Lowering the technical threshold to participate

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

Species occurrence records accessible through GBIF over time

Number of species having occurrence records accessible through GBIF over time

https://www.gbif.org/analytics/global
GROWTH IN PUBLISHERS AND DATASETS

Source: GBIF Registry
GROWTH IN USE

Users of GBIF.org per week
~150k per month now

Citations of GBIF mediated data

Source: Google Analytics and GBIF tracking
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
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
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

https://registry.gbif.org
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

GBIF Data Portal

Hello?

Server Not Available

Hi!

Bird Observations

Insect Collection

Herbarium

Herbarium

Graphic reused from old GBIF presentation
IPT: The Integrated Publishing Toolkit

A free open source software tool used to publish and share biodiversity datasets through the GBIF network.
CO-LOCATED REPOSITORIES (EXAMPLES)
PARTNERSHIPS USING OPEN REPOSITORIES

Recent example of connecting infrastructure

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
EXAMPLE 2: PEOPLE

Claim the natural history specimens you collected or identified, track their use in new science, and help acknowledge your peers, mentors, and organizations.

Get started

Sample Profiles

Verrill, Addison Emery  
(b. February 06, 1839 – d. December 10, 1928)  
United States of America  
Collected Pleuraxa  
21 specimens claimed

Henry, Augustine  
(b. July 02, 1857 – d. March 23, 1930)  
Ireland  
Collected Rosaceae  
2,488 specimens claimed

Geitler, Lothar  
(b. May 18, 1899 – d. May 01, 1990)  
Austria  
Identified Batrachospermaceae and collected Batrachospermaceae  
24 specimens claimed

Mayr, Ernst  
(b. July 03, 1904 – d. February 03, 2005)  
United States of America; Germany  
Identified Pelleticidae and collected Pelleticidae  
4,392 specimens claimed

Anderson, William  
(b. December 28, 1750 – d. August 03, 1778)  
Kingdom of Great Britain  
Collected Nyctaginaceae  
19 specimens claimed

Cabrera, Ángel Lulio  
(b. October 19, 1908 – d. July 08, 1999)  
Spain; Argentina  
Identified Scrophulariaceae and collected Poaceae  
29 specimens claimed

See https://labs.gbif.org/visual-concepts
EXAMPLE 3: DISCOVERING SAMPLE BASED DATA

See https://labs.gbif.org/visual-concepts
EXAMPLE 4: LINKING SEQUENCES TO SPECIMENS

See https://www.gbif.no/news/2017/bold.html
EXAMPLE 5: MEDIATING TAXONOMIC TREATMENTS

A taxonomic review of the genus Triscaedecia (Lepidoptera: Alucitidae) in the world fauna

PETR USTUZHANIN, VASILIY KVTUNOVICH, DONALD HOBERN

Abstract

The article describes four new species of Alucitidae in the genus Triscaedecia from the Malay and Polynesian regions. Triscaedecia sulawesi Ustuzhanin, Kvtunovich & Hobern sp. nov., T. saviti Ustuzhanin, Kvtunovich & Hobern sp. nov., T. svetleae Ustuzhanin, Kvtunovich & Hobern sp. nov., and T. suva Ustuzhanin, Kvtunovich & Hobern sp. nov. The female genitalia of Triscaedecia dactyloptera Hamsson, 1908, and the adult and male genitalia of Triscaedecia septemductyla (Pagenstecher, 1901) are described and illustrated for the first time. Diagnostic characters are specified for each new taxon. Genitalia are described and illustrated, and images are provided of adults. A spreadsheet of all known localities and specimens is included as a supplementary file.

Keywords

Lepidoptera, Many-plumed moths, new species, new data, Malay and Polynesian regions
“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

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

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%
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
Range-expansion effects on the belowground plant microbiome

Range-expansion effects on the belowground plant microbiome

Range-expansion effects on the belowground plant microbiome
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

About

Open data for science every observation and recording of species can contribute to

What is GBIF?

GRIF—the Global Biodiversity Information Facility—is an international

Applying simple* branding

*font, logo, header, link color
21,660,144 results

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 preparation 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.

Mixed data sources
There are 182 datasets in your search results. Each with its own qualities.
- 18 Grided datasets
- 6 eDNA datasets
- 2 Tracking datasets

See more
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!