



Global Biodiversity Information Facility a FAIR data infrastructure and network

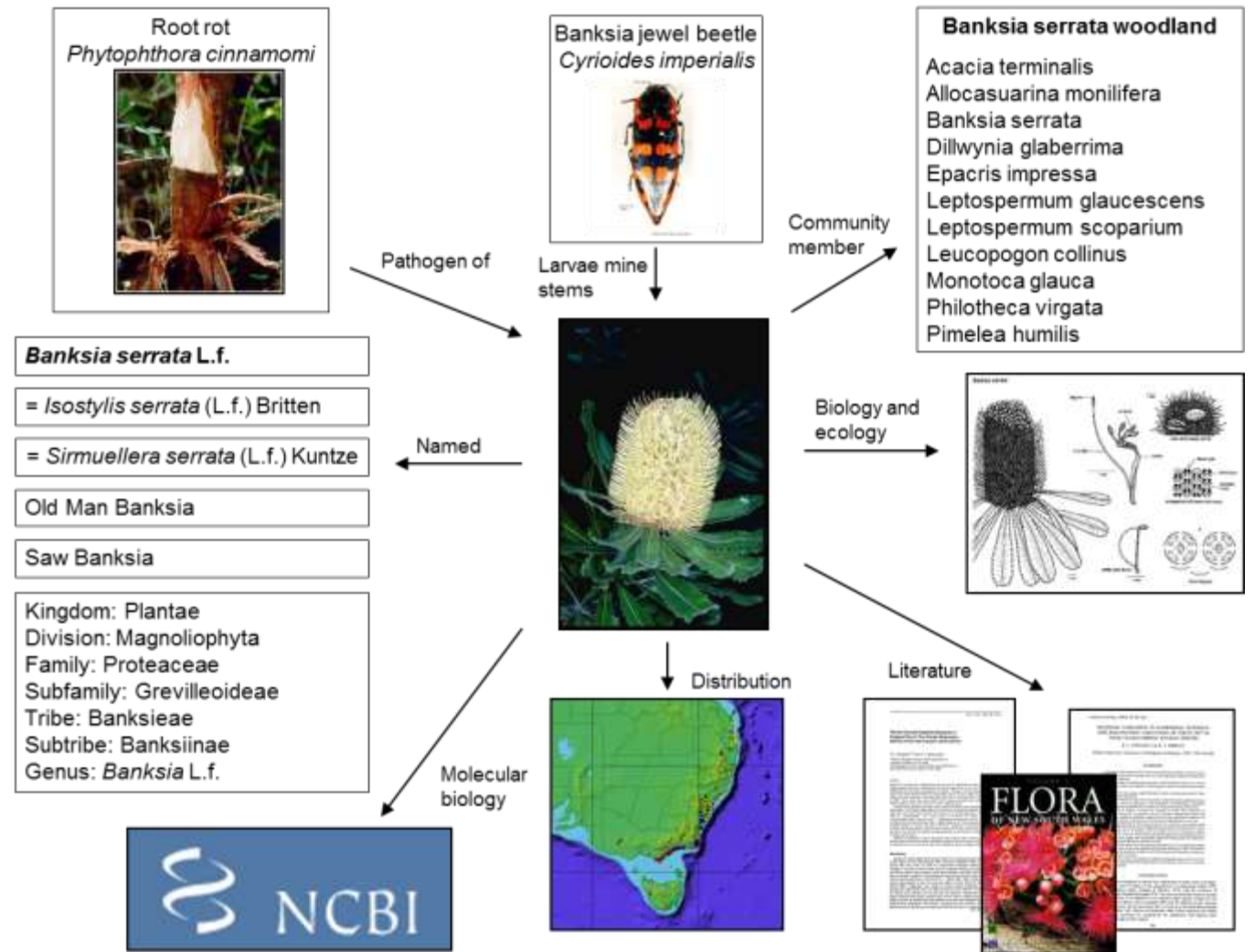
Dmitry Schigel | Scientific officer

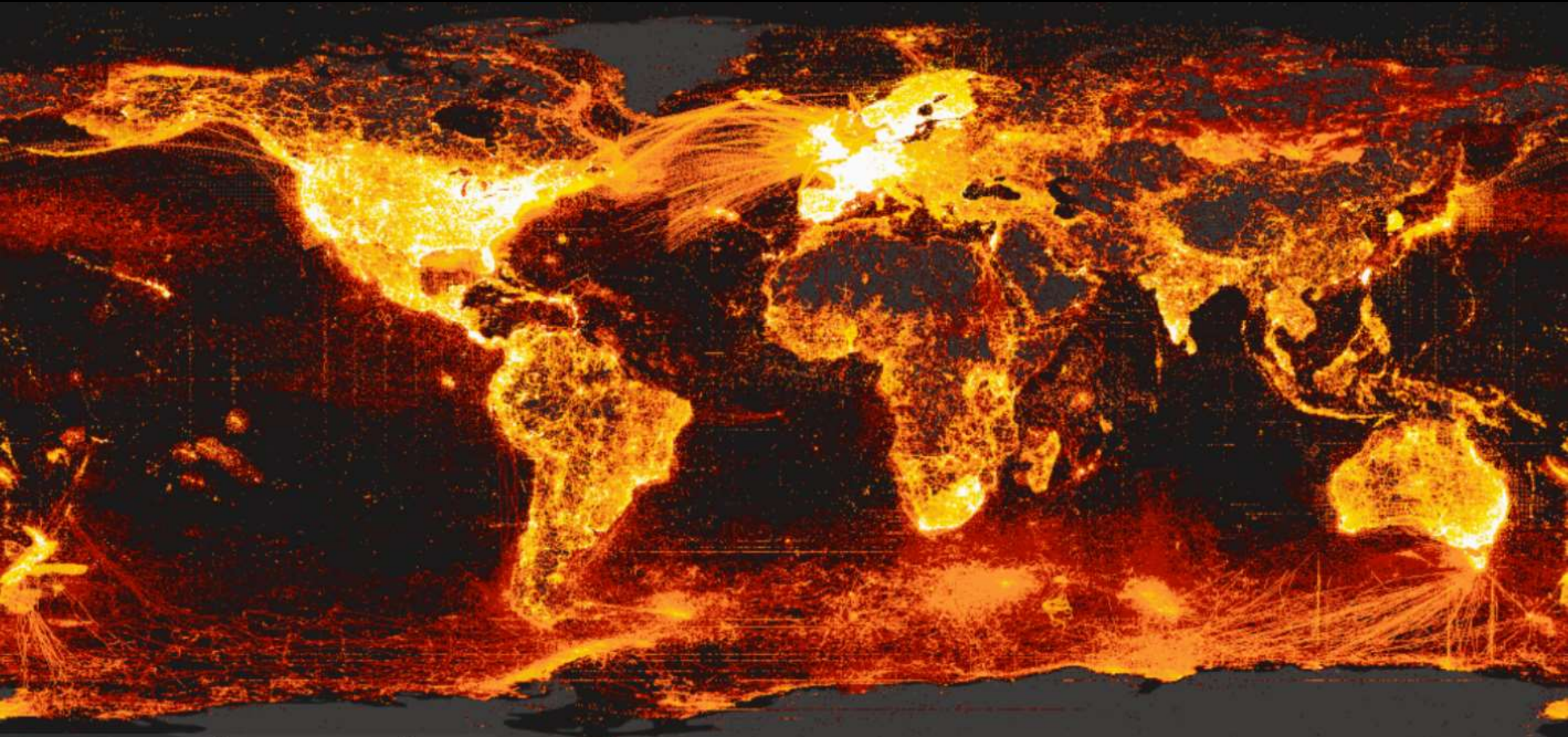


Outline

- **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

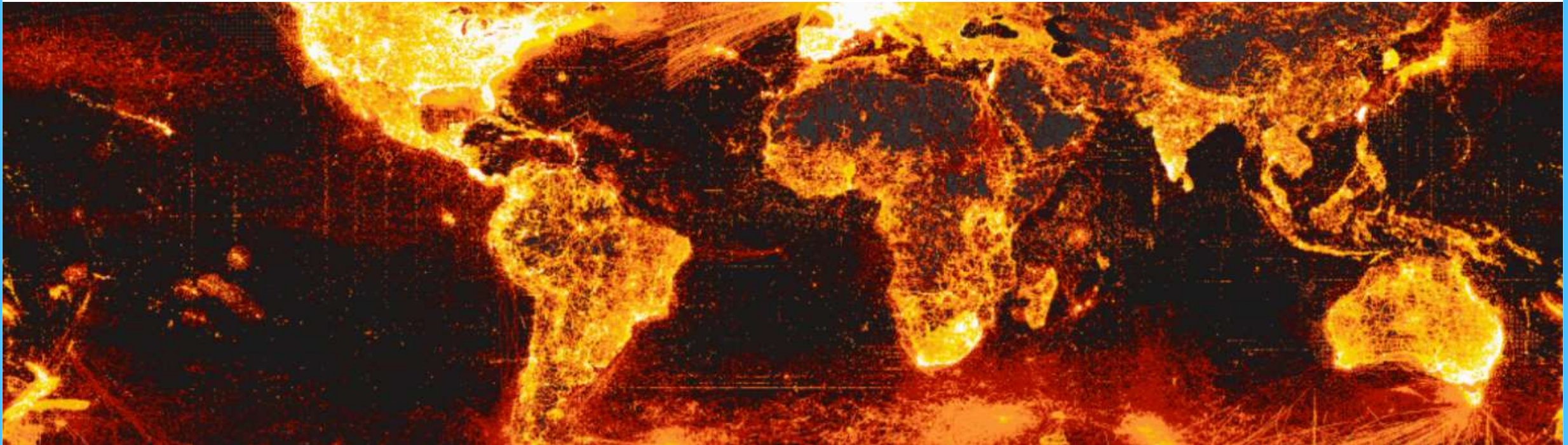
Biodiversity information







GBIF: Global Biodiversity Information Facility



GBIF: GLOBAL BIODIVERSITY INFORMATION FACILITY

Key facts

Intergovernmental open data infrastructure

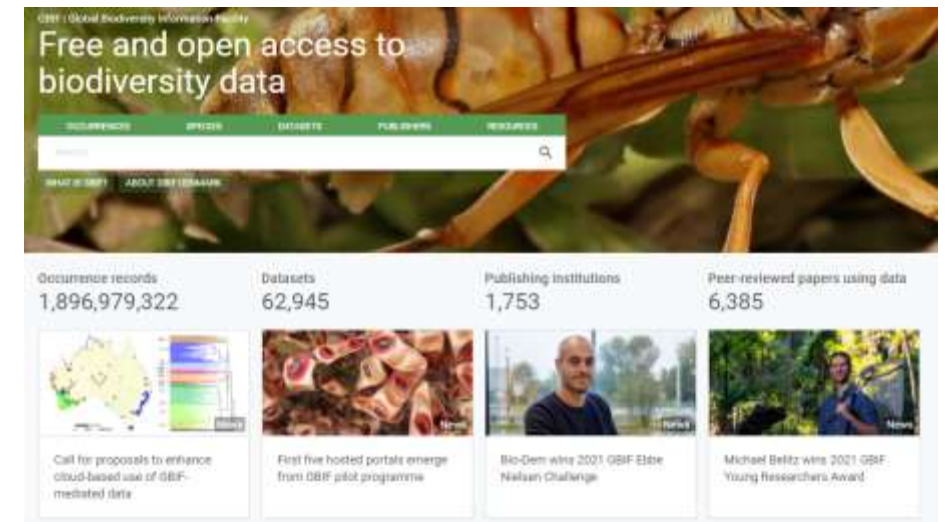
Established in 2001, OECD recommendation

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

Voluntary memorandum of understanding

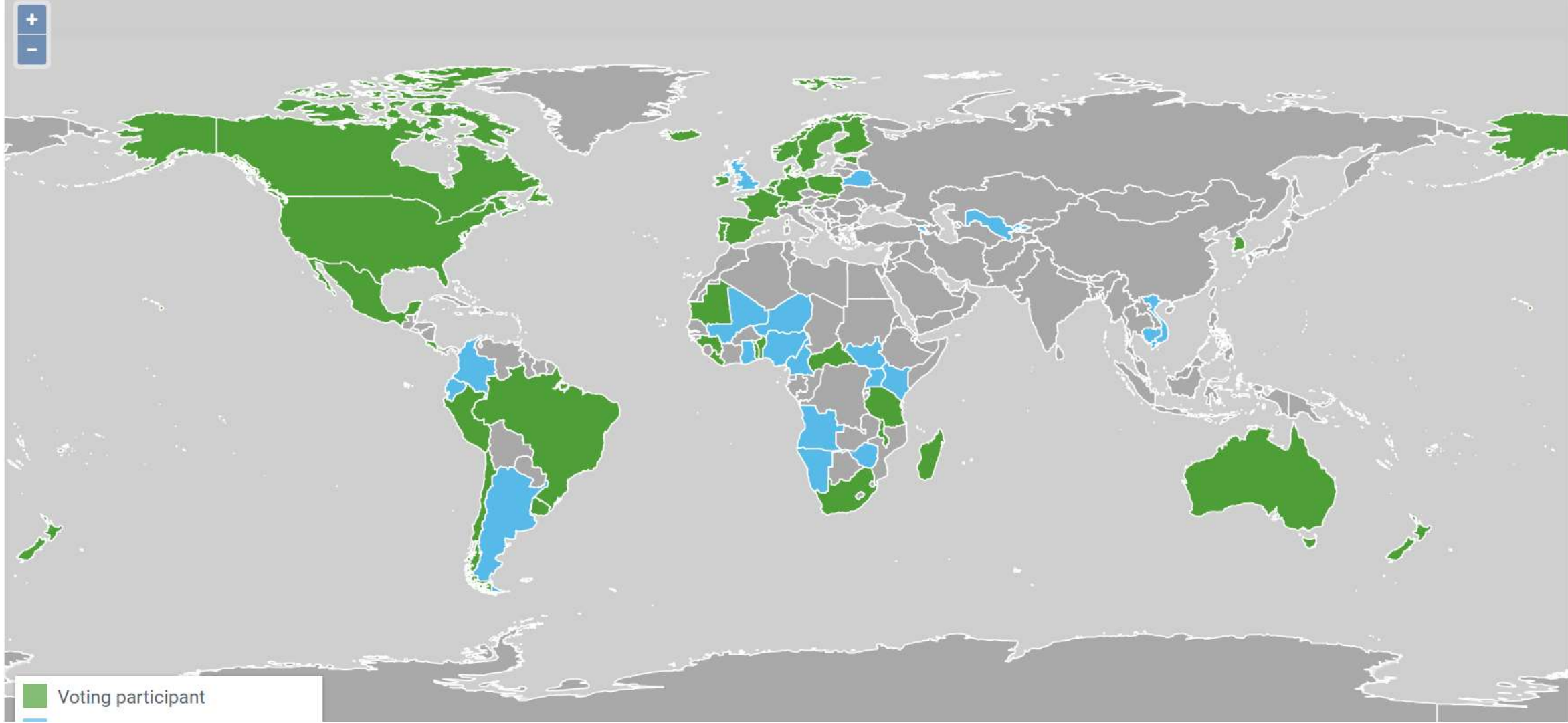
Funds: governments of the participant countries

102 participants

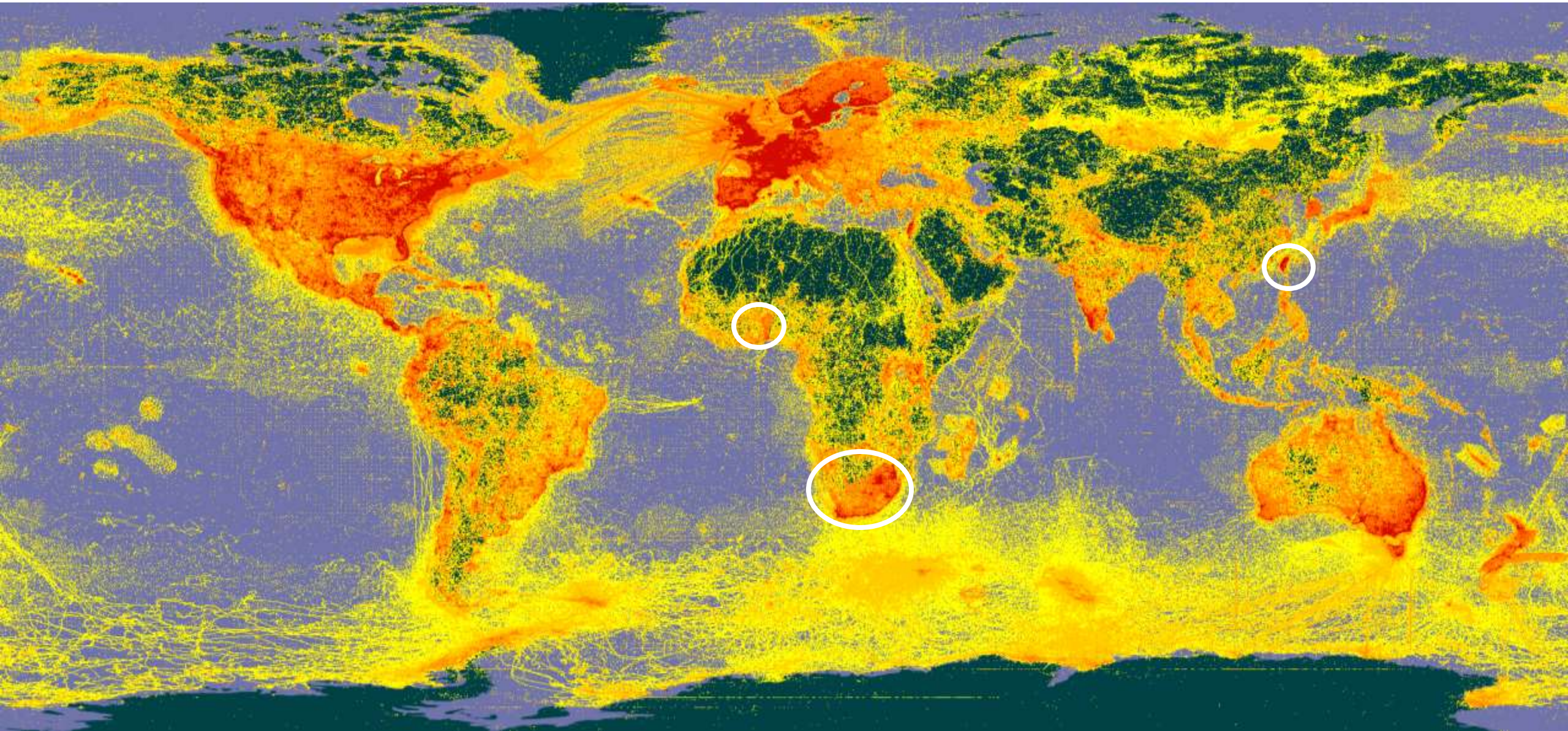


GBIF PARTICIPANT NETWORK

23 September 2021

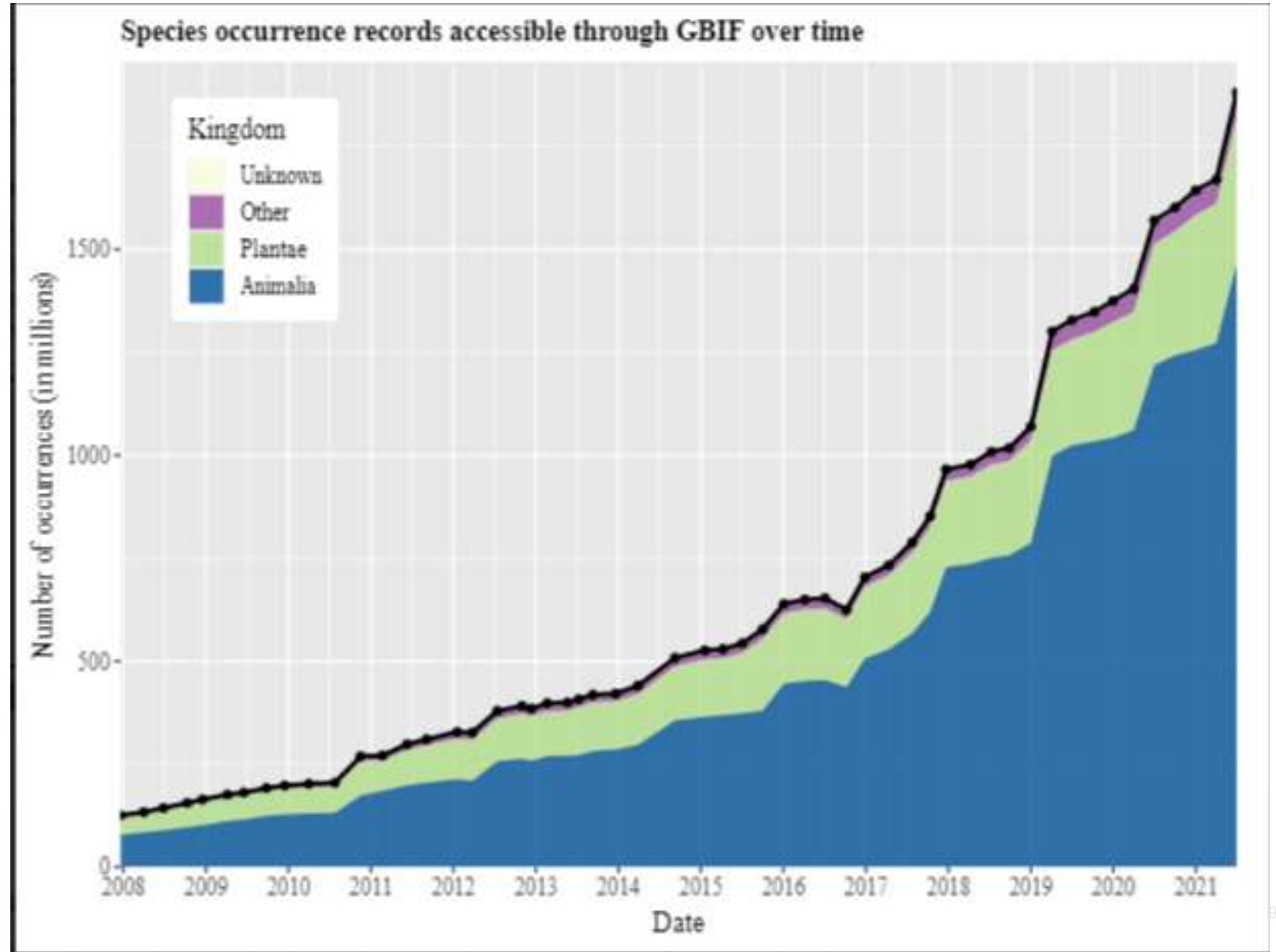


DATA FROM THE GBIF NETWORK 30 June 2021



Data published through GBIF.org

data availability



BY THE NUMBERS | 2 NOVEMBER 2021

Species occurrence records

1,896,979,322



Country
Participants

61

Organizational
Participants

41

Avg records downloaded per month (2021)



86.6 billion

Datasets

62,945



Data-publishing
institutions

1,753

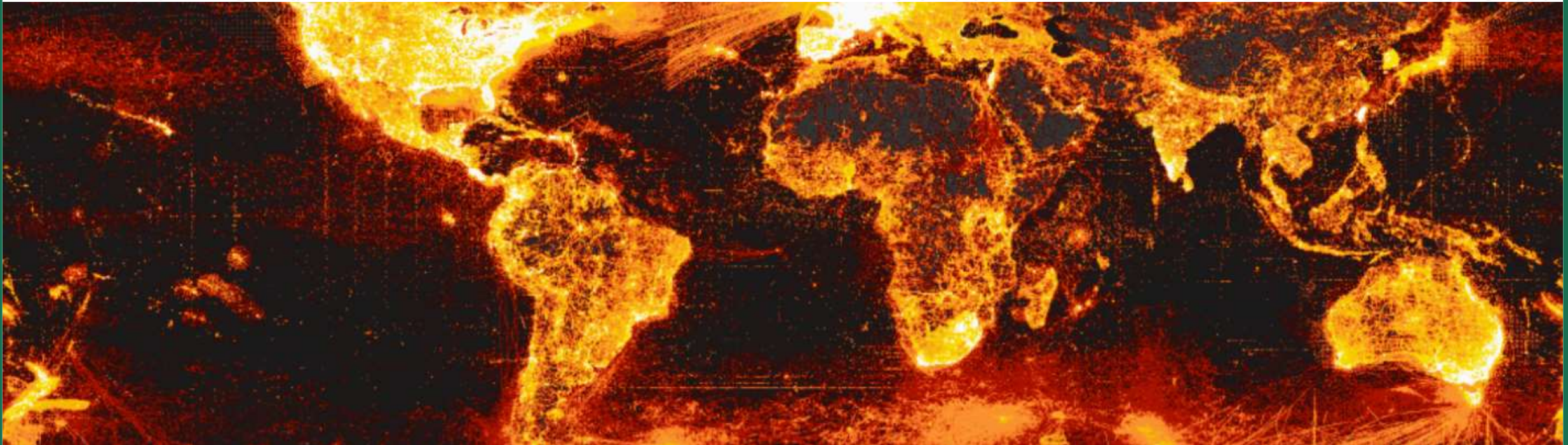
Peer-review papers
using data

6,385





FINDABLE: MANY FACES OF DISCOVERABILITY



LET'S GOOGLE IT



endomychus coccineus



All Images Videos Shopping Maps More

Tools

About 43.200 results (0,56 seconds)

https://en.wikipedia.org/wiki/Endomychus_coccineus

Endomychus coccineus - Wikipedia

Endomychus coccineus, common name scarlet endomychus or false ladybird, is a species of beetles in the family Endomychidae.

<https://www.naturbasen.dk/art/ska...> - Translate this page

Skarlagensvampehøne (Endomychus coccineus) - Naturbasen

: Længde: 4-6 mm. En karakteristisk bille der er kendetegnet ved den skarlagenerøde farve med 4 store sorte pletter på dækvingerne. Følehornene er sorte og ret ...

<https://www.ukbeetles.co.uk/endomychus-coccineus>

Endomychus coccineus | uk beetles

Endomychus coccineus (Linnaeus, 1758) ... This widely distributed and generally common species occurs from Spain eastwards to Central Asia and from Italy north to ...

<https://species.nbnatlas.org/NBNSYS0000024711>

Endomychus coccineus (Linnaeus, 1758) - NBN Atlas

Endomychus coccineus, common name scarlet endomychus or false ladybird, is a species of beetles in the family Endomychidae.

<https://www.gbif.org/species>

Endomychus coccineus (Linnaeus, 1758) - GBIF

Citation (for citing occurrences, please see guidelines). Endomychus coccineus (Linnaeus, 1758) in GBIF Secretariat (2021). GBIF Backbone Taxonomy. Checklist ...

<https://allearter-databasen.dk/takson...> - Translate this page

Skarlagensvampehøne - Endomychus coccineus :: allearter.dk

Skarlagensvampehøne (Endomychus coccineus). Firepletlet svampebille. (Linnaeus, 1758). Dyreriget > Leddyr > Insekter > Biller > Svampehøns > .



#1 Wikipedia

#2 naturbasen.dk

#3 ukbeetles.co.uk

#4 nbnatlas.org

#5 gbif.org

#6 allearter-databasen.dk



More images

Endomychus coccineus

Endomychus coccineus, common name scarlet endomychus or false ladybird, is a species of beetles in the family Endomychidae. [Wikipedia](#)

Species: E. coccineus

Endomychus coccineus (Linnaeus, 1758):
Endomychus coccineus; (Linnaeus, 1758)

People also search for

View 5+ more

GBIF.ORG

GBIF | Global Biodiversity Information Facility

Free and open access to biodiversity data

OCCURRENCES SPECIES DATASETS PUBLISHERS RESOURCES

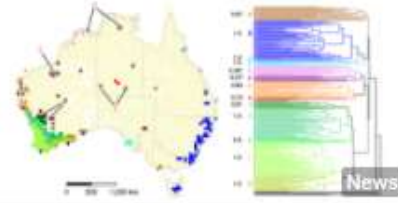
Search

WHAT IS GBIF? ABOUT GBIF DENMARK


Occurrence records
1,896,979,322

Datasets
62,945

Publishing institutions
1,753

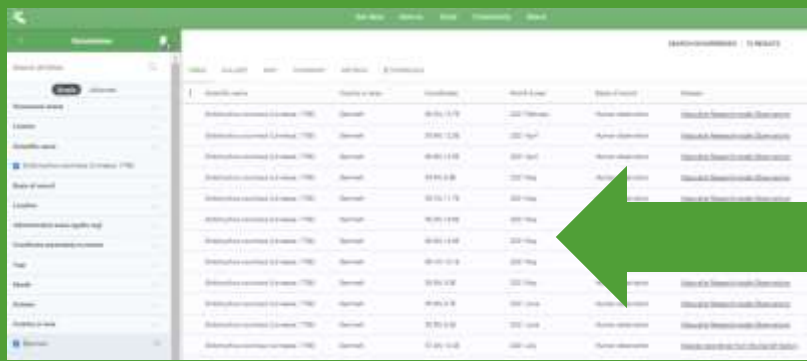


Call for proposals to enhance cloud-based use of GBIF-mediated data

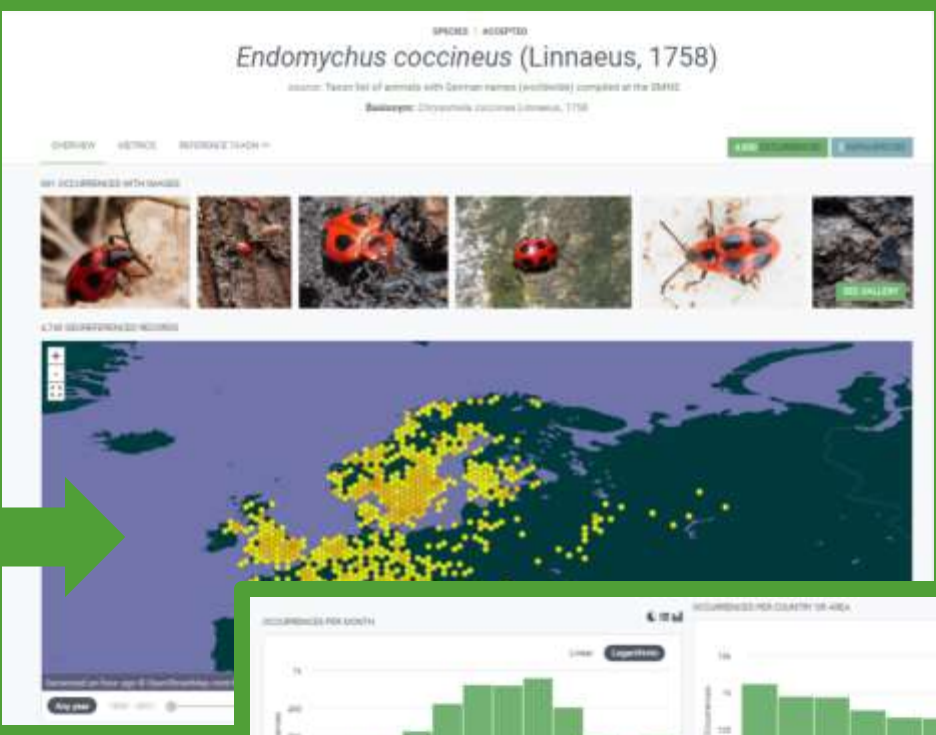


First five hosted from GBIF pilot

plitz wins 2 researchers



Dataset	Year	Occurrence	Latitude	Longitude	Max Elevation	Max Depth	Max Slope
Endomychus coccineus (Linnaeus, 1758)	2010	1	55.56	15.19	200	None	None
Endomychus coccineus (Linnaeus, 1758)	2010	1	55.56	15.19	200	None	None
Endomychus coccineus (Linnaeus, 1758)	2010	1	55.56	15.19	200	None	None
Endomychus coccineus (Linnaeus, 1758)	2010	1	55.56	15.19	200	None	None
Endomychus coccineus (Linnaeus, 1758)	2010	1	55.56	15.19	200	None	None



SPECIES | ACCEPTED

Endomychus coccineus (Linnaeus, 1758)

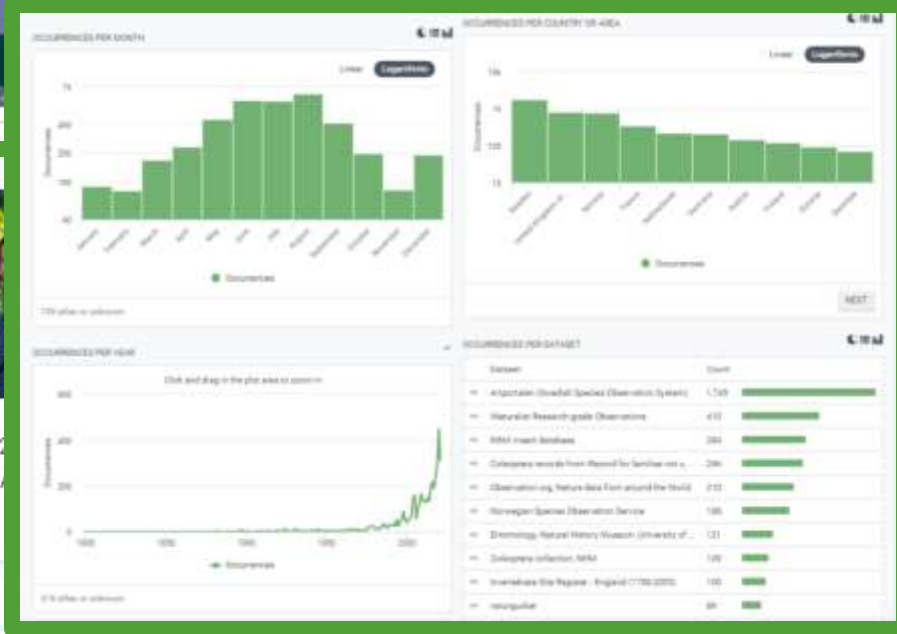
Source: Taxon list of animals with German names (available) compiled at the GBIF

Subspecies: *Endomychus coccineus* (Linnaeus, 1758)

OVERVIEW METRICS INFORMATION

4741 OCCURRENCES WITH IMAGES

4741 OCCURRENCES WITH RECORDS



DATA PAPERS

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



The screenshot shows the top navigation bar of the RIO journal website. The logo features an elephant and the text 'RIO'. Navigation links include 'Home', 'Articles', 'About', 'About Pensoft', 'Books', and 'Journals'. A 'Guidelines' button is highlighted in teal. The article title 'Data Published in Data Papers' is displayed, along with the issue information 'Research Ideas and Outcomes 3: e12431' and the DOI 'https://doi.org/10.3897/rio.3.e12431 (28 Feb 2017)'. A 'Reviewed v1' badge is visible in the top right corner.

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:

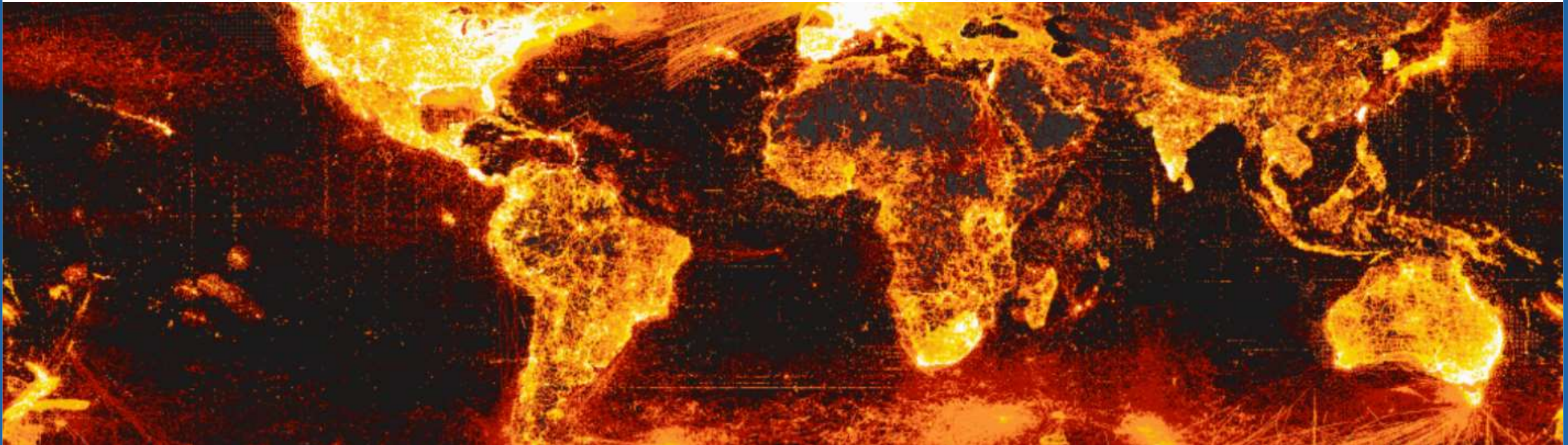
- to provide a citable journal publication that brings scholarly credit to data creators,
- to describe the data in a structured human-readable form, and
- to bring the existence of the data to the attention of the scholarly community.

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.

A: DATA ACCESS FOR NEWCOMERS AND FOR PRO



Free and open access to biodiversity data

OCCURRENCES

SPECIES

DATASETS

PUBLISHERS

RESOURCES

Search



WHAT IS GBIF?

ABOUT GBIF DENMARK

Polistes olivaceus (DeG. 1773) observed in Cocon (Keele)

Occurrence records
1,896,979,563

Datasets
62,949

Publishing institutions
1,753

Peer-reviewed papers using data
6,390

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 coordinatcleaner 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	X	✓	X	✓ (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	X	✓	X	X	Tab-delimited CSV ?	

DATA DOWNLOADS

Data can be downloaded in three formats

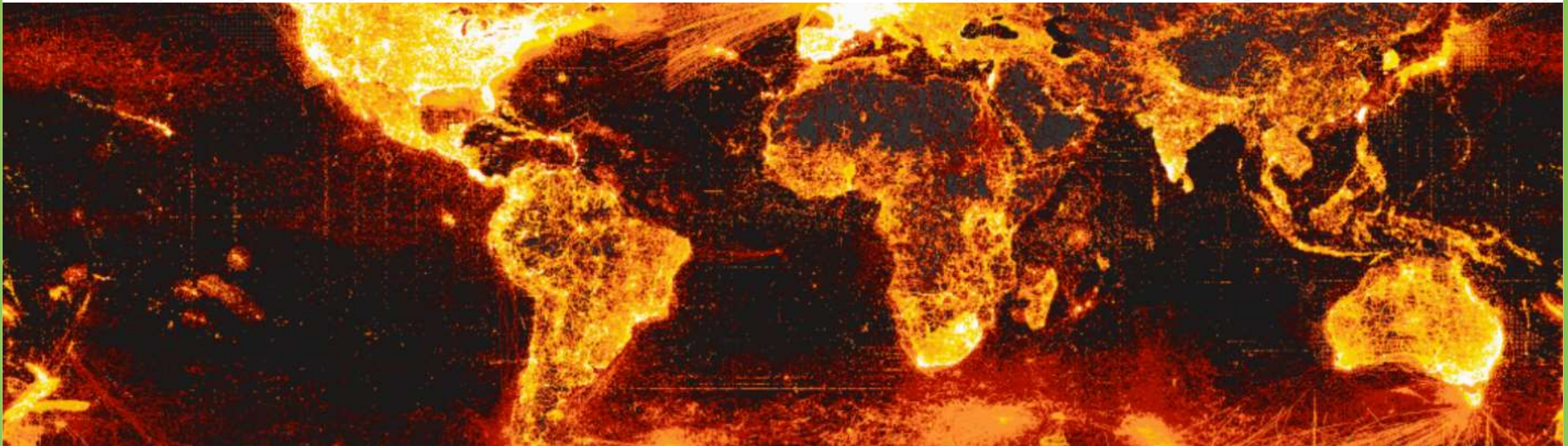
Simple: Tab delimited CSV. Only contains the data after GBIF interpretation. No multimedia included. [More information about CSV](#)

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](#)

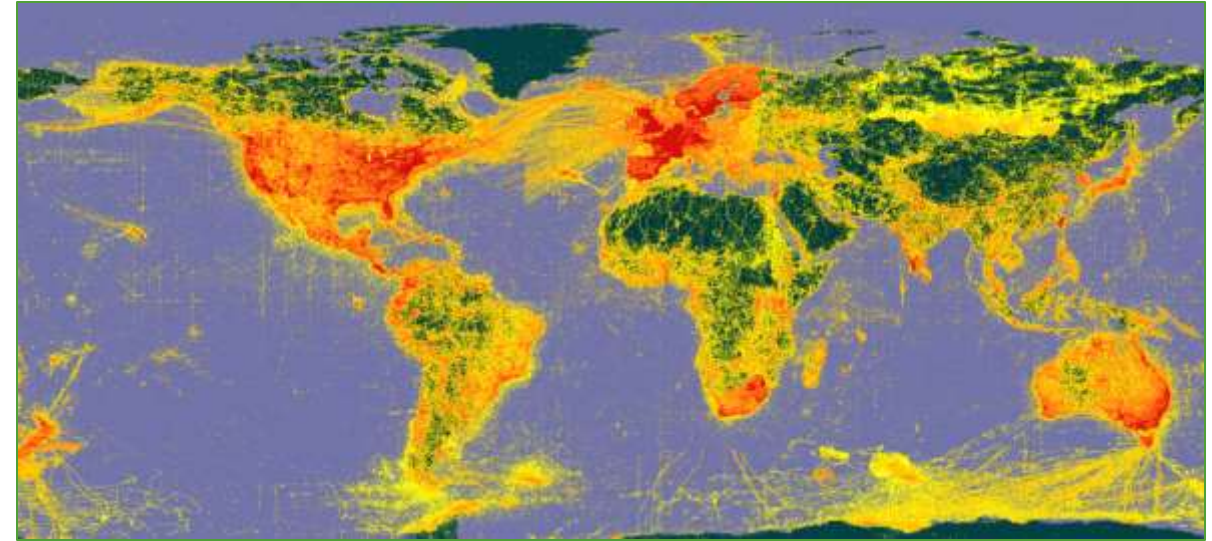
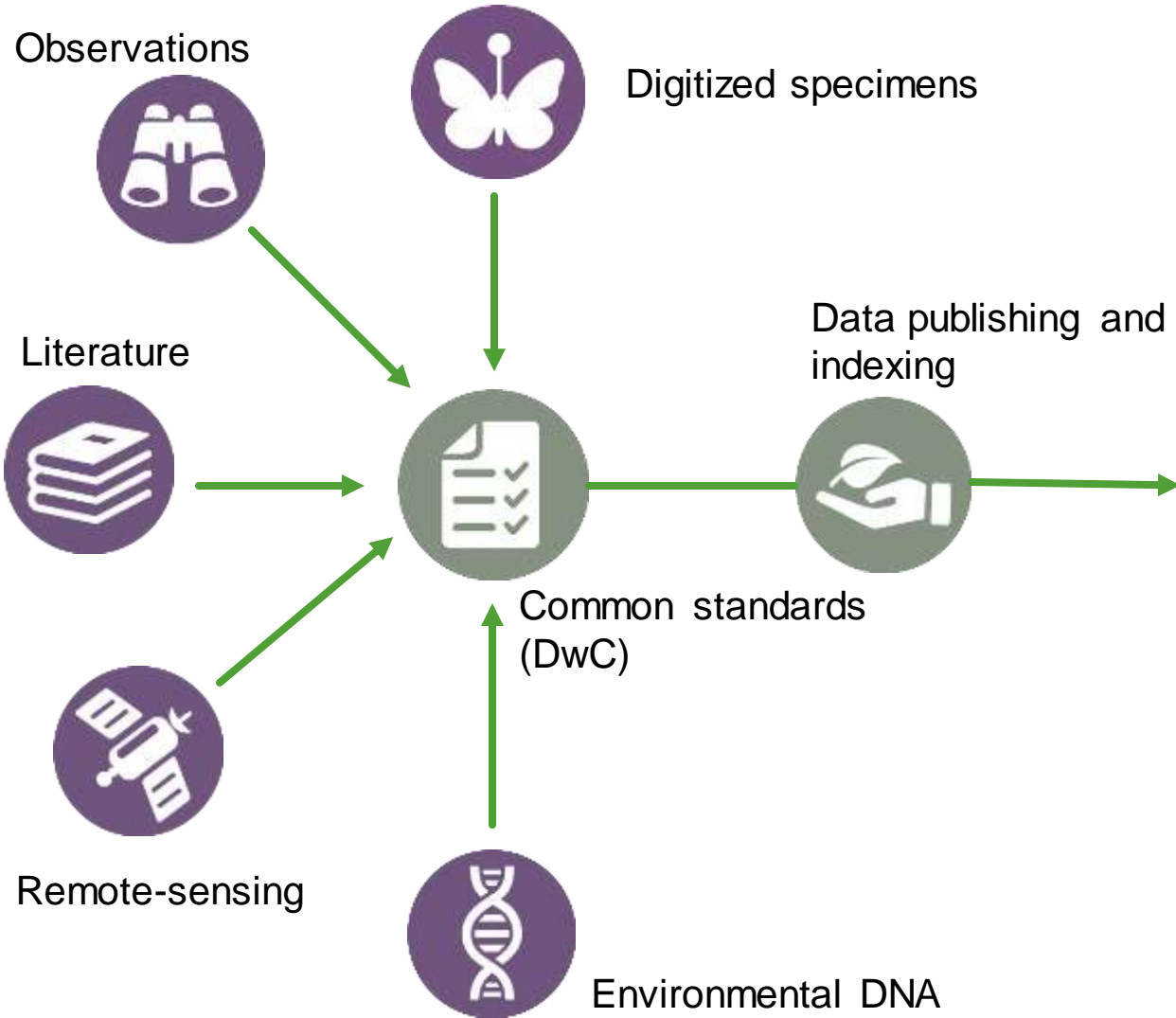
Species list: Tab delimited CSV with the distinct list of names in the search result.



I: DATA STANDARDS AND DATA MODEL



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



Data discovery and use

INTERNATIONAL DATA STANDARDS: DARWIN CORE

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 sta

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/2878937", "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	#MNH#name:1234, NPS_YELL08778 #BG_33424	

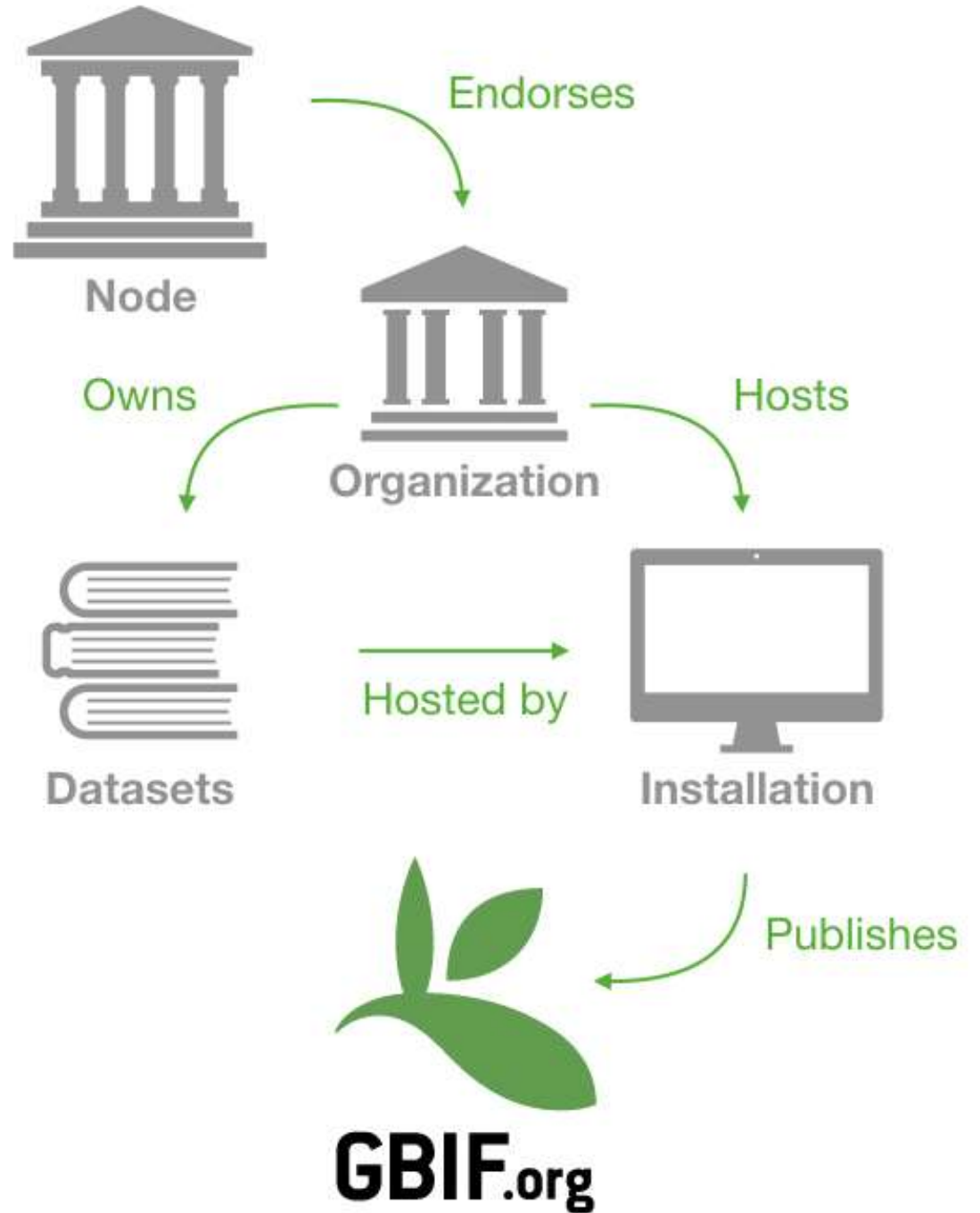
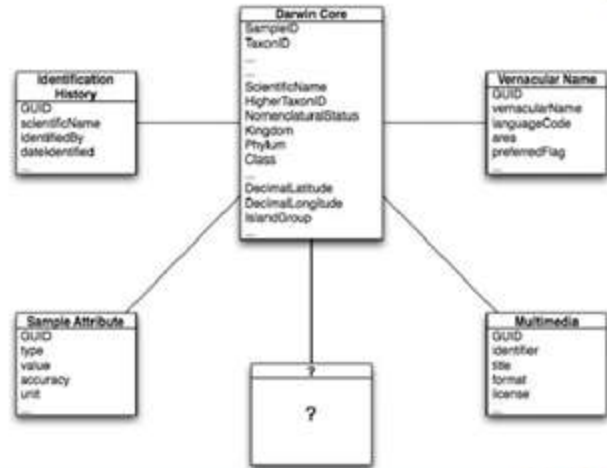
occurrenceRemarks		Property
Identifier	http://rs.tdwg.org/dwc/terms/occurrenceRemarks	
Definition	Comments or notes about the Occurrence.	
Comments		
Examples	found dead on road	

PUBLISHING TO GBIF, THE BASICS

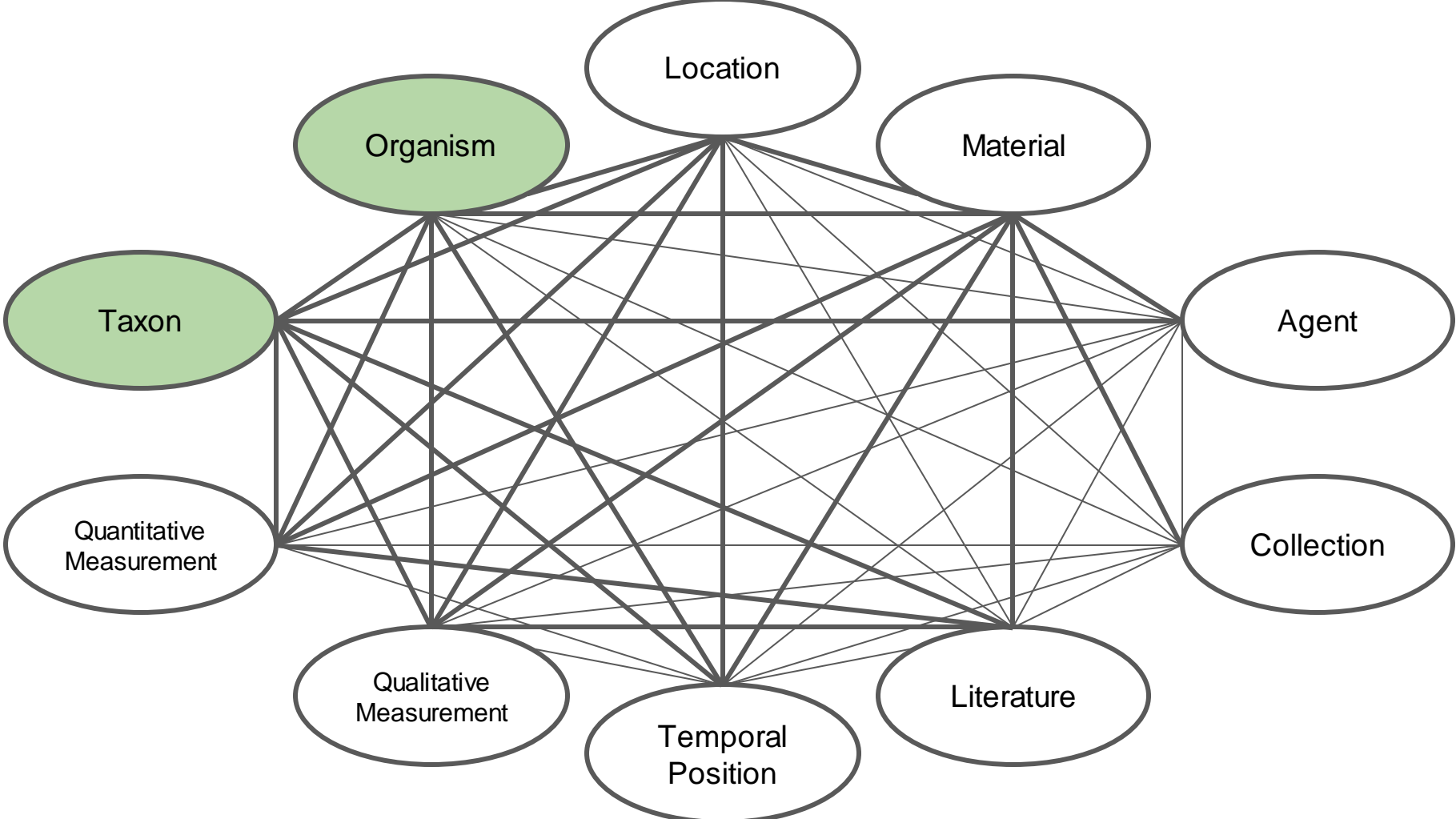
DwC star schema



- ❖ Star schema model
- ❖ Can relate elements one-to-many

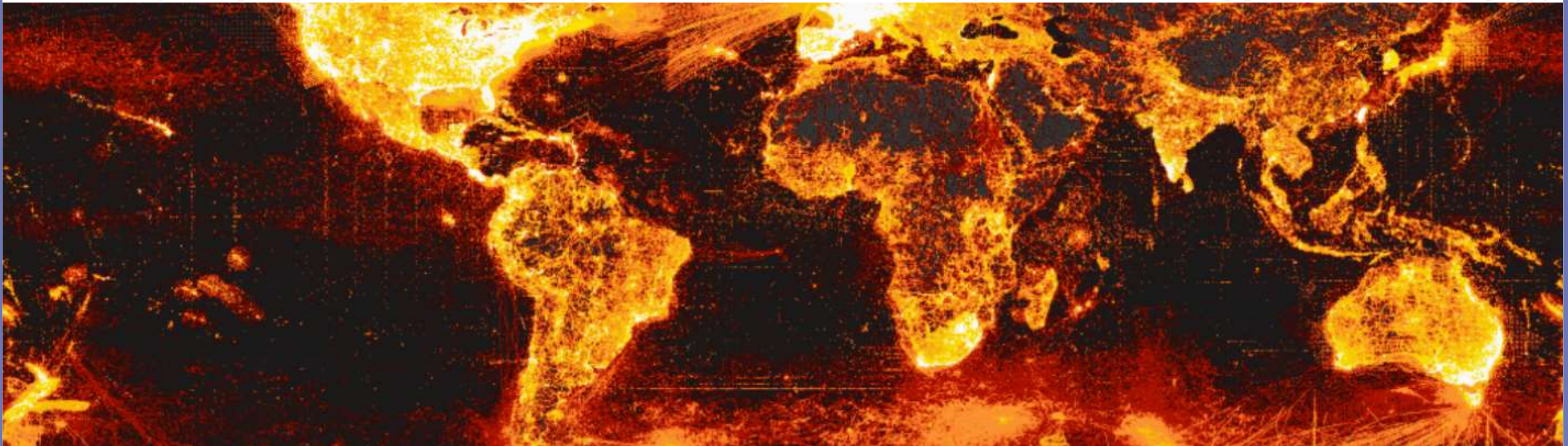


Darwin core thinking

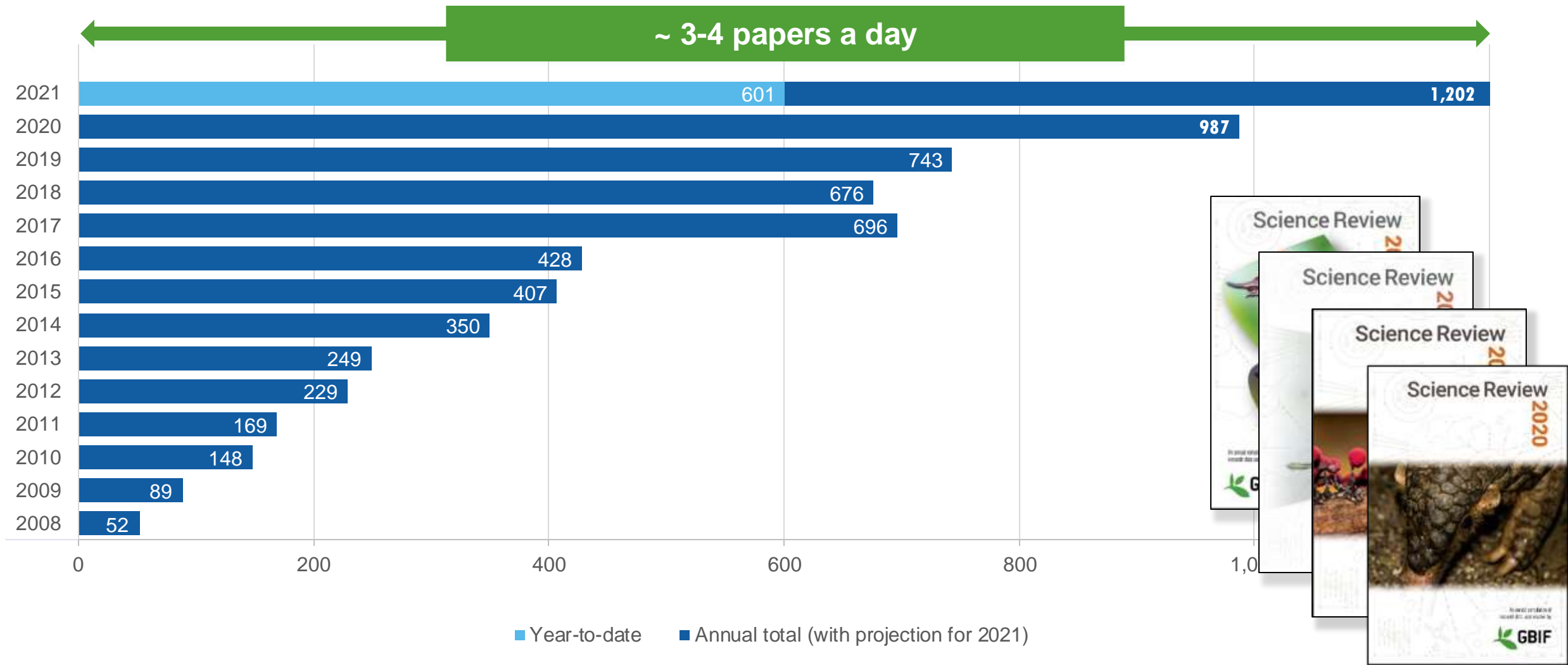




R: DOI BASED DATA CITATION



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





Article

Mapping Disease Transmission Risk of Malaria in South and Southeast Asia

Mark A. Deka * and Niaz Morshed

Department of Geography, Texas State University, 601 University Drive, San Marcos, TX 78682, USA; Email: 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



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.

<https://doi.org/10.1088/1748-9326/aaf3db>

Read Citation

Global Warming

An IPCC special report on the impacts of global warming of 1.5°C above pre-industrial levels and related global energy transformation pathways in the context of strengthening the global sustainable development.



IOP Publishing

Environ. Res. Lett. 14 (2019) 025005

<https://doi.org/10.1088/1748-9326/aaf3db>

Environmental Research Letters

LETTER

Research infrastructure for biodiversity valuation

Alex R Hardisty, Kristen J Williams

- ¹ School of Computer Science
- ² The Atlas of Living Australia
- ³ Global Biodiversity Information
- ⁴ School of Biological Sciences
- ⁵ Biodiversity and Ecosystem
- ⁶ Institute for Biodiversity and
- ⁷ Netherlands



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

SCIENCE ADVANCES | REVIEW

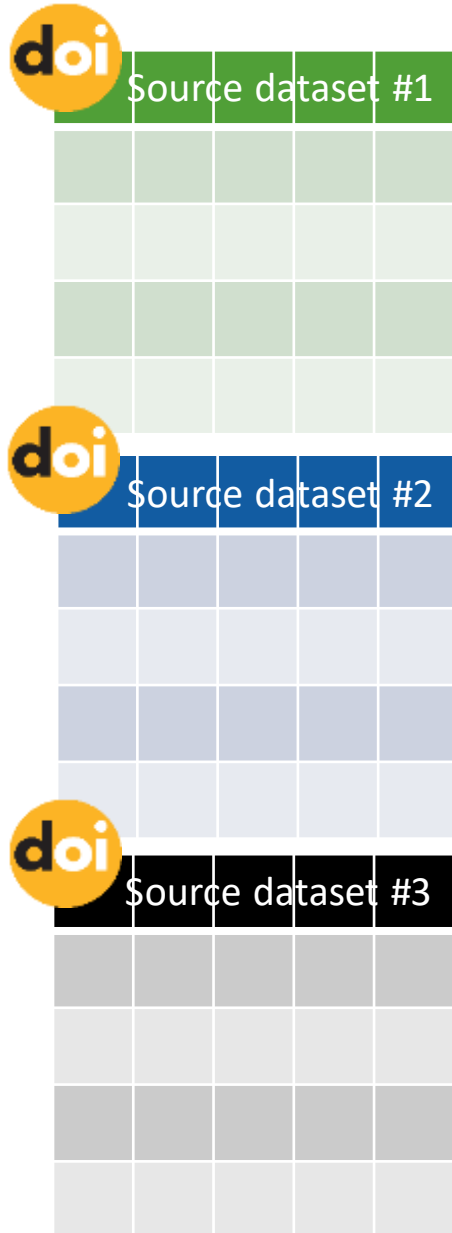
ECOLOGY

Standards for distribution models in biodiversity assessments

Miguel B. Araújo^{1,2,3*}, Robert P. Anderson^{4,5,6}, A. Márcia Barbosa³, Colin M. Beale⁷, Carsten F. Dormann⁸, Regan Early⁹, Raquel A. Garcia^{2,3,10,11}, Antoine Guisan^{12,13}, Luigi Maiorano^{14,15}, Babak Naimi², 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.

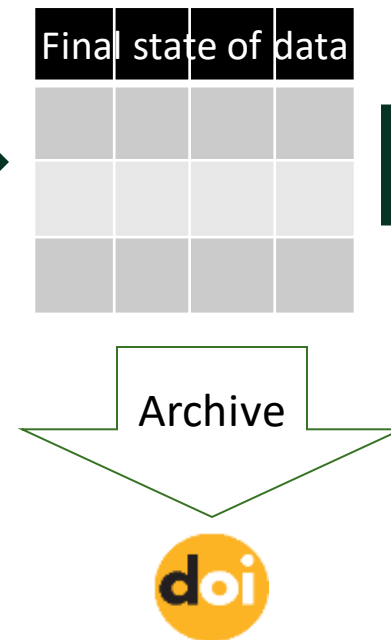
Dataset DOIs



Download DOI



Archive DOI



Bibliographic DOI

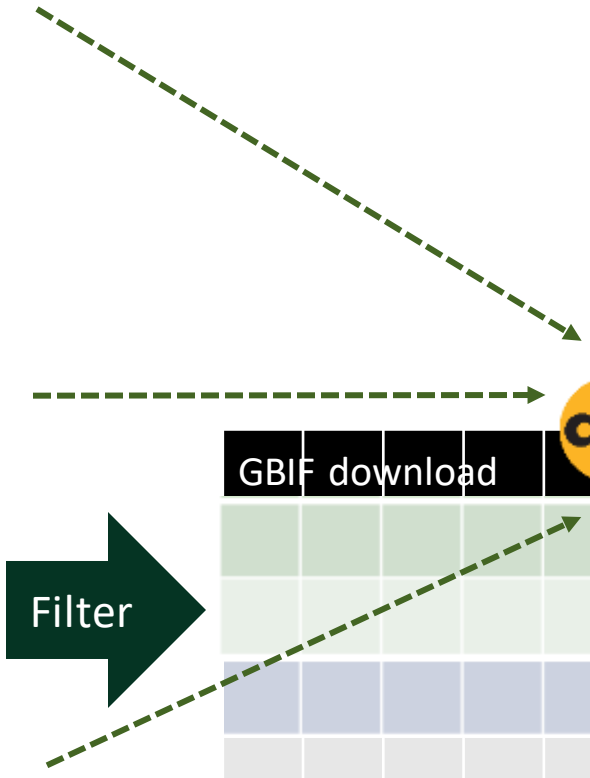
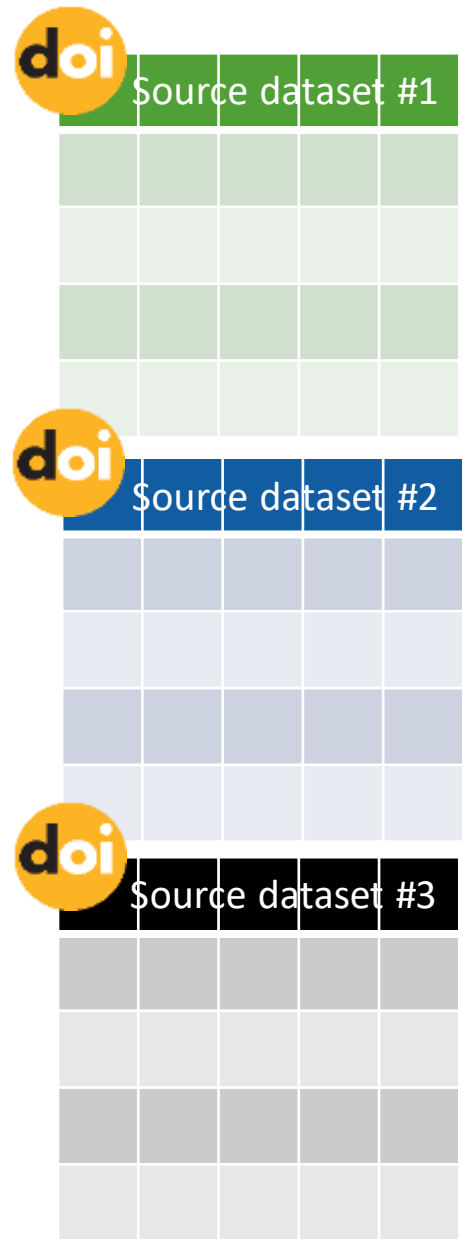


Dataset DOIs

Download DOI

Archive DOI

Bibliographic DOI



DOWNLOAD READY: DATA LINK AND DOI CITATION



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 citation with DOI

DOI BASED DATA CITATION AT GBIF.ORG

Get data How-to Tools Community About

OCCURRENCE DATASET | REGISTERED JULY 2, 2018

Birds and Mammals Collections of the Zoological Museum of M.V. Lomonosov Moscow State University

Published by Lomonosov Moscow State University

Andrey Klyukvin • Anton Morkovin • Mikhail Kalyakin • Pavel Tomkovich • Yaroslav Red'kin • Sergei Kruskop • Vladimir Lebedev

DATASET PROJECT METRICS ACTIVITY DOWNLOAD HOME PAGE

The dataset contains information about Birds and Mammals collections stored in the Zoological Museum of M.V. Lomonosov Moscow State University. In total, these collections include about 220K mammal and 150K bird specimens. A part of its catalogues is available in an electronic form on the museum website (http://zmmu.msu.ru/dbs/search_start.php; type specimen) and on the online portal of the National Depository Bank of Live Systems (<http://animal.depo.msu.ru>). In the course of the project "Support... More

Project ID: Russia-01
Metadata last modified: November 20, 2018
Hosted by: Lomonosov Moscow State University
License: CC BY 4.0
How to cite: DOI 10.15468/164hg0

54,120 Occurrences
99% With taxon match
99.9% With coordinates
98% With year

54,050 GEOREFERENCED RECORDS

Landscape Analysis for the Spicespin Data Refinery

Waters, S. L., ... (2022) Research Ideas and Outcomes

This report reviews the current state of the art applied approaches on ecosystem data, services and workflows for extracting information from images of natural history specimens and their labels. We consider the potential for repurposing existing tools, including workflow management systems, and AI.

Abstract: ...

Journal article | Open access | Peer-reviewed

Date released in GBIF: 10.15468/164hg0

15 CITATIONS

54,120 OCCURRENCES

SpringerLink

Original Paper | Published: 10 March 2021

Modelling risks posed by wind turbines and power lines to soaring birds: the black stork (*Ciconia nigra*) in Italy as a case study

Stano Stanković, Luciano Sacco, ...

Abstract and Conservation: 25 | 1939–1976(2021) | Cit this article

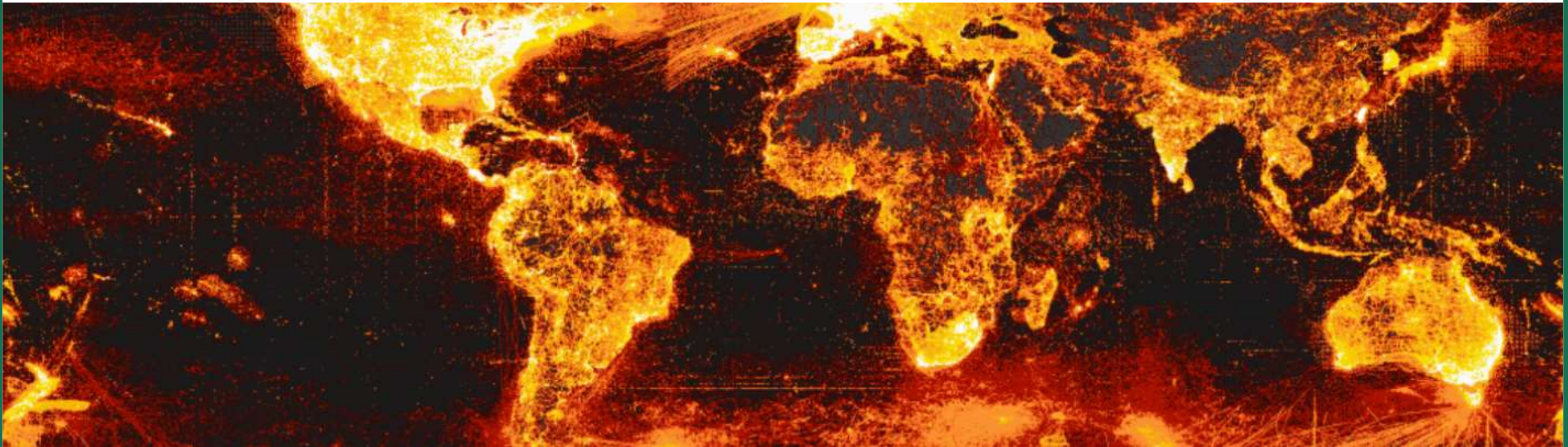
387 Accesses | 9 Citations | 14 Abstracts | Metrics

Abstract

Recent growth of investments in wind energy and power industries has increased concerns about the associated adverse impacts on wildlife. In particular, flying vertebrates are especially at risk, both directly, through an extra mortality rate due to collision with turbines and electrocution, and indirectly through habitat loss or fragmentation. In this study, we propose a modelling approach that combines species distribution models and data managed in geographic information systems to predict and quantify the effects of wind turbines and power lines on the breeding habitat of a soaring migratory bird, the black stork *Ciconia nigra*, in Italy. The species is recolonizing the country, where it had been driven to extinction in the Middle Age by human persecution. Today, infrastructures such as those considered in our study might in fact hamper this recolonization. Our results show a high probability of presence of the species in several areas in Italy. The most important variables in influencing habitat suitability for *C. nigra* are the mean temperature of May followed by the distance from urban areas, inland wetlands and hydrographic network. Exposure to wind turbine collision and electrocution resulted to be potentially high. In particular, in Northern Italy the main potential risk of mortality for *C. nigra* is posed by power lines, whereas in southern regions the species

An abstract background featuring a network of thin grey lines and circles of various sizes. A prominent sunburst-like shape is centered behind the text. To the right, there are faint outlines of a ruler and a microscope. A solid green horizontal line is positioned above the text.

CONCLUSION



Data sharing: the **when** choices

SHORT TERM

LONG TERM



**RESEARCH
PHASE**

- file formats
- ownership
- metadata
- storage
- backups

**DISSEMINATION
PHASE**

- share with whom?
- embargo?
- licensing
- metadata

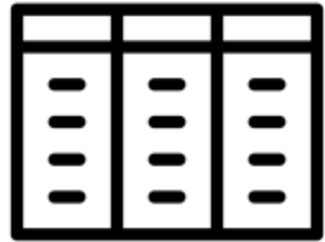
**PRESERVATION
PHASE**

- repository?
- long-term manager?

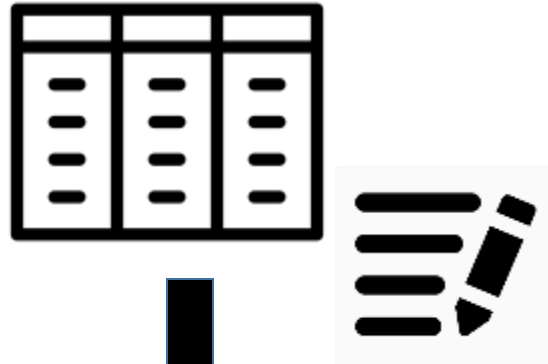
Data sharing: the **where** choices



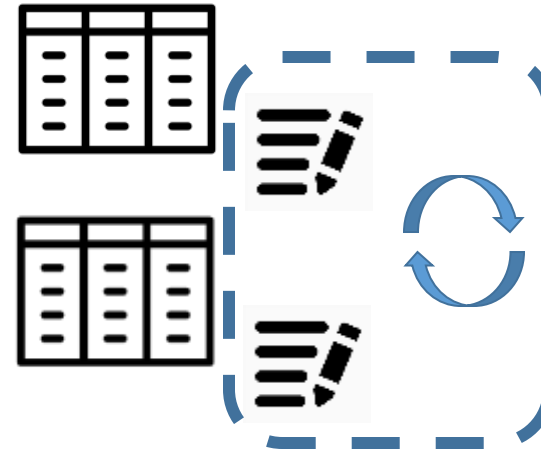
Archive



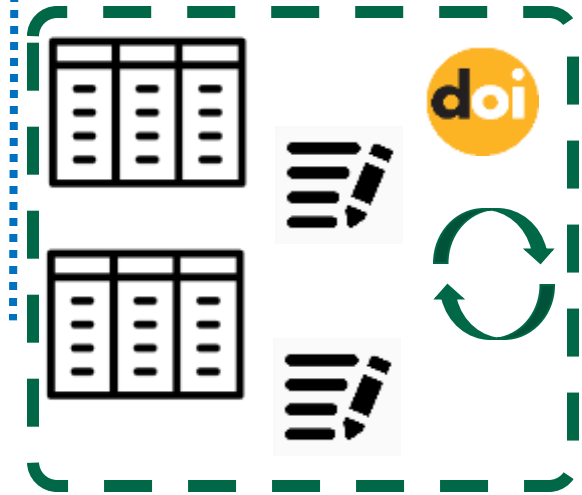
Generalist repository



Data catalogue



Data index



Preservation

Open data

FAIR data

Save

Minimum description

Metadata standartization

Data standartization

Open data and FAIR data

F
Findable



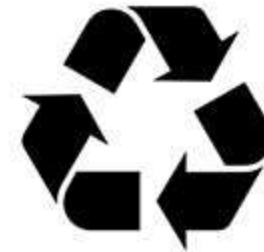
A
Accessible



I
Interoperable



R
Reusable



Dmitry Schigel
dschigel@gbif.org

@dschigel
adlignum.org

