytes (National Data- and Information Center of Swiss Bryophytes, formerly "National Inventory of Swiss bryophytes", NISM). We are a member of InfoSpecies. The dataset includes records of Bryophytes (Anthocerotophyta, Bryophyta, Marchantiophyta) from Switzerland and the adjacent area. Data sources include official herbaria and private collections from a large network of volunteer collaborators, inventories (National Inventory of Swiss bryophytes NISM, Red ... [More](#))

**Swissbryophytes**

Publication date: March 8, 2024  
Metadata last modified: March 8, 2024  
Hosted by: GBIF Swiss Node  
Licence: CC BY 4.0  
How to cite | DOI: 10.15468/ajkhha

236,552 Occurrences | 100% With taxon match | 100% With coordinates | 100% With year

236,552 GEOREFERENCED RECORDS

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<sup>1,2\*</sup>, Gustaf Granath<sup>2</sup> and Håkan Rydin<sup>2</sup>

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

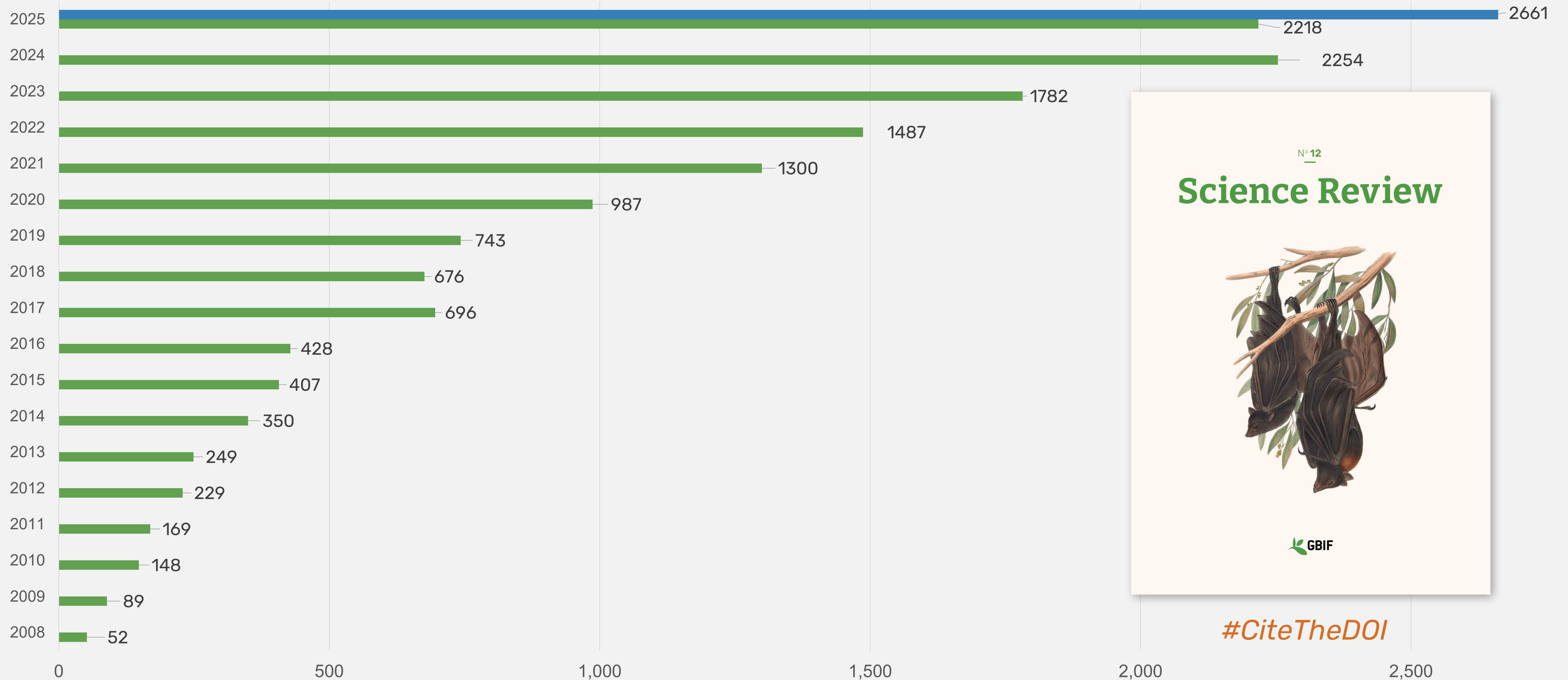
**Abstract**  
Peat mosses (genus *Sphagnum*) dominate most Northern mires and show distinct distributional limits in Europe despite having efficient dispersal and few dispersal barriers. This pattern indicates that *Sphagnum* species distributions are strongly linked to climate. *Sphagnum*-dominated 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.

# Peer-reviewed publications using GBIF-mediated data

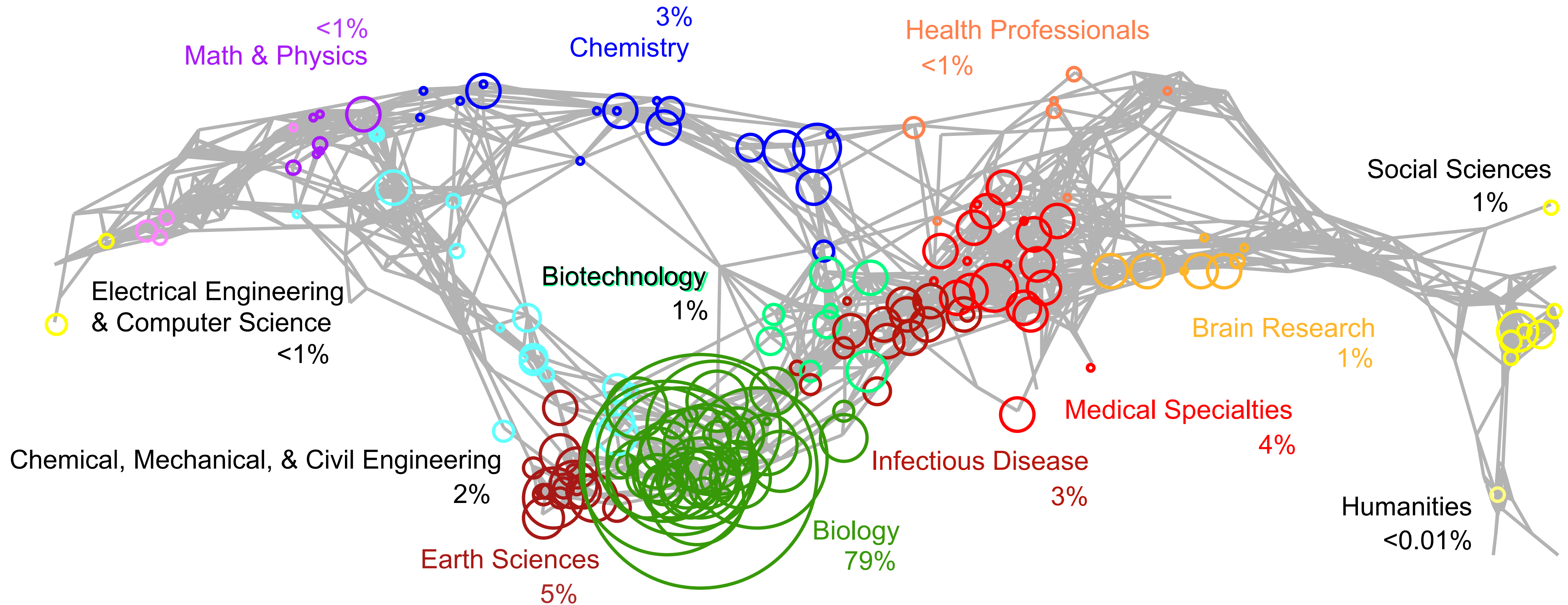


#CiteTheDOI

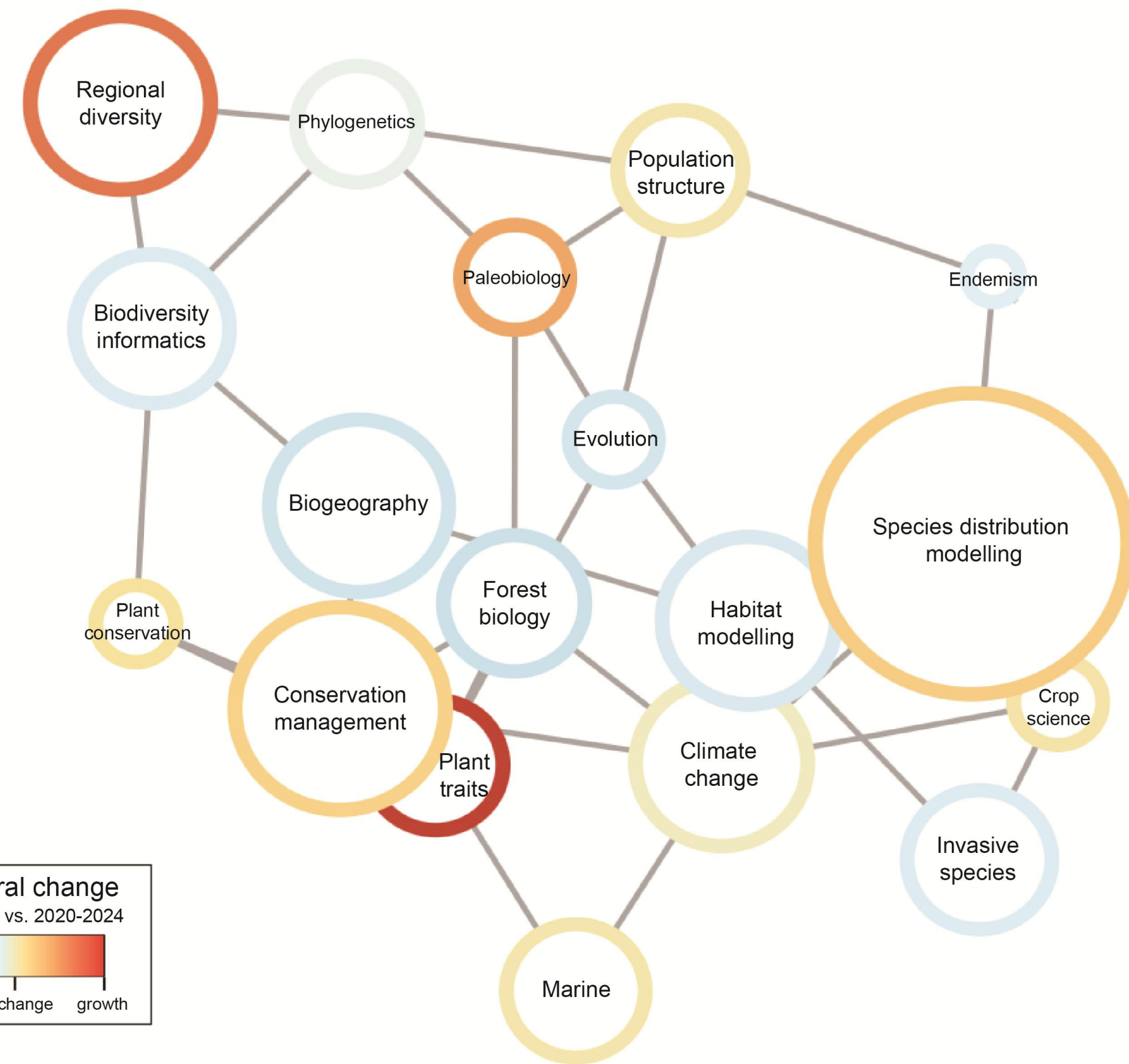
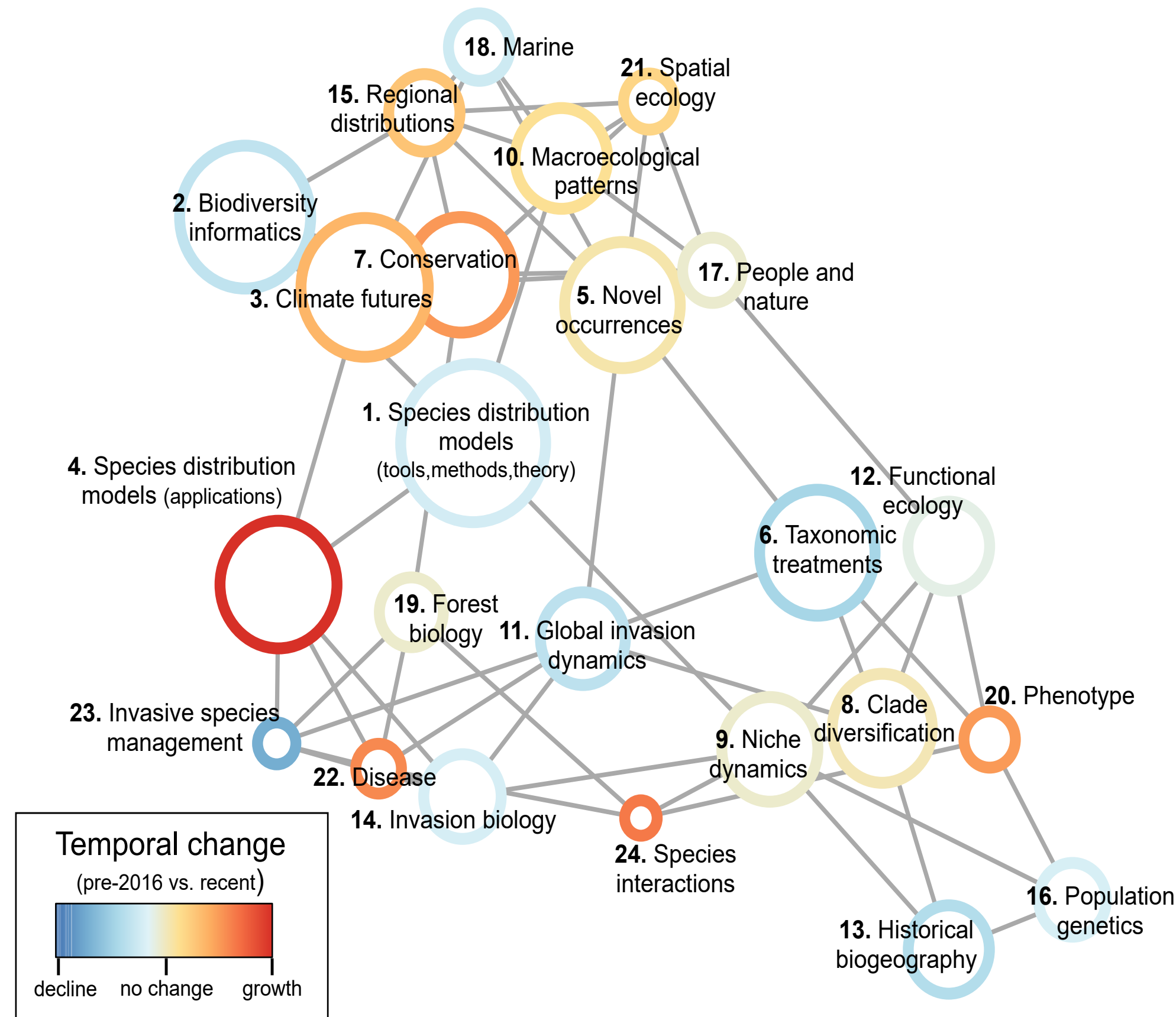
■ Projection ■ Annual totals



# Map of GBIF-enabled science



# Systematic reviews of GBIF-enabled science



*Heberling et al. 2021*  
Landscaping study I  
2016 – 2019

DOI: [10.1073/pnas.2018093118](https://doi.org/10.1073/pnas.2018093118)

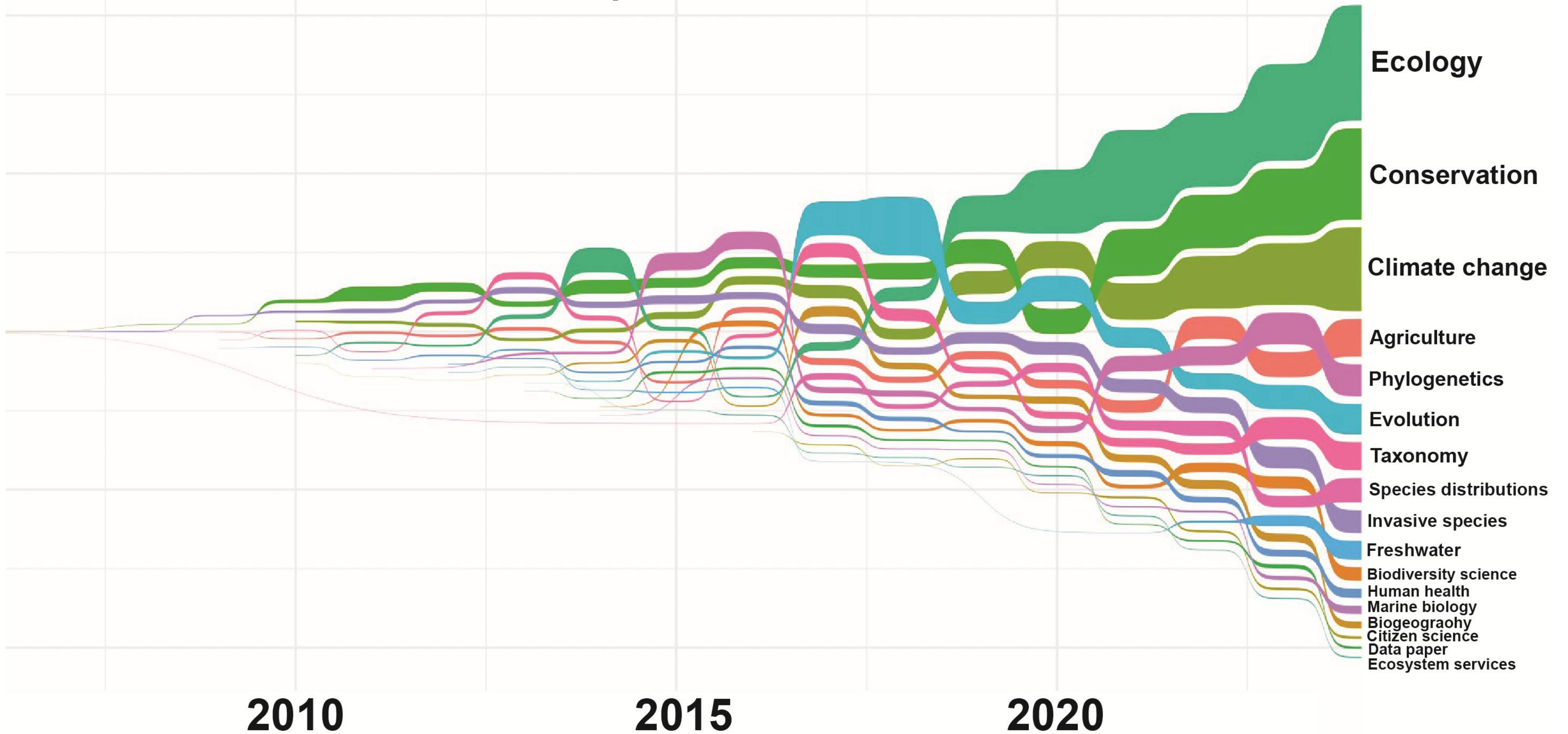


*Steinke et al. 2025*  
Landscaping study II  
2020 – 2024

DOI: [10.1073/pnas.2519119122](https://doi.org/10.1073/pnas.2519119122)



# GBIF-enabled science: research topics over time: 2010 – 2024



# GBIF thematic relevance



[www.gbif.org/science-review](http://www.gbif.org/science-review)



## Thematic communities



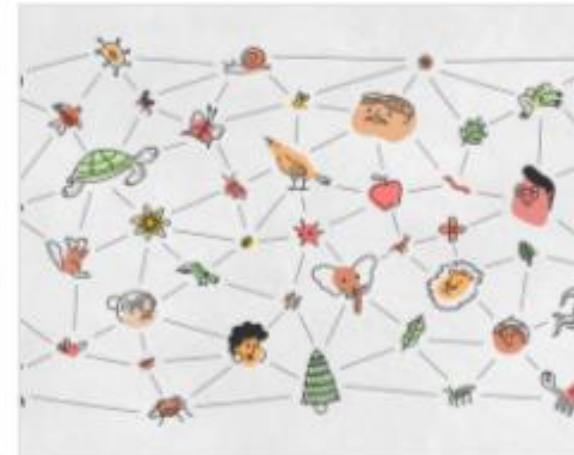
Soil



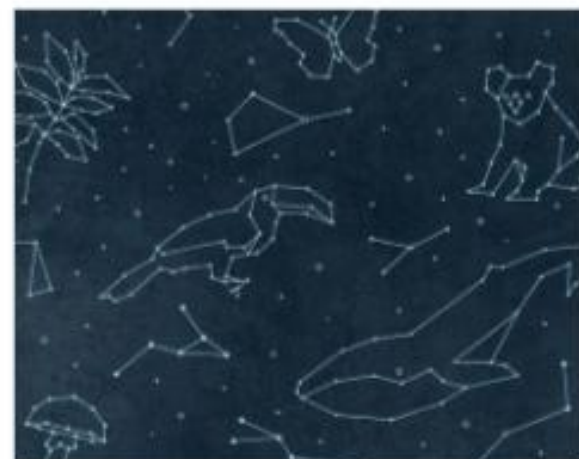
Conservation



Climate change



Human health and One Health



DNA barcoding and metagenomics



Agriculture and food security



Business sector



Scientific collections

Get data How-to Tools Community About

### GBIF and food

GBIF mediates a link to agricultural usage and agricultural policy by helping to identify areas for agricultural crop wild relatives

Data from the GBIF network provides and contributes to various dimensions of knowledge, research and policy in agriculture and food security.

Much of this work focuses on the relationships between domesticated crops and their wild relatives, or crop wild relatives, and how they can shape and inform farming practices in the face of global climate change. The protection of wild species can improve food security and agricultural policy by increasing crop yields and making cultivated species more adaptable and resilient, free from pests and pathogens, and better able to adapt to changes in temperature or precipitation.

Combining climate, soil, and other data with GBIF-mediated data also helps identify changes in regions suitable for different crops in future, high-risk areas for agricultural pests, possible benefits and risks of biofuel production, and relationships and interactions between pollinators.

#### Featured datasets and publishers related to food security and agriculture

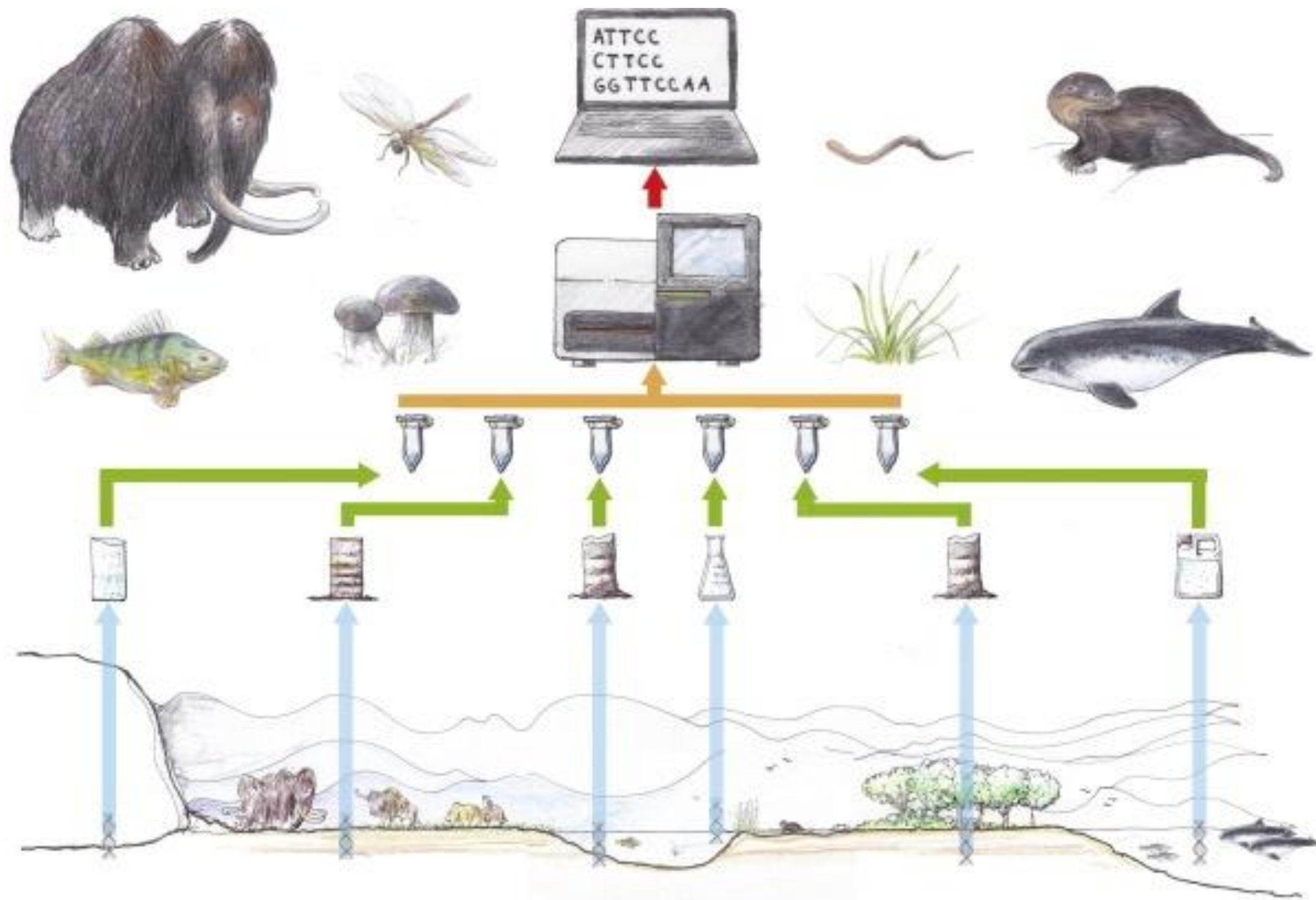
A global database for the distributions of crop wild relatives	Bee Biology and Systematics Laboratory	Malawi Plant Genetic Resources Centre (NPGRC)	Categorization of the wild food plants and their importance for primate development in the natural forest of Alogu (Lama) in southern Bahr.



# eDNA and metabarcoding



# What is eDNA?

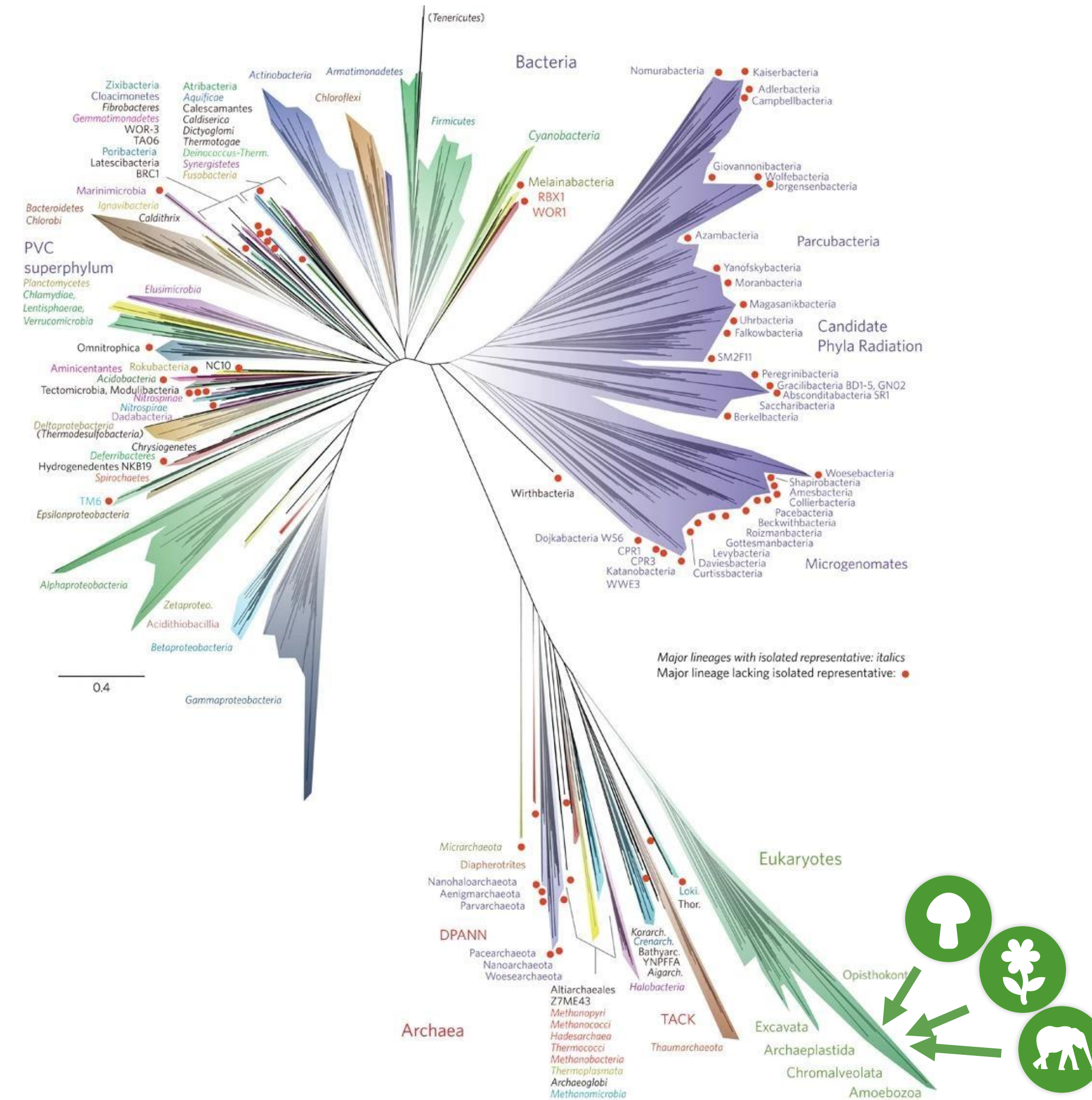


Uncovering “invisible footprints” of organisms



# Why is eDNA Data Essential?

Filling critical taxonomic and geographic gaps for comprehensive biodiversity data



# Collaborating for a global eDNA data future




# The Metabarcoding Data Toolkit

Metabarcoding Data Toolkit

New dataset FAQ thomasgbif

## Metabarcoding Data Toolkit

bridging metabarcoding and biodiversity



How to use it? Got feedback? FAQ

The most recently updated resources are available as an RSS feed

GBIF Metabarcoding Data Toolkit

About the DNA Programme How to use it? FAQ Report a bug Request a feature

Metabarcoding Data Toolkit

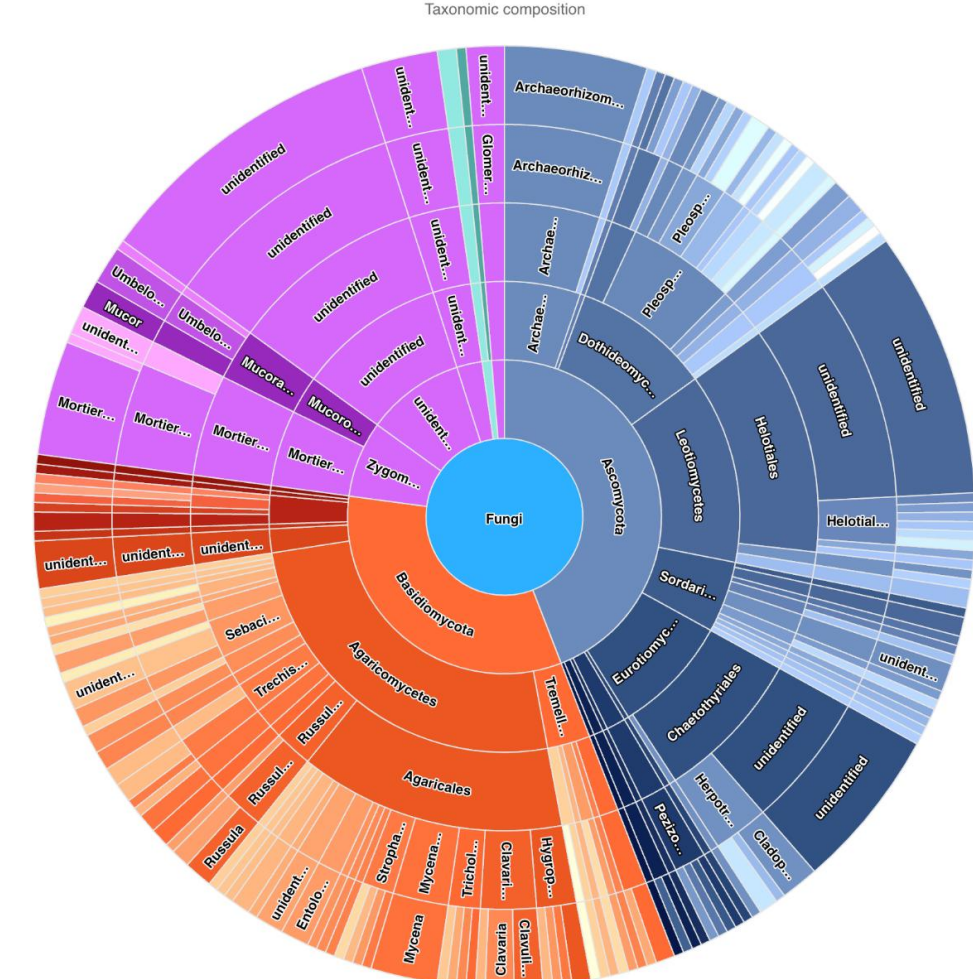
New dataset FAQ thomasgbif

Upload data Map terms Process data **4 Review** Edit metadata Export Publish

Proceed

EM057

Taxonomic composition



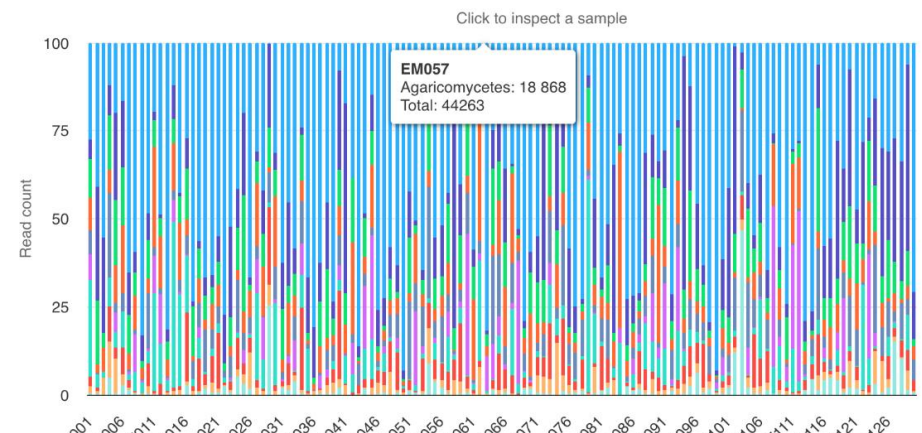
Taxonomy barplot PCoA/MDS plot Most frequent OTUs Least frequent non-singleton OTUs

Relative read abundance

Taxon rank: class

Click to inspect a sample

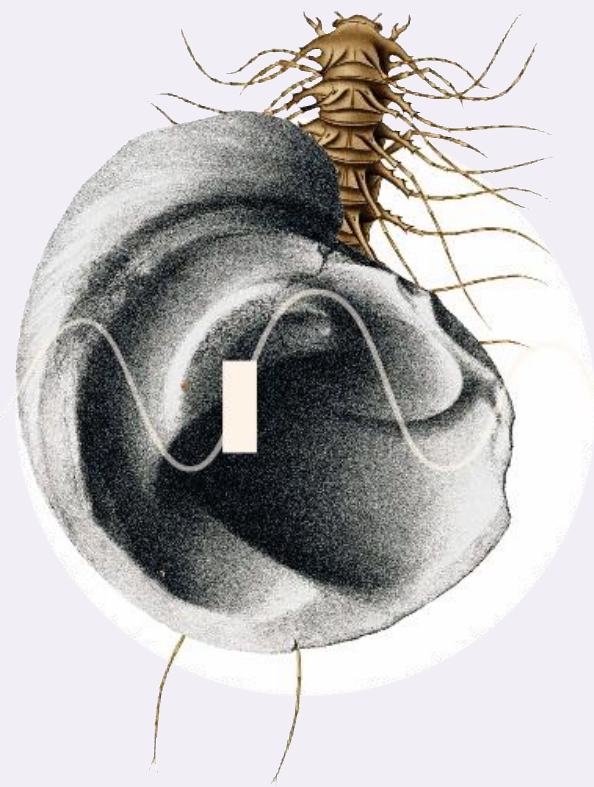
EM057  
Agaricomycetes: 18 868  
Total: 44263



Highcharts.com

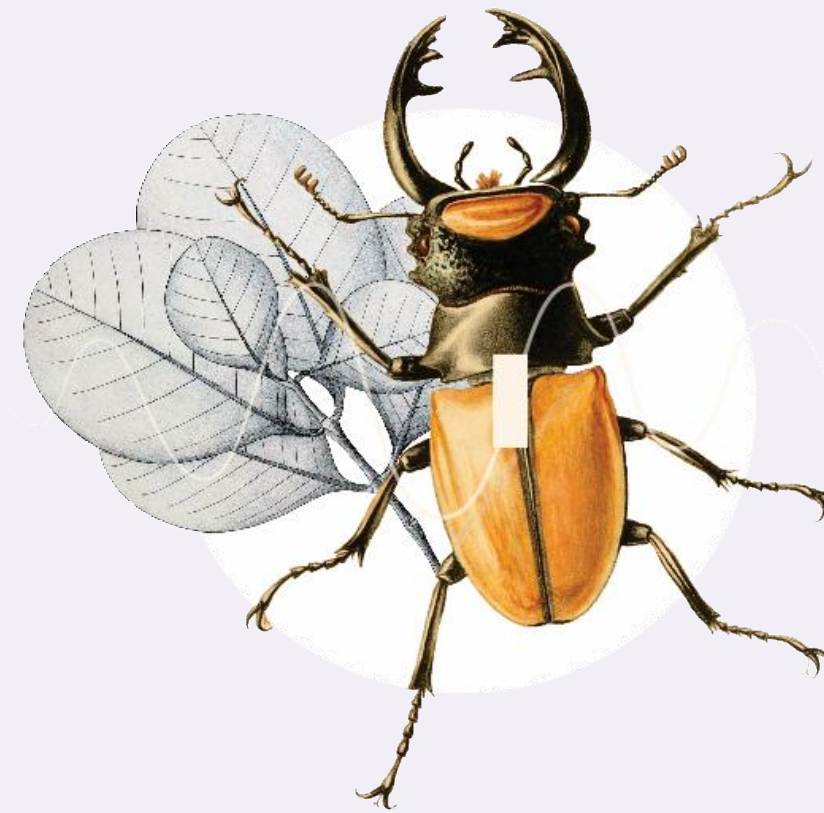


# Metabarcoding data programme



**63**

Datasets



**109,822,958**

Records



**22**

Installations

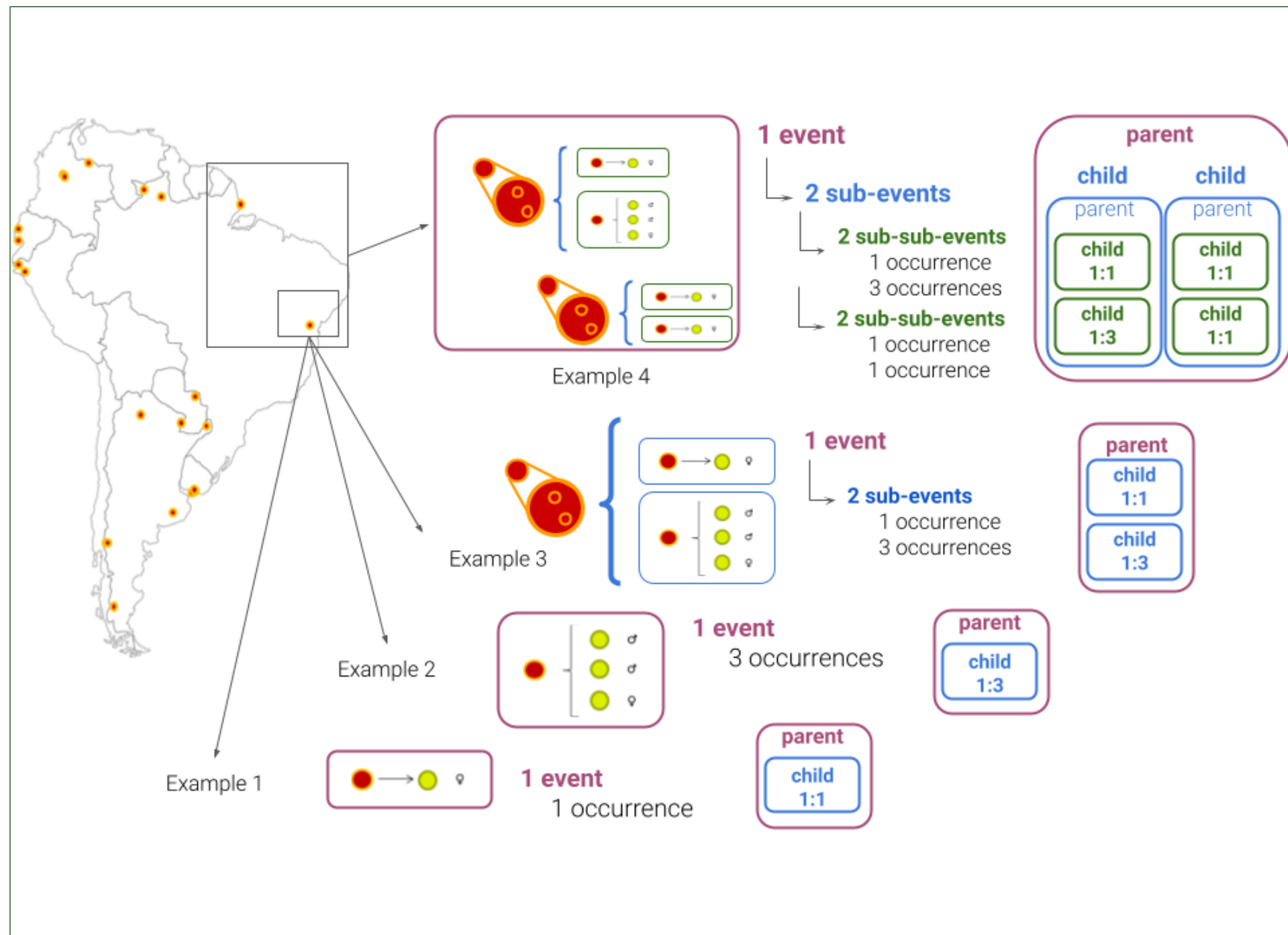


# Surveys and monitoring

Humboldt extension and data model



# What are survey and monitoring data?



- Reported sampling protocol
- Hierarchical w/ variable complexity
- Effort types
  - Bioblitzes, museum expeditions, long-term monitoring initiatives, ...

Figure 4, Sica et al. 2023. Humboldt Extension for Ecological Inventories: User Guide w/ Test IPT Instructions. <https://tinyurl.com/HEuserguide>



# Why are survey and monitoring data important to GBIF?

## 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



- Assessment of biodiversity trends and rates of change
- Impacts
  - Scientific research
  - Policy and management

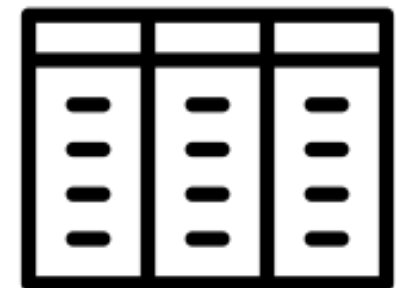


# Science and research

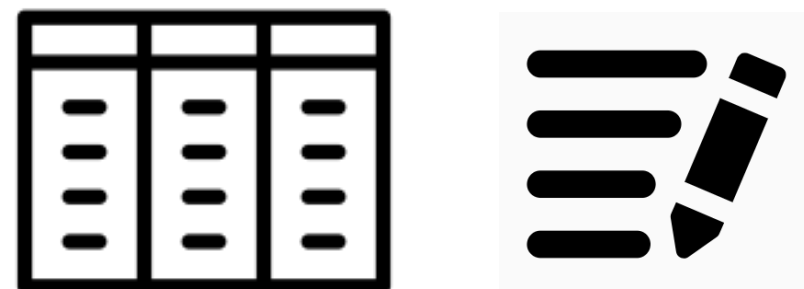


# Data sharing in science: the WHERE choices

Archive



Generalist repository



Data catalogue



Data index



Preservation

Open data

FAIR data

**Save**

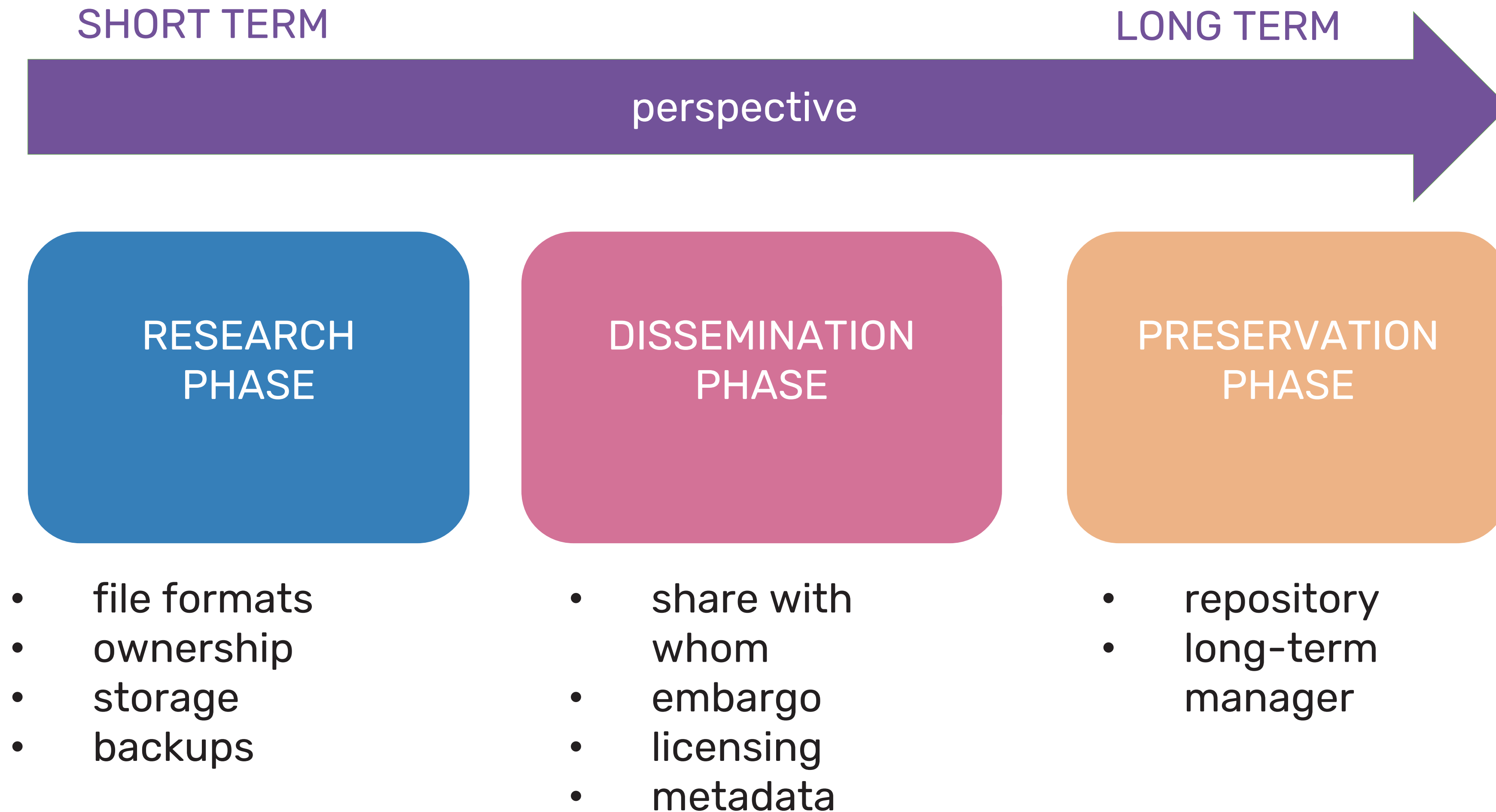
**Minimum  
description**

**Metadata  
standardization**

**Data  
standardization**



# Data sharing in science: the WHEN choices



# Sharing Detailed Research Data Is Associated with Increased Citation Rate

Heather A. Piwowar\*, Roger S. Day, Douglas B. Fridsma

Department of Biomedical Informatics, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania, United States of America

**Background.** Sharing research data provides benefit to the general scientific community, but the benefit is less obvious for the investigator who makes his or her data available. **Principal Findings.** We examined the citation history of 85 cancer microarray clinical trial publications with respect to the availability of their data. The 48% of trials with publicly available microarray data received 85% of the aggregate citations. Publicly available data was significantly ( $p=0.006$ ) associated with a 69% increase in citations, independently of journal impact factor, date of publication, and author country of origin using linear regression. **Significance.** This correlation between publicly available data and increased literature impact may further motivate investigators to share their detailed research data.

Citation: Piwowar HA, Day RS, Fridsma DB (2007) Sharing Detailed Research Data Is Associated with Increased Citation Rate. PLoS ONE 2(3): e308. doi:10.1371/journal.pone.0000308

## INTRODUCTION

Sharing information facilitates science. Publicly sharing detailed research data—sample attributes, clinical factors, patient outcomes, DNA sequences, raw mRNA microarray measurements—with other researchers allows these valuable resources to contribute far beyond their original analysis[1]. In addition to being used to confirm original results, raw data can be used to explore related or new hypotheses, particularly when combined with other publicly available data sets. Real data is indispensable when investigating and developing study methods, analysis techniques, and software implementations. The larger scientific community also benefits: sharing data encourages multiple perspectives, helps to identify errors, discourages fraud, is useful for training new researchers, and increases efficient use of funding and patient population resources by avoiding duplicate data collection.

Believing that that these benefits outweigh the costs of sharing research data, many initiatives actively encourage investigators to make their data available. Some journals, including the *PLoS* family, require the submission of detailed biomedical data to publicly available databases as a condition of publication[2–4]. Since 2003, the NIH has required a data sharing plan for all large funding grants. The growing open-access publishing movement will perhaps increase peer pressure to share data.

However, while the general research community benefits from shared data, much of the burden for sharing the data falls to the study investigator. Are there benefits for the investigators themselves?

A currency of value to many investigators is the number of times their publications are cited. Although limited as a proxy for the scientific contribution of a paper[5], citation counts are often used in research funding and promotion decisions and have even been assigned a salary-increase dollar value[6]. Boosting citation rate is thus a potentially important motivator for publication authors.

In this study, we explored the relationship between the citation rate of a publication and whether its data was made publicly available. Using cancer microarray clinical trials, we addressed the following questions: Do trials which share their microarray data receive more citations? Is this true even within lower profile trials? What other data-sharing variables are associated with an increased citation rate? While this study is not able to investigate causation, quantifying associations is a valuable first step in understanding these relationships. Clinical microarray data provides a useful environment for the investigation: despite being valuable for reuse and extremely costly to collect, is not yet universally shared.

## RESULTS

We studied the citations of 85 cancer microarray clinical trials published between January 1999 and April 2003, as identified in a systematic review by Ntzani and Ioannidis[7] and listed in Supplementary Text S1. We found 41 of the 85 clinical trials (48%) made their microarray data publicly available on the internet. Most data sets were located on lab websites (28), with a few found on publisher websites (4), or within public databases (6 in the Stanford Microarray Database (SMD)[8], 6 in Gene Expression Omnibus (GEO)[9], 2 in ArrayExpress[10], 2 in the NCI GeneExpression Data Portal (GEDP)(gedp.nci.nih.gov); some datasets in more than one location). The internet locations of the datasets are listed in Supplementary Text S2. The majority of datasets were made available concurrently with the trial publication, as illustrated within the WayBackMachine internet archives (www.archive.org/web/web.php) for 25 of the datasets and mention of supplementary data within the trial publication itself for 10 of the remaining 16 datasets. As seen in Table 1, trials published in high impact journals, prior to 2001, or with US authors were more likely to share their data.

The cohort of 85 trials was cited an aggregate of 6239 times in 2004–2005 by 3133 distinct articles (median of 1.0 cohort citation per article, range 1–23). The 48% of trials which shared their data received a total of 5334 citations (85% of aggregate), distributed as shown in Figure 1.

**Academic Editor:** John Ioannidis, University of Ioannina School of Medicine, Greece

**Received:** December 13, 2006; **Accepted:** February 26, 2007; **Published:** March 21, 2007

**Copyright:** © 2007 Piwowar et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** HAP was supported by NLM Training Grant Number 5T15-LM007059-19. The NIH had no role in study design, data collection or analysis, writing the paper, or the decision to submit it for publication. The publication contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIH.

**Competing Interests:** The authors have declared that no competing interests exist.

\* To whom correspondence should be addressed. E-mail: hpiwowar@cbmi.pitt.edu

# The research data lifecycle

Generate / **Access**

(re)Organize

**Modify**

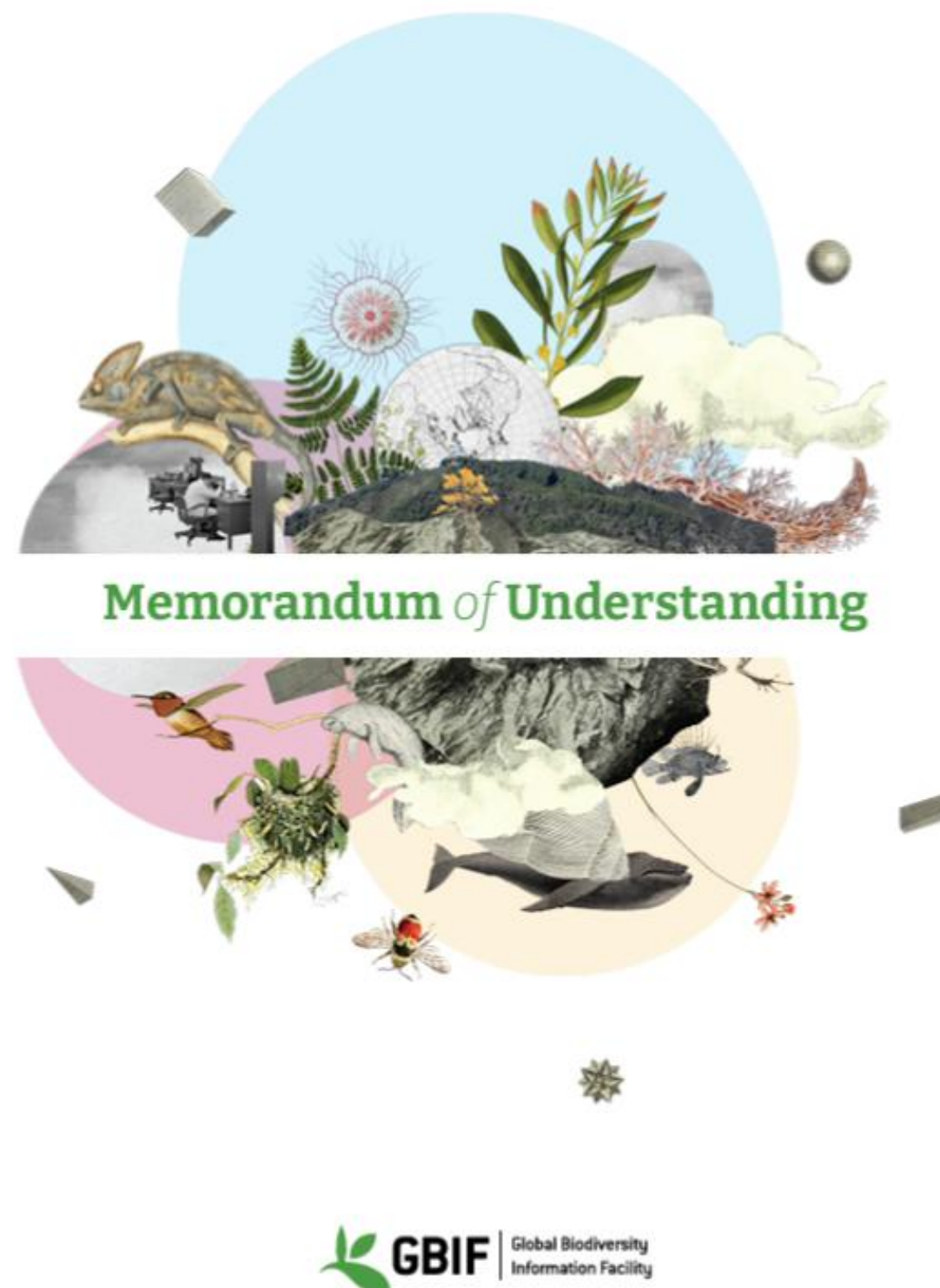
**Analyze**

**Archive**

**Cite**



# GBIF Memorandum of Understanding



The signatories to this non-binding Memorandum of Understanding (MOU), being countries, economies, inter-governmental or international organizations, other organizations with an international scope, or entities designated by them, have decided that a **co-ordinated international scientific effort** is needed to enable users throughout the world to openly share and put to use vast quantities of global biodiversity data, thereby **advancing scientific research** in many disciplines, promoting technological and sustainable development, facilitating the conservation of biodiversity and the equitable sharing of its benefits, and enhancing the quality of life of members of society. The importance of making biodiversity data openly available to all countries and individuals is underscored by various international agreements.

Noting that GBIF was established in March 2001, and that the first and second MOU for GBIF had each a duration of five years (2001-2006/2007-2011), the signatories to this Memorandum of Understanding hereby express their intention either to continue their existing Participation in GBIF or to become new Participants in **GBIF as a mechanism of technical and scientific international co-operation.**

The Participants intend to encourage co-operation amongst themselves in the implementation of GBIF and in the development of joint work programmes in areas of mutual interest with the Secretariat of the Convention on Biological Diversity and other appropriate bodies and initiatives to avoid duplication and to benefit from existing resources and expertise.



# 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



# THANK YOU

Dmitry Schigel | Lead scientific officer

@dschigel

adlignum.com



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



