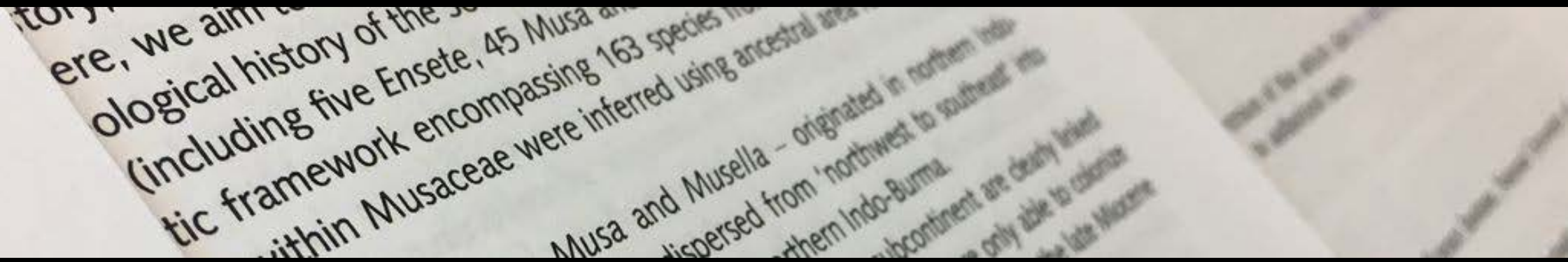




Data integration enables biodiversity synthesis

A literature review of GBIF-enabled studies



*Mason Heberling*¹, Joe Miller², Daniel Noesgaard², Scott Weingart³, Dmitry Schigel²

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

²Global Biodiversity Information Facility Secretariat, Copenhagen, Denmark

³Carnegie 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.org

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

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. *ugandensis* 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 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.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 species...

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!

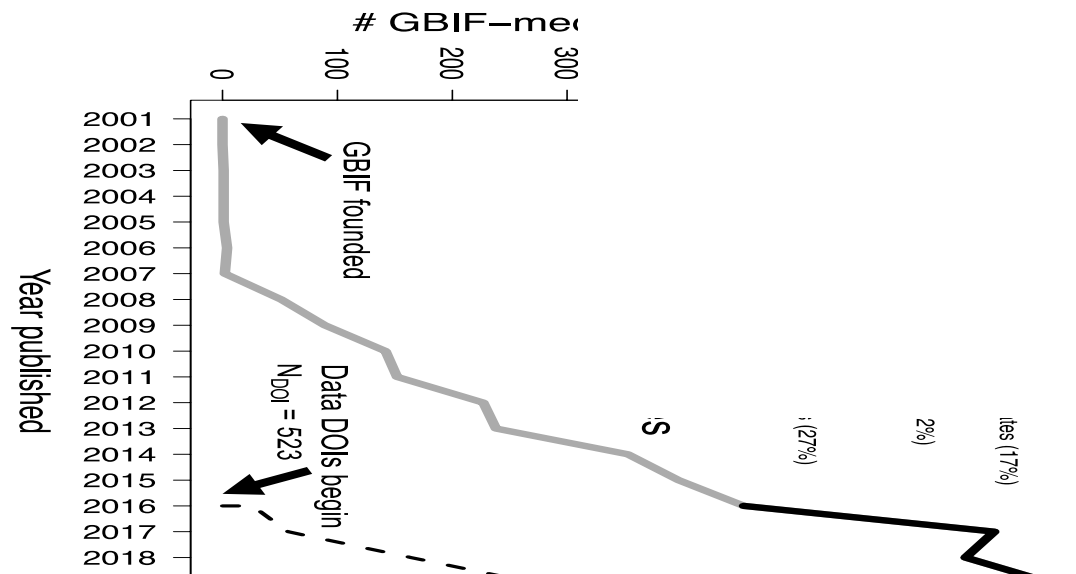
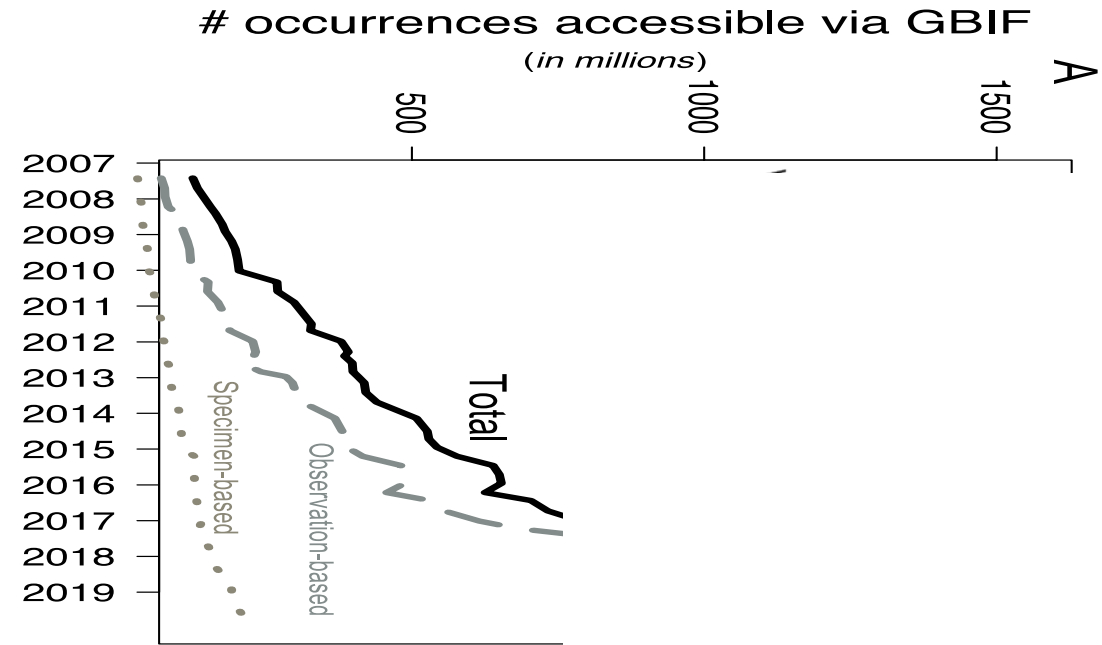
Rapid growth in data use.

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!

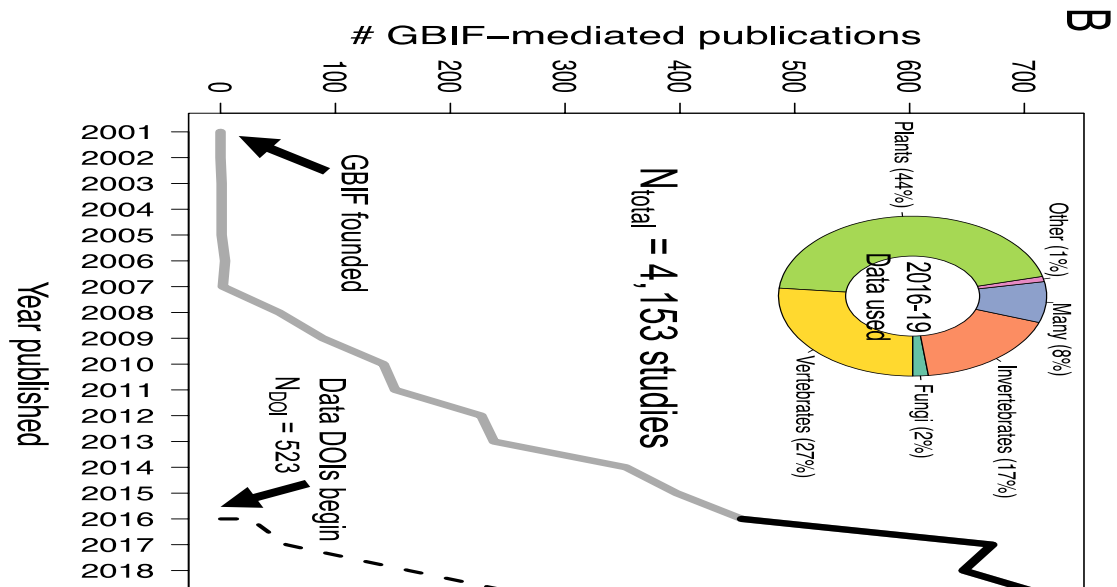
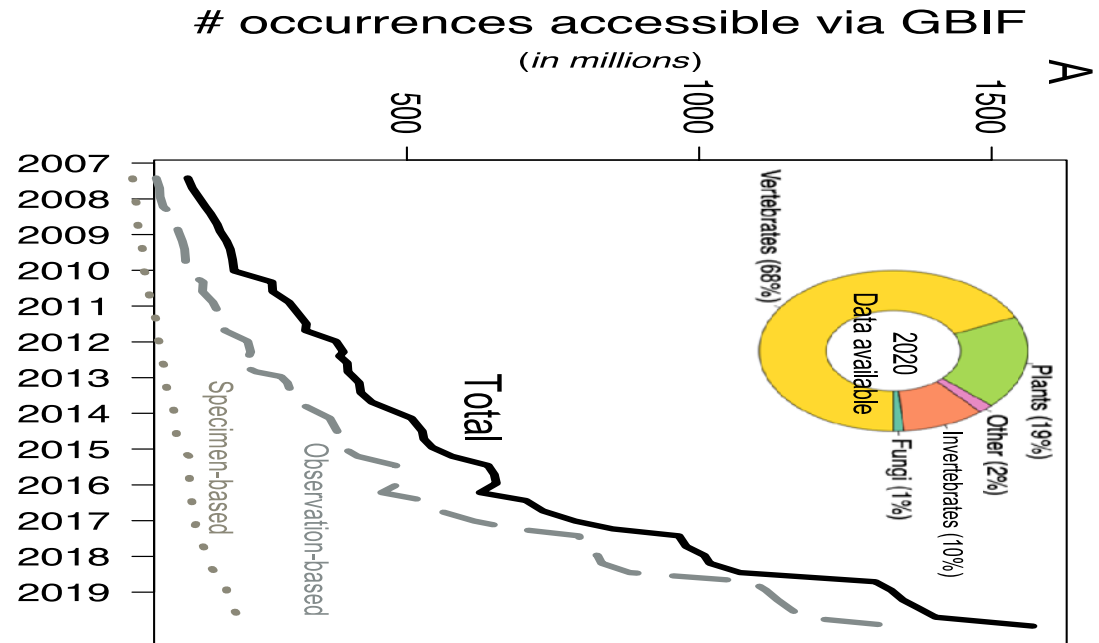
Rapid growth in data use.

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:

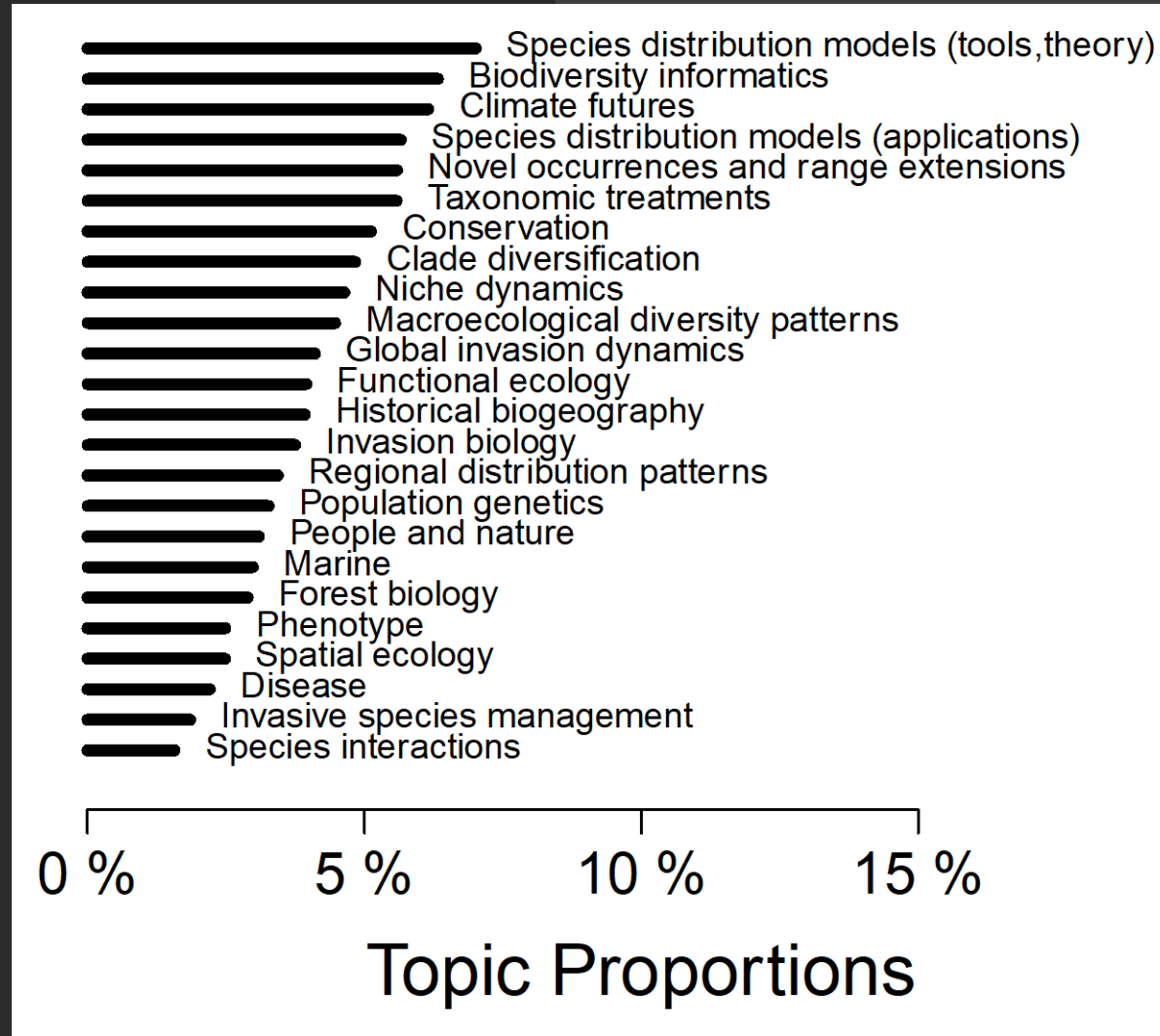
Computational text analysis of 4,035 studies using GBIF.

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:

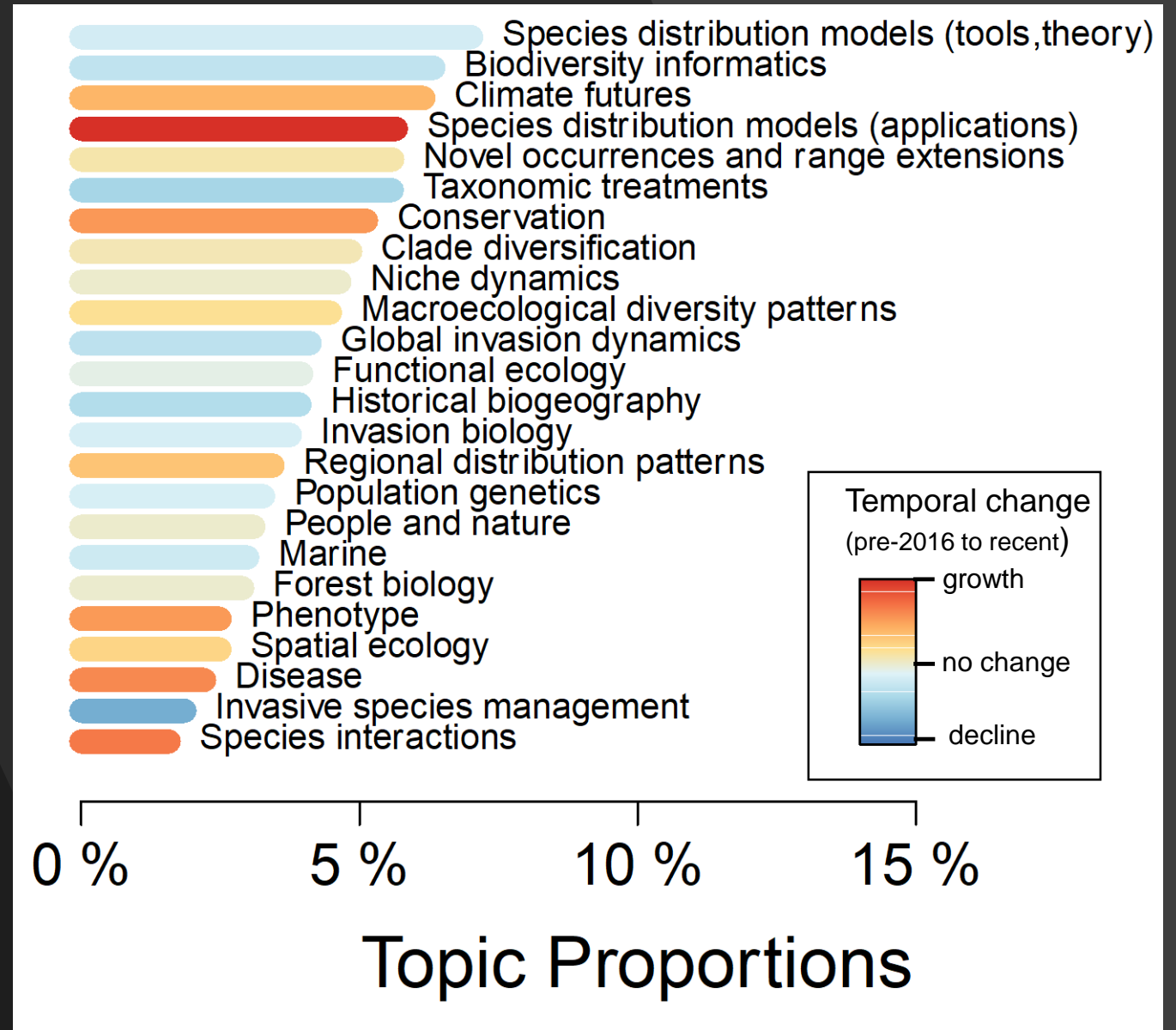
Increasing topics +:

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!

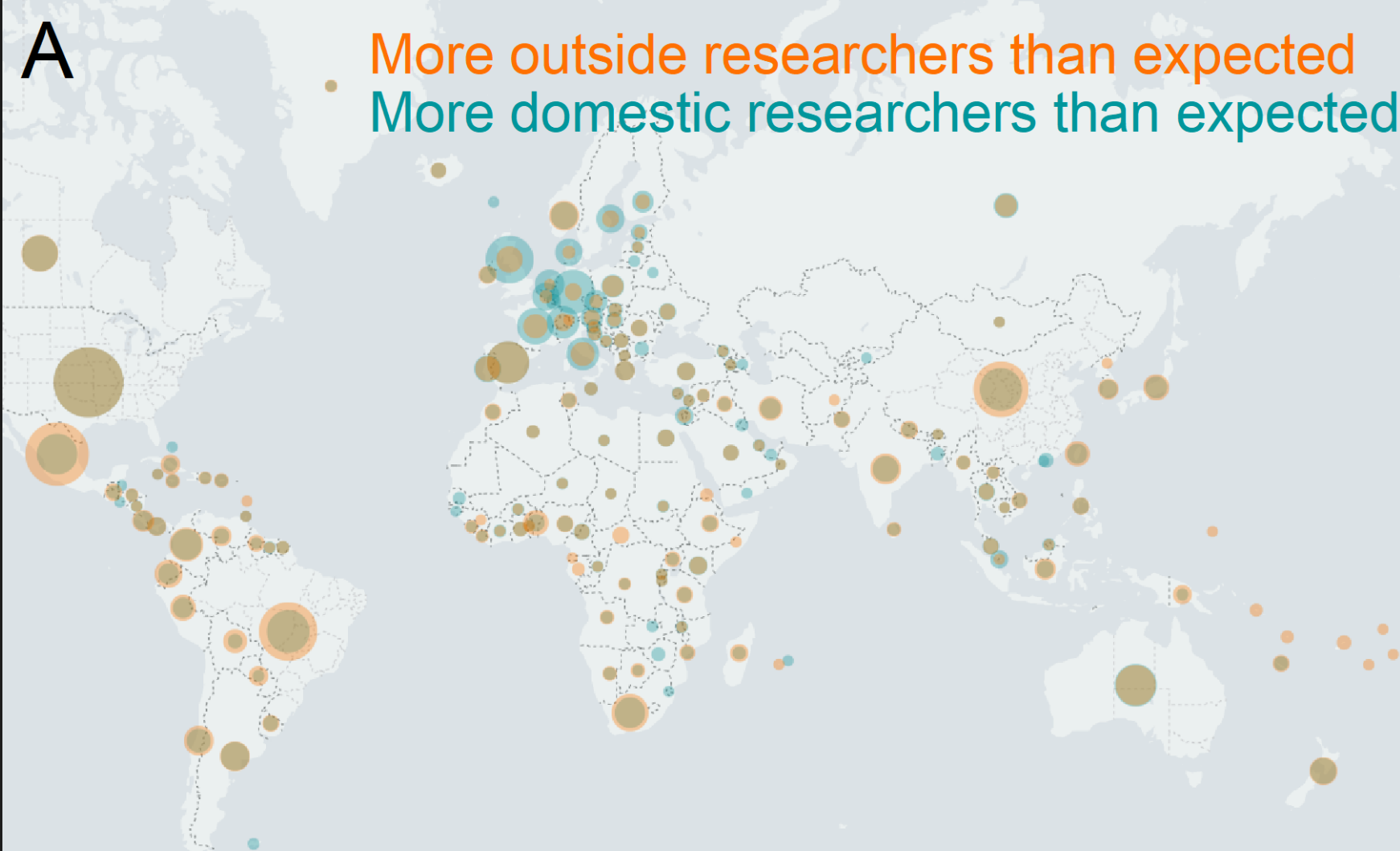


The who and where

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

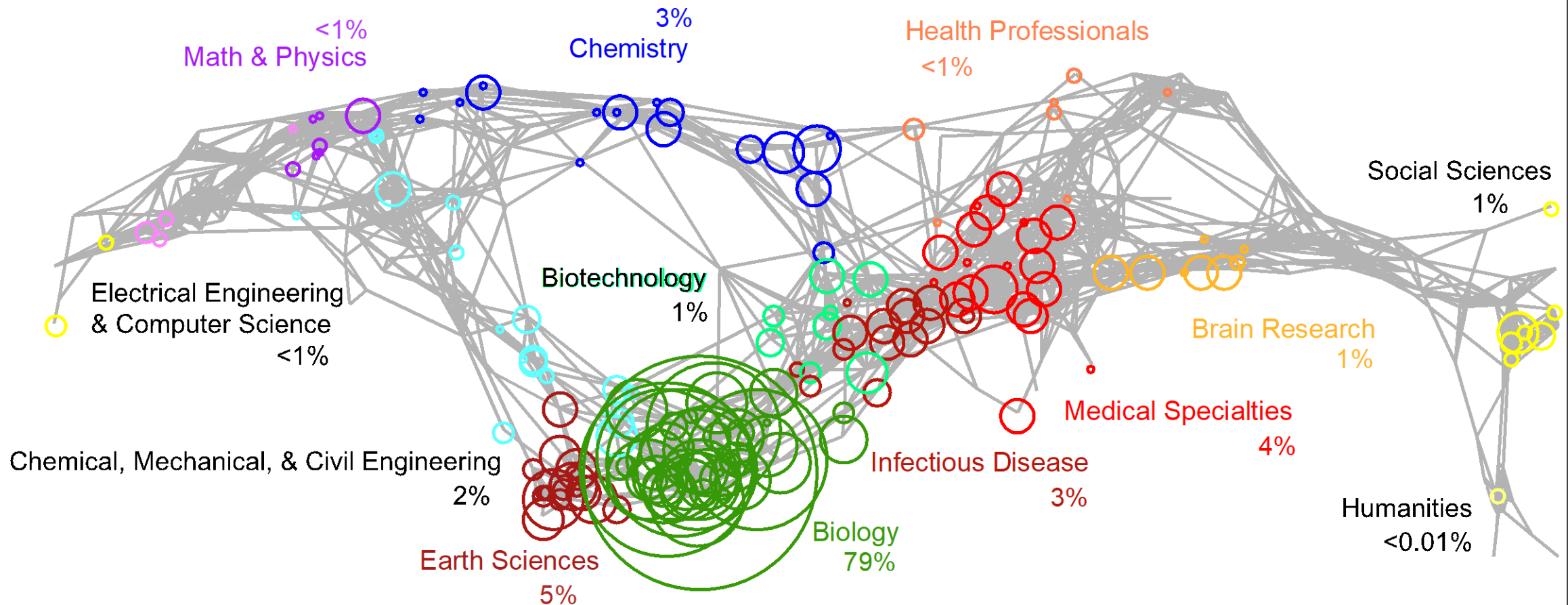
A

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 Heberling¹, Joe Miller², Daniel Noesgaard², Scott Weingart³, Dmitry Schigel²

¹Carnegie Museum of Natural History Pittsburgh, PA USA

²Global Biodiversity Information Facility Secretariat, Copenhagen, Denmark

³Carnegie Mellon University, Pittsburgh, PA USA

GB27 Meeting, October 2020