Global Biodiversity Information Facility a FAIR data infrastructure and network

Dmitry Schigel | Scientific officer

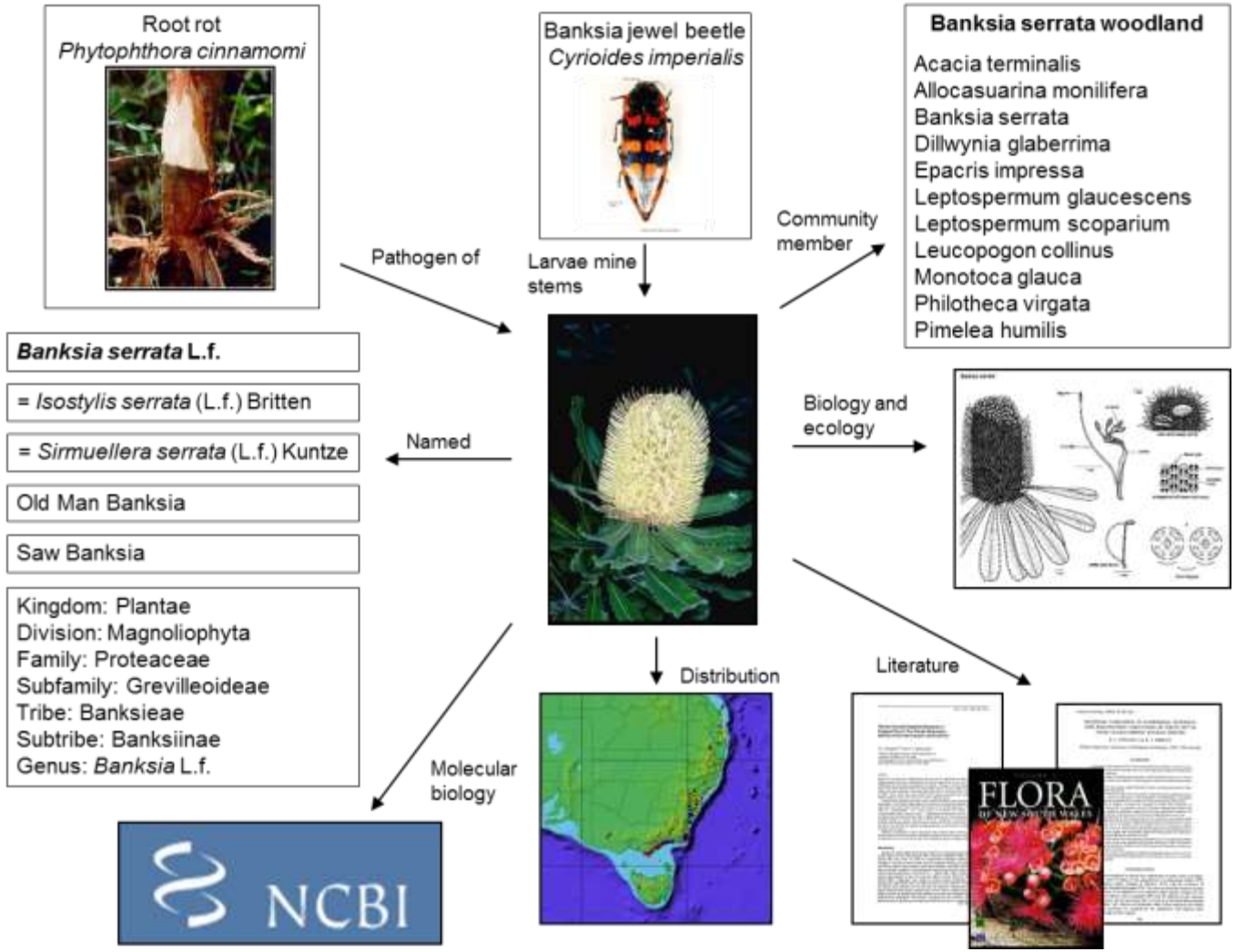
GBIF Global Biodiversity Information Facility

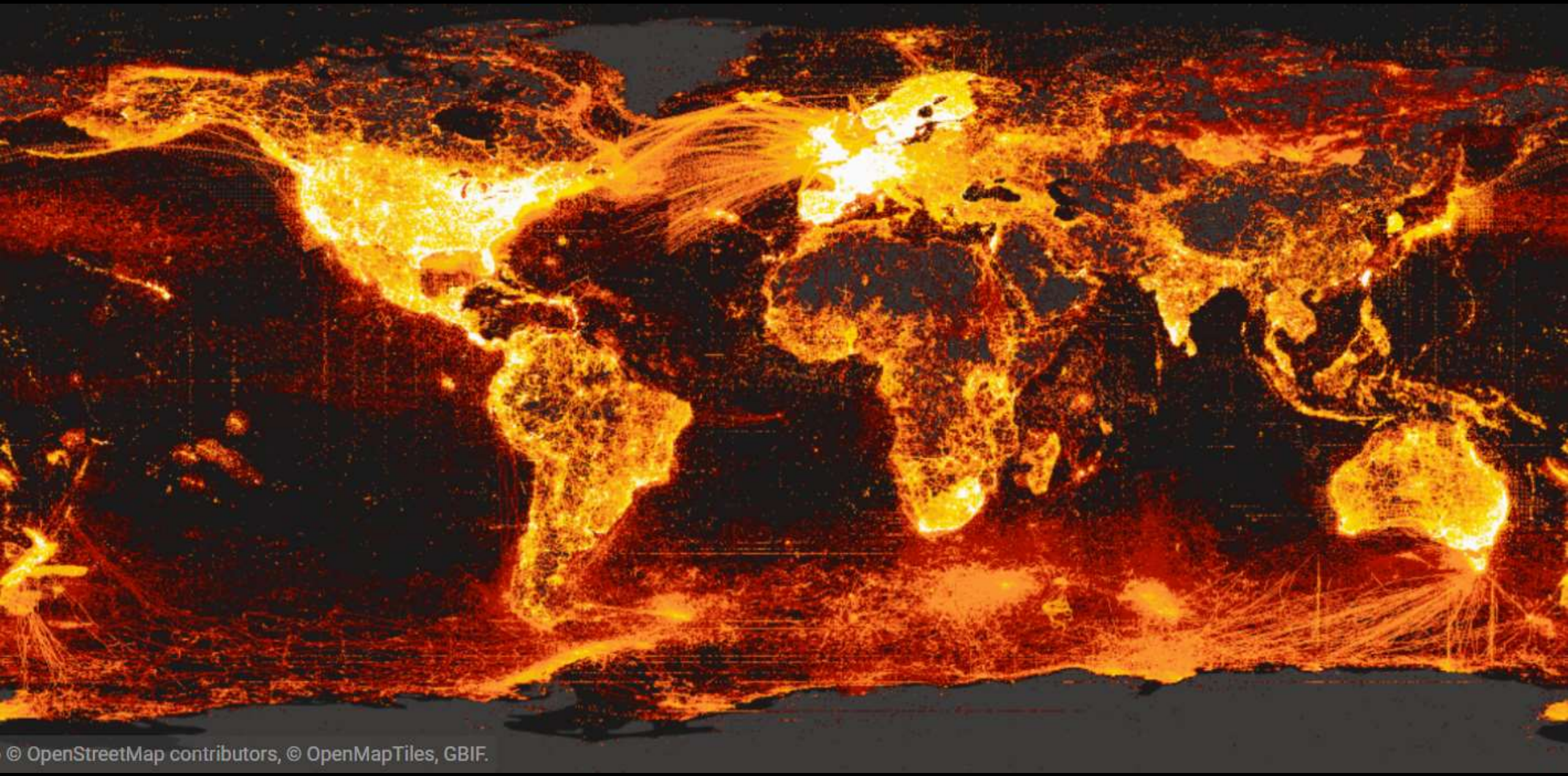
DeiC konference 2021

Outline

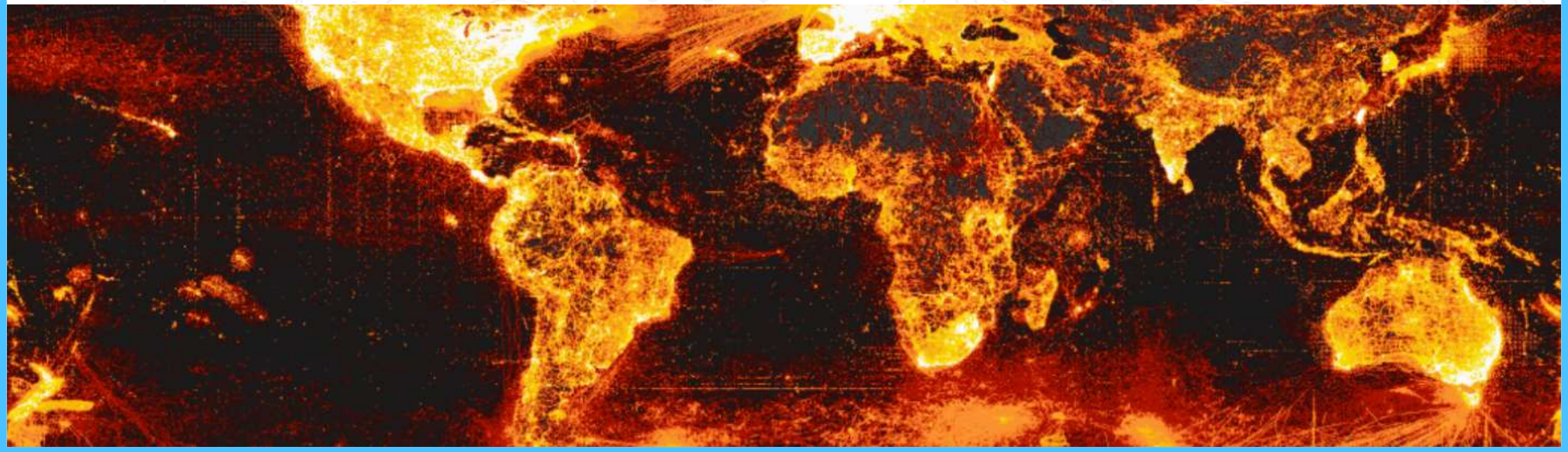
- **GBIF**: Global Biodiversity Information Facility
- **F:** many faces of discoverability
- **A:** data access for newcomers and for pro
- **I:** data standards and data model
- **R:** DOI based data citation

Biodiversity information




© OpenStreetMap contributors, © OpenMapTiles, GBIF.

GBIF: Global Biodiversity Information Facility



GBIF: GLOBAL BIODIVERSITY INFORMATION FACILITY

Key facts

Intergovernmental open data infrastructure

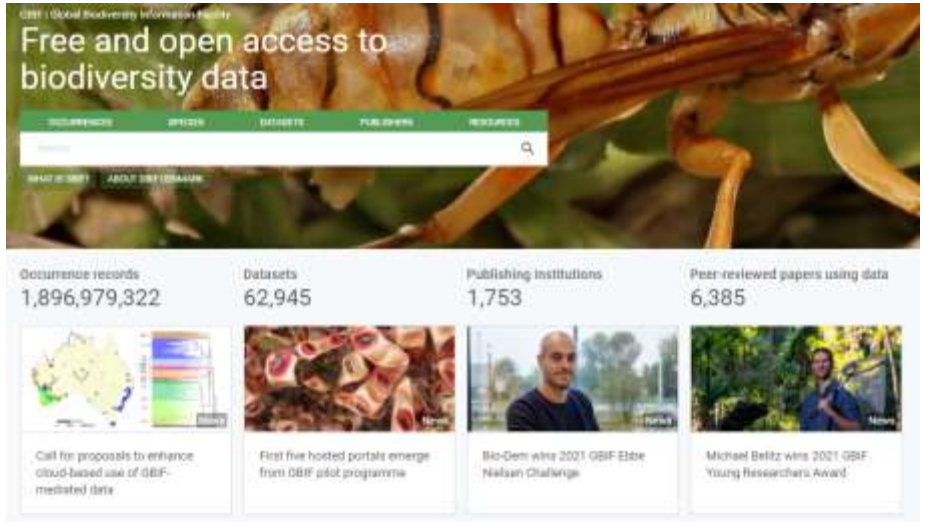
Established in 2001, OECD recommendation

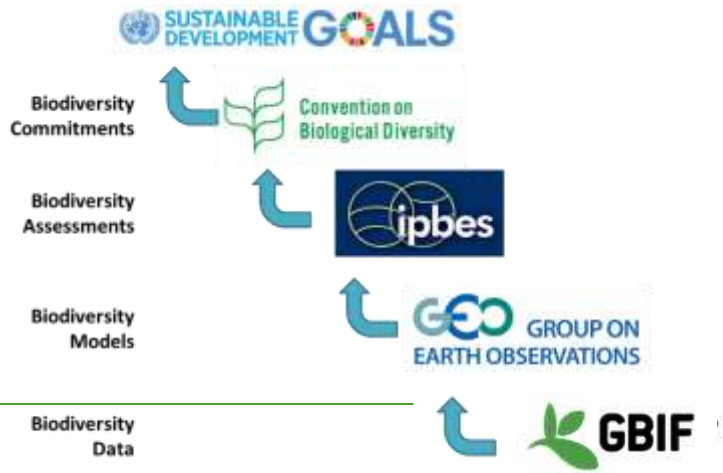
Network for **free and open access** to biodiversity data from all countries

Voluntary memorandum of understanding

Funds: governments of the participant countries

102 participants

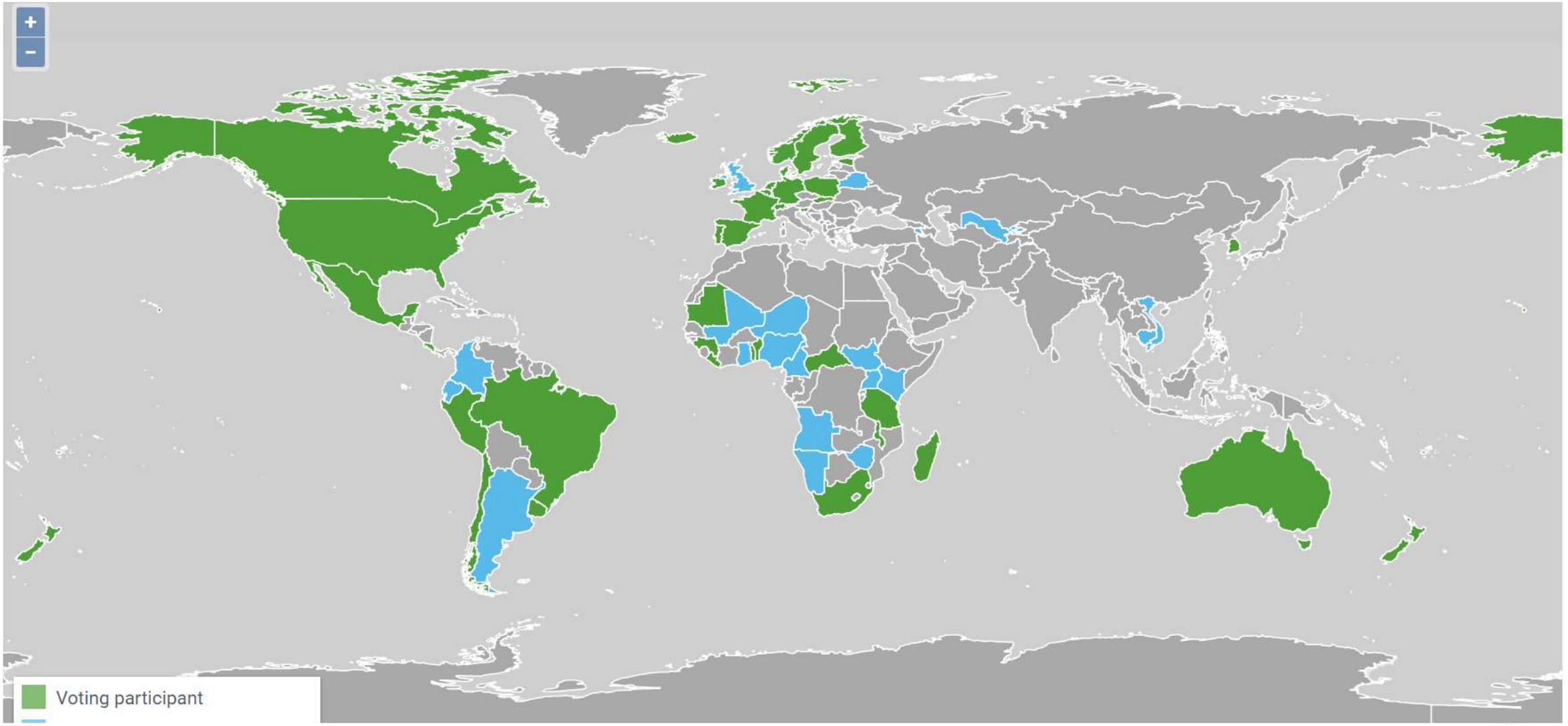




https://www.gbif.org/the-gbif-network

GBIF PARTICIPANT NETWORK

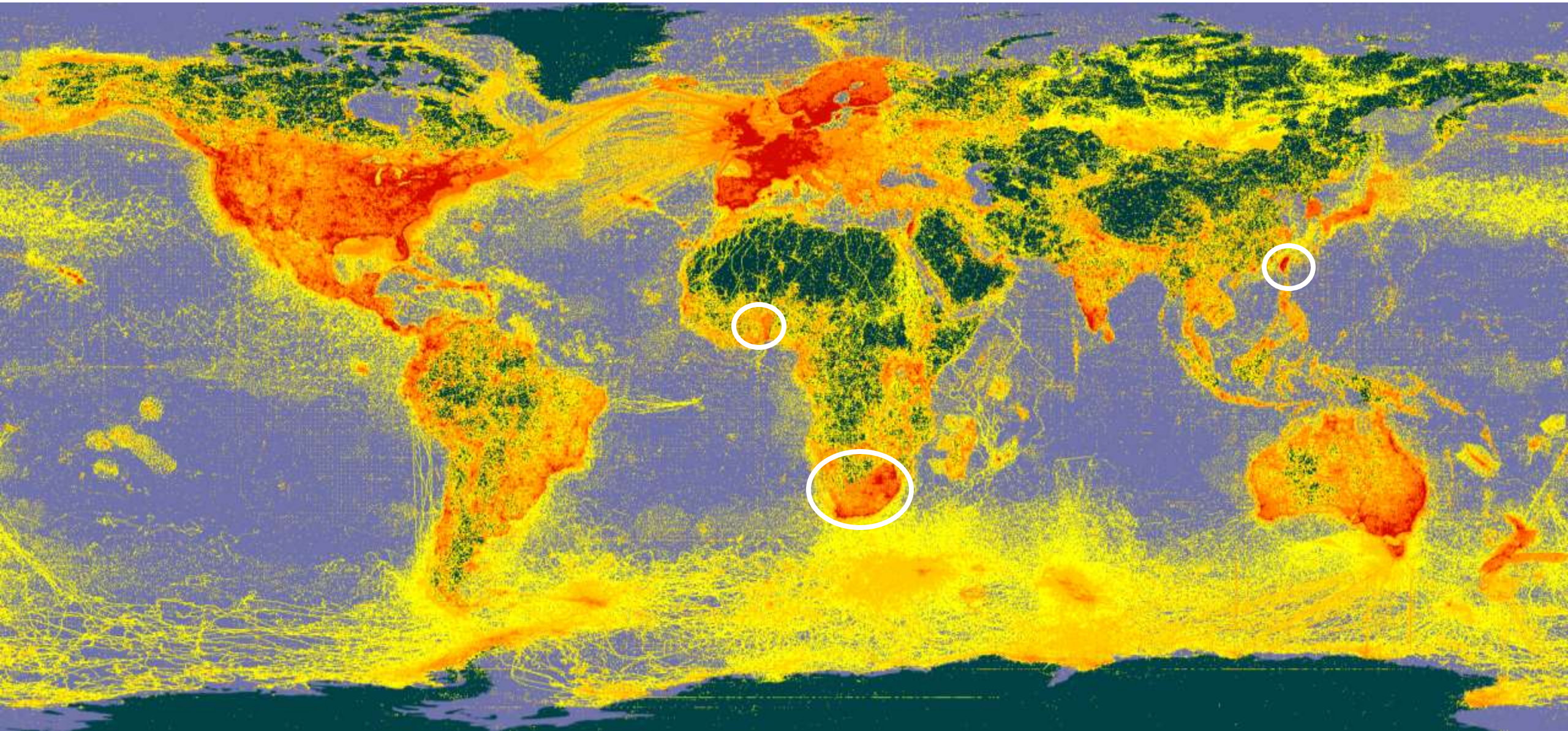
23 September 2021



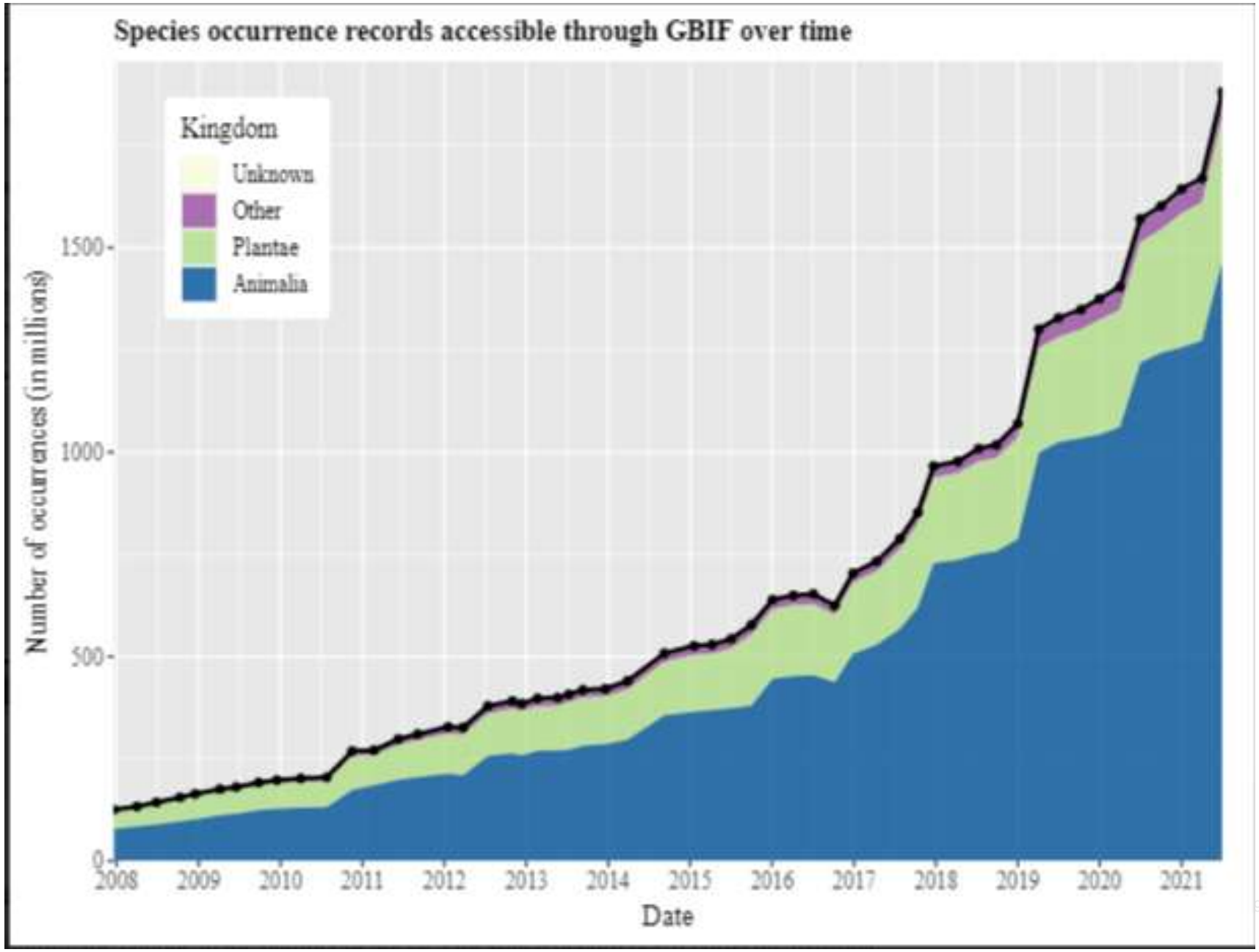
https://www.gbif.org/the-gbif-network

GBIF

DATA FROM THE GBIF NETWORK 30 June 2021



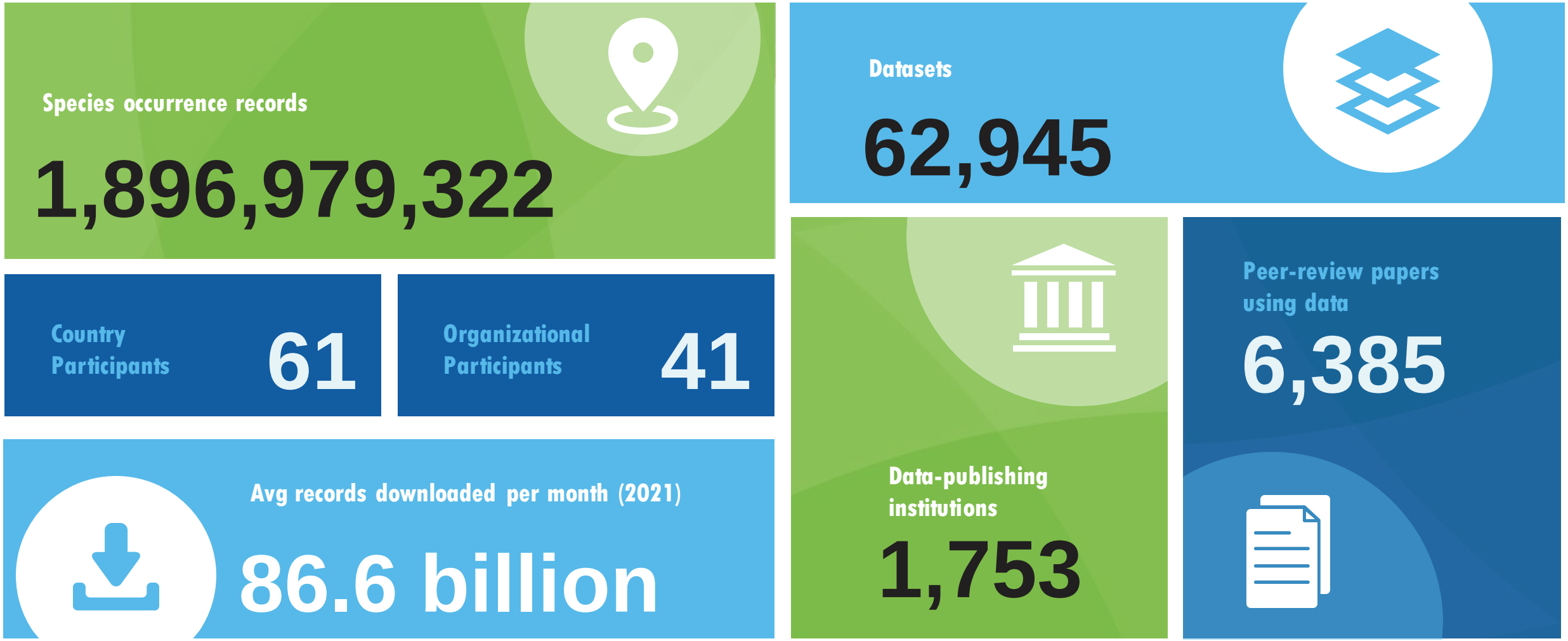
Data published through GBIF.org



data availability

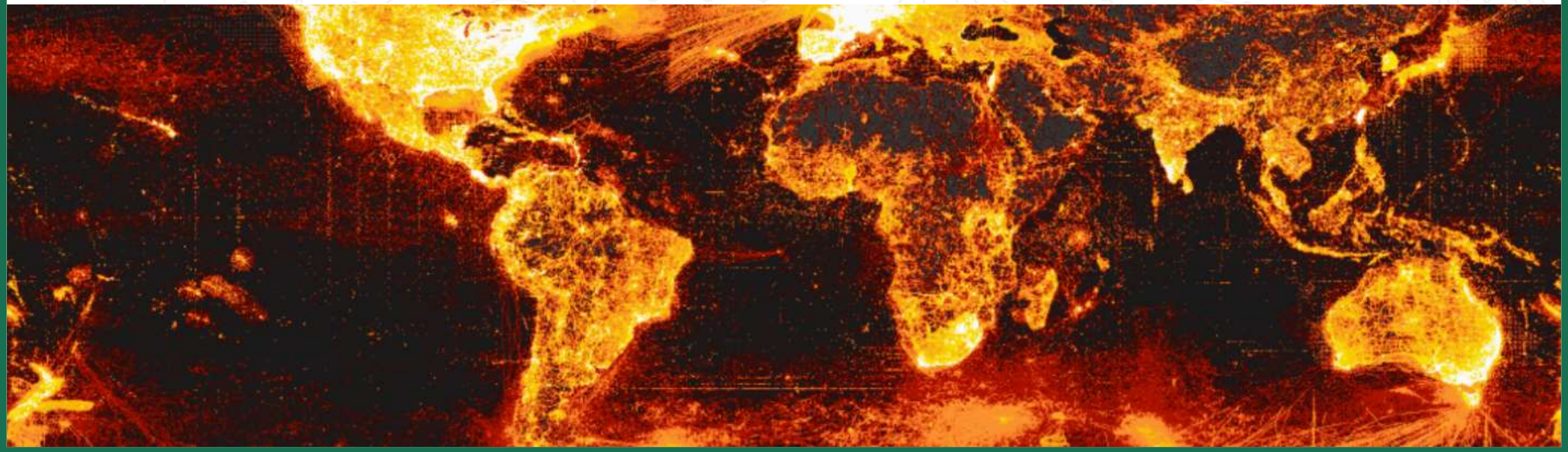
www.gbif.org/analytics/global

BY THE NUMBERS | 2 NOVEMBER 2021



GBIF

FINDABLE: MANY FACES OF DISCOVERABILITY



LET'S GOOGLE IT

#1 Wikipedia

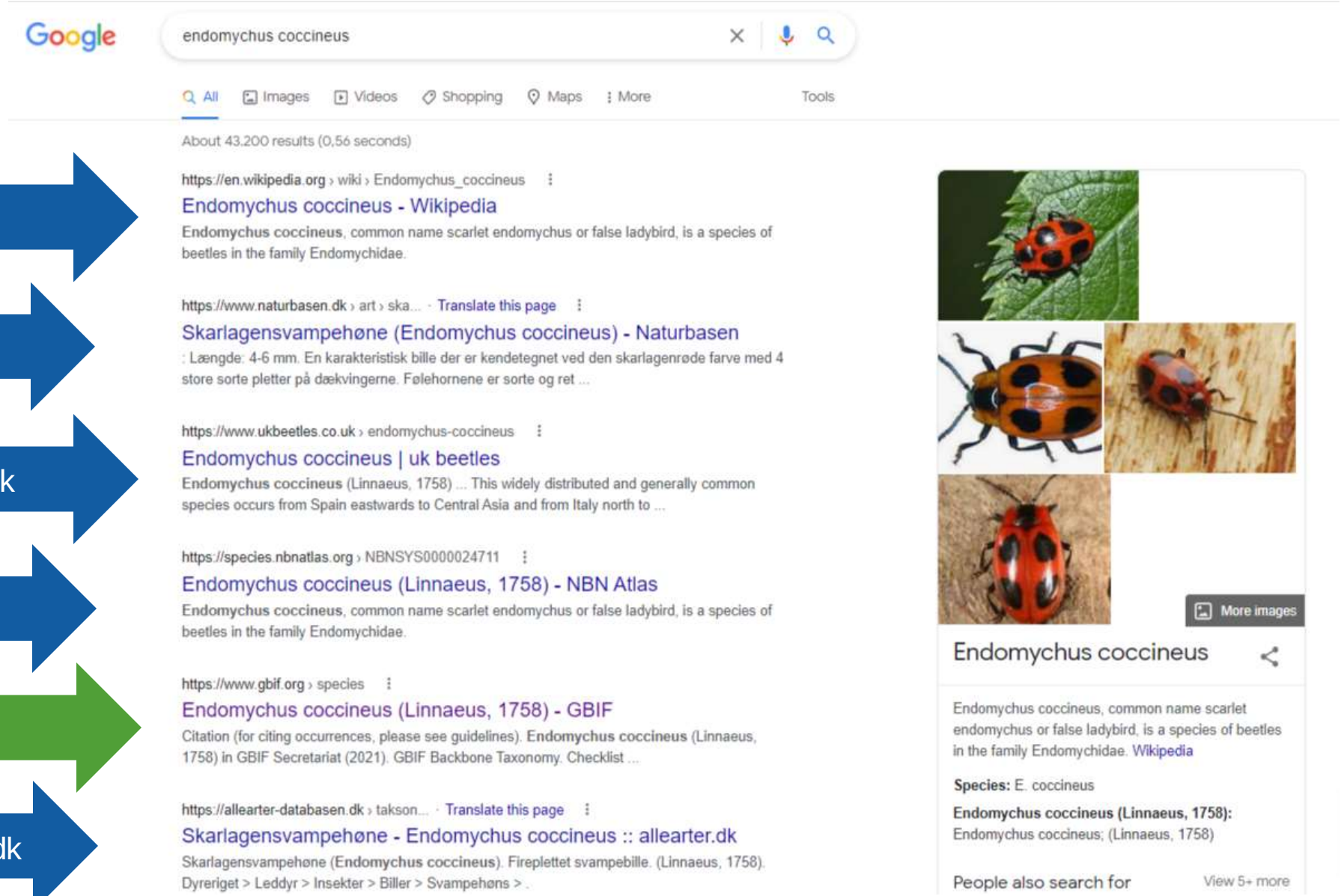
#2 naturbasen.dk

#3 ukbeetles.co.uk

#4 nbnatlas.org

#5 gbf.org

#6 allearter-databasen.dk

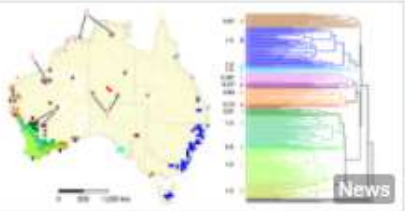


GBIF.ORG

Occurrence records
1,896,979,322

Datasets
62,945

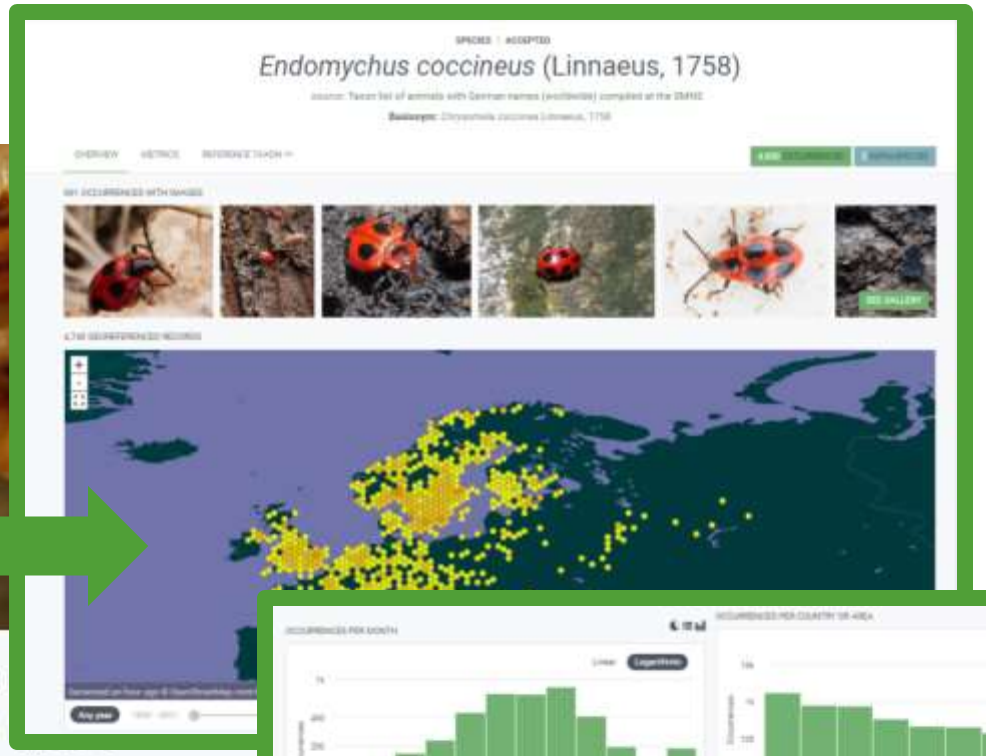
Publishing institutions
1,753

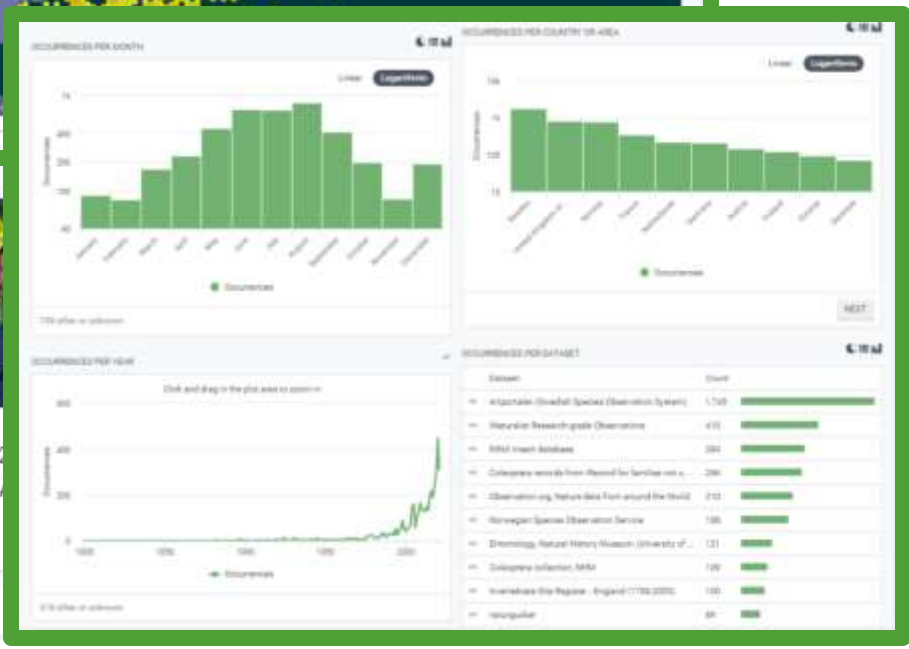


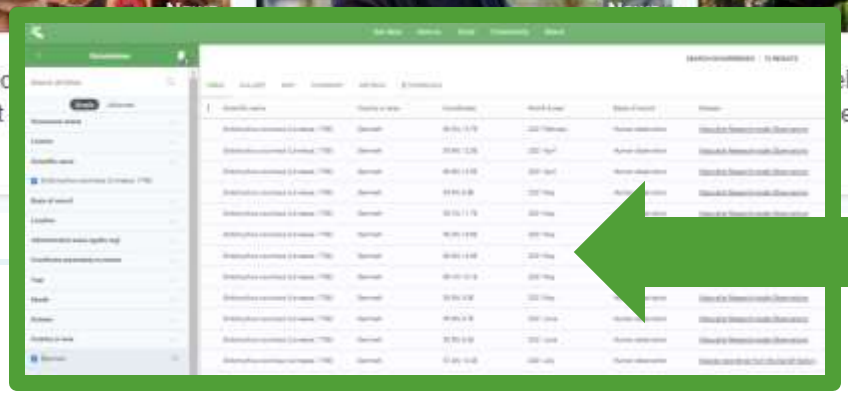
Call for proposals to enhance cloud-based use of GBIF-mediated data

First five hosted
from GBIF pilot

litz wins 2
earchers



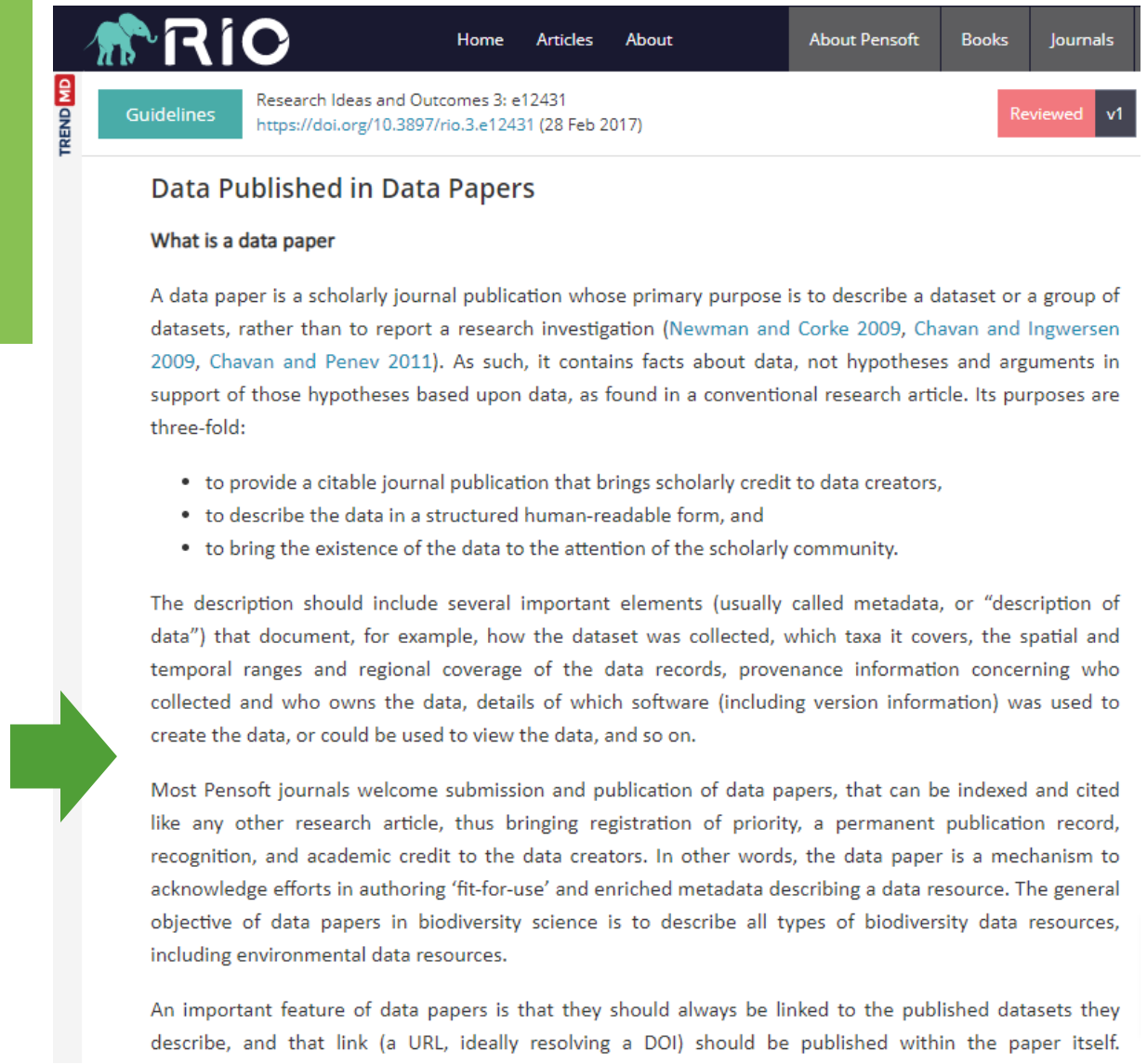




GBIF

DATA PAPERS

- citable journal publication, scholarly credit to data creators
- structured human-readable data descriptor
- brings data existence to scholarly community



RIO Home Articles About About Pensoft Books Journals

TRENDMD

Guidelines

Research Ideas and Outcomes 3: e12431
https://doi.org/10.3897/rio.3.e12431 (28 Feb 2017)

Reviewed v1

Data Published in Data Papers

What is a data paper

A data paper is a scholarly journal publication whose primary purpose is to describe a dataset or a group of datasets, rather than to report a research investigation (Newman and Corke 2009, Chavan and Ingwersen 2009, Chavan and Penev 2011). As such, it contains facts about data, not hypotheses and arguments in support of those hypotheses based upon data, as found in a conventional research article. Its purposes are three-fold:

- to provide a citable journal publication that brings scholarly credit to data creators,
- to describe the data in a structured human-readable form, and
- to bring the existence of the data to the attention of the scholarly community.

The description should include several important elements (usually called metadata, or "description of data") that document, for example, how the dataset was collected, which taxa it covers, the spatial and temporal ranges and regional coverage of the data records, provenance information concerning who collected and who owns the data, details of which software (including version information) was used to create the data, or could be used to view the data, and so on.

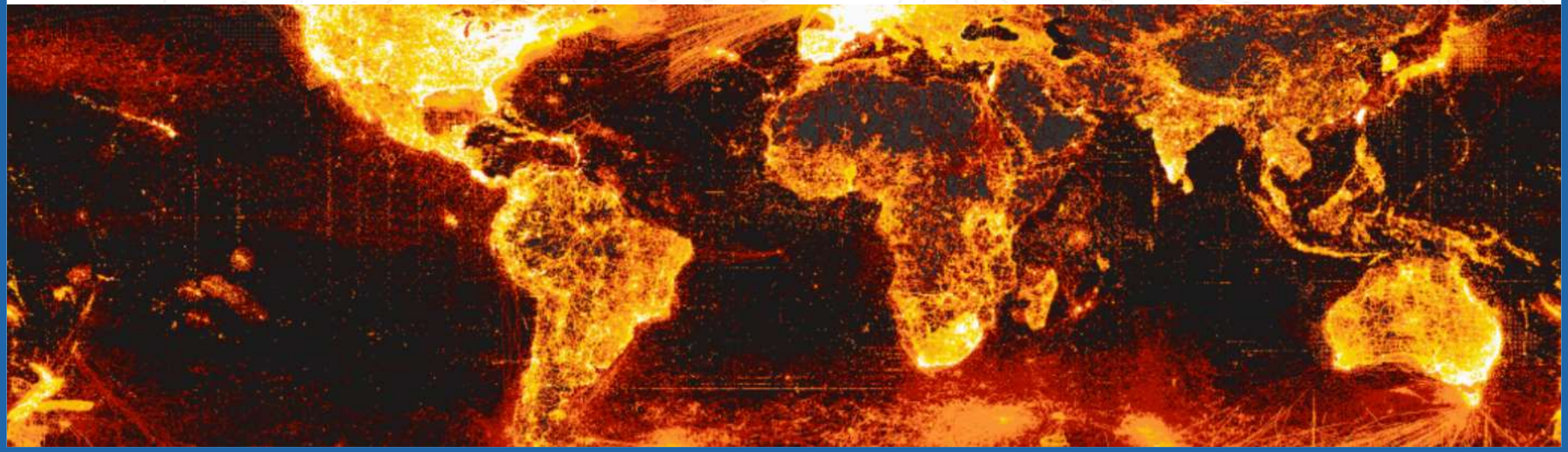
Most Pensoft journals welcome submission and publication of data papers, that can be indexed and cited like any other research article, thus bringing registration of priority, a permanent publication record, recognition, and academic credit to the data creators. In other words, the data paper is a mechanism to acknowledge efforts in authoring 'fit-for-use' and enriched metadata describing a data resource. The general objective of data papers in biodiversity science is to describe all types of biodiversity data resources, including environmental data resources.

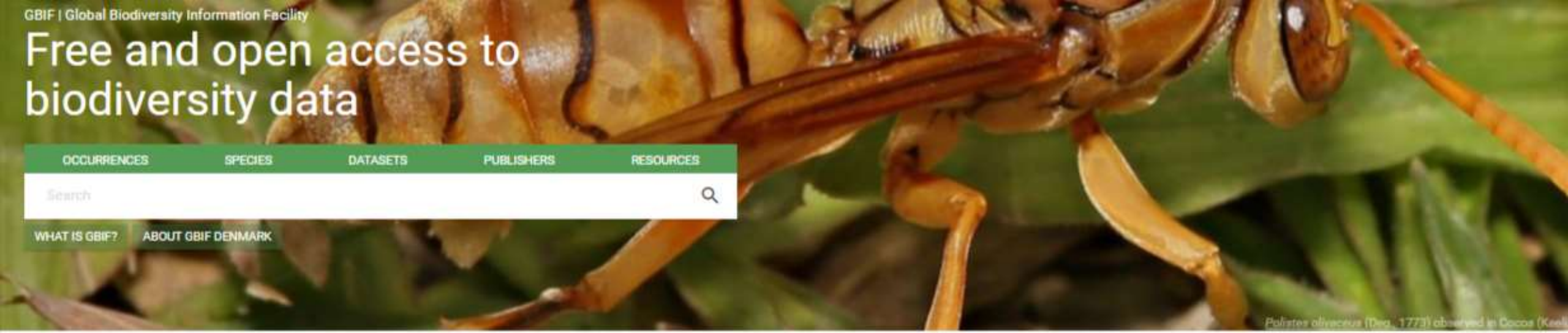
An important feature of data papers is that they should always be linked to the published datasets they describe, and that link (a URL, ideally resolving a DOI) should be published within the paper itself.

Penev et al 2017: Strategies and guidelines for scholarly publishing of biodiversity data https://doi.org/10.3897/rio.3.e12431

GBIF

A: DATA ACCESS FOR NEWCOMERS AND FOR PRO

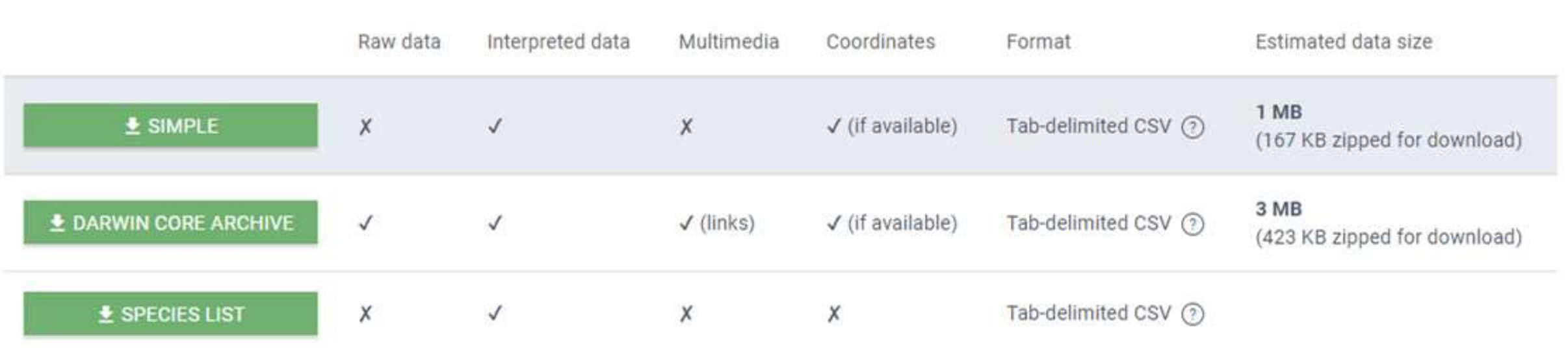




DATA ACCESS

User access points to GBIF-mediated data

1. www.gbif.org = user friendly with large number of filters and readily available metrics. Lacks some functionality. No programming skills necessary.
2. R - https://cloud.r-project.org/ - a number of packages including rgbif and coordinatecleaner for data analysis, processing and visualisation
3. Application Programming Interface (API) – Provides access to GBIF databases in a safe way. Allows GBIF.org and r packages to function.
4. Microsoft Azure– hosting by Microsoft AI for Earth allows for use of occurrences in combination with other environmental layers and not need to upload any of it to the Azure.



	Raw data	Interpreted data	Multimedia	Coordinates	Format	Estimated data size
SIMPLE	✗	✓	✗	✓ (if available)	Tab-delimited CSV	**1 MB** (167 KB zipped for download)
DARWIN CORE ARCHIVE	✓	✓	✓ (links)	✓ (if available)	Tab-delimited CSV	**3 MB** (423 KB zipped for download)
SPECIES LIST	✗	✓	✗	✗	Tab-delimited CSV	

DATA DOWNLOADS

Data can be downloaded in three formats

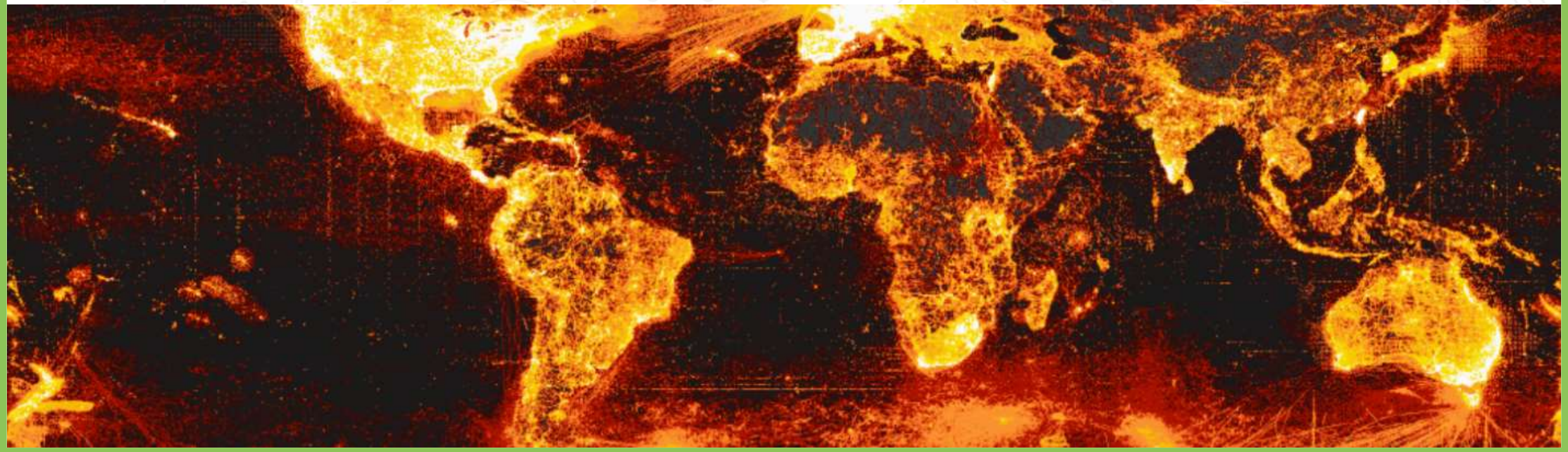
Simple: Tab delimited CSV. Only contains the data after GBIF interpretation. No multimedia included. More information about CSV

Darwin Core Archive: The Darwin Core Archive (DwC-A) contains both the original data as publisher provided it and the GBIF interpretation. Links (but not files) to multimedia included. More information about DwC-A

Species list: Tab delimited CSV with the distinct list of names in the search result.

GBIF

I: DATA STANDARDS AND DATA MODEL



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

Observations

Digitized specimens

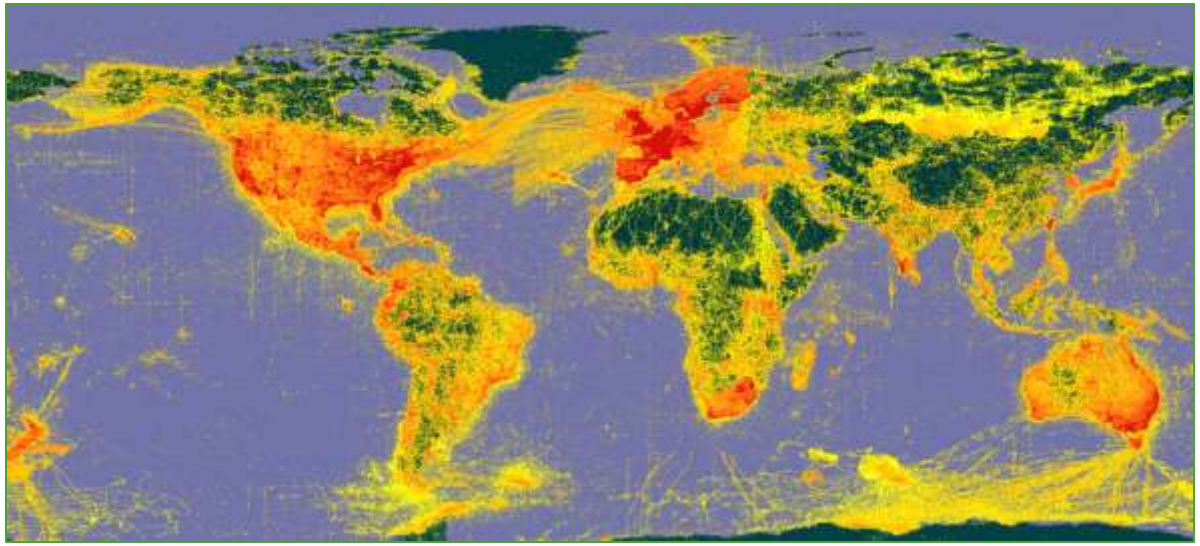
Literature

Data publishing and indexing

Common standards (DwC)

Remote-sensing

Environmental DNA



Data discovery and use

https://www.gbif.org/zh/occurrence/search

GBIF

INTERNATIONAL DATA STANDARDS: DARWIN CORE

TDWG Home Terms Guides GitHub

Darwin Core

Darwin Core is a standard maintained by the Darwin Core Maintenance Interest Group. It includes a glossary of terms (in other contexts these might be called properties, elements, fields, columns, attributes, or concepts) intended to **facilitate the sharing of information about biological diversity** by providing identifiers, labels, and definitions. Darwin Core is primarily based on taxa, their occurrence in nature as documented by observations, specimens, samples, and related information.

Getting started

- Quick reference guide
- Usage guides: how to use Darwin Core
- GitHub repository: where Darwin Core
- Normative term list: the document con
- Distribution files: convenient files to st

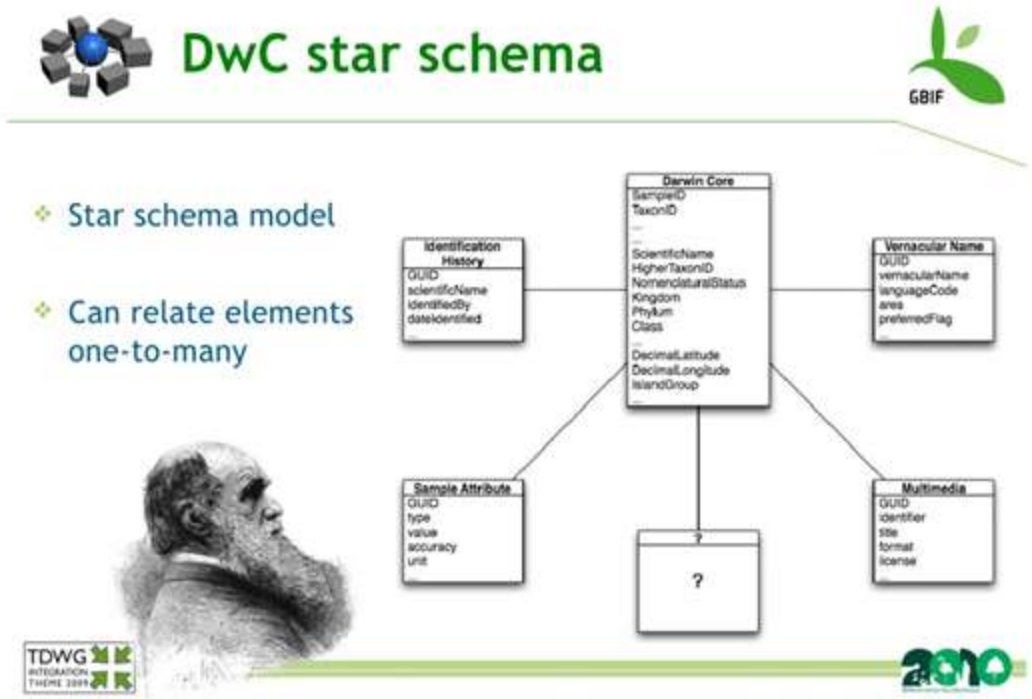
associatedTaxa	Property
Identifier	http://rs.tdwg.org/dwc/terms/associatedTaxa
Definition	A list (concatenated and separated) of identifiers or names of taxa and the associations of this Occurrence to each of them.
Comments	This term can be used to provide a list of associations to Taxa other than the one defined in the Occurrence. Note that the ResourceRelationship class is an alternative means of representing associations, and with more detail. This term is not apt for establishing relationships between Taxa, only between specific Occurrences of an Organism with other Taxa. Recommended best practice is to separate the values in a list with space vertical bar space (|).
Examples	"host":"Quercus alba", "host":"gbif.org/species/2879737", "parasitoid of":"Cyclocephala signaticollis" | "predator of":"Apis mellifera"

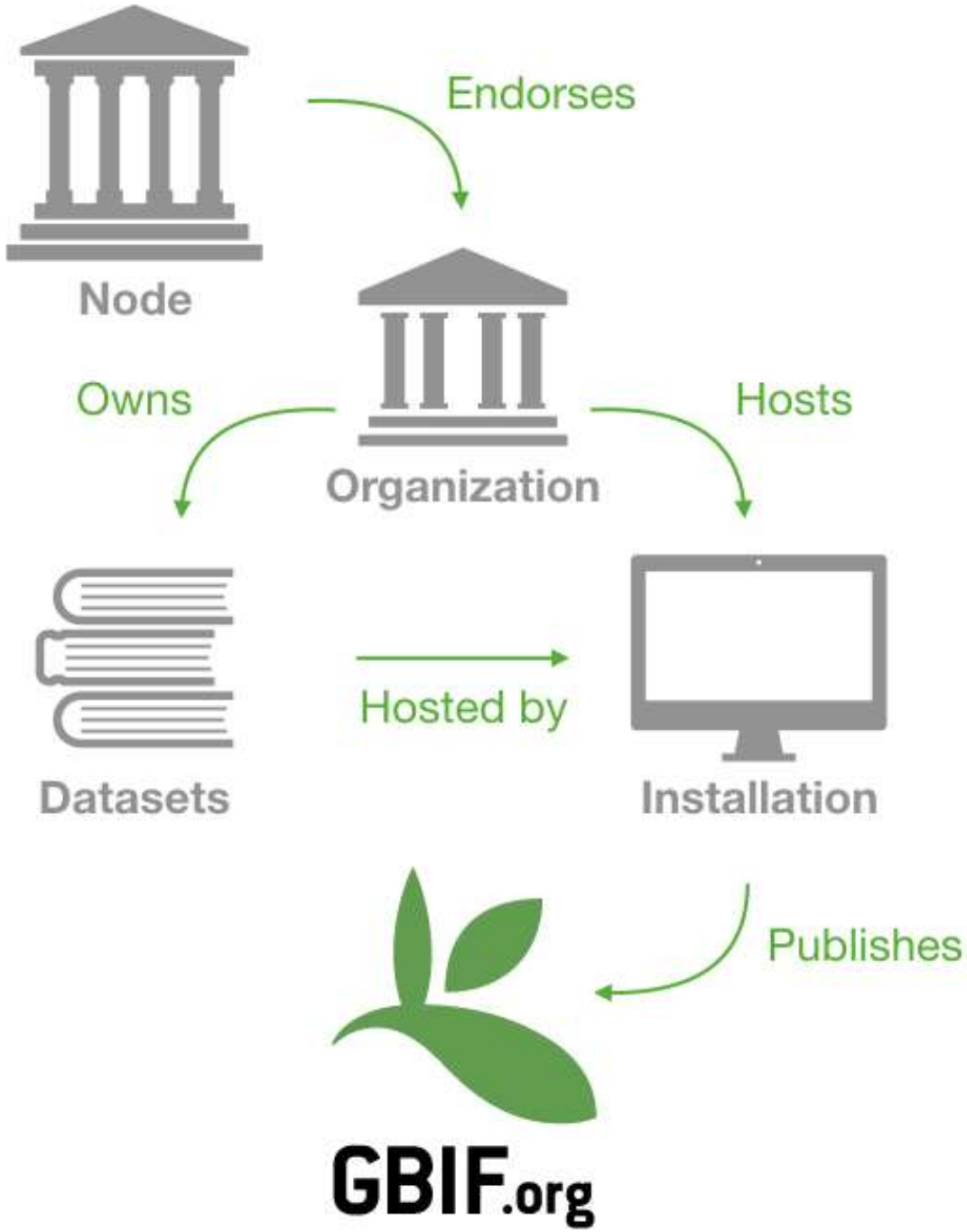
otherCatalogNumbers	Property
Identifier	http://rs.tdwg.org/dwc/terms/otherCatalogNumbers
Definition	A list (concatenated and separated) of previous or alternate fully qualified catalog numbers or other human-used identifiers for the same Occurrence, whether in the current or any other data set or collection.
Comments	Recommended best practice is to separate the values in a list with space vertical bar space (|).
Examples	FMNH:Mammal:1234 , NPS YELLO6778 | MBG 33424

occurrenceRemarks	Property
Identifier	http://rs.tdwg.org/dwc/terms/occurrenceRemarks
Definition	Comments or notes about the Occurrence.
Comments	
Examples	found dead on road

GBIF

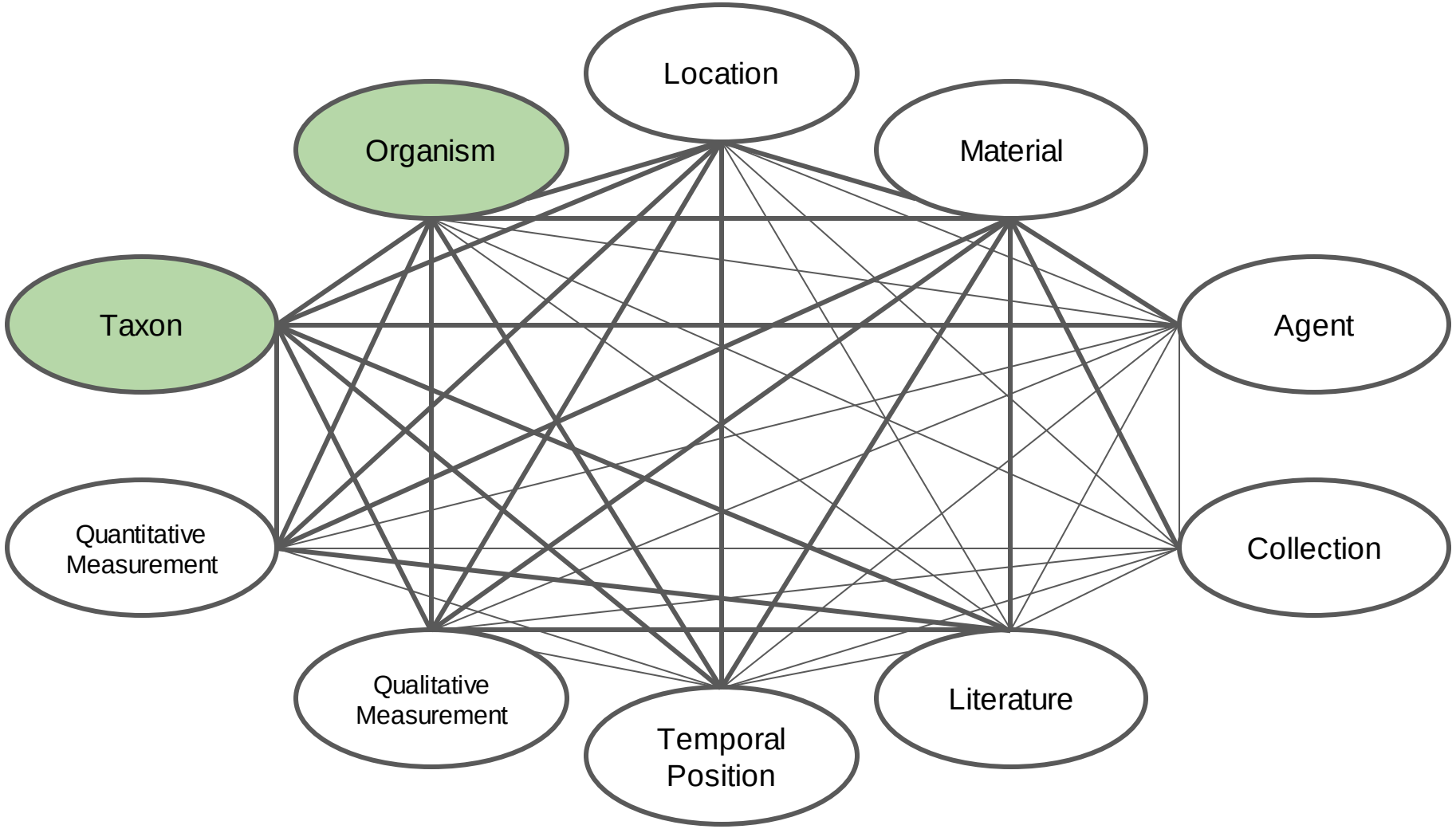
PUBLISHING TO GBIF, THE BASICS



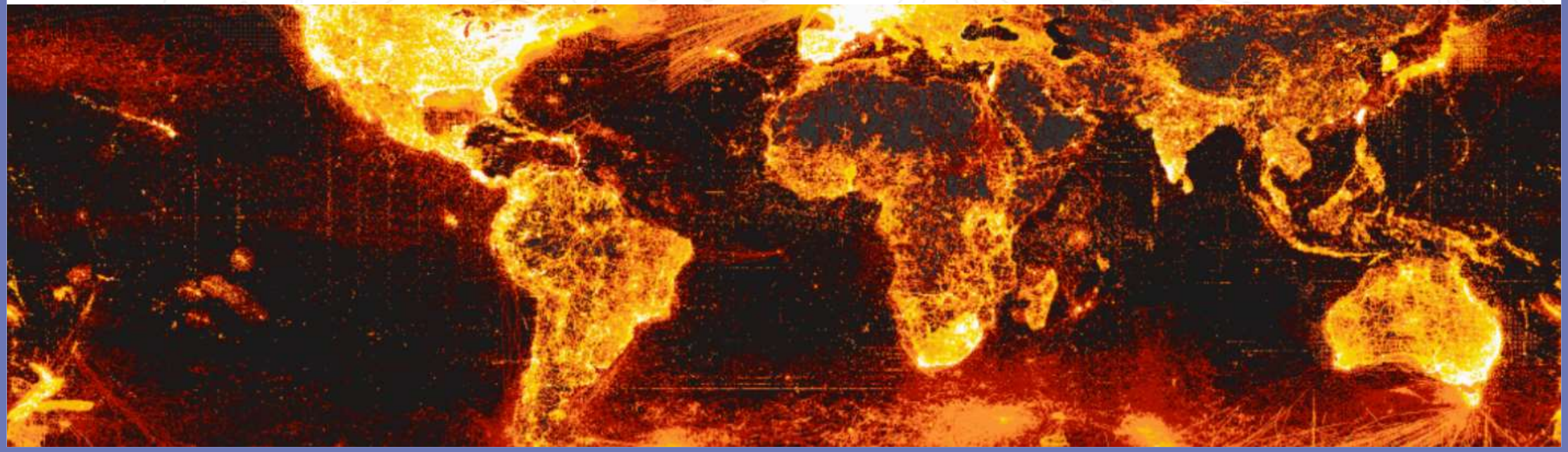


https://data-blog.gbif.org/post/installations-and-hosting-solutions-explained/

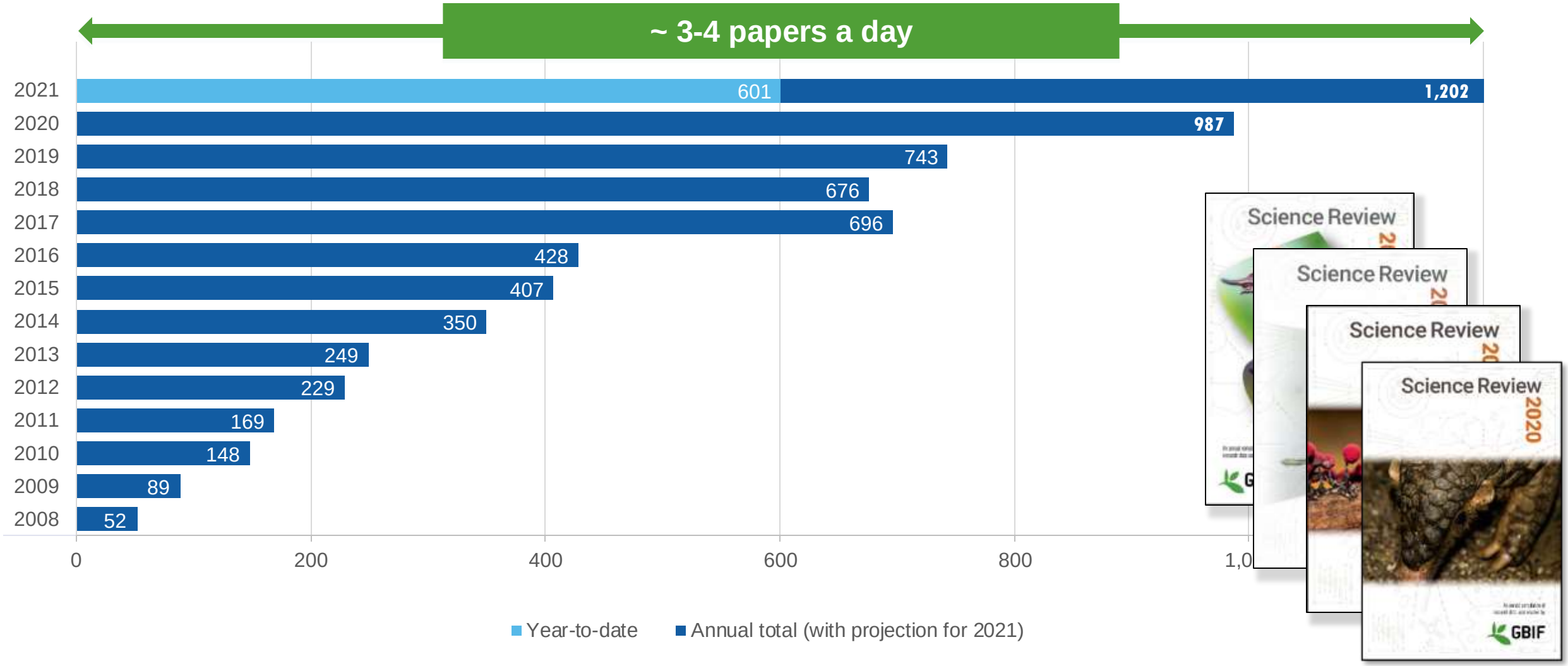
Darwin core thinking



R: DOI BASED DATA CITATION



PEER-REVIEWED PUBLICATIONS USING GBIF-MEDIATED DATA 30 June 2021



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

#CiteTheDOI

GBIF

Tropical Medicine and Infectious Disease

Article

Mapping Disease Transmission Risk of N
in South and Southeast Asia

Mark A. Deka * and Niaz Morshed

Department of Geography, Texas State University, 601 University Drive, San Marco
m_m617@txstate.edu
* Correspondence: mad214@txstate.edu; Tel.: +1-512-557-5647

Received: 3 May 2018; Accepted: 25 May 2018; Published: 30 May 2018

International Semantic Web Conference

ISWC 2018: The Semantic Web – ISWC 2018 pp 217-232 | Cite as

Constructing a Recipe Web from Historical Newspapers

nature > nature ecology & evolution > articles > article

MENU

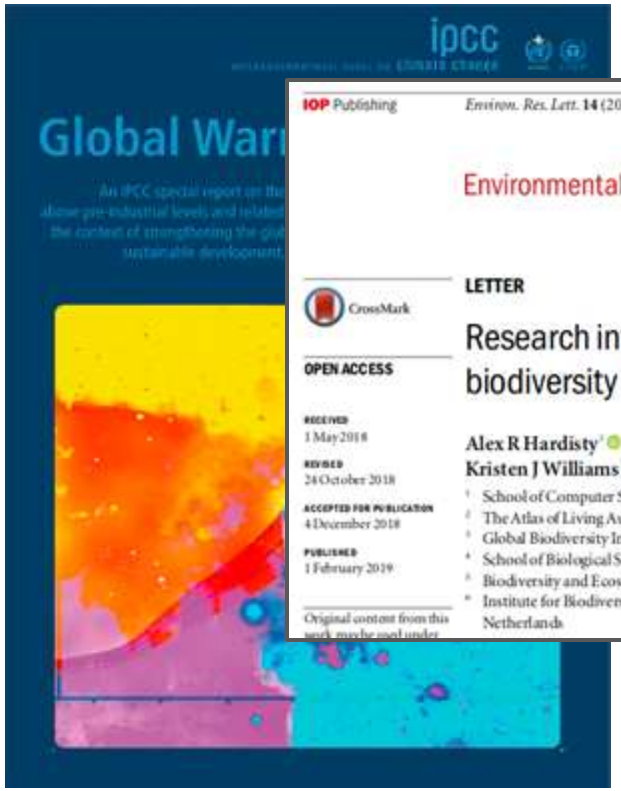
nature ecology & evolution

Article | Published: 05 November 2018

A global test of ecoregions

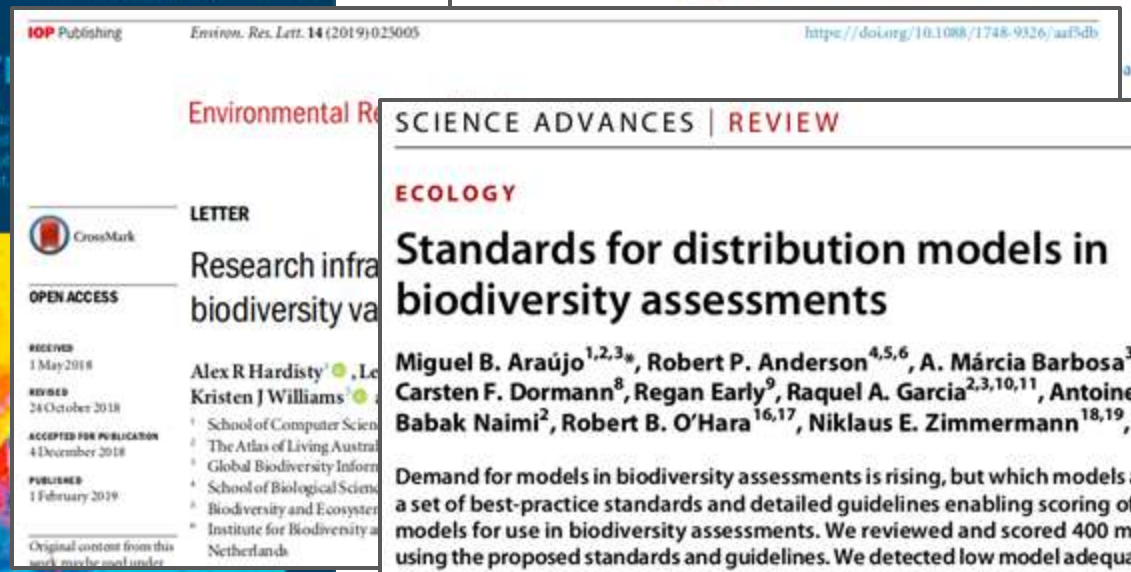
Jeffrey R. Smith, Andrew D. Letten, Po-Ju Ke, Christopher B. Anderson, J. Nicholas Hendershot, Manpreet K. Dhami, Glade A. Dlott, Tess N. Grainger, Meghan E. Howard, Beth M. L. Morrison, Devin Routh, Priscilla A. San Juan, Harold A. Mooney, Erin A. Mordecai, Thomas W. Crowther & Gretchen C.

ad Citation



ipcc

Global War



IOP Publishing Environ. Res. Lett. 14 (2019) 025005 https://doi.org/10.1088/1748-9326/aaf5db

Environmental R

LETTER

Research infra
biodiversity va

OPEN ACCESS

RECEIVED
1 May 2018

REVISED
24 October 2018

ACCEPTED FOR PUBLICATION
4 December 2018

PUBLISHED
1 February 2019

Original content from this work may be used under

Alex R Hardisty, Le
Kristen J Williams

School of Computer Scien
The Atlas of Living Austral
Global Biodiversity Inform
School of Biological Scien
Biodiversity and Ecosyster
Institute for Biodiversity a
Netherlands

SCIENCE ADVANCES | REVIEW

ECOLOGY

Standards for distribution models in biodiversity assessments

Miguel B. Araújo[1,2,3]*, Robert P. Anderson[4,5,6], A. Márcia Barbosa[3], Colin M. Beale[7], Carsten F. Dormann[8], Regan Early[9], Raquel A. Garcia[2,3,10,11], Antoine Guisan[12,13], Luigi Maiorano[14,15], Babak Naimi[2], Robert B. O'Hara[16,17], Niklaus E. Zimmermann[18,19], Carsten Rahbek[2,20]

Demand for models in biodiversity assessments is rising, but which models are adequate for the task? We propose a set of best-practice standards and detailed guidelines enabling scoring of studies based on species distribution models for use in biodiversity assessments. We reviewed and scored 400 modeling studies over the past 20 years using the proposed standards and guidelines. We detected low model adequacy overall, but with a marked tendency of improvement over time in model building and, to a lesser degree, in biological data and model evaluation. We argue that implementation of agreed-upon standards for models in biodiversity assessments would promote transparency and repeatability, eventually leading to higher quality of the models and the inferences used in assessments. We encourage broad community participation toward the expansion and ongoing development of the proposed standards and guidelines.

Dataset DOIs

Download DOI

Archive DOI

Bibliographic DOI

Source dataset #1

Source dataset #2

Source dataset #3

Filter

GBIF download

Process

Final state of data

Archive

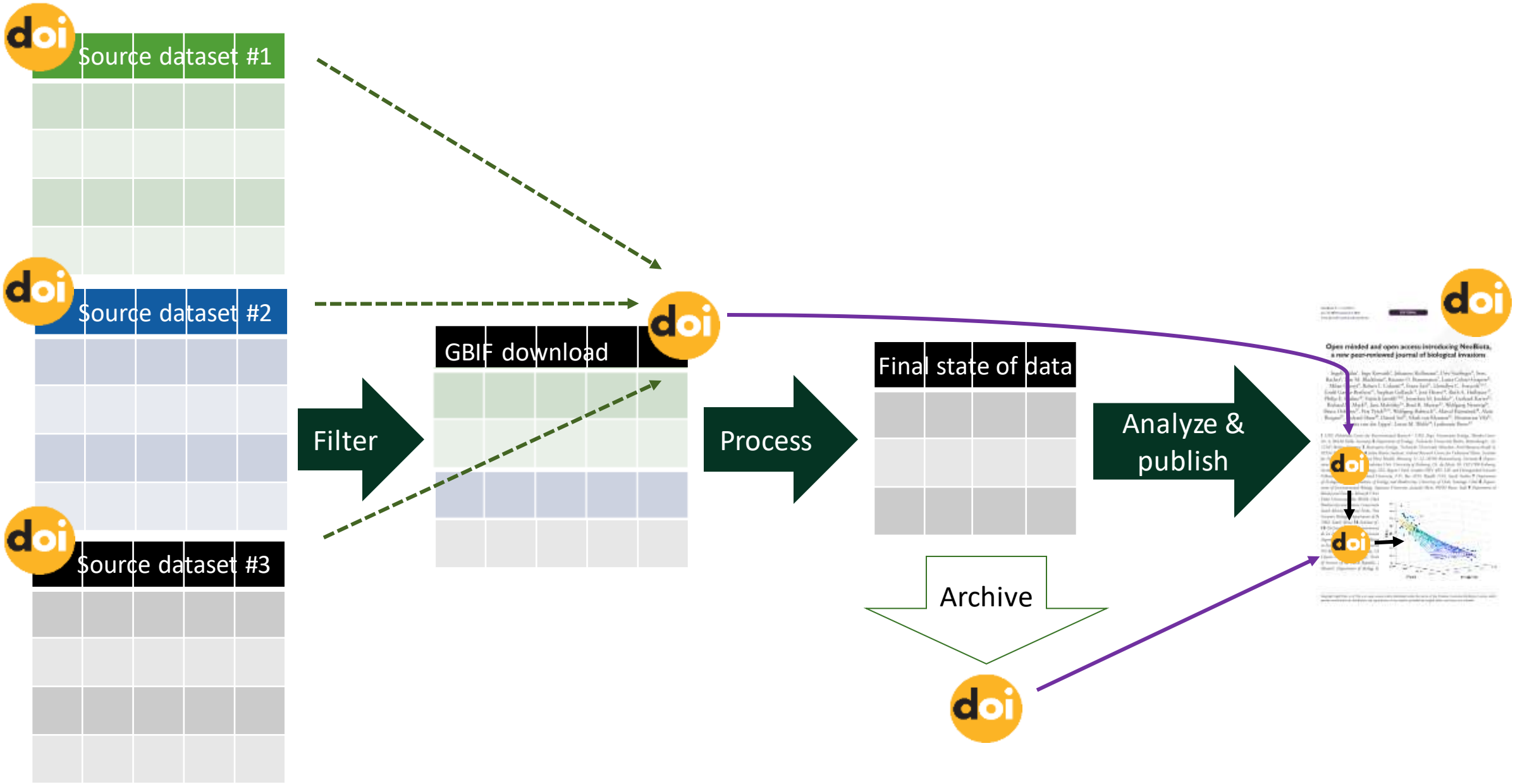
Analyze & publish

Dataset DOIs

Download DOI

Archive DOI

Bibliographic DOI



DOWNLOAD READY: DATA LINK AND DOI CITATION

GBIF Global Biodiversity Information Facility

Hello dschigel,

Your download is available at the following address:
https://api.gbif.org/v1/occurrence/download/request/0304929-200613084148143.zip

Citation

When using this dataset **please use the following citation:**

GBIF.org (18 June 2021) GBIF Occurrence Download https://doi.org/10.15468/dl.yqxh3d

Download Information

DOI: https://doi.org/10.15468/dl.yqxh3d (may take some hours before being active)
Creation Date: 08:54:39 18 June 2021
Records included: 15 records from 1 published datasets
Compressed data size: 2.2 kB
Download format: simple tab-separated values (TSV)
Filter used:

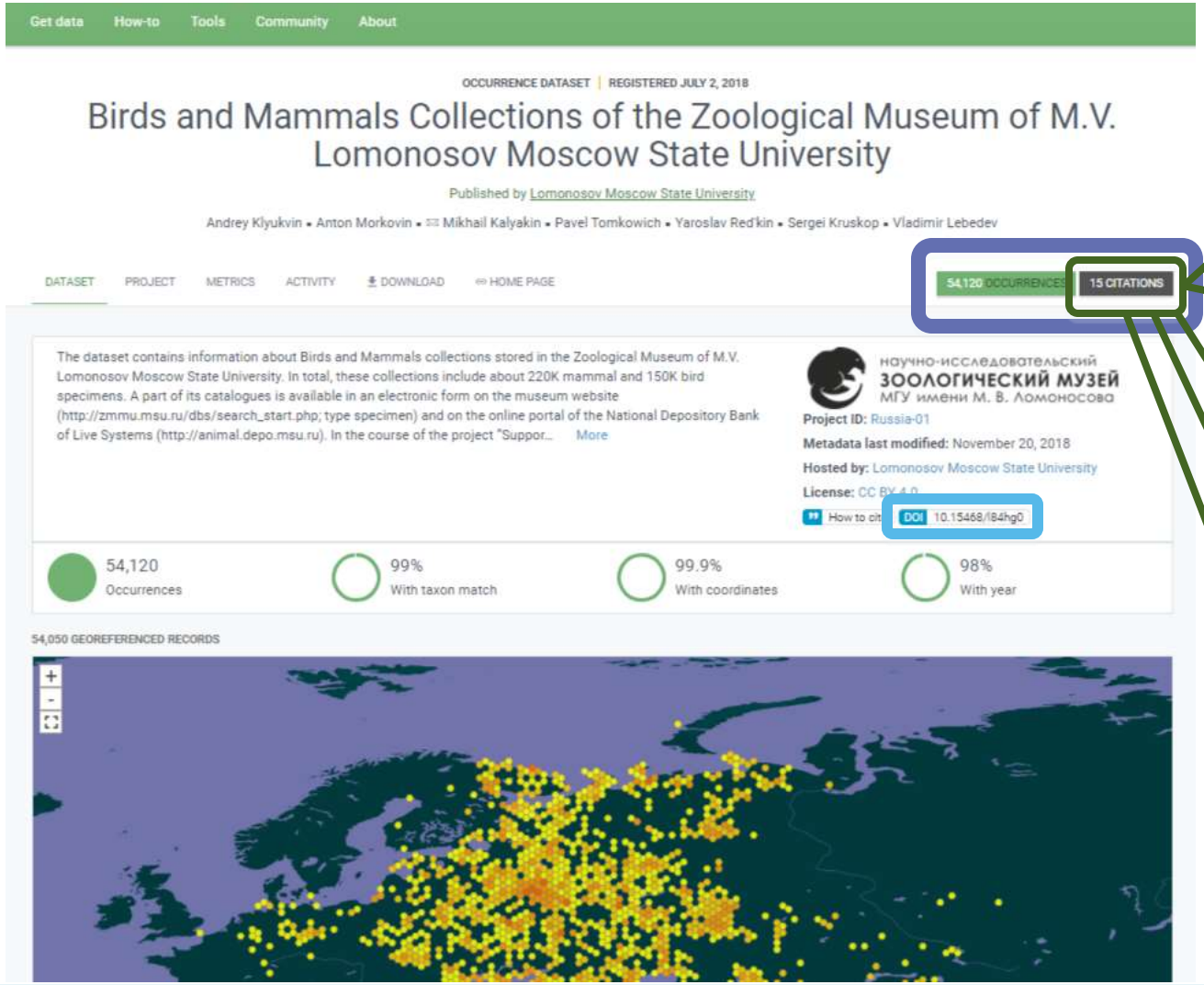
```
{
  "and" : [
    "DatasetKey is International Barcode of Life project (iBOL)",
    "MediaType is Image",
    "TaxonKey is Mycetina Mulsant, 1846"
  ]
}
```

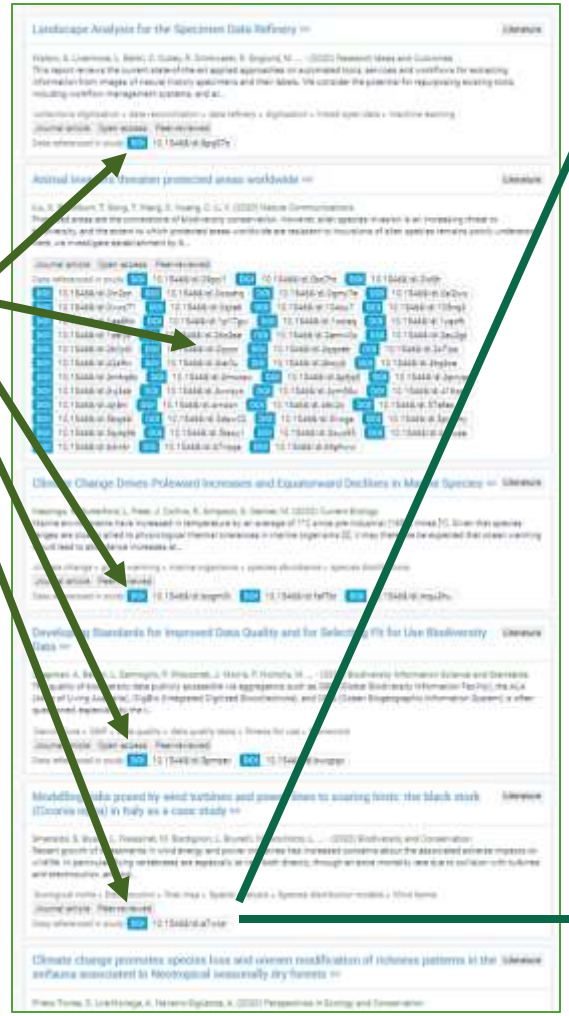
<- Data link you asked for

<- Data citaton with DOI

GBIF

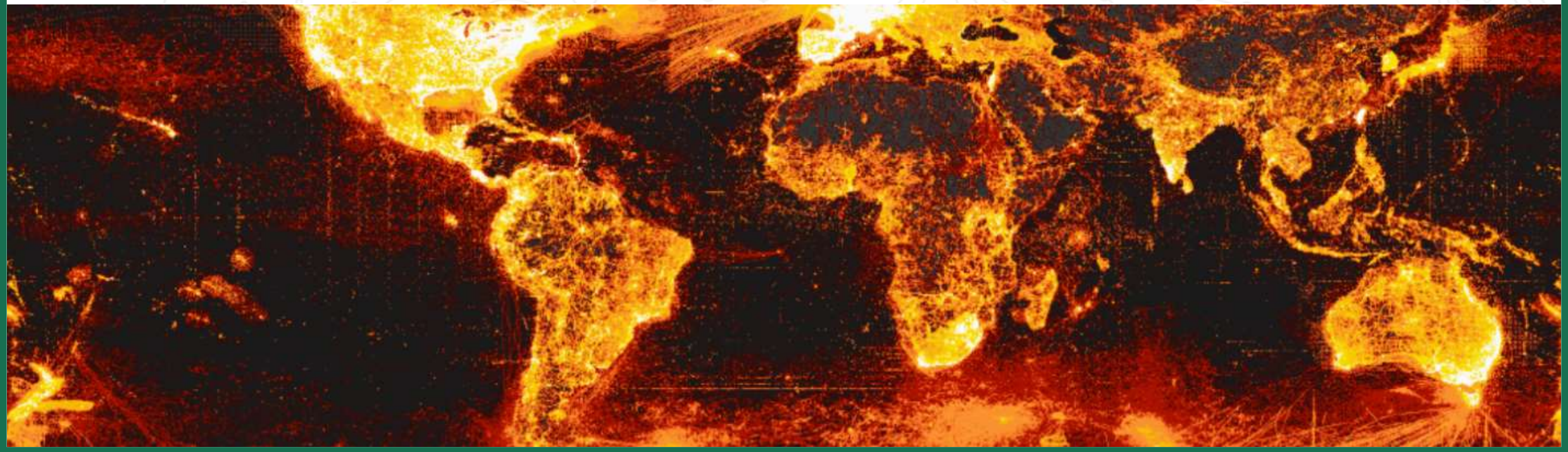
DOI BASED DATA CITATION AT GBIF.ORG



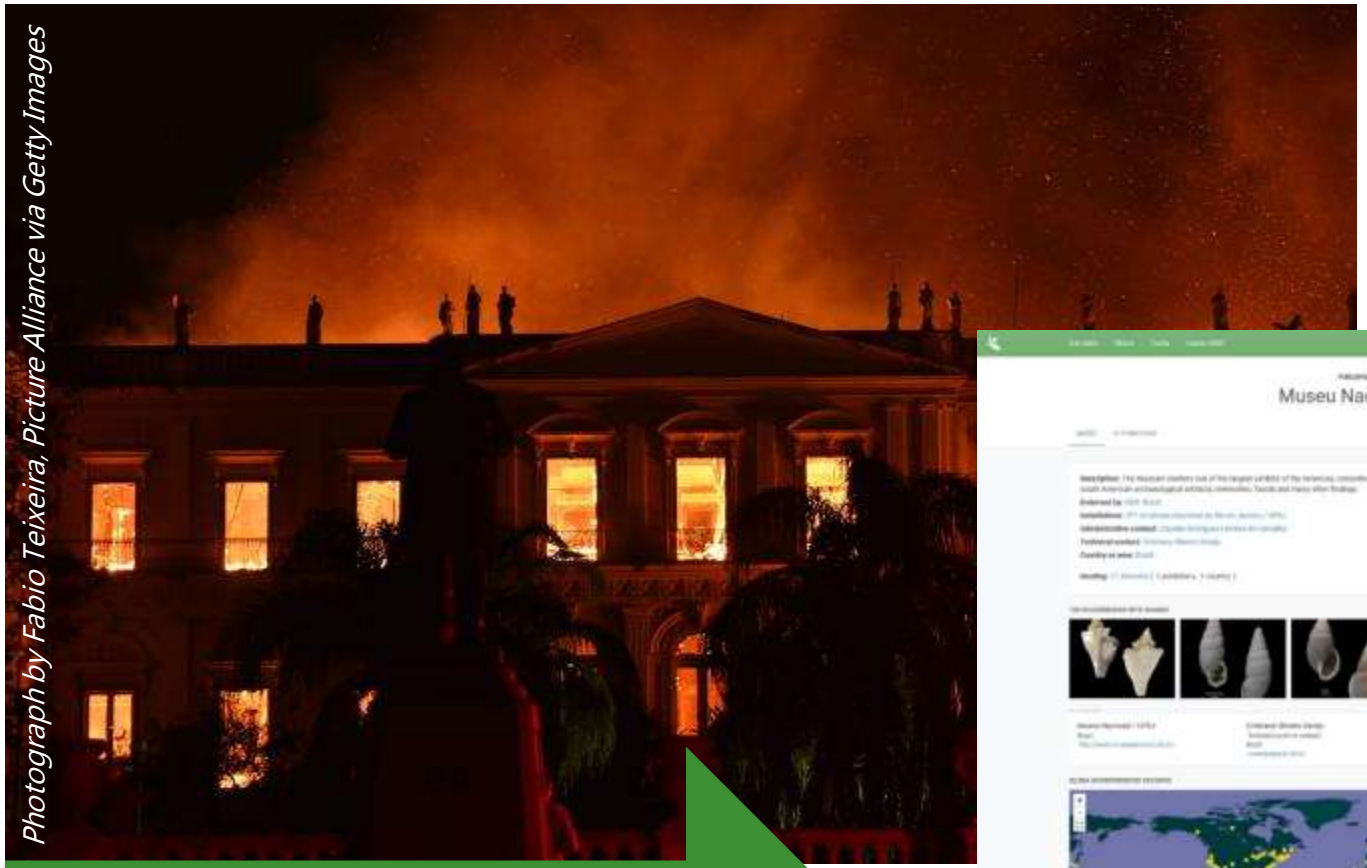


https://www.gbif.org/dataset/9beef55a-b4a4-4031-a1e0-0d2bdeea2c21; https://doi.org/10.1007/s10531-020-01961-3

CONCLUSION

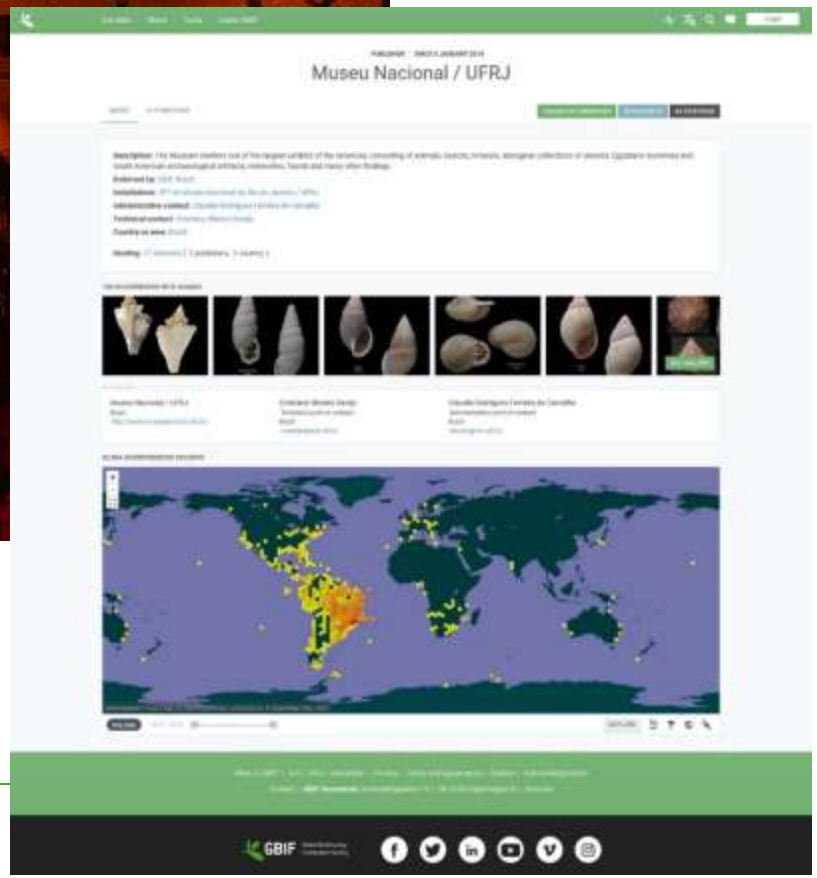


All specimens will be destroyed at some point



Photograph by Fabio Teixeira, Picture Alliance via Getty Images

… but digitization and data publishing is also form of digital **security** for heritage knowledge



GBIF

Data sharing: the **when** choices

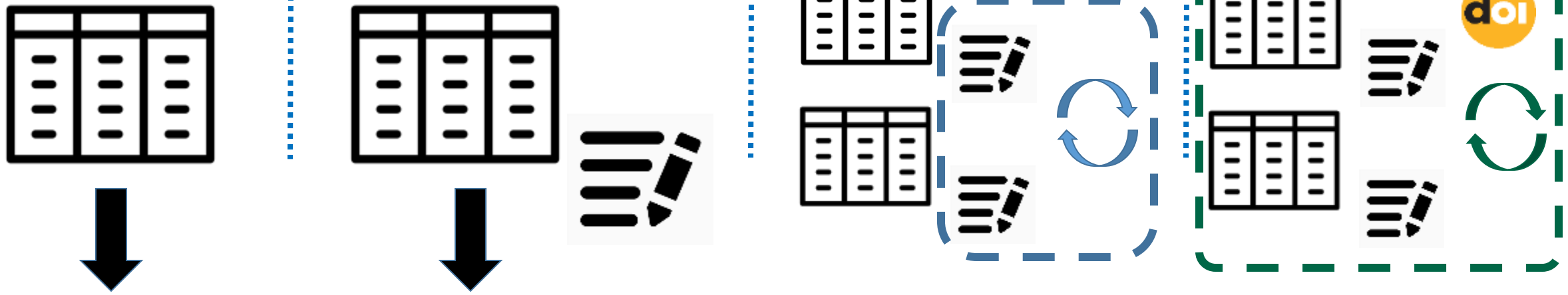
Data sharing: the **where** choices

Archive

Generalist repository

Data catalogue

Data index



Preservation

Open data

FAIR data

Save

Minimum description

Metadata standartization

Data standartization

Open data and FAIR data

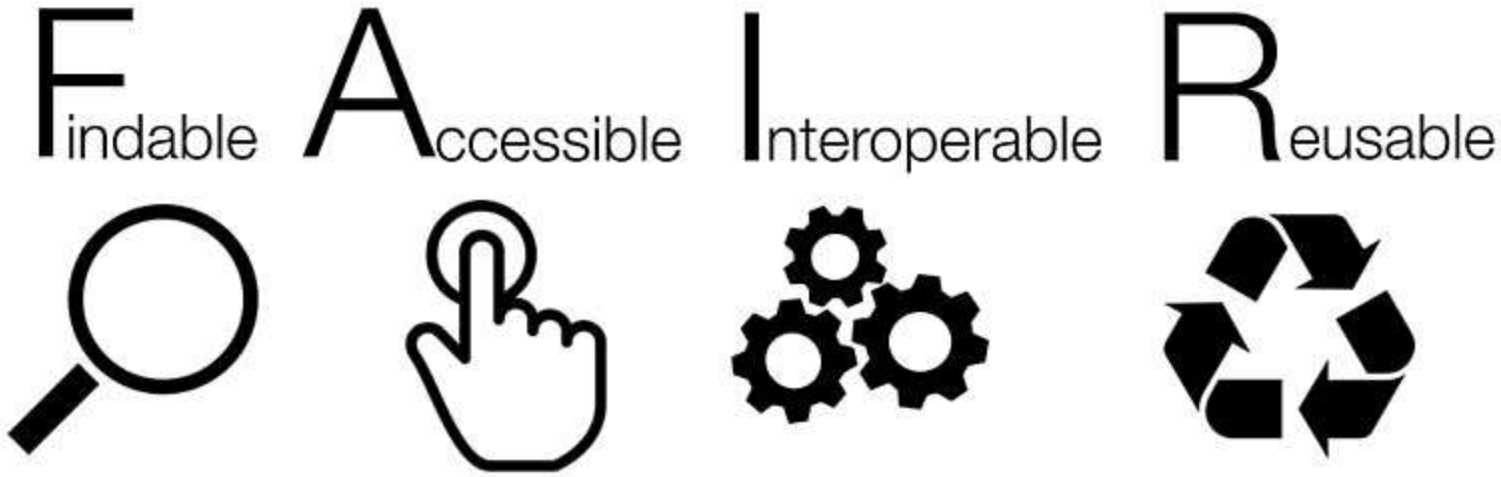


Image by Sangya Pundir

Dmitry Schigel
dschigel@gbif.org

@dschigel
adlignum.org

GBIF