

SEINet AZ-NM Chapter Portal Campaign

July 11, 2023



iDigBio
Integrated Digitized Biocollections



What we did

- Unmatched taxonomic names
- Unmatched geographic names
- Negated lat/longs / swapped lat/longs
- Inverted minimum and maximum elevation values
- Invalid Basis Of Record



What we did

Taxonomic names:

- Used statistics from taxonomic name cleaner

Geographic names:

- Used statistics from the geography cleaner
- You can also determine whether there are states and counties with unmatched names!

Negated lat/longs:

- Looked for negative lat/long values in a list of countries that should only have positive lat/long values (and vice versa)
- Tested out what changing the sign of these coordinates would do (ran them through a GBIF tool that checked whether they landed in the correct country).

Inverted minimum and maximum elevation values:

- Flagged any occurrences where minimum elevation > maximum elevation

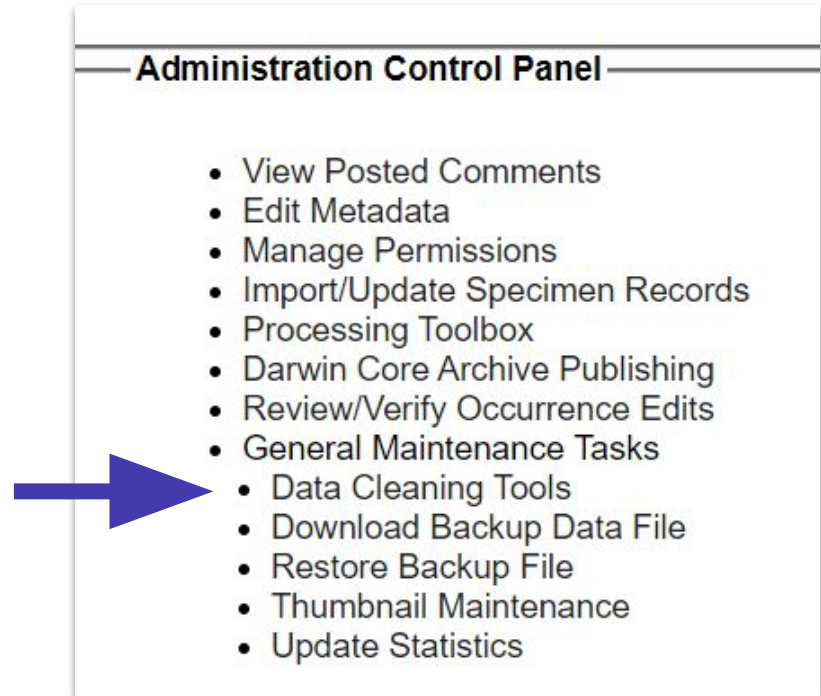
Invalid Basis Of Record:

- Looked for any non-standard values in the Basis of Record field
 - explained here: <https://dwc.tdwg.org/terms/#dwc:basisOfRecord>



Data cleaning in a Symbiota portal

- Taxonomic Cleaning Tool
- Geography Cleaning Tool
- Batch editing






Administration Control Panel

- View Posted Comments
- Edit Metadata
- Manage Permissions
- Import/Update Specimen Records
- Processing Toolbox
- Darwin Core Archive Publishing
- Review/Verify Occurrence Edits
- General Maintenance Tasks
 - Data Cleaning Tools
 - Download Backup Data File
 - Restore Backup File
 - Thumbnail Maintenance
 - Update Statistics

Geography Cleaning Tool

- Like the taxonomic thesaurus, it is intended to be a resource for data discoverability, not an authority

U.S.S.R (1) 	<input type="text" value="Replace with..."/>	<input type="button" value="Replace Country"/>
U.S.S.R. (8) 	<input type="text" value="Replace with..."/>	<input type="button" value="Replace Country"/>
USSR (1) 	<input type="text" value="Replace with..."/>	<input type="button" value="Replace Country"/>

Batch Editing

- Use caution when using!
- Edits will be tracked in the Review/Verify Edits table

Data Editor Control Panel

- [Add New Occurrence Record](#)
- [Create New Records Using Image](#)
- [Add Skeletal Records](#)
- [Edit Existing Occurrence Records](#)
- [Add Batch Determinations/Nomenclatural Adjustments](#)
- [Print Specimen Labels](#)
- [Print Annotations Labels](#)
- [Occurrence Trait Coding Tools](#)
- [Batch Georeference Specimens](#)
- [Loan Management](#)

Administration Control Panel

- [View Posted Comments](#)
- [Edit Metadata](#)
- [Manage Permissions](#)
- [Import/Update Specimen Records](#)
- [Processing Toolbox](#)
- [Darwin Core Archive Publishing](#)
- [Review/Verify Occurrence Edits](#)
- [Duplicate Clustering](#)
- **General Maintenance Tasks**
 - [Data Cleaning Tools](#)
 - [Download Backup Data File](#)
 - [Restore Backup File](#)
 - [Thumbnail Maintenance](#)
 - [Update Statistics](#)

Batch Editing

- Use caution when using!
- Edits will be tracked in the Review/Verify Edits table

Data Editor Control Panel

- [Add New Occurrence Record](#)
- [Create New Records Using Image](#)
- [Add Skeletal Records](#)
- [Edit Existing Occurrence Records](#)
- [Add Batch Determinations/Nomenclatural Adjustments](#)
- [Print Specimen Labels](#)
- [Print Annotations Labels](#)
- [Occurrence Trait Coding Tools](#)
- [Batch Georeference Specimens](#)
- [Loan Management](#)

Administration Control Panel

- [View Posted Comments](#)
- [Edit Metadata](#)
- [Manage Permissions](#)
- [Import/Update Specimen Records](#)
- [Processing Toolbox](#)
- [Darwin Core Archive Publishing](#)
- [Review/Verify Occurrence Edits](#)
- [Duplicate Clustering](#)
- **General Maintenance Tasks**
 - [Data Cleaning Tools](#)
 - [Download Backup Data File](#)
 - [Restore Backup File](#)
 - [Thumbnail Maintenance](#)
 - [Update Statistics](#)

Safety!



Batch Editing



California Polytechnic State University, Robert F. Hoover Herbarium (OBI)



Record Search Form

Collector: Number: Date:

Catalog Number: Other Catalog Numbers:

Entered by: Date Entered: Date Modified:

Processing Status: with images without images

Custom Field 1:

Sort By: Record output:

Display as dynamic table

[Home](#) >> [Collection Management](#) >> **Occurrence Table View**

| 1-1000 of 94700 records | [>>](#) [>|](#)

Symbiota ID	Catalog Number	Other Catalog #	Family	Scientific Name	Author	Collector
24202980	OBI100071		Scrophulariaceae	<i>Pedicularis rigginsiae</i>	D.J. Keil	Rhonda Riggins
24202981	OBI100072		Scrophulariaceae	<i>Pedicularis rigginsiae</i>	D.J. Keil	David Keil

SEINet Portal Campaign Arizona-New Mexico Chapter

Final Office Hours - July 18, 2023



iDigBio
Integrated Digitized Biocollections



Agenda

- Portal campaign accomplishments: a recap
- Publishing to GBIF & iDigBio
- Further opportunities

1,025

taxonomic names indexed to the
central thesaurus

1,984

specimens made searchable





599

country values standardized



146

records with swapped minimum & maximum elevation values fixed

Newly publishing to GBIF



(@ **ASU**)



Tonto National Forest

13,887
specimens
added!

Other Success Stories

- Synonymies and other taxonomic issues fixed (thanks, Sue Carnahan!)
- Several contacts updated
- New collections invited to the portal (JMH, CMH, ASUF)
- In the future: adding *in situ* images from DES



Anything else to share?

Agenda

- Portal campaign accomplishments: a recap
- **Publishing to GBIF & iDigBio**
- Further opportunities

What is GBIF?

“an **international network** of country and organizational Participants that exists to enable **free and open access to biodiversity data** from all sources and to support biodiversity science, environmental research, and evidence based decision-making.”



GBIF data portal

GBIF | Global Biodiversity Information Facility

Free and open access to biodiversity data

OCCURRENCES SPECIES DATASETS PUBLISHERS RESOURCES

Search 🔍

What is GBIF? About GBIF United States of America

Delonix regia (Bojer ex Hook.) Raf. observed in Guinea by Anne-Helene Paradis (CC BY-NC 4.0)



1,927,619,110

Occurrence records



65,549

Datasets



1,789

Publishing institutions



6,806

Peer-reviewed papers using data

Why Publish to GBIF?

- **Greater access** to your data → **greater visibility** for your collection



PUBLISHER | SINCE MAY 3, 2010

Duke University Herbarium

ABOUT METRICS ↻ HOME PAGE

313,960 OCCURRENCES

4 DATASETS

308 CITATIONS

Description: <http://www.biology.duke.edu/herbarium>

Endorsed by: [U.S. Geological Survey](#)

Administrative contact: [Blanka Aguero](#)

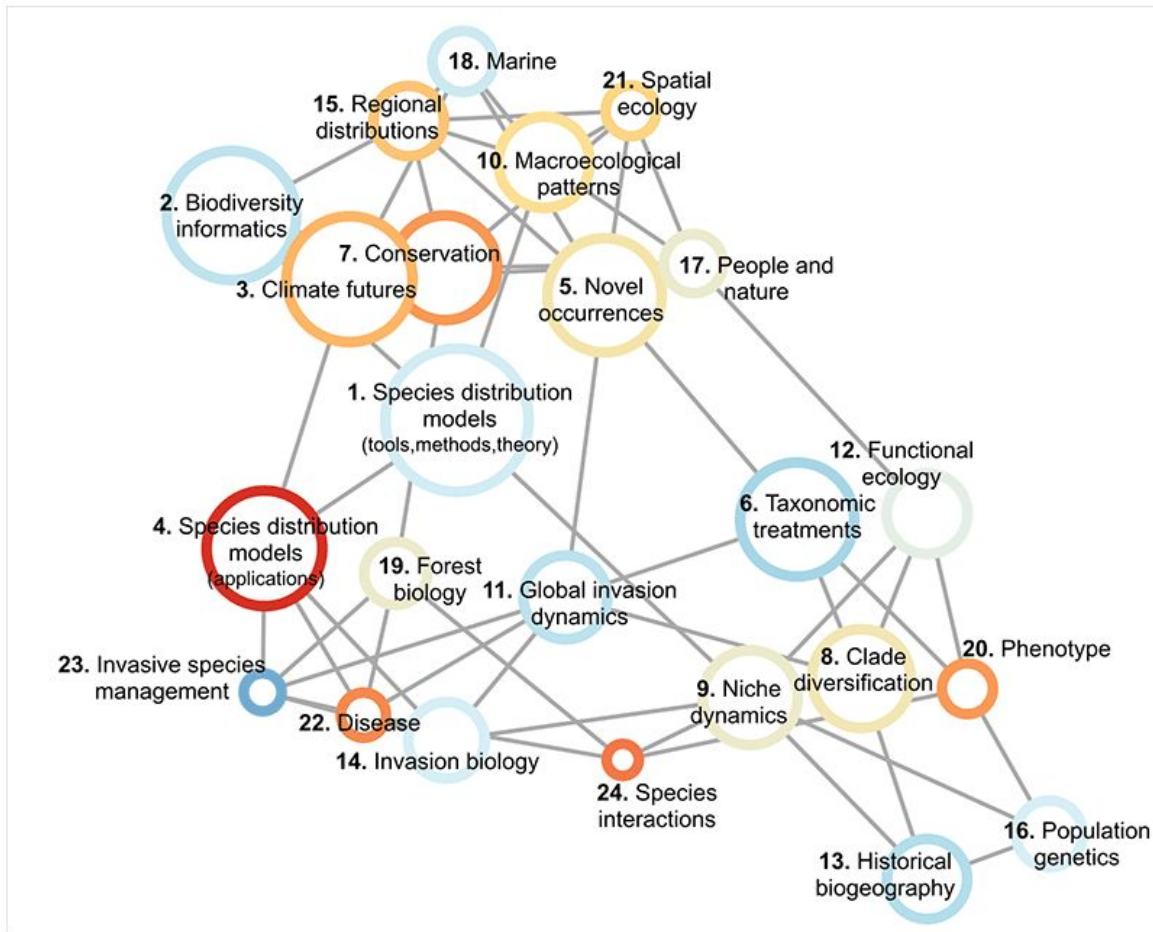
Country or area: [United States of America](#)

Download activity report



Why Publish to GBIF?

- Greater access to your data → greater visibility for your collection
- Allows **researchers** from a broader **range of disciplines** to compile **more comprehensive datasets**



Structural topic model results from 4,035 studies that used GBIF-mediated data published between 2003 and 2019.

<https://docs.gbif.org/course-introduction-to-gbif/en/how-is-gbif-mediated-data-used.html>

Why Publish to GBIF?

- Greater access to your data → greater visibility for your collection
- Allows researchers from a broader range of disciplines to compile more comprehensive datasets
- **Citation tracking**



PUBLISHER | SINCE MAY 3, 2010

Duke University Herbarium

ABOUT METRICS ↻ HOME PAGE

313,960 OCCURRENCES

4 DATASETS

308 CITATIONS

Description: <http://www.biology.duke.edu/herbarium>

Endorsed by: U.S. Geological Survey

Administrative contact: Blanka Aguero

Country or area: United States of America

Download activity report



© OpenStreetMap contributors

Why Publish to GBIF?

- Greater access to your data → greater visibility for your collection
- Allows researchers from a broader range of disciplines to compile more comprehensive datasets
- Citation tracking
- **Data quality flags**



OCCURRENCE DATASET | REGISTERED MARCH 26, 2019

Duke University Herbarium Bryophyte Collection

Published by [Duke University Herbarium](#)

DATASET **METRICS** ACTIVITY [DOWNLOAD](#)

173,869 OCCURRENCES

73 CITATIONS

OCCURRENCE METRICS



173,869
Occurrences



99.9%
With taxon match



49%
With coordinates

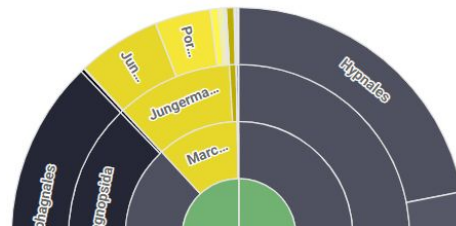


93%
With year

TAXONOMIC DISTRIBUTION OF OCCURRENCES

Explore	Major groups
Plantae	173,759
incertae sedis	110

TAXONOMIC DISTRIBUTION OF OCCURRENCES



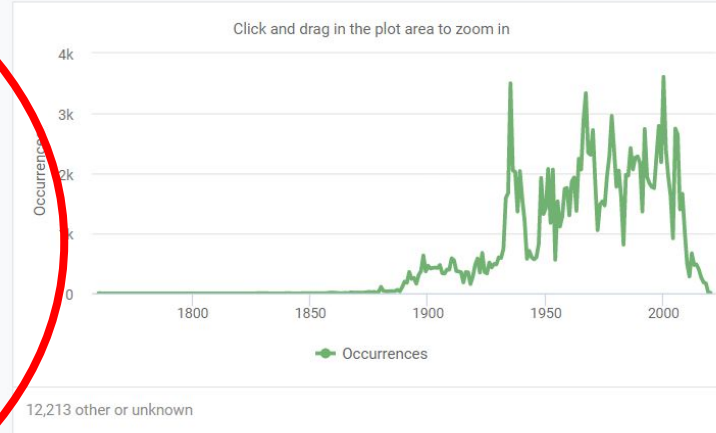


OCCURRENCES PER ISSUES AND FLAGS

Issues and flags	Count
Collection match none	173,869
Institution match fuzzy	173,869
Geodetic datum assumed WGS84	70,631
Coordinate rounded	42,669
Recorded date invalid	10,220
Coordinate reprojected	1,549
Taxon match higherrank	743
Taxon match fuzzy	314
Taxon match none	110
Country coordinate mismatch	67

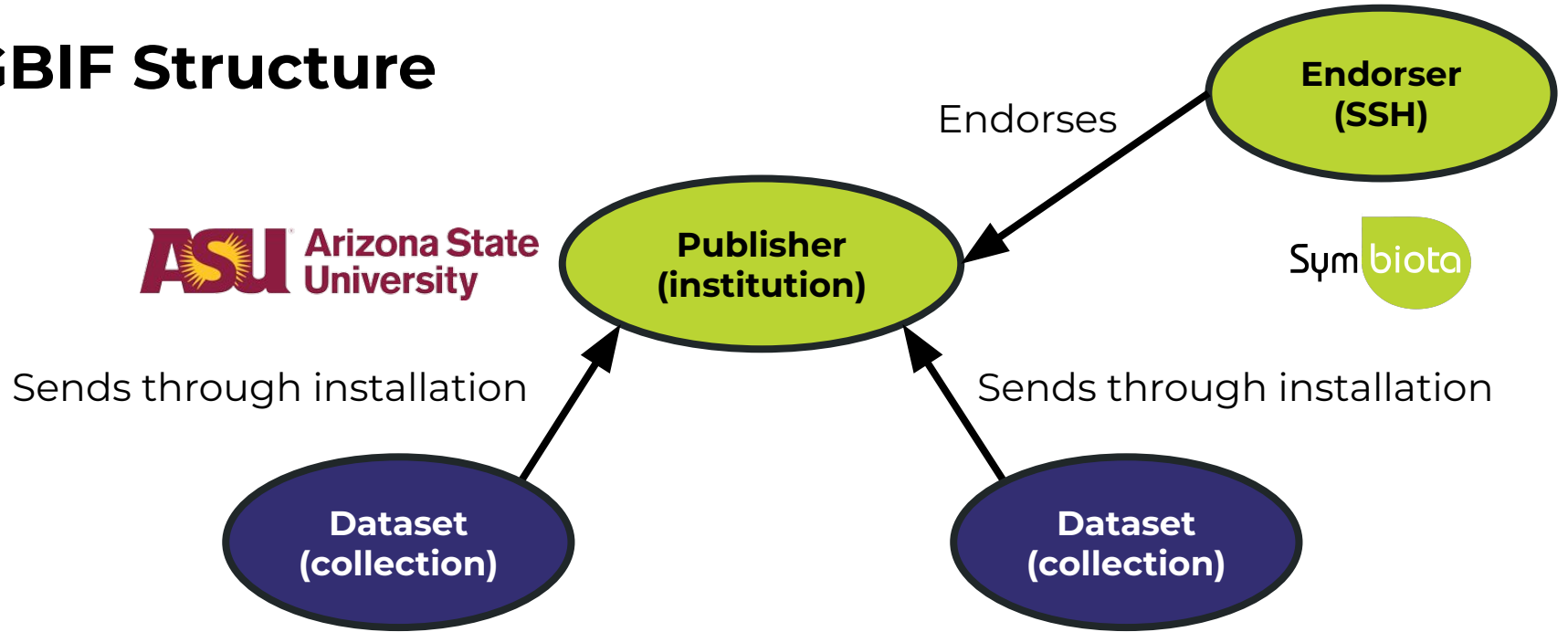
NEXT

OCCURRENCES PER YEAR



How do I publish to
GBIF?

GBIF Structure



Arizona State University Mammalogy Collection

Arizona State University Vascular Plant Herbarium

Consortium of
Small
Vertebrate
Collections



Arizona State University Biocollections

[ABOUT](#) [METRICS](#) [HOME PAGE](#)

673,214 OCCURRENCES

18 DATASETS

691 CITATIONS

Description: The Arizona State University Biocollections comprise two sets of distinct collections: (1) the Arizona State University Natural History Collections - currently with nine collections focused on documenting Greater Sonoran and New World biodiversity; and (2) the NEON Biorepository at Arizona State University, with a unique constellation of organismal and environmental samples generated in the context of monitoring and forecasting long-term ecological change in the North American subcontinent, including Alaska, Hawaii, and Puerto Rico.

Endorsed by: [Symbiota Support Hub](#)

Administrative contact: [Nico Franz](#)

Technical contact: [Laura Rocha Prado](#)

Country or area: [United States of America](#)



ALL OCCURRENCE CHECKLIST SAMPLING EVENT METADATA

DOWNLOAD AS TSV

Arizona State University Vascular Plant Herbarium

Occurrence dataset

The Arizona State University Vascular Plant Herbarium (ASU) is among the most important in the greater Sonoran Desert region with over 315,000 specimens. We are particularly proud of our holdings of C...

Published by Arizona State University Biocollections

283,279 occurrences 407 citations



Arizona State University Hasbrouck Insect Collection

Occurrence dataset

The ASU Frank F. Hasbrouck Insect Collection contains approximately 1,000,000 insect specimens, representing at least 25 orders, 390 families, 4,000 genera, 12,000 species and 1,240 subspecies. Most s...

Published by Arizona State University Biocollections

167,107 occurrences 155 citations



Arizona State University Lichen Herbarium

Occurrence dataset

No description available

Published by Arizona State University Biocollections

123,620 occurrences 105 citations



Arizona State University Herpetology Collection

Occurrence dataset

The Herpetology Collection at the ASU Natural History Collections contains approximately 38,000 specimens representing more than 900 species, with a geographic concentration in the western United Stat...

Published by Arizona State University Biocollections



Publishing to GBIF from a Symbiota portal

1. Make sure your **metadata are accurate**

[Home](#) >> [Collection Search Page](#) >> [Collection Profile](#)

Arizona State University Vascular Plant Herbarium (ASU-Plants)

← 548 citations

The Arizona State University Vascular Plant Herbarium (ASU) is among the most important in the greater Sonoran Desert region with over 315,000 specimens. We are particularly proud of our holdings of Cactaceae which include over 1,100 chromosome counts.

An herbarium is a collection of pressed, dried, and archived plants that are systematically arranged - each specimen a physical record of a plant growing at a particular place and at a particular time. Like most herbaria, we seek to document the geographical and ecological distribution of the regional flora, facilitate research, support teaching, and promote conservation. Vouchers from floristic studies have resulted in comprehensive collections of many important geographic regions in Arizona.

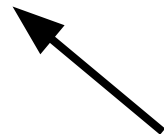
ASU Type Specimens: <http://swbiodiversity.org/seinet/checklists/checklist.php?cl=2638>

Administration Control Panel

- View Posted Comments
- Edit Metadata
- Manage Permissions

Publishing to GBIF from a Symbiota portal

1. Make sure your metadata are accurate
2. *(If not already done)* **Request to become a publisher** in GBIF:
<https://www.gbif.org/become-a-publisher>



This page allows you to search for your institution to see if it's already registered.

Publishing to GBIF from a Symbiota portal

1. Make sure your metadata are accurate
2. (If not already done) **Request to become a publisher** in GBIF:
<https://www.gbif.org/become-a-publisher>

Endorsing node

To support publishers and review data quality all publishers are associated with a GBIF node. Please check the suggestion below, and correct it if needed:

Help me with endorsement

Marine data publishers: request endorsement for OBIS (Ocean Biogeographic Information System) related data

If endorsement through the country node suggested above is not the right option, please check this list of associated participants for multinational or thematic networks:

- Symbiota Support Hub
- Amazon Cooperation Treaty Organization

SSH can (and wants to) endorse you!



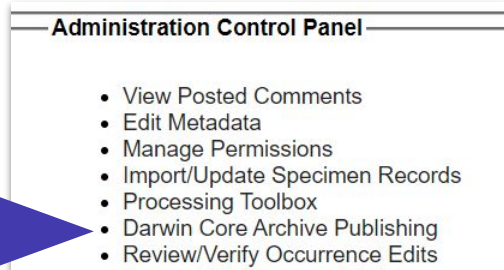
Publishing to GBIF from a Symbiota portal

1. Make sure your metadata are accurate
2. *(If not already done)* Request to become a publisher in GBIF:
<https://www.gbif.org/become-a-publisher>
3. Once approved, enter your **dataset key** in “Darwin Core Archive Publishing” in your portal
4. **Email GBIF** (helpdesk@gbif.org) to notify them that the portal has your permission to push your dataset to GBIF (an example email is provided!)
5. Once approved, **publish** the data!

**We can help with Steps 3-5
(Email us after Step 2)**


Important notes

- **GBIF doesn't automatically harvest your data**
- Push the button to refresh:
“Create/Refresh Darwin Core Archive”
 - Portal managers will do this regularly
 - Data that are redacted in your Symbiota portal will also be redacted in GBIF



Administration Control Panel

- View Posted Comments
- Edit Metadata
- Manage Permissions
- Import/Update Specimen Records
- Processing Toolbox
- Darwin Core Archive Publishing
- Review/Verify Occurrence Edits



Darwin Core Archive Publishing

OBI - Robert F. Hoover Herbarium, Cal Poly State University

Use the controls below to publish occurrence data from this collection as a Darwin Core document that describes the content. The occurrence data file is required, but id exchange standard. We recommend that you also review instructions for Publish

RSS Feed: <https://cch2.org/portal/webservices/dwc/rss.xml>

Title: OBI DwC-Archive ✗

Description: Darwin Core Archive for OBI - Robert F. Hoover Herbarium, Cal Poly

EML: <https://cch2.org/portal/collections/datasets/emlhandler.php?collid=12>

DwC-Archive File: https://cch2.org/portal/content/dwca/OBI_DwC-A.zip

Pub Date: Tue, 01 Feb 2022 09:53:45

Publishing Information

GUID source: symbiotaUUID

GBIF Dataset page: <http://www.gbif.org/dataset/f56df26e-73f5-4d37-bfed->

Publish/Refresh DwC-A Data

- Include Determination History
- Include Image URLs
- Redact Sensitive Localities

Create/Refresh Darwin Core Archive

OBI - Robert F. Hoover Herbarium, Cal Poly State University (OBI)

393 citations

The Hoover Herbarium houses 85,000+ specimens of vascular plants, algae, lichens, and bryophytes. The geographic focus is San Luis Obispo areas of California, other states of the US, particularly Arizona, and some from other regions of the world, especially Mexico. Emphasis areas in collections include Robert F. Hoover (1946–1969), David J. Keil (1966–present), Rhonda Riggins (1970s–2000), Tracy Call (mostly Apiaceae—extensively in undergraduate teaching and training).

Important Collections: Robert F. Hoover (1946–1969), David J. Keil (1966–present), Rhonda Riggins (1970s–2000), Tracy Call (mostly Apiaceae)

Director and Associate Professor: Jenn Yost, jyost@calpoly.edu

Curator: Katie Pearson, kdpearso@calpoly.edu

Homepage: <http://bio.calpoly.edu>

Collection Type: Preserved Specimens

Management: Live Data managed directly within data portal

Global Unique Identifier: 3818d95b-b6a4-11e8-b408-001a64db2964

DwC-Archive Access Point: https://cch2.org/portal/content/dwca/OBI_DwC-A.zip

Live Data Download: [DwC-Archive File](#)

Digital Metadata: [EML File](#)

Usage Rights: [CC BY-NC \(Attribution-Non-Commercial\)](#)

GBIF Dataset page: <http://www.gbif.org/dataset/f56df26e-73f5-4d37-bfed-3d46c0834e82>

Address: Robert F. Hoover Herbarium 
Biological Sciences Department, California Polytechnic State University
San Luis Obispo, CA 93407-0401
USA
(805) 756-5869

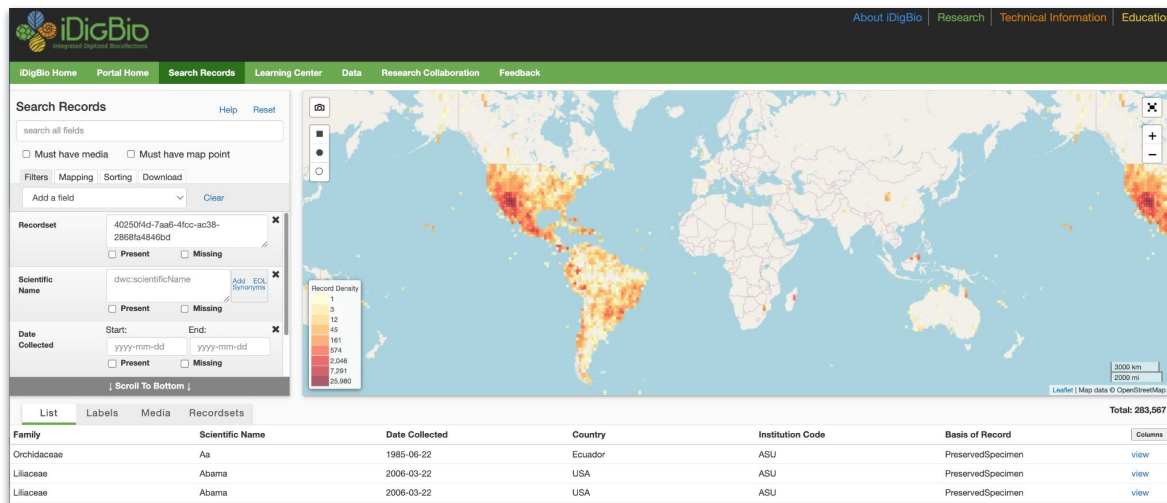
Collection Statistics

- 93,149 specimen records
- 64,830 (70%) georeferenced
- 81,168 (87%) with images (81,417 total images)
- 89,512 (96%) identified to species
- 400 families
- 2,673 genera
- 11,464 species
- 14,186 total taxa (including subsp. and var.)



What about iDigBio?

- Once you've published to GBIF, publishing to iDigBio is **very easy**
- The SSH can facilitate publishing to iDigBio for everyone who publishes to GBIF. **Please let us know if you would like to opt out of publishing to iDigBio.**



The screenshot displays the iDigBio website interface. At the top, there is a navigation bar with links for "About iDigBio", "Research", "Technical Information", and "Education". Below this is a secondary navigation bar with "iDigBio Home", "Portal Home", "Search Records", "Learning Center", "Data", "Research Collaboration", and "Feedback".

The main content area is divided into two sections. On the left, the "Search Records" section includes a search box, filter options (e.g., "Must have media", "Must have map point"), and a "Recordset" table. The "Recordset" table has columns for "Recordset", "Scientific Name", and "Date Collected". The "Recordset" column contains a long alphanumeric string. The "Scientific Name" column contains "dwc:scientificName". The "Date Collected" column contains "1985-06-22", "2006-03-22", and "2006-03-22".

On the right, there is a world map showing record density. A legend indicates record density ranges from 1 to 25,380. The map shows high density in North and South America.

At the bottom, there is a table with columns: "List", "Labels", "Media", "Recordsets", "Family", "Scientific Name", "Date Collected", "Country", "Institution Code", "Basis of Record", and "Columns". The table lists records for Orchidaceae, Lilaceae, and another Lilaceae entry, all from ASU.

Family	Scientific Name	Date Collected	Country	Institution Code	Basis of Record	Columns
Orchidaceae	Aa	1985-06-22	Ecuador	ASU	PreservedSpecimen	View
Lilaceae	Abama	2006-03-22	USA	ASU	PreservedSpecimen	View
Lilaceae	Abama	2006-03-22	USA	ASU	PreservedSpecimen	View

Questions?

Agenda

- Portal campaign accomplishments: a recap
- Publishing to GBIF & iDigBio
- **Further opportunities**

Curating Comments

The comment function is a way that the public can interact with your specimens. It is often used to:

- Flag dubious identifications
- Point out problematic georeferences

[OBI150564](#) [David Keil #12939](#) [_1978-07-23](#)

Hanofee, Shane posted on 2022-11-08 10:22:19

This looks much like *Aconogonon phytolaccifolium* which is abundant where this collection was made.

Hide Comment from Public

Mark as Reviewed

Delete Comment

Curating Comments

- Accessed through Administration Control Panel
- Can **delete** (goes away forever) or **mark as reviewed** (still visible on your record but no longer a notification)

Harvesting georeferences from duplicates

- SEINet contains over 22 million herbarium specimens
- Some of them might be duplicates of your specimens
- Some of those duplicates might have georeferences!

Harvesting georeferences from duplicates

We have a script that we can run to identify all potential georeferences that you might want to add to your specimen records.

We would send this spreadsheet to you, and you could choose which georeferences to upload.



Baylor University

48,616 specimens without coordinates



3,318 potential duplicate coordinates



**2,342 specimens that could be georeferenced
using the duplicate coordinates**



Baylor University

48,616 specimens without coordinates



3,318 potential duplicate coordinates



**2,342 specimens that could be georeferenced
using the duplicate coordinates**

At 5 minutes per specimen, that saves **195 hours** of georeferencing!

Harvesting georeferences from duplicates

Interested? Fill out this form:

<https://forms.gle/9Cg6jrUmxucST5po6>

Standing issues and opportunities

- Adding or transferring bryophyte, lichen, and fungi specimens to their respective portals

CONSORTIUM OF BRYOPHYTE HERBARIA

- building a Consortium of Bryophytes and Lichens as keystones of cryptobiotic communities -

CONSORTIUM OF LICHEN HERBARIA

- building a Global Consortium of Bryophytes and Lichens as keystones of cryptobiotic communities -

Adding data to other taxonomic-focused portals

Interested? Fill out this form:

<https://forms.gle/9Cg6jrUmxucST5po6>

Ideas for further digitization funding

- NSF grants
- [IMLS grants](#)
- Institutional grants

symbiota.org/funding-ideas

Infrastructure Capacity for Biology (Capacity)

- **Synopsis:**

Support the implementation of, scaling of, or major improvements to research tools, products, and services that advance contemporary biological research.

- **Programmatic Areas:**

- **Capacity: Cyberinfrastructure**
- **Capacity: Biological Collections**
- **Capacity: Field Stations & Marine Labs (FSML)**

- Proposals accepted anytime

- NSF 21-501

<https://www.nsf.gov/pubs/2021/nsf21501/nsf21501.htm>

Ideas for further digitization funding

Funding opportunities for scientific collections:

<https://www.youtube.com/watch?v=3PvsZI8spJ0>

Research Resources (RR)

Innovation

Capacity

Sustaining

Major Research
Instrumentation (MRI)

Human Resources (HR)

Postdoctoral Fellowships
(PRFB)

RCN for Undergraduate
Biology Education (RCN-UBE)

Research Experiences for
Undergraduates (REU Sites)

Mid-Career Advancement
(MCA)

Centers, Facilities, and Additional Research Infrastructure Cluster (CFARI)

Biology Integration Institutes
(BII)

Center for Open
Environmental Data (OED)

Management of Operations
and Maintenance of NEON

Mid-scale Research
Infrastructure-1 and 2

Portal Development

The Symbiota Support Hub was funded to support existing digitization and mobilization activities, not to develop new tools...

Portal Development

The Symbiota Support Hub was funded to support existing digitization and mobilization activities, not to develop new tools...

...however, new tool development can be initiated with outside support.

Portal Development

The Symbiota Support Hub was funded to support existing digitization and mobilization activities, not to develop new tools...

...however, new tool development can be initiated with outside support.

So, if you want a specific tool, include funds for its development in your next digitization grant!

Including Symbiota development into your grant

1. Determine need(s)/want(s)
2. Meet with Symbiota Support Hub team to discuss possibilities and necessary funding.
3. With SSH help, include development in budget

Including Symbiota development into your grant

1. Determine need(s)/want(s)
2. Meet with Symbiota Support Hub team to discuss possibilities and necessary funding.
3. With SSH help, include development in budget. Three options:
 - a. Collaborating institution
 - b. Subaward
 - c. Contract with BioKIC Services

No grant?

- Contract with BioKIC Services

services.biokic.asu.edu

Standing Opportunities (summary)

To consider:

- Curate comments & engage with the community
- Harvest georeferences from duplicates
- Host data in other taxonomically-focused portals
- Develop the SEINet portal (and Symbiota code!) by including the Symbiota Support Hub in a grant or contracting through BioKIC Services
- Promotional poster for the SEINet community

What about the Symbiota Support Hub?

We're still here to help!
help@symbiota.org