GBIF Global Biodiversity Information Facility

Ectatomma tuberculatum observed in Peru by manimiranda.
Photo via iNaturalist (CC BY-NC 4.0)

Lowering the technical threshold to participate

Tim Robertson

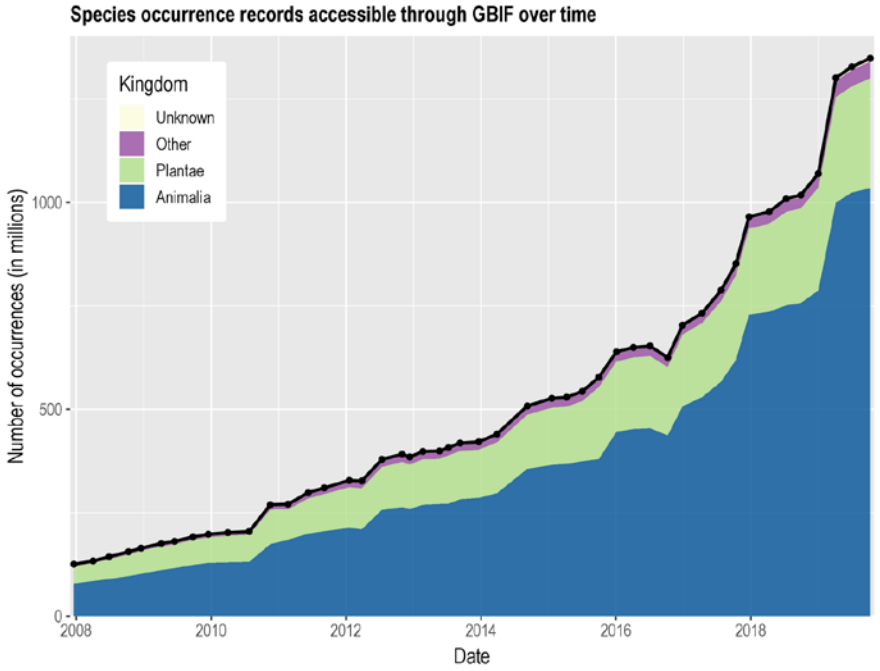
GLOBAL NODES MEETING | LEIDEN | OCTOBER 2019

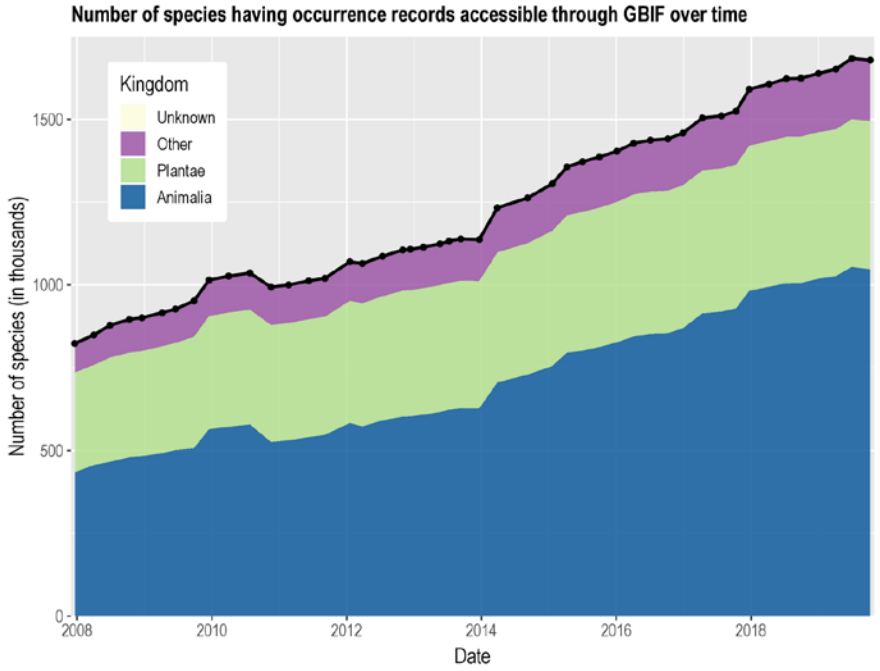
SESSION CONTENTS

1. Growth (and some reflections)
2. Managing content in GBIF
3. Repositories
4. Broadening our data model
5. “Powered by GBIF”: Making the most use of our infrastructure

Growth (and some reflections)

GROWTH IN VOLUME AND VARIETY OCCURRENCE DATA

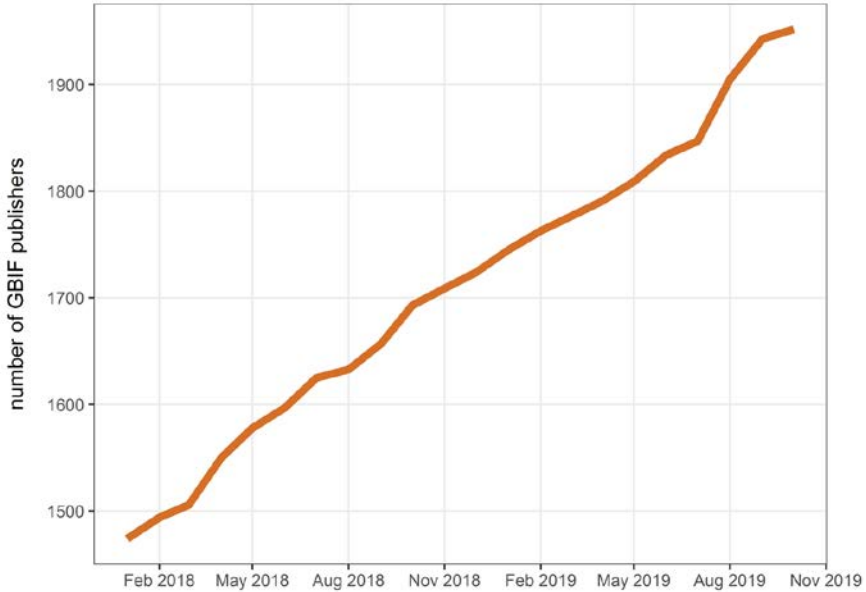


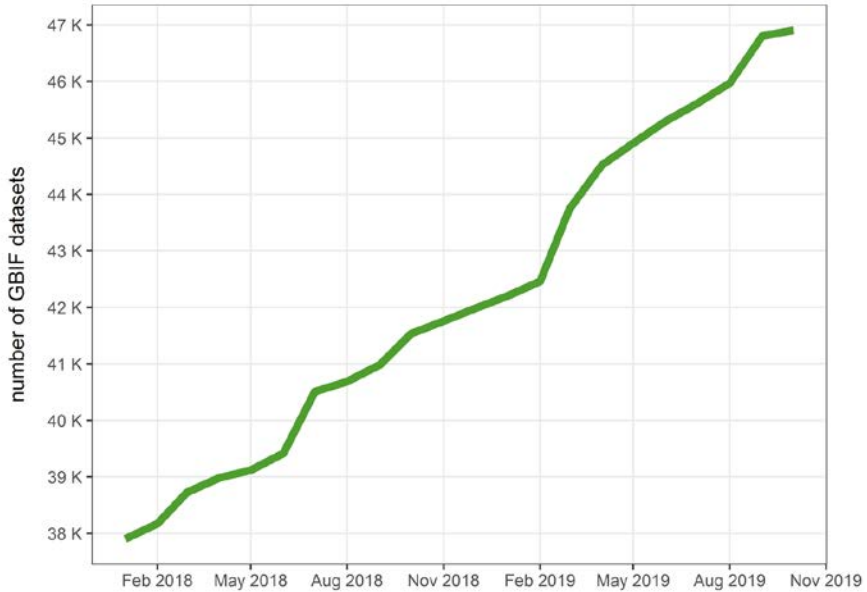


https://www.gbif.org/analytics/global

GBIF

GROWTH IN PUBLISHERS AND DATASETS





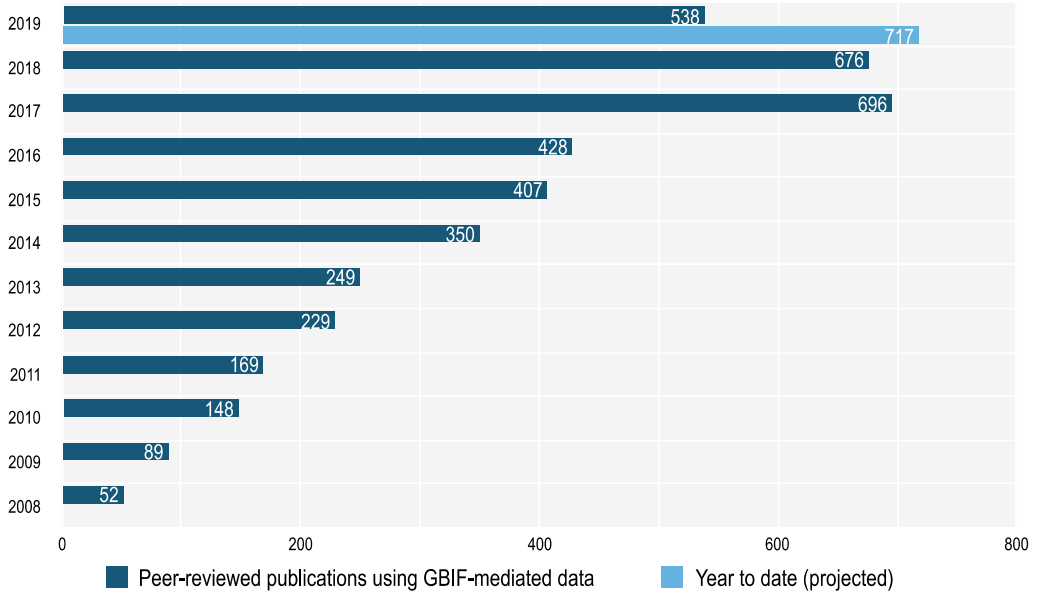
Source: GBIF Registry

GBIF

GROWTH IN USE

Users of GBIF.org per week

~150k per month now

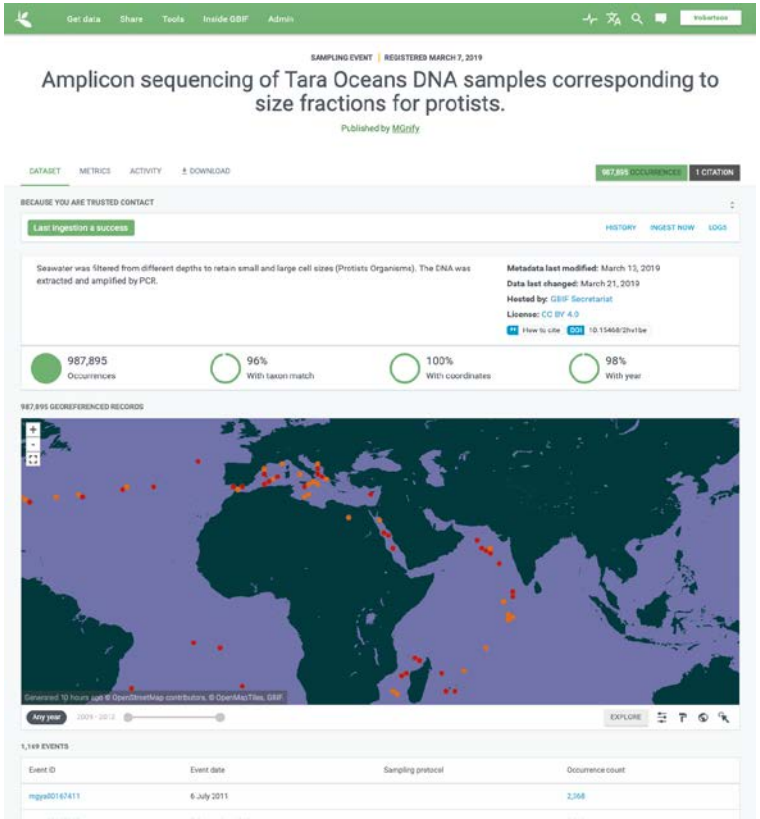


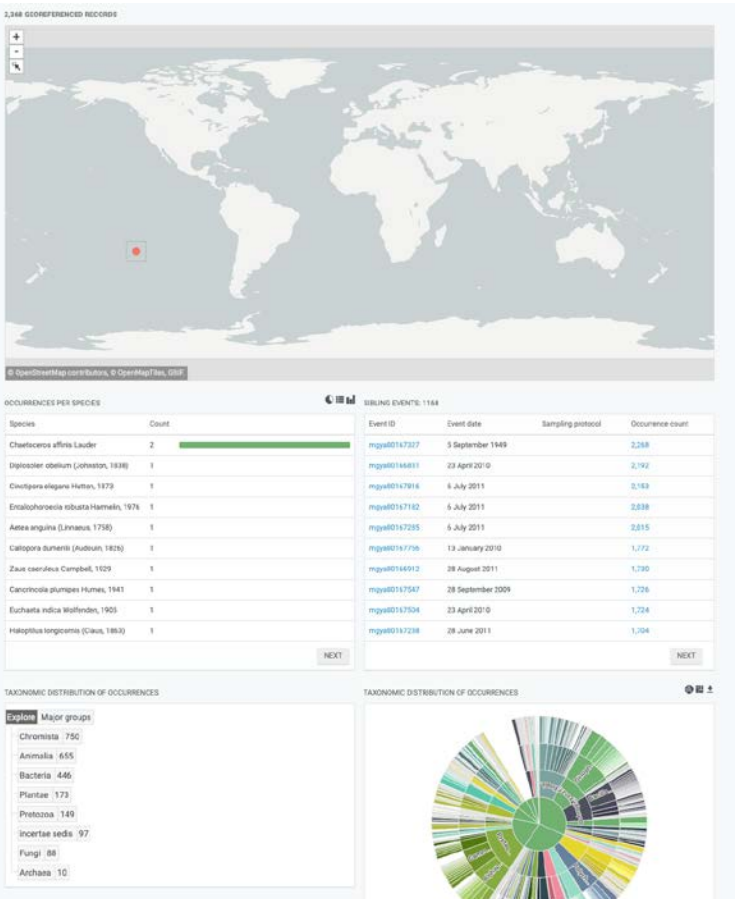
Citations of GBIF mediated data

Source: Google Analytics and GBIF tracking

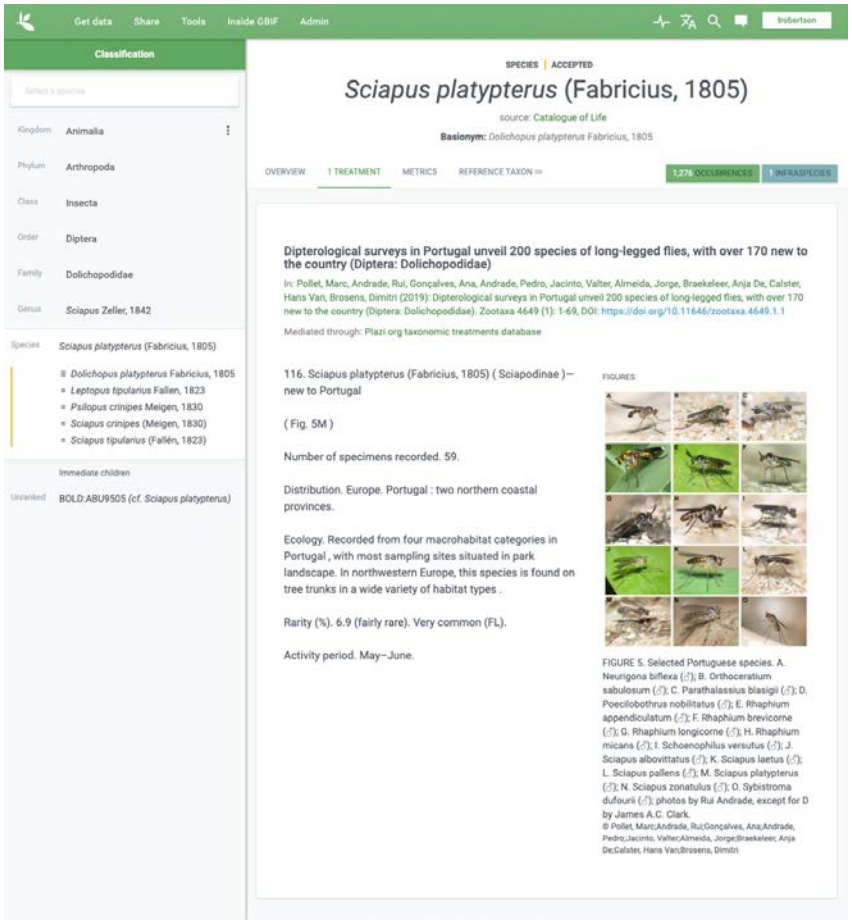
GBIF

METAGENOMIC DATA





TAXONOMIC TREATMENTS

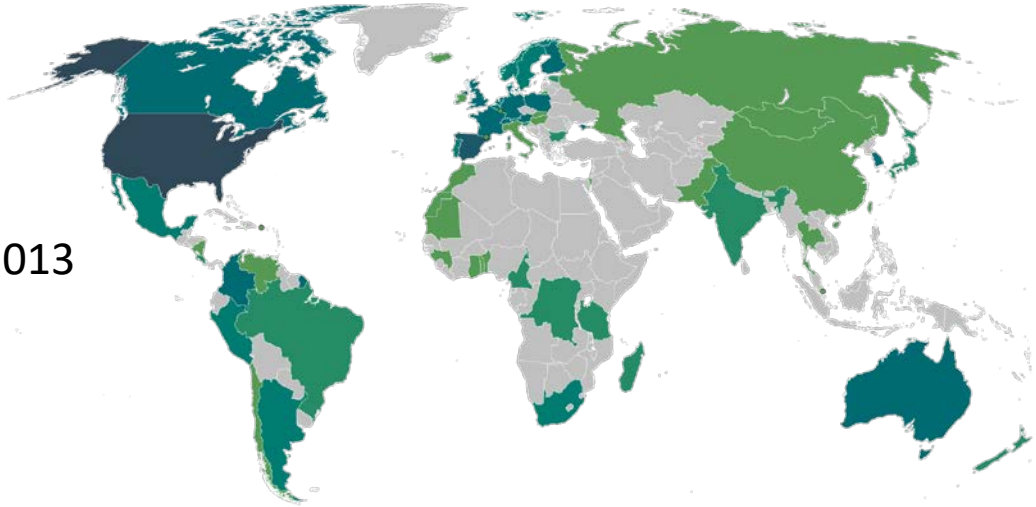


Mediated through Plazi:

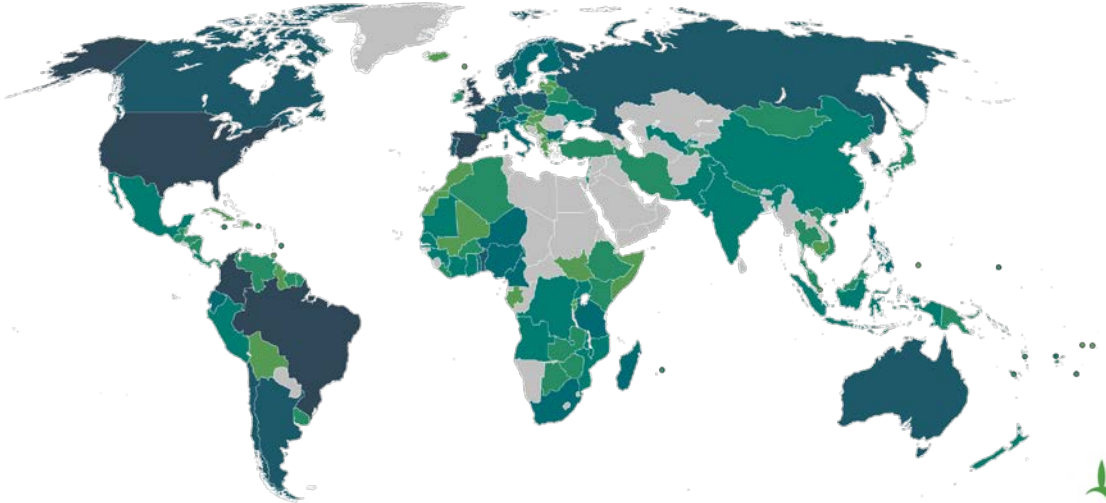
1. 286k treatments at genus or more specific rank
2. connect to 212k concepts in the GBIF backbone
3. 48k concepts **are only supplied by Plazi** - largely representing newly described species

COUNTRIES WITH DATA PUBLISHERS

September 2013

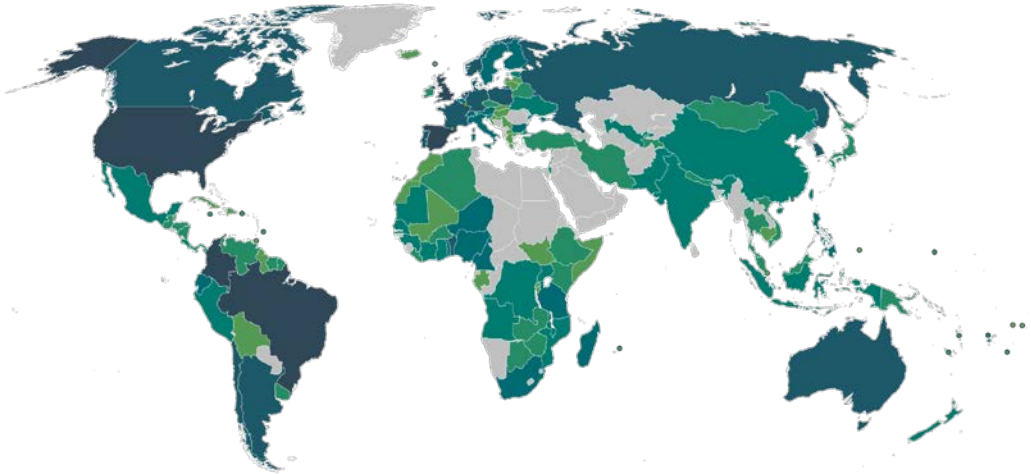


Today



GBIF

REFLECTIONS

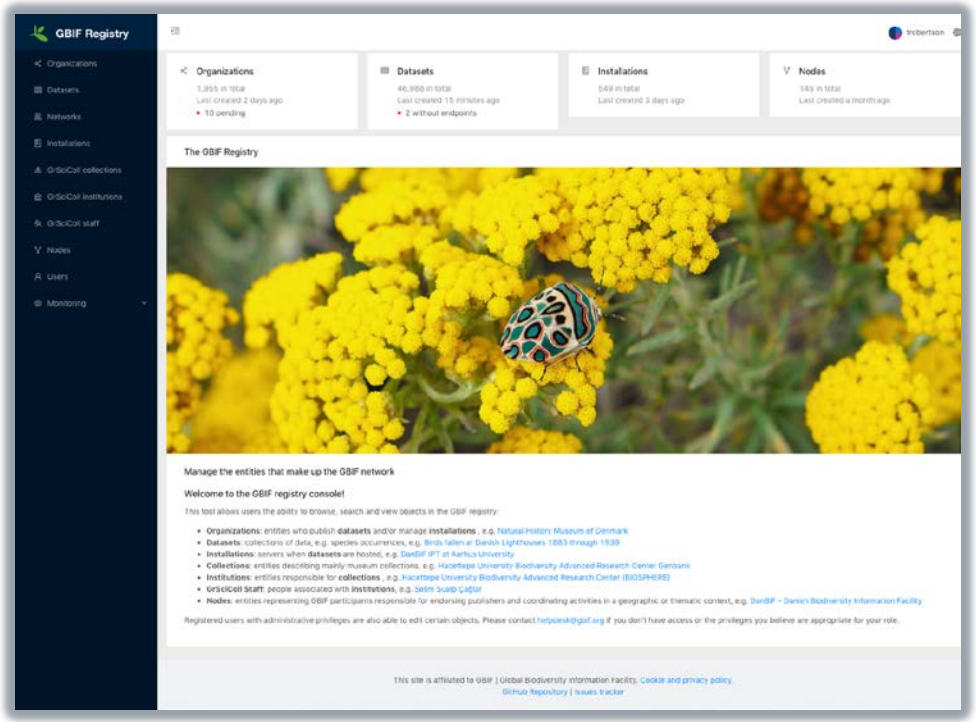


1. Very positive trends
2. Increased costs
    i. Helpdesk and support
    ii. More powerful hardware
    iii. Data hosting
    iv. Development effort
    v. Engagement activities
    vi. Translations

There are growing calls to lower the technical threshold to operate

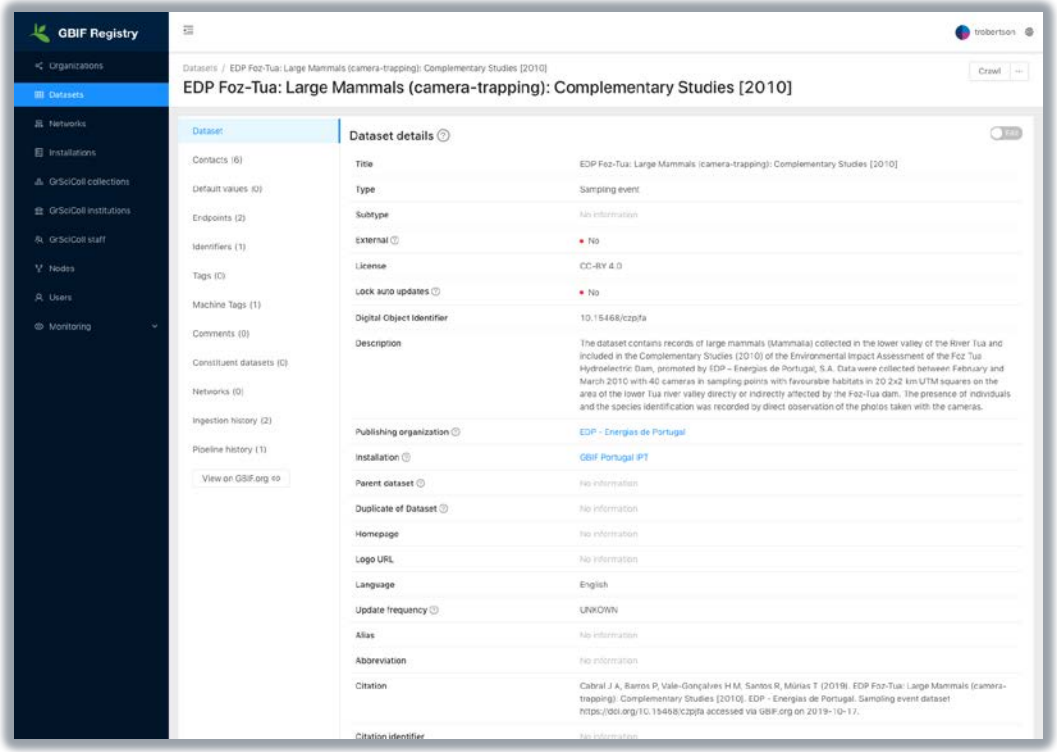
Managing content in GBIF

OPEN REGISTRY



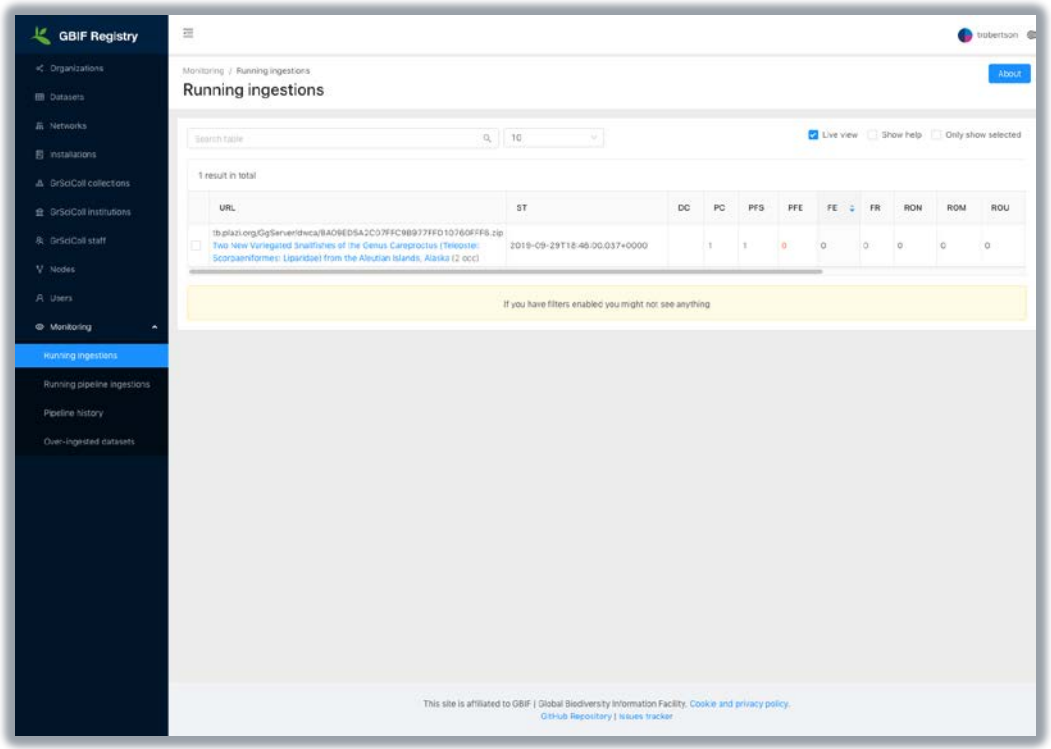
https://registry.gbif.org

OPEN EDITING



1. Manage Institutions, Datasets, Collections etc
2. Editor permission scoped to Country, Organisation or Installation
3. All data exposed through an API

OPERATE THE CRAWLING



1. Control data ingestions
2. View all data ingestion diagnostic logs
3. Apply default values (e.g. *kingdom=Plantae*)
4. Future:
   i. Manage the vocabularies to which we organise data (e.g. *habitat*)
   ii. Richer metadata, such as Collection Description

LOWER TECHNICAL THRESHOLD

1. Opportunity to reuse infrastructure
   i. “Use GBIF” rather than “connect to GBIF”
   ii. Optimise (future) development effort
   iii. Efficiencies in data management (e.g. more effort on data “cleaning”)

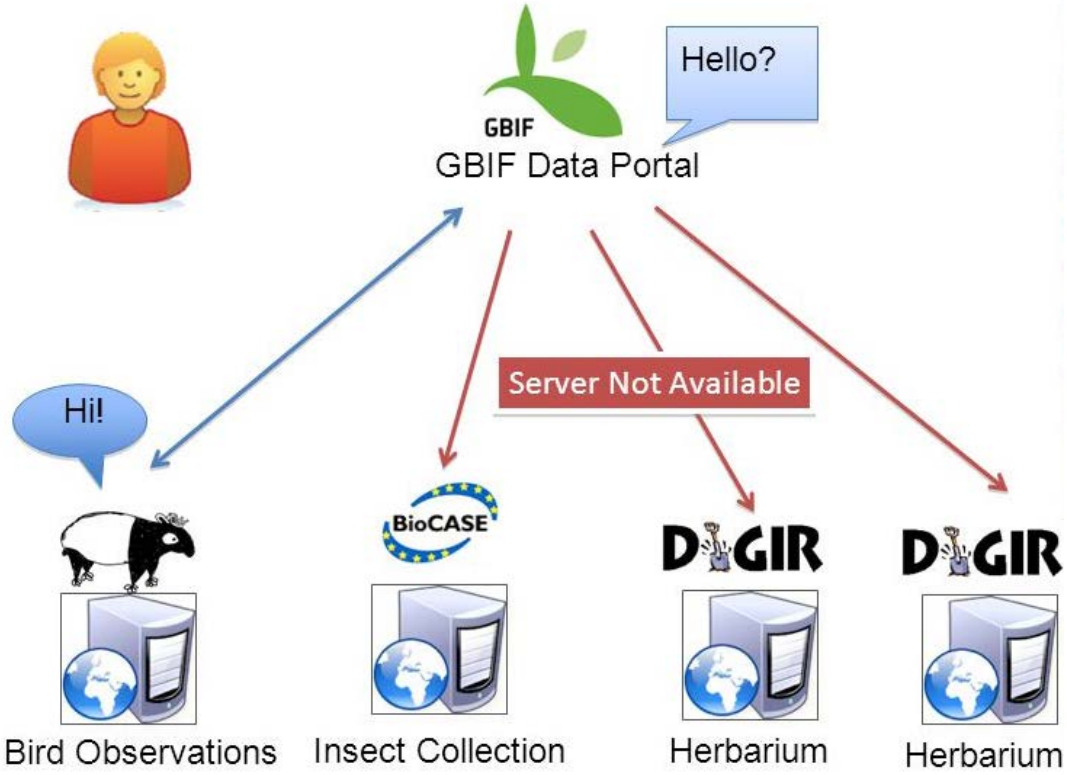
2. Commitment from Australia to explore shared infrastructure:

   “ALA will adopt the GBIF registry. As part of this we suggest the functionality of the GBIF registry be expanded to provide hosted IPT style”

Repositories

BACK IN THE DAY…

Circa 2008



Graphic reused from old GBIF presentation

“MICRO REPOSITORIES”

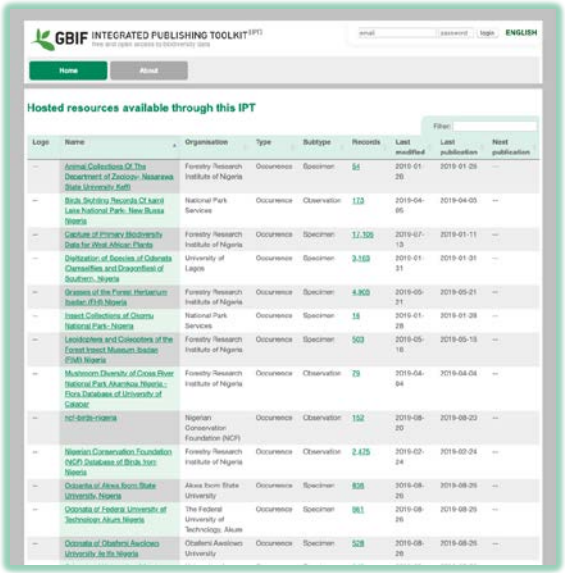
IPT: The Integrated Publishing Toolkit

A free open source software tool used to publish and share biodiversity datasets through the GBIF network.

Don't see your IPT? Send GBIF your coordinates.

GBIF

CO-LOCATED REPOSITORIES (EXAMPLES)



SiB Colombia

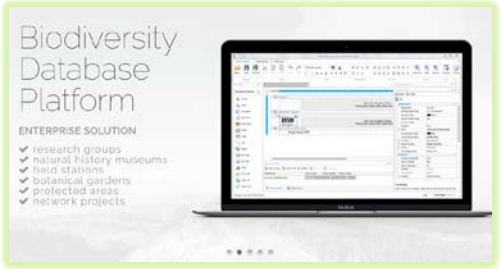
gbif France

Nigeria

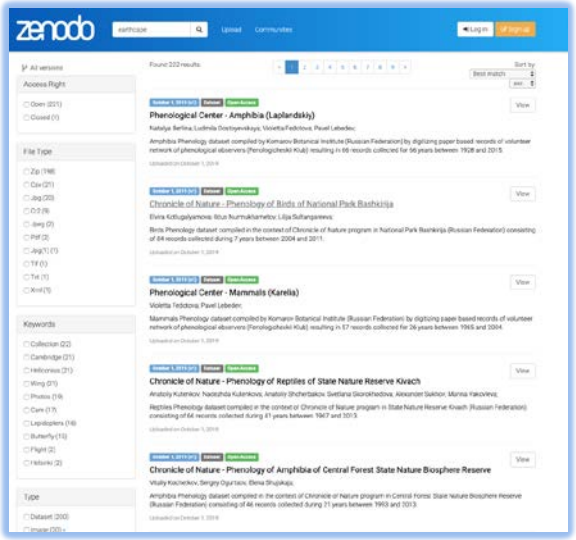
GBIF

PARTNERSHIPS USING OPEN REPOSITORIES

Recent example of connecting infrastructure



earthcape



GBIF

Dedicated,
domain-specific tools

General purpose open
science repository

Understanding this trend is important for our activities…

1. For the long-term future of the IPT

2. Should GBIF seek strong partnerships (e.g. Zenodo)?

3. Do you wish GBIF to more prominently offer repository services?

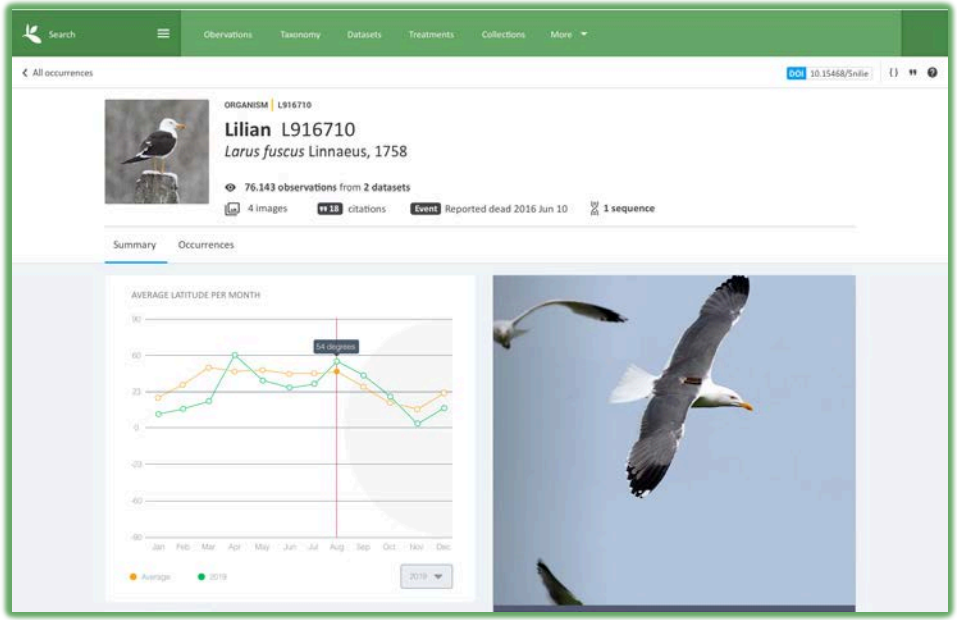
4. How does your node see this landscape?

Broadening our data model

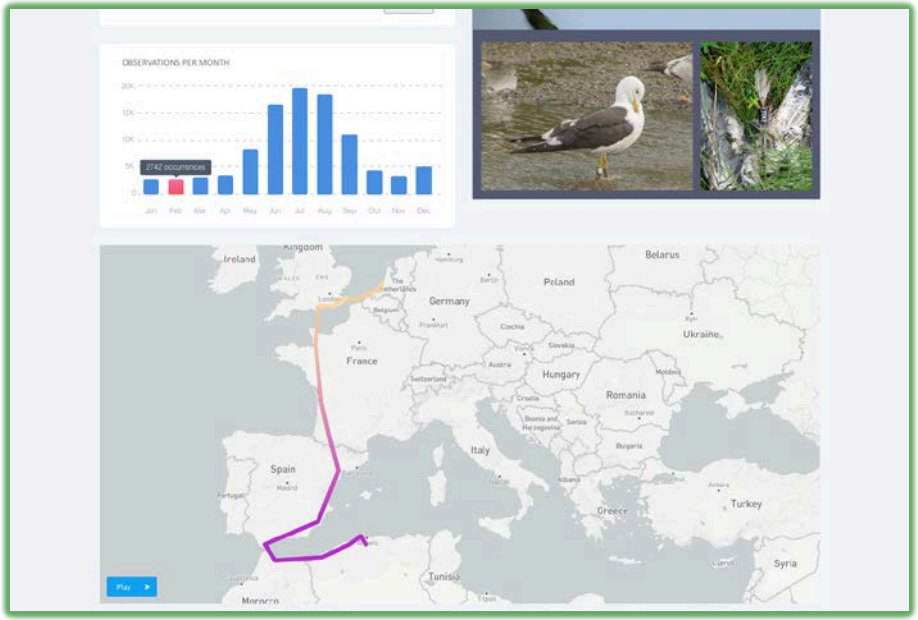
Images on https://labs.gbif.org/visual-concepts

”During 2020 we will focus on advancing and refining data models for Collections, Taxonomic treatments, Sampling Events, Organisms, Specimens, Citations and the **linkages between them**.”

EXAMPLE 1: ORGANISM

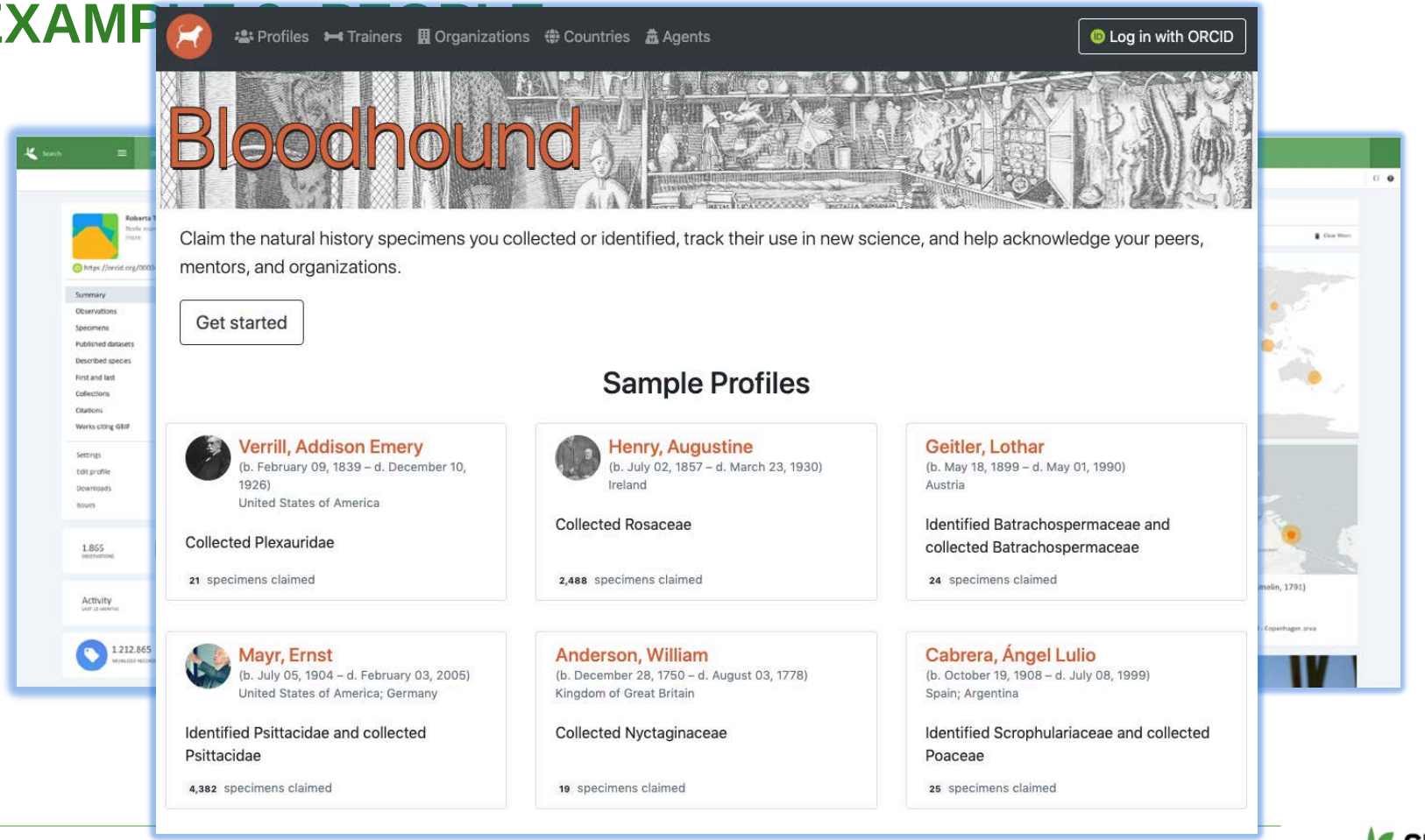


Simpler navigation to entity types



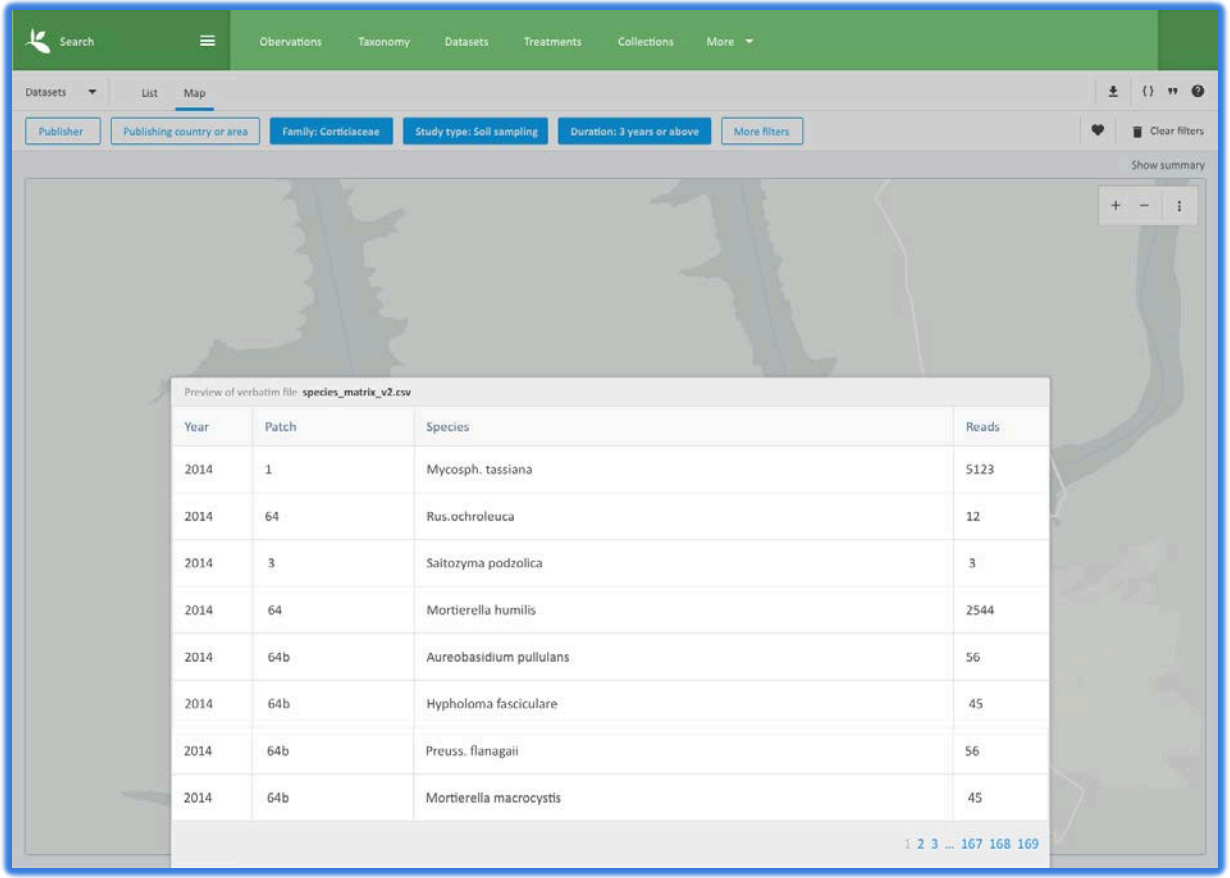
https://labs.gbif.org/visual-concepts

EXAMP



See https://labs.gbif.org/visual-concepts

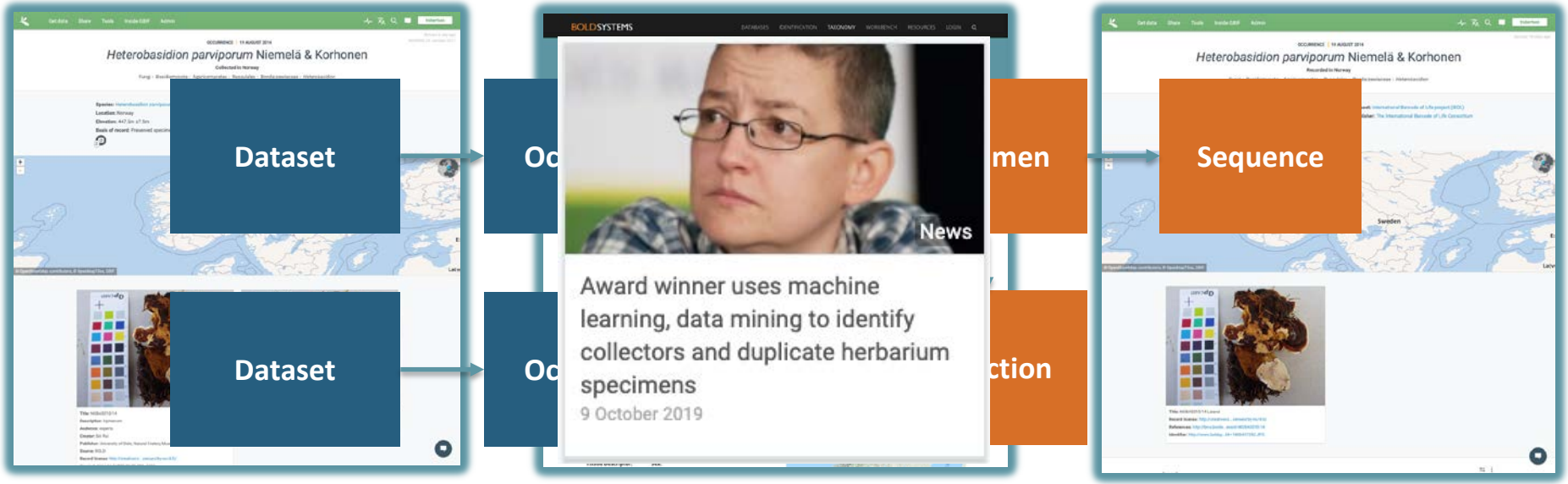
EXAMPLE 3: DISCOVERING SAMPLE BASED DATA



See https://labs.gbif.org/visual-concepts

GBIF

EXAMPLE 4: LINKING SEQUENCES TO SPECIMENS



Museum record

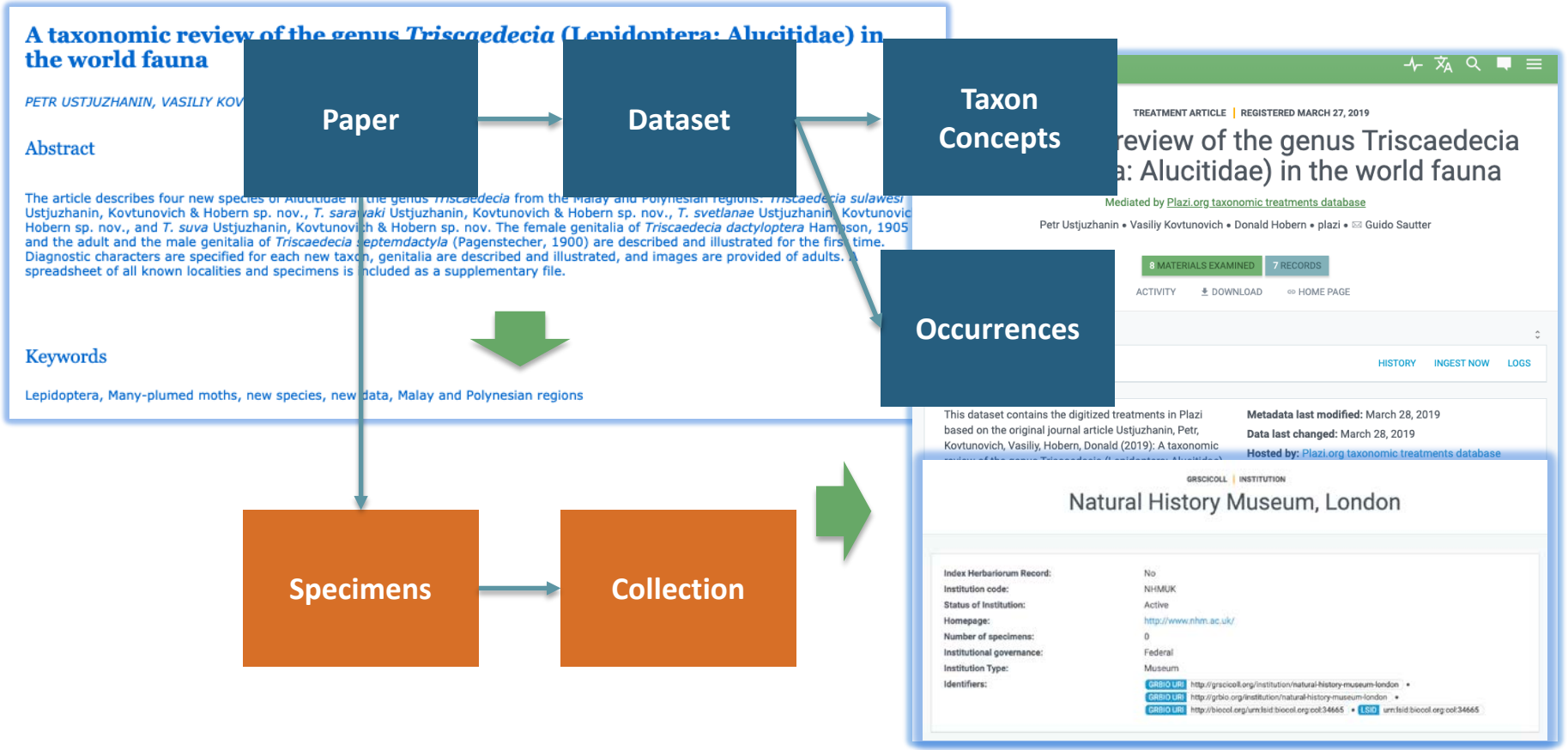
Sequenced in BOLD

Republished

See https://www.gbif.no/news/2017/bold.html

GBIF

EXAMPLE 5: MEDIATING TAXONOMIC TREATMENTS



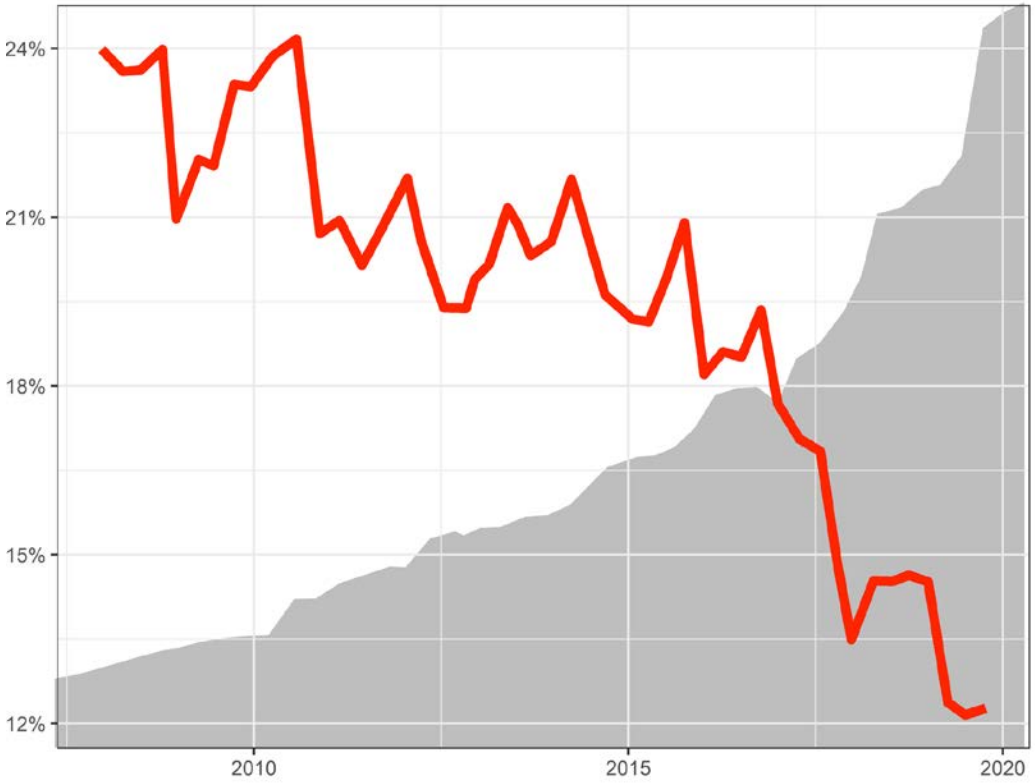
“Powered by GBIF”:
Making the most use of our infrastructure

Images on https://labs.gbif.org/visual-concepts

”…design and implement user interfaces and services … to support simple hosted portals.

BID programme portal
exploration of country portals
an exemplar virtual natural history collection…”

DECLINE IN VISIBILITY OF SPECIMEN DATA



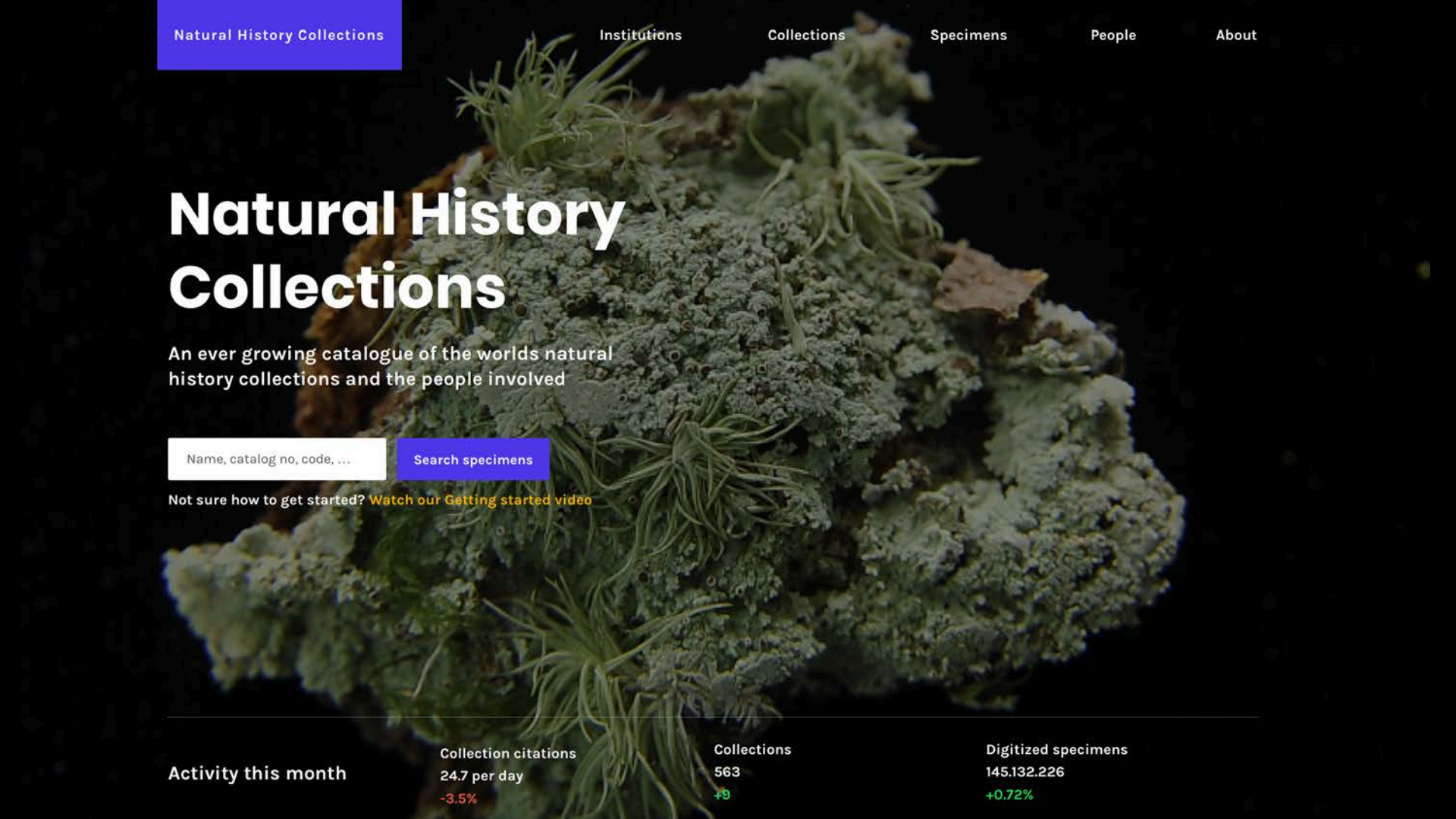
Percentage of GBIF.org originating from specimens

2010: 23%

2015: 21%

2017: 15%

2019: 12%



Natural History Collections

Institutions Collections Specimens People About

Natural History Collections

An ever growing catalogue of the worlds natural history collections and the people involved

Name, catalog no, code, ... Search specimens

Not sure how to get started? Watch our Getting started video

Activity this month

Collection citations
24.7 per day
-3.5%

Collections
563
+9

Digitized specimens
145.132.226
+0.72%

Natural History Collections | Institutions | Collections | Specimens | People | About

Collections | List | Map

Hacet | Species or group | Institution | France | Digitization state | More filters | Clear filters

1.302 results

Hacettepe University Biodiversity Advanced Research Center Zoology collection

From Hacettepe University
71.266 specimens (32% digitized)
Major groups **Nudibranchia • Polychaeta • Malacostraca**

Canadian Museum of Nature Mollusc Collection

From Hacettepe University
71.266 specimens (32% digitized)
Major groups **Nudibranchia • Polychaeta • Malacostraca**

Nematoda Collection - Instituto Nacional de Pesquisas da Amazônia (INPA)

From Hacettepe University
71.266 specimens (32% digitized)
Major groups **Nudibranchia • Polychaeta • Malacostraca**

Distribution de quelques espèces végétales du Bénin

From Hacettepe University
71.266 specimens (32% digitized)
Major groups **Nudibranchia • Polychaeta • Malacostraca**

Hacettepe University Biodiversity Advanced Research Center Zoology collection

From Hacettepe University
71.266 specimens (32% digitized)
Major groups **Nudibranchia • Polychaeta • Malacostraca**

DOI 10.15468/5nille

COLLECTION | ZOOLOGY

Hacettepe University Biodiversity Advanced Research ...

... Center Zoology Collection

From Hacettepe University

- 71.266 specimens (32% digitized)
- Major groups **Magnolia • Protozoa**
- Jane Franklin • Viggo Mortensen

Journal articles
Treatments
Downloads
How to cite this collection

2 results | Add filter

Range-expansion effects on the belowground plant microbiome
Ramirez, K. Snoek, L. Koorem, K. Geisen, S. Bloem, L. ten Hooven, F. ... - (2019) Nature Ecology & Evolution

Plant range expansion is occurring at a rapid pace, largely in response to human-induced climate warming. Although the movement of plants along latitudinal and altitudinal gradients is well-documented, effects on belowground ...

Cites collection | Cites 2 specimens

Range-expansion effects on the belowground plant microbiome
Ramirez, K. Snoek, L. Koorem, K. Geisen, S. Bloem, L. ten Hooven, F. ... - (2019) Nature Ecology & Evolution

Plant range expansion is occurring at a rapid pace, largely in response to human-induced climate warming. Although the movement of plants along latitudinal and altitudinal gradients is well-documented, effects on belowground ...

Cites collection | Cites 2 specimens

Range-expansion effects on the belowground plant microbiome
Ramirez, K. Snoek, L. Koorem, K. Geisen, S. Bloem, L. ten Hooven, F. ... - (2019) Nature Ecology & Evolution

Plant range expansion is occurring at a rapid pace, largely in response to human-induced climate warming. Although the movement of plants along latitudinal and altitudinal gradients is well-documented, effects on belowground ...

Cites collection | Cites 2 specimens

DOI 10.15468/5nilie

COLLECTION | ZOOLOGY

Hacettepe University Biodiversity Advanced Research ...

... Center Zoology Collection

From Hacettepe University

- 71.266 specimens (32% digitized)
- Major groups **Magnolia • Protozoa**
- Jane Franklin • Viggo Mortensen

About | People | Citations | Digitized specimens | More

Afroboletus Pegler & T.W.K.Young | Where | When | Evidence | Quality | More filters | List

2.660.144 results

HORNED PONDWEED

***Lepilaena cylindrocarpa* Benth.**

Holotype

Plantae › Tracheophyta › Liliopsida › Alismatales › Potamogetonaceae › Lepilaena

13 July 2015 | Taiwan | 2 images | 16 citations | Is sequenced

DOI 10.15468/5nilie | L : Botany : L.1196509 | + 2 more

HORNED PONDWEED

***Hyphaene guineensis* Schumach. & Thonn.**

Plantae › Tracheophyta › Liliopsida › Alismatales › Potamogetonaceae › Lepilaena

13 July 2015 | Estonia | 16 citations

DOI 10.15468/5nilie | L : Botany : L.1196509 | + 2 more

HORNED PONDWEED

***Hyphaene guineensis* Schumach. & Thonn.**

Paratype

Plantae › Tracheophyta › Liliopsida › Alismatales › Potamogetonaceae › Lepilaena

13 July 2015 | British Indian Ocean Territory | 2 images | 16 citations | Is sequenced

DOI 10.15468/5nilie | L : Botany : L.1196509 | + 2 more

All specimens

DOI 10.15468/5nilie

SPECIMEN | BOTANY

Sciodaphyllum P. Browne

Collected by David A. Neill • Porter P. Lowry II • Gregory M. Plunkett • Efrén Merino
Identified by (S. recaldiorum)

Collected in **Ecuador**

Types **Holotype • 2 isotypes**

Clustered with **3 other records** | 12 images | 16 citations | **2 sequences** | Held in **3 herbaria**

Specimen | People | Treatments | Cluster | More

Images
Label
Treatments
Where is it housed
Sequences
Material samples
Organisation according to Catalogue of Life
How to cite

Example 2:
Simple hosted portals

Images on https://labs.gbif.org/visual-concepts

Biodiversity of Tecala

GBIF TECALA LETS YOU EXPLORE SPECIES IN OUR WONDERFUL COUNTRY

START EXPLORING →

About

OPEN DATA FOR SCIENCE EVERY OBSERVATION AND RECORDING OF SPECIES CAN CONTRIBUTE TO

What is GBIF?

GBIF—THE GLOBAL BIODIVERSITY INFORMATION FACILITY—IS AN INTERNATIONAL

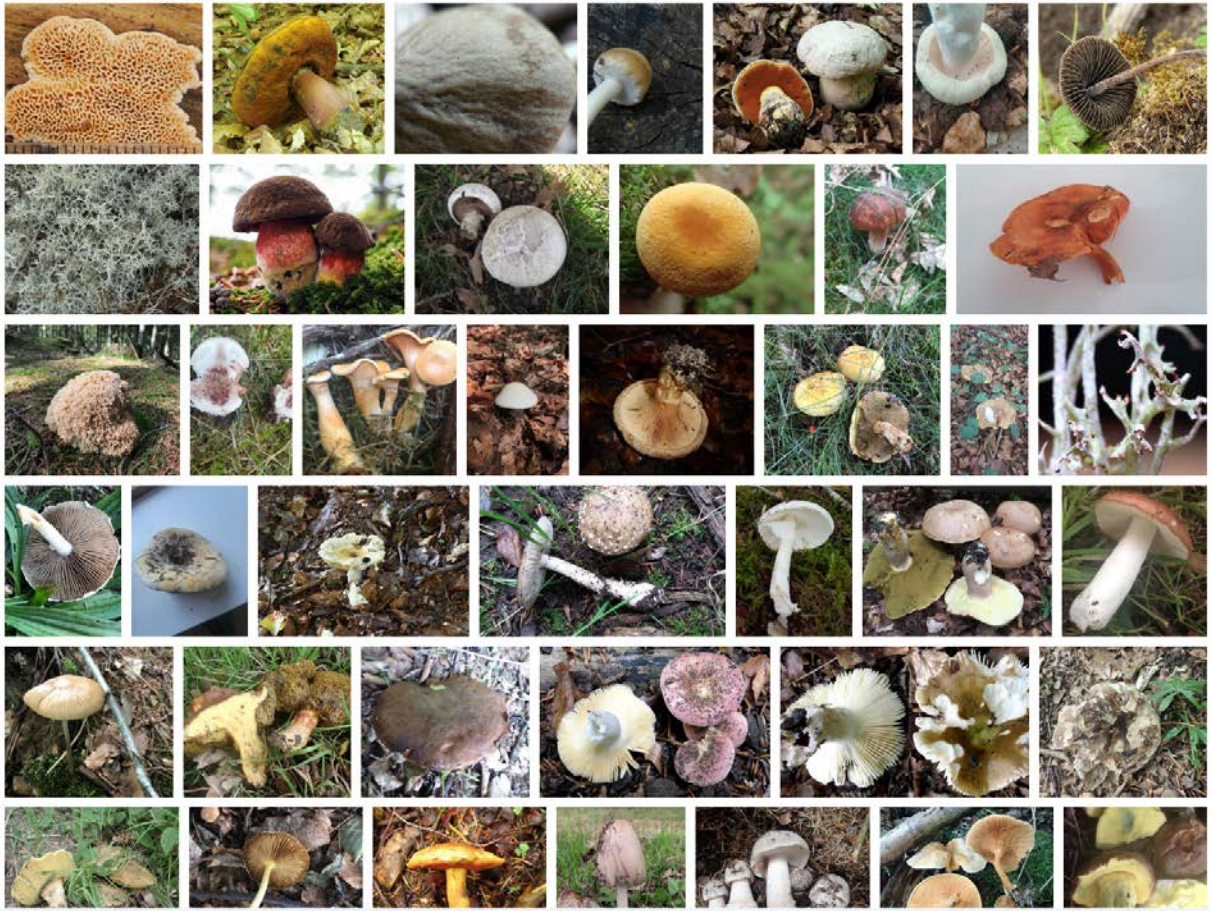
GBIF TECALA

MENU

Search

Table | Gallery | Map

2 Substrates | Clear filters



Occurrences

319,979

Basis of record

human observation	8.2M
preserved specimen	6.5M
unknown	425.8k
observation	274.2k
living specimen	165.4k
machine observation	45.6k
material sample	23.8k
fossil specimen	3.1k

Dataset

Search

All selected

Danish Mycological Society, fungal r... 320k

Recorded by

Search

Substrate

Search

2 selected | Select all

soil	198.4k
wood	121.6k

319,979 occurrences

Applying simple* branding

*font, logo, header, link color

21.660.144 results

Hide summary

Data sources and quality

You are browsing a normalized view across multiple datasets. To understand this data it is best to read about data sources and quality flags.

Remarks and issues

During data preperation we try to detect potential issues, but you decide which matters to you.

To ease filtering we have Fitness for use profiles that preset filters.

No profile selected

Exlclude selected

- [ ] Taxon match none
- [x] Taxon match higherrank
- [ ] Taxon match fuzzy

See more

Mixed data sources

There are 182 datasets in your search results. Each with its own qualities.

18 Gridded datasets

6 eDNA datasets

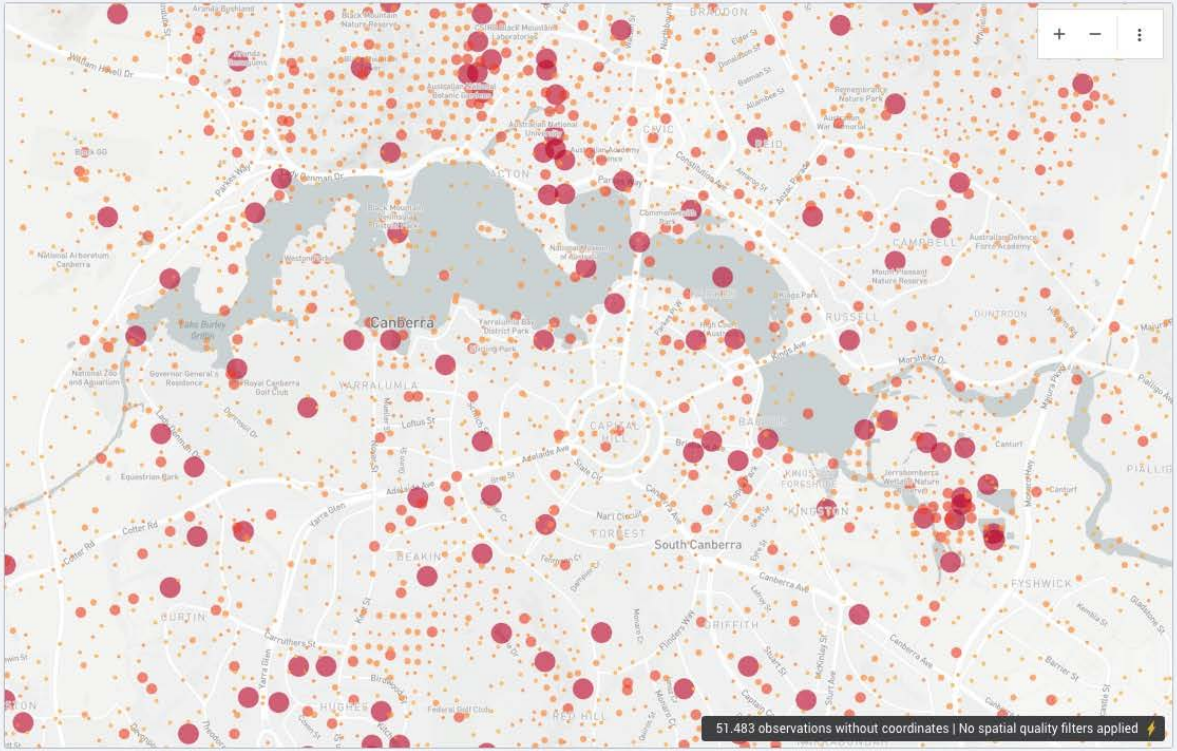
2 Tracking datasets

See more

Afroboletus Pegler & T.W.K.Young | Where | When | Evidence | Quality | More filters | Clear filters

21.660.144 results

Hide summary



Data sources and quality

You are browsing a normalized view across multiple datasets. To understand this data it is best to read about data sources and quality flags.

Remarks and issues

During data preperation we try to detect potential issues, but you decide which matters to you.

To ease filtering we have Fitness for use profiles that preset filters.

No profile selected

Exlclude selected

- [ ] Taxon match none
- [x] Taxon match higherrank
- [ ] Taxon match fuzzy

See more

Mixed data sources

There are 182 datasets in your search results. Each with its own qualities.

18 Gridded datasets

6 eDNA datasets

2 Tracking datasets

CBIF

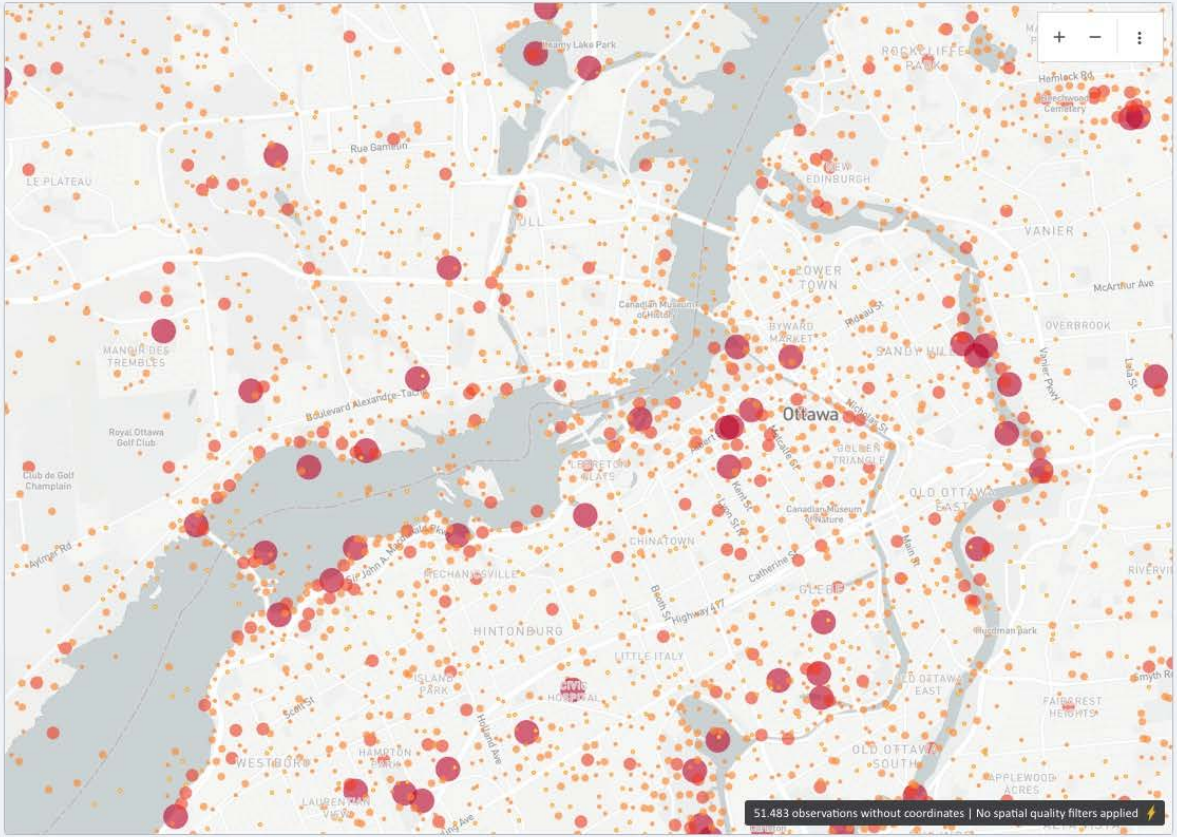
Obervations Taxonomy Datasets Treatments Collections More

Observations | Map Table Gallery Dashboard

Afroboletus Pegler & T.W.K.Young | Where | When | Evidence | Quality | More filters | Clear filters

21.660.144 results

Hide summary



Data sources and quality

You are browsing a normalized view across multiple datasets. To understand this data it is best to read about data sources and quality flags.

Remarks and issues

During data preperation we try to detect potential issues, but you decide which matters to you.

To ease filtering we have Fitness for use profiles that preset filters.

No profile selected

Exlclude selected

- [ ] Taxon match none
- [x] Taxon match higherrank
- [ ] Taxon match fuzzy

See more

Mixed data sources

There are 182 datasets in your search results. Each with its own qualities.

18 Gridded datasets

6 eDNA datasets

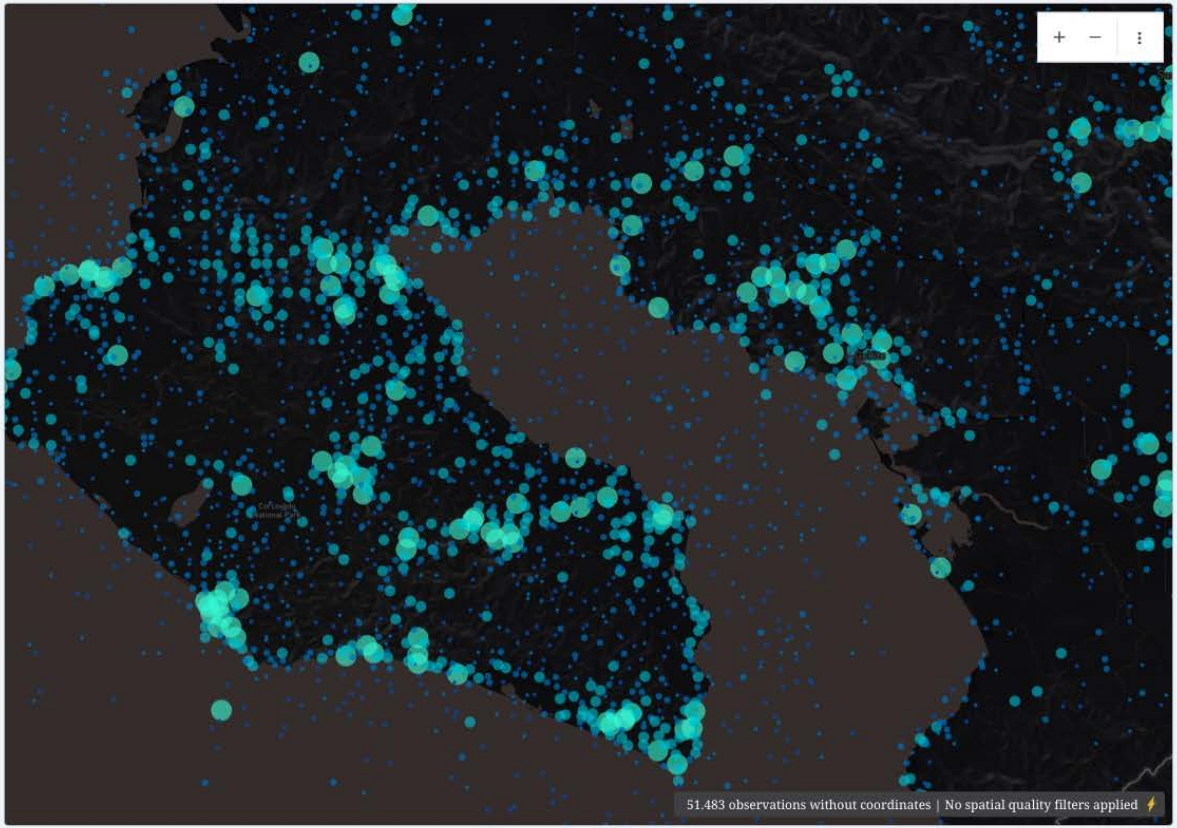
2 Tracking datasets

See more

Mixed datasources | 12 errors | 35 warnings | Mixed occurrence type: Hide data summary

21.660.144 results

Hide summary



Data sources and quality

You are browsing a normalized view across multiple datasets. To understand this data it is best to read about data sources and quality flags.

Remarks and issues

During data preperation we try to detect potential issues, but you decide which matters to you.

To ease filtering we have Fitness for use profiles that preset filters.

No profile selected

Exlclude selected

- [ ] Taxon match none
- [x] Taxon match higherrank
- [ ] Taxon match fuzzy

See more

Mixed data sources

There are 182 datasets in your search results. Each with its own qualities.

18 Gridded datasets

6 eDNA datasets

OPPORTUNITIES

1. Efficiencies in data management
2. Consistent data handling
3. Improved citation tracking
4. Optimize development effort
5. Expand helpdesk
6. Share infrastructure (“powered by GBIF”)
7. Lower total cost of running

Thank you!

GBIF