SEINet AZ-NM Chapter Portal Campaign

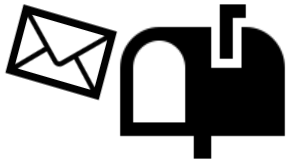
July 11, 2023

iDigBio
Integrated Digitized Biocollections

Symbiota

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

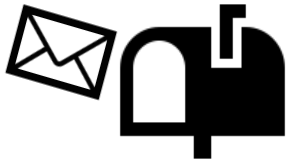
Taxonomic names:
- Used statistics from taxonomic name cleaner

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).

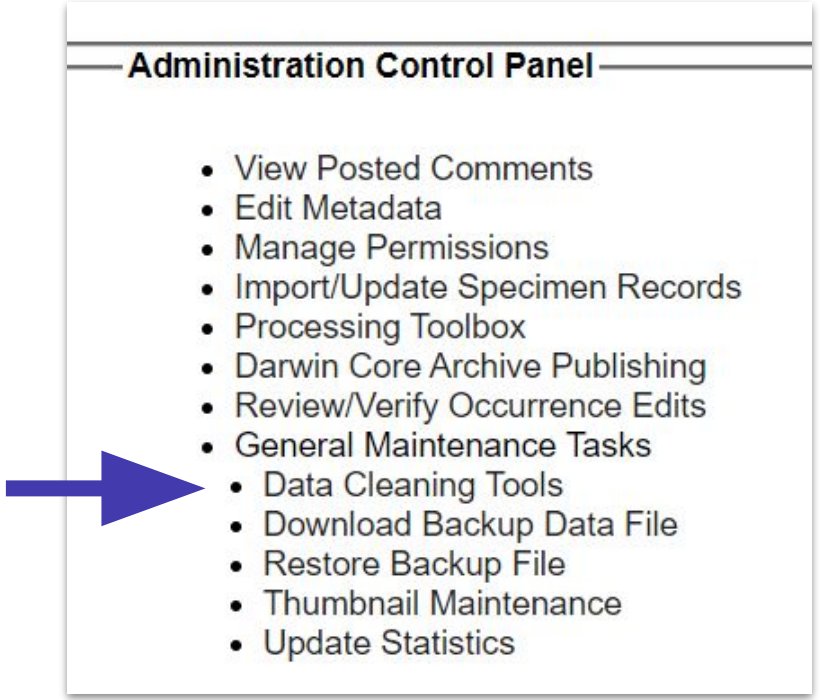
Inverted minimum and maximum elevation values:
- Flagged any occurrences where minimum elevation > maximum elevation

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



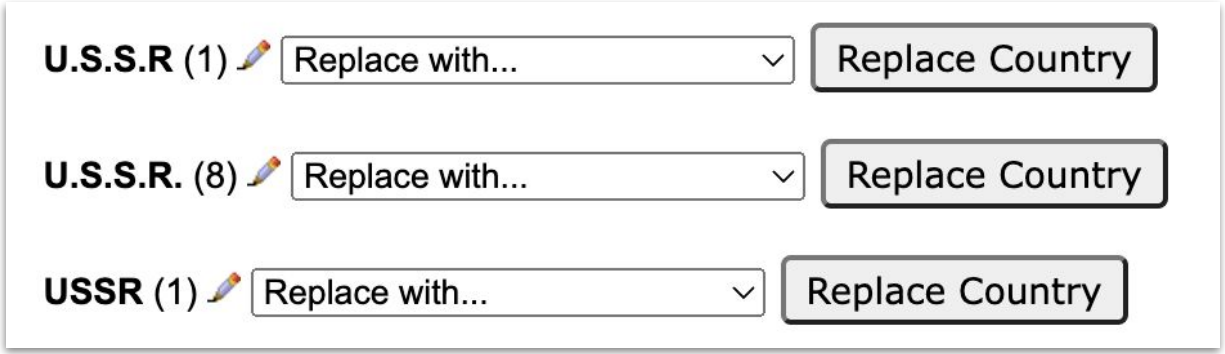
Data cleaning in a Symbiota portal

- Taxonomic Cleaning Tool
- Geography Cleaning Tool
- Batch editing



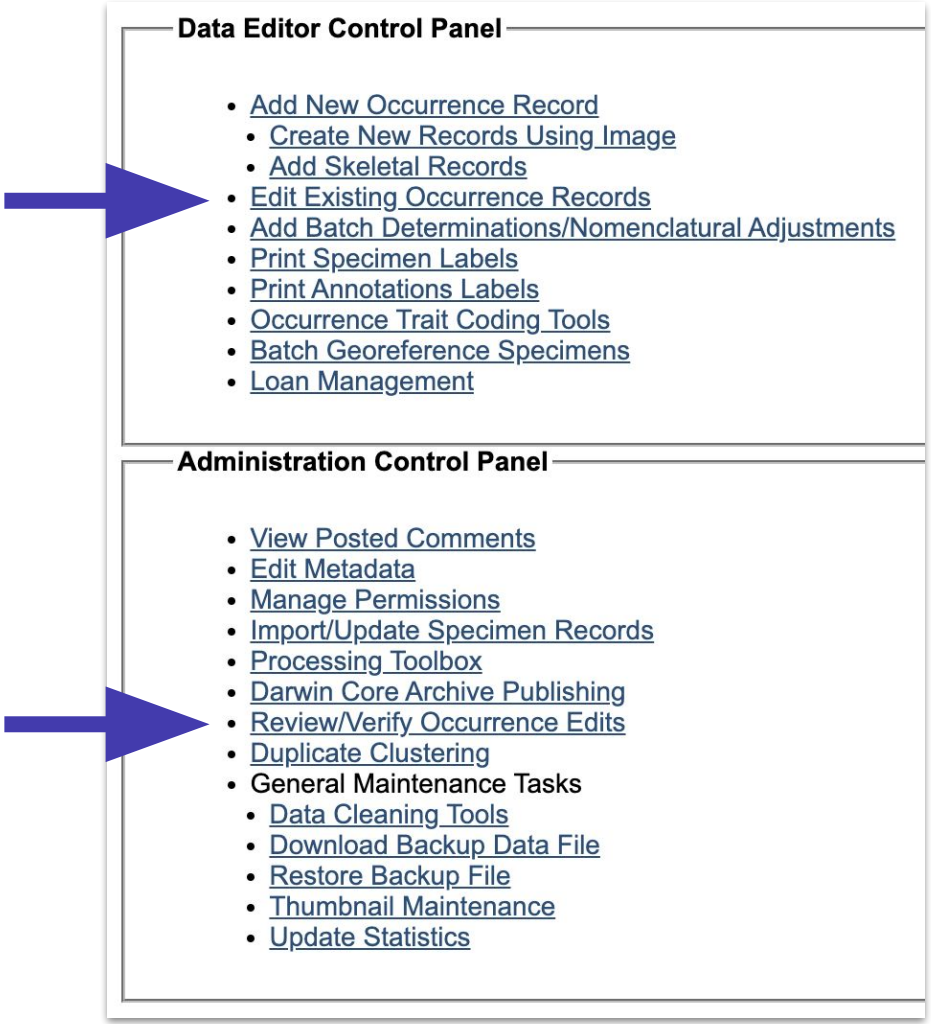
Geography Cleaning Tool

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



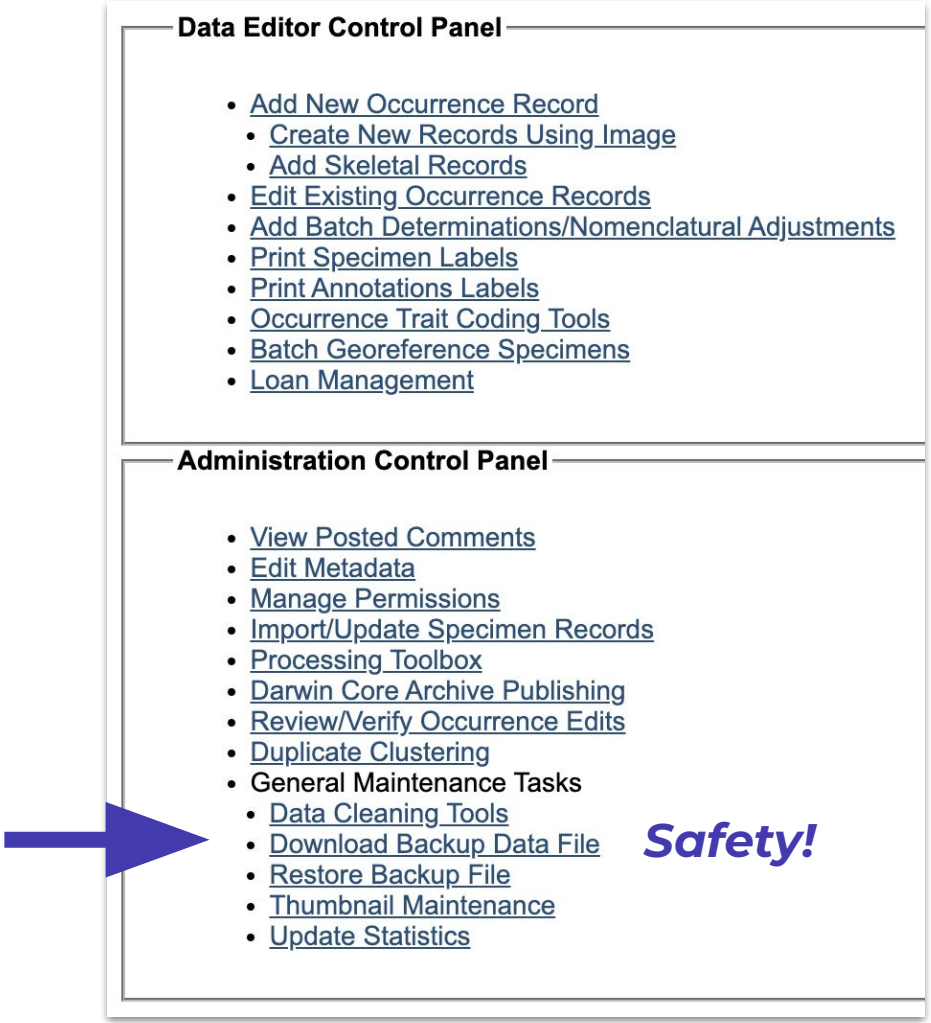
Batch Editing

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

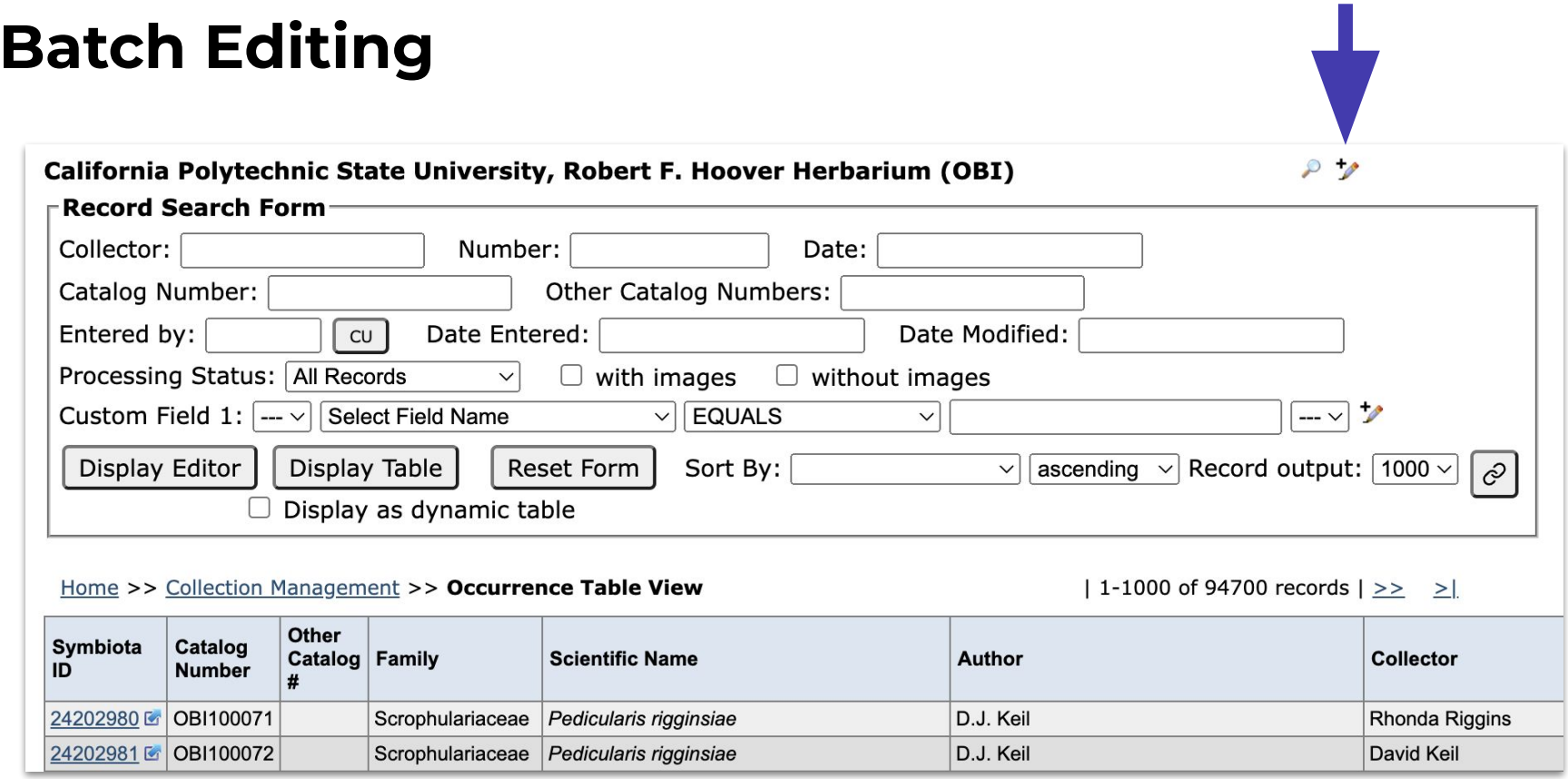


Batch Editing

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



Batch Editing



SEINet Portal Campaign Arizona-New Mexico Chapter

Final Office Hours - July 18, 2023

iDigBio
Integrated Digitized Biocollections

Symbiota

Agenda

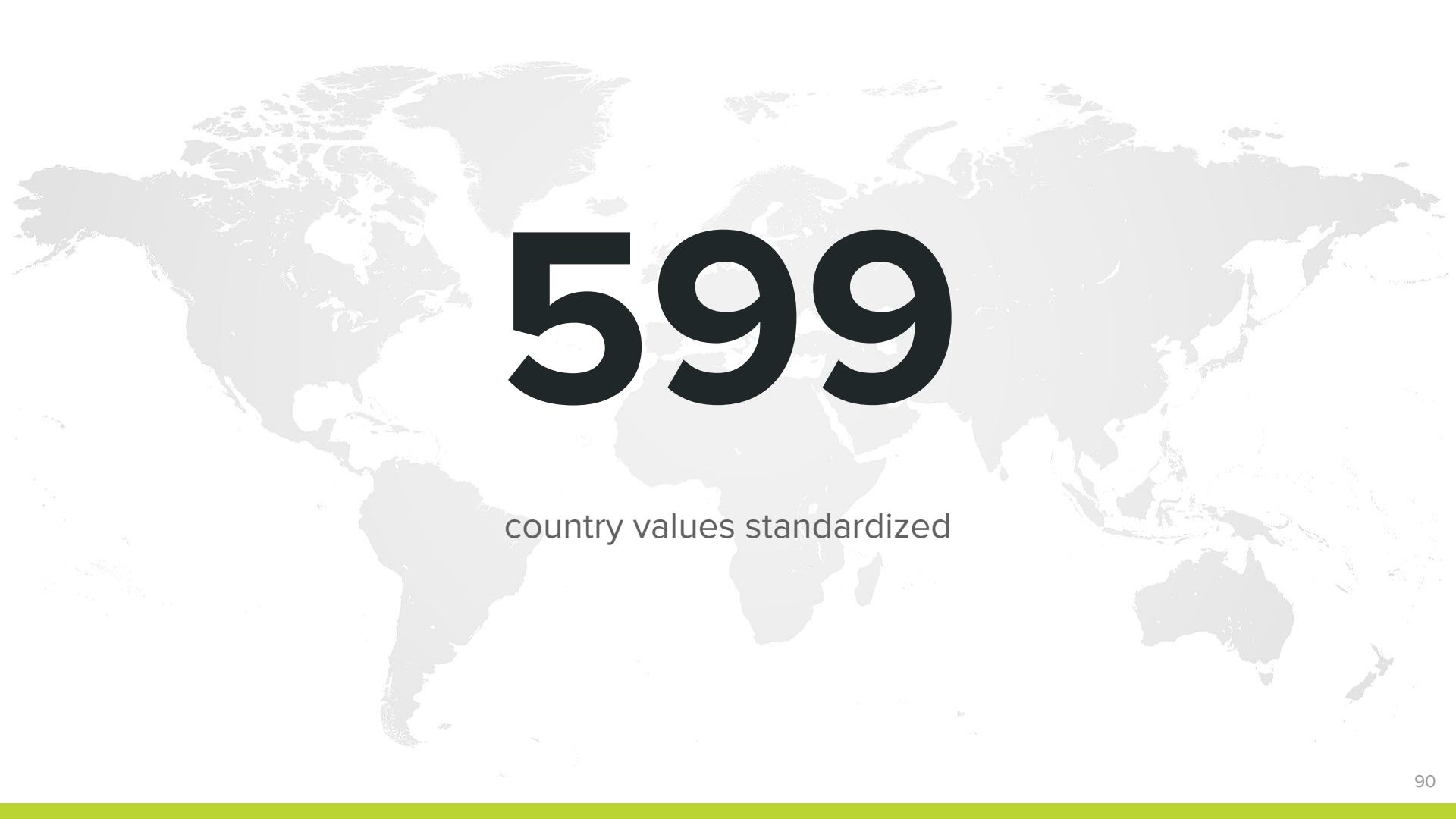
- Portal campaign accomplishments: a recap
- Publishing to GBIF & iDigBio
- Further opportunities

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

13,887 specimens added!

(@ ASU)

Tonto National Forest

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

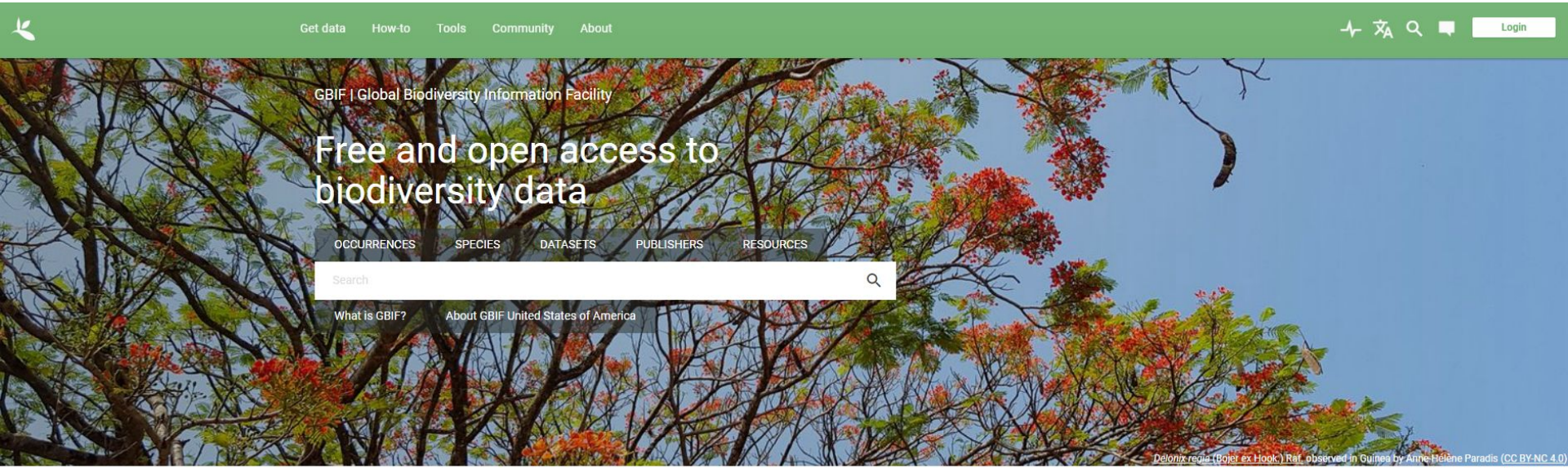
- Portal campaign accomplishments: a recap
- **Publishing to GBIF & iDigBio**
- Further opportunities

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 Global Biodiversity Information Facility

GBIF data portal



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

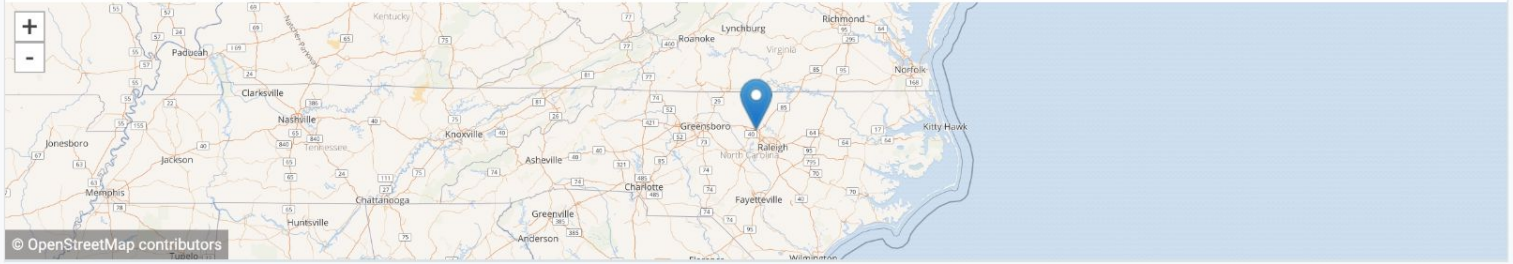
Duke University Herbarium

ABOUT METRICS HOME PAGE

313,960 OCCURRENCES 4 DATASETS 308 CITATIONS

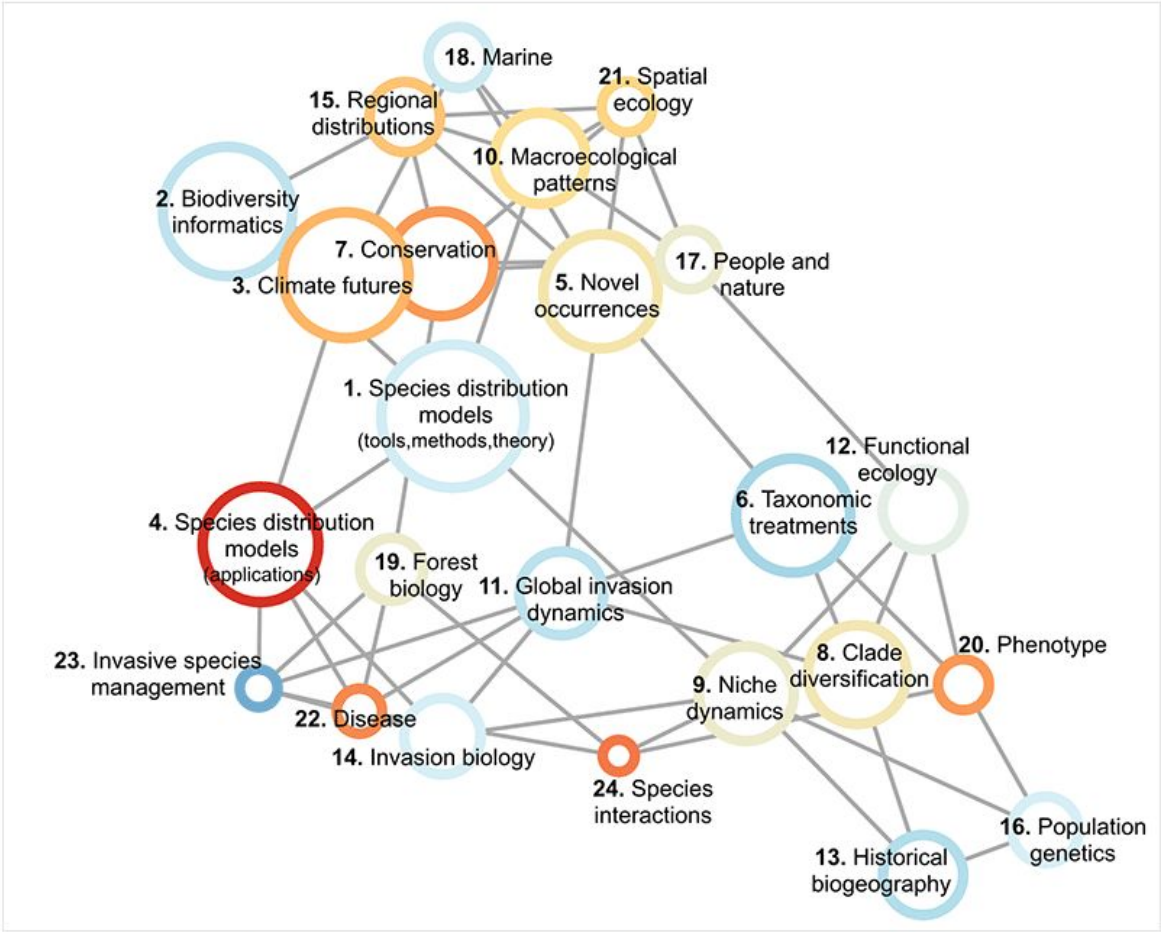
Description: http://www.biology.duke.edu/herbarium
Endorsed by: U.S. Geological Survey
Administrative contact: Blanka Aguero
Country or area: United States of America

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**

Get data How-to Tools Community About Login

PUBLISHER | SINCE MAY 3, 2010

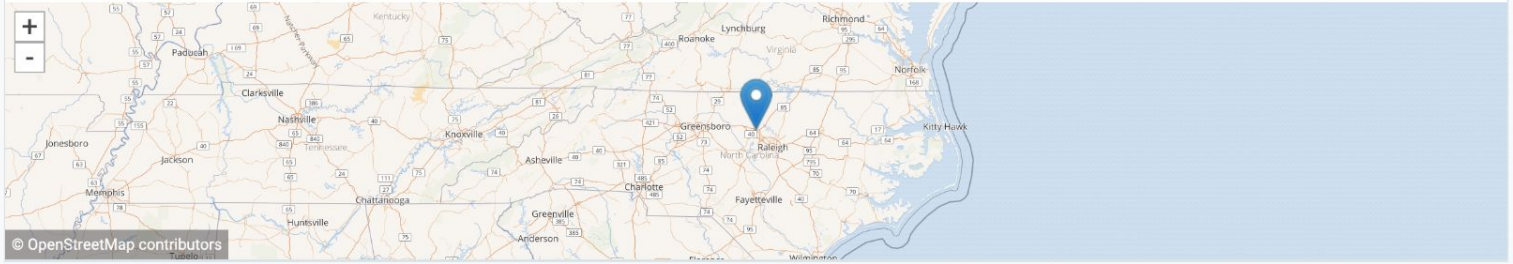
Duke University Herbarium

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
- Citation tracking
- **Data quality flags**

Get data How-to Tools Community About Login

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

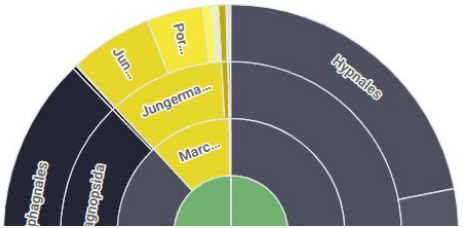
TAXONOMIC DISTRIBUTION OF OCCURRENCES

Explore Major groups

Plantae 173,759

incertae sedis 110

TAXONOMIC DISTRIBUTION OF OCCURRENCES



Get data How-to Tools Community About Login

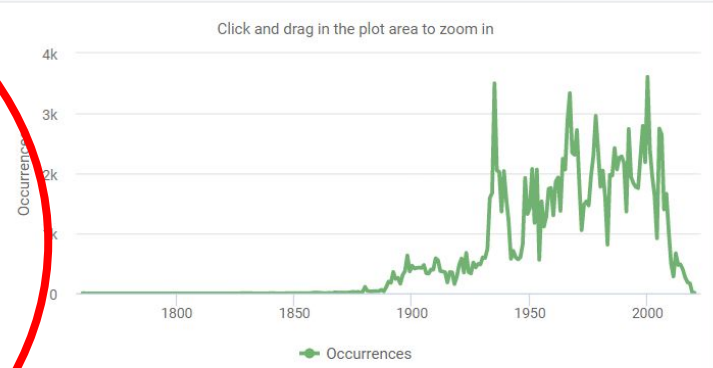
DATASET METRICS ACTIVITY DOWNLOAD 173,869 OCCURRENCES 73 CITATIONS

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

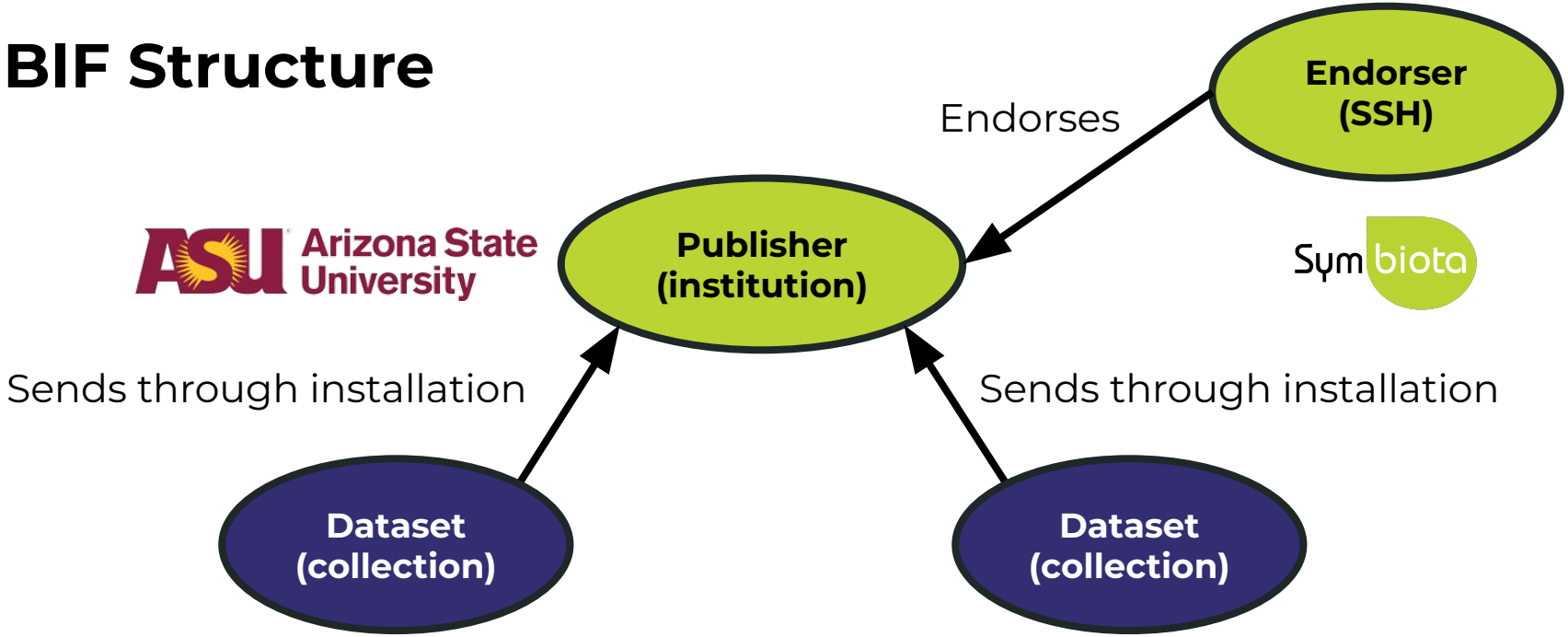
OCCURRENCES PER YEAR



12,213 other or unknown

How do I publish to GBIF?

GBIF Structure



Arizona State University

Symbiota

Arizona State University Mammalogy Collection

Arizona State University Vascular Plant Herbarium

Consortium of Small Vertebrate Collections

SEINet Arizona - New Mexico Chapter

PUBLISHER | SINCE APRIL 4, 2017

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

CONTACTS

SEARCH DATASETS | 18 RESULTS

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

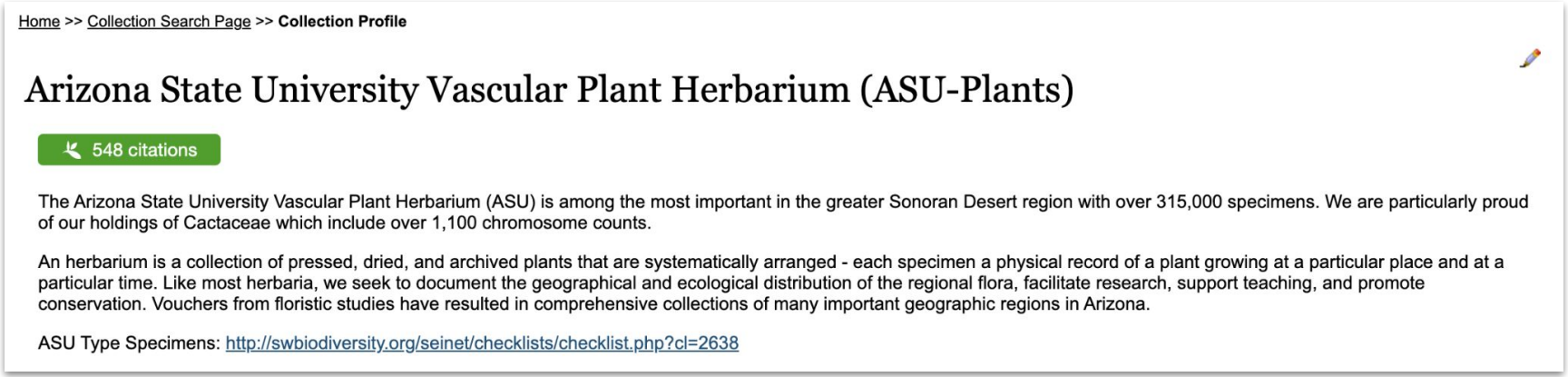
Arizona State University Herpetology Collection

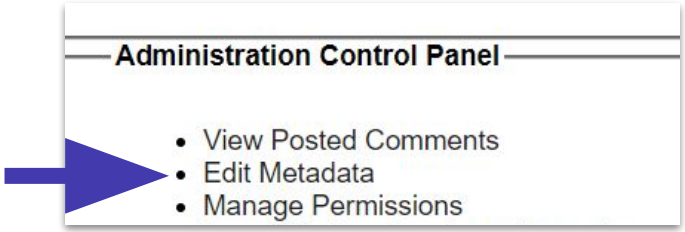
Occurrence dataset

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...

Publishing to GBIF from a Symbiota portal

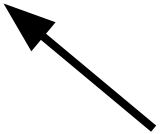
1. Make sure your **metadata are accurate**





Publishing to GBIF from a Symbiota portal

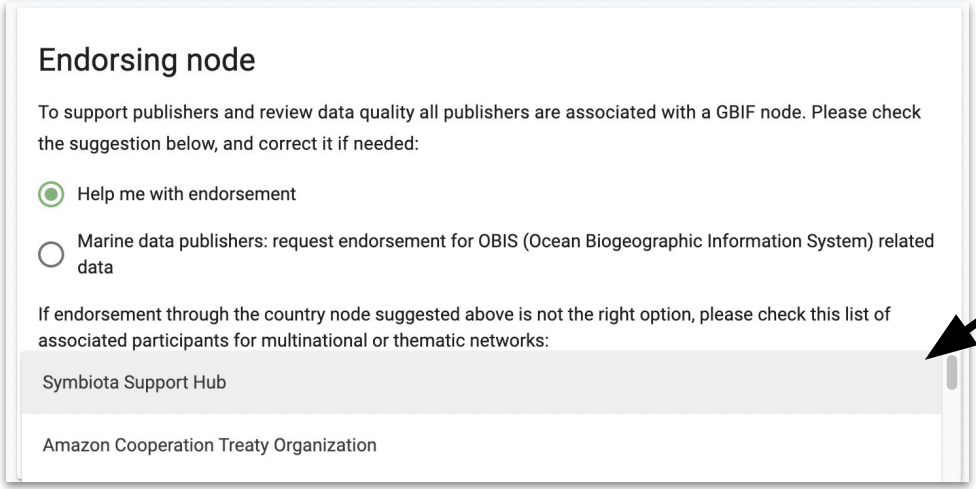
1. Make sure your metadata are accurate
2. *(If not already done)* **Request to become a publisher** in GBIF: https://www.gbif.org/become-a-publisher



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



SSH can (and wants to) endorse you!

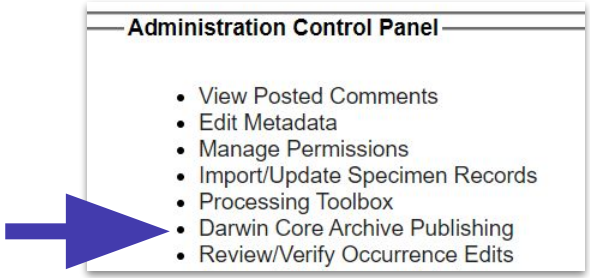
Publishing to GBIF from a Symbiota portal

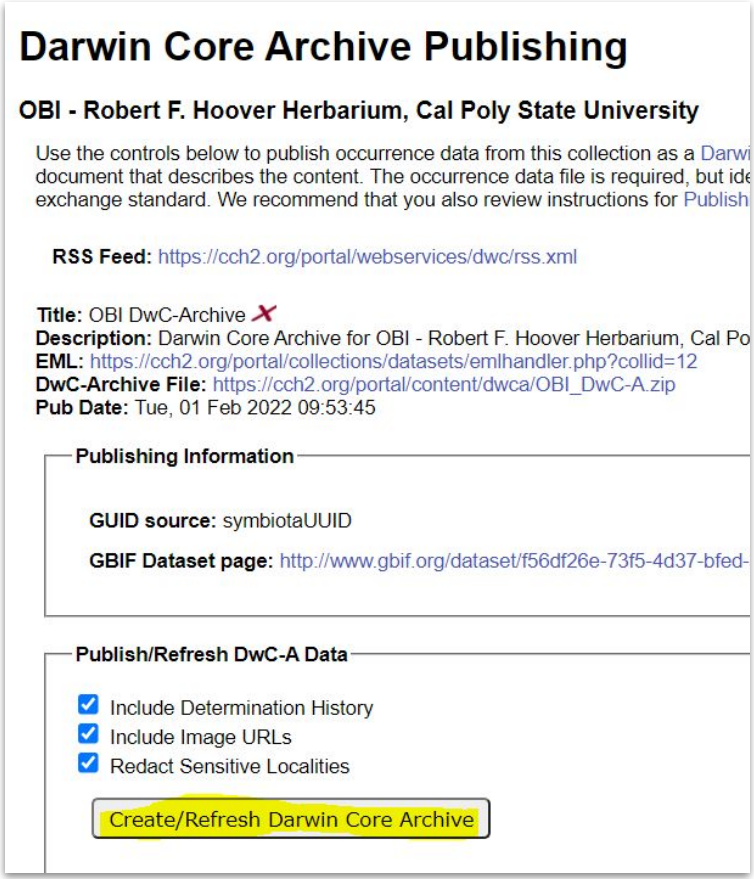
1. Make sure your metadata are accurate
2. *(If not already done)* Request to become a publisher in GBIF: https://www.gbif.org/become-a-publisher
3. Once approved, enter your **dataset key** in “Darwin Core Archive Publishing” in your portal
4. **Email GBIF** (helpdesk@gbif.org) to notify them that the 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





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—l 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

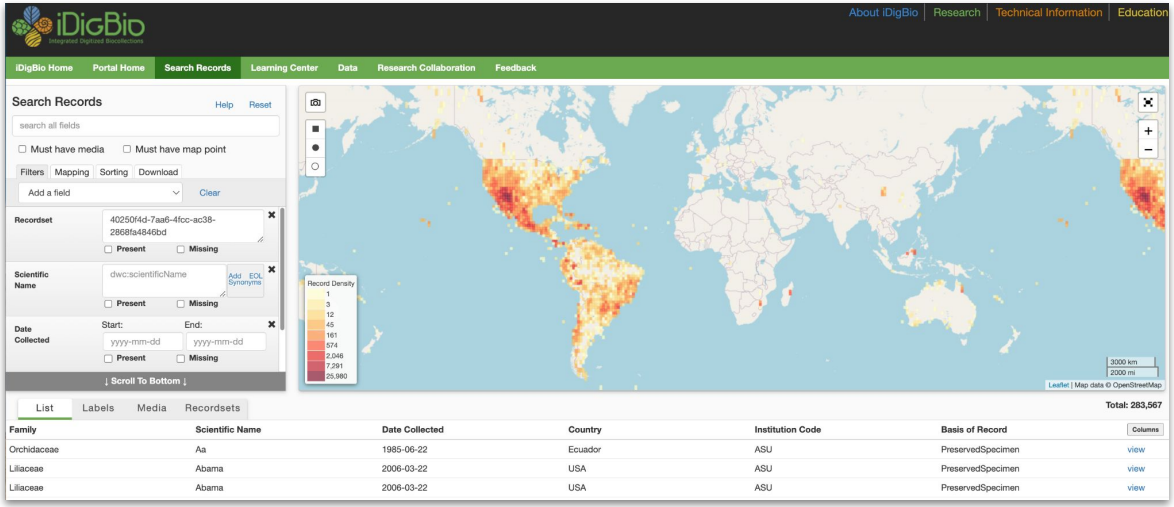
Address: Robert F. Hoover Herbarium
Biological Sciences Department, California Polytechnic State University
San Luis Obispo, CA 93407-0401
USA
(805) 756-5869

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.**



Questions?

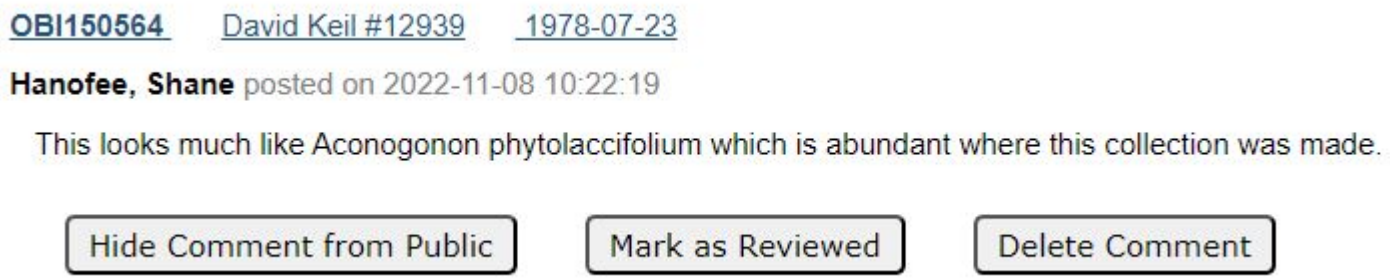
Agenda

- 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



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.

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!

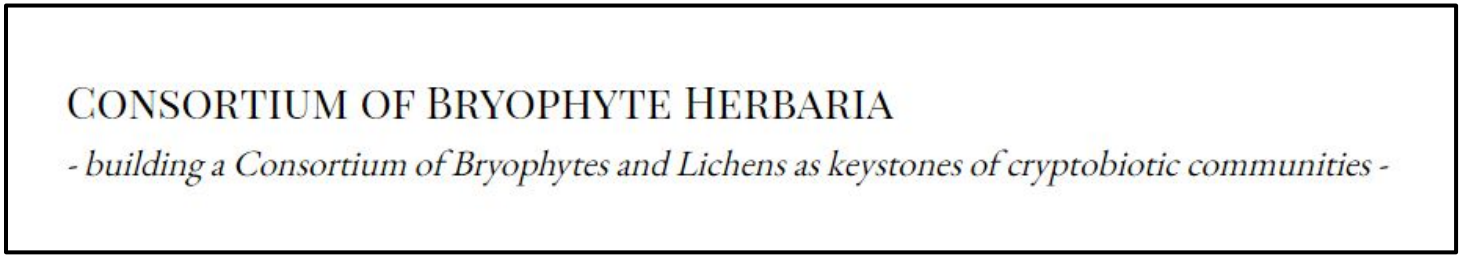
Harvesting georeferences from duplicates

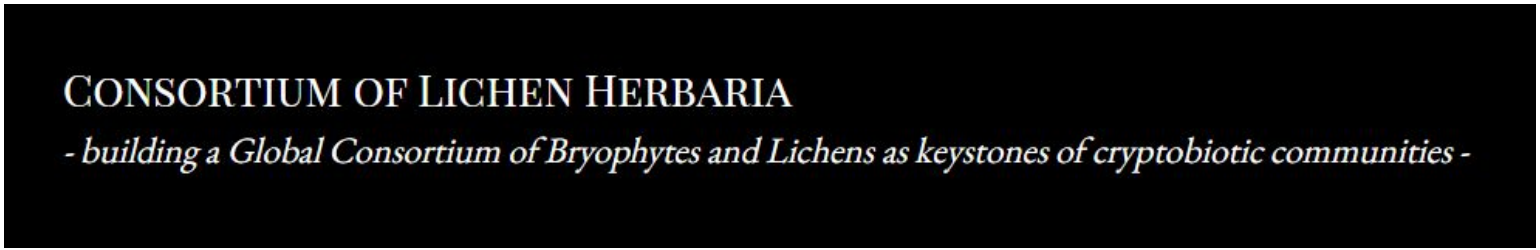
Interested? Fill out this form:

https://forms.gle/9Cg6jrUmxucST5po6

Standing issues and opportunities

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





Adding data to other taxonomic-focused portals

Interested? Fill out this form:

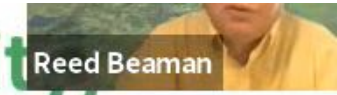
https://forms.gle/9Cg6jrUmxucST5po6

Ideas for further digitization funding

- NSF grants
- IMLS grants
- Institutional grants

symbiota.org/funding-ideas

Infrastructure Capacity for Biology (Capacity)



- **Synopsis**:

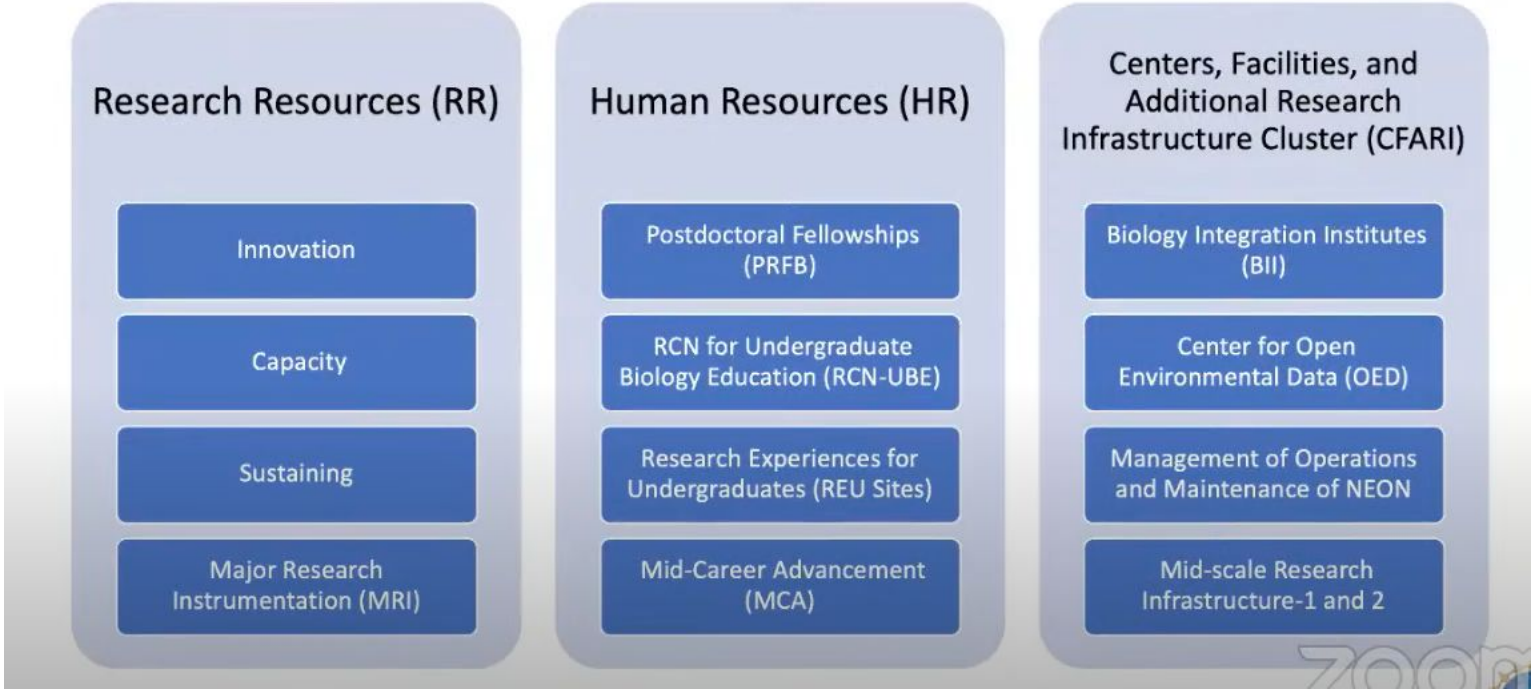
  Support the implementation of, scaling of, or major improvements to research tools, products, and services that advance contemporary biological research.

- **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



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

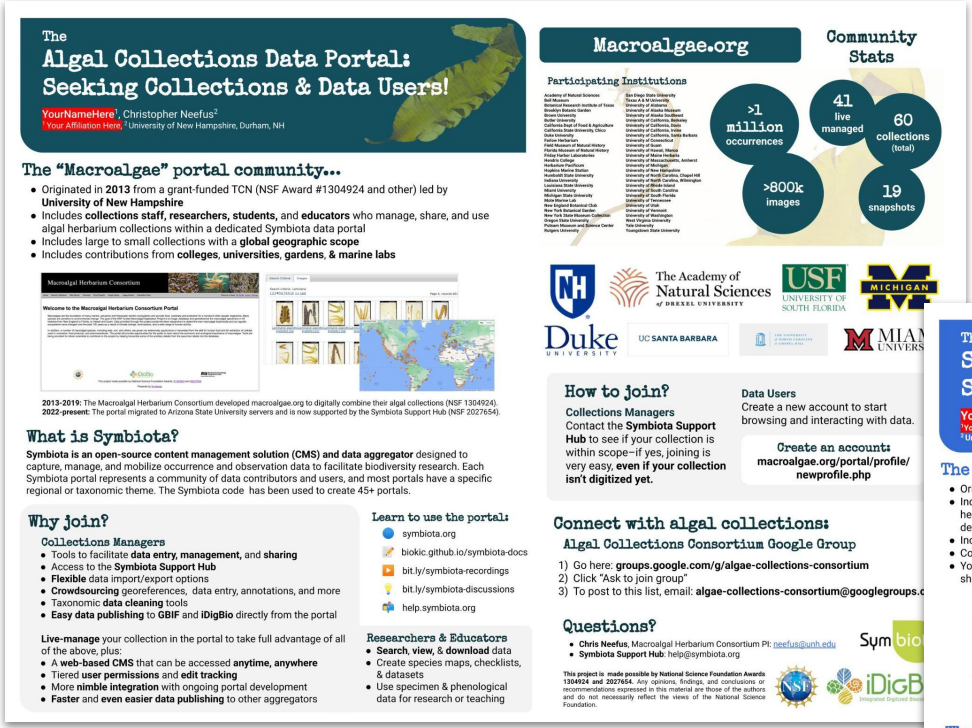
1. Determine need(s)/want(s)
2. Meet with Symbiota Support Hub team to discuss possibilities and necessary funding.
3. With SSH help, include development in budget. Three options:
    a. Collaborating institution
    b. Subaward
    c. Contract with BioKIC Services

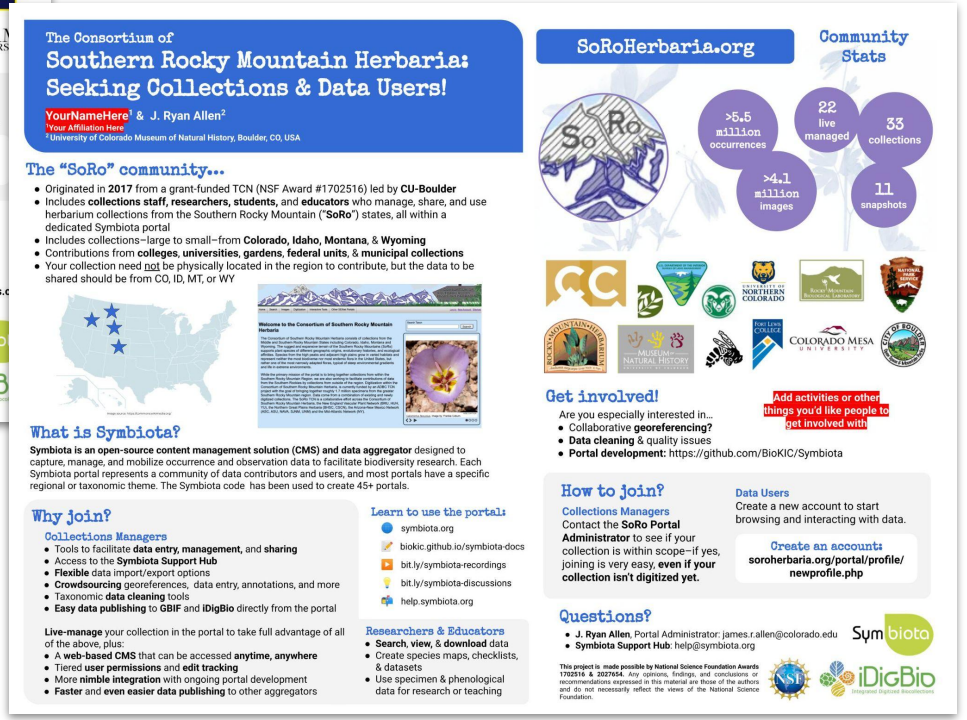
No grant?

- Contract with BioKIC Services

services.biokic.asu.edu

Promotional Poster





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