SPNHC 2025 | SYM08

Things to Know Before Publishing Your Data

# How the Symbiota Support Hub Can Help Publish Your (Extended) Specimen Data





Lindsay Walker, Ed Gilbert, Mark Fisher, Samanta Orellana, Katie Pearson, Gregory Post, Nikita Salikov, Logan Wilt, Jenn Yost, & Nico Franz

Symbiota Support Hub
University of Kansas
Biodiversity Institute & Natural History Museum



SAN LUIS OBISPO

# Symbiota / Symbiota Support Hub What is it? Who are we?



## What is Symbiota?

Open source software for creating themed data portals that help people share and actively manage biodiversity data.

- Taxonomic
- Geographic
- Institutional
- Project-based
- ★ Based on the Darwin Core data standard



Photo: Natalie Vondrak/University of Kansas

# What is the Symbiota Support Hub?

Team dedicated to the technical and social implementation of Symbiota-based data portals

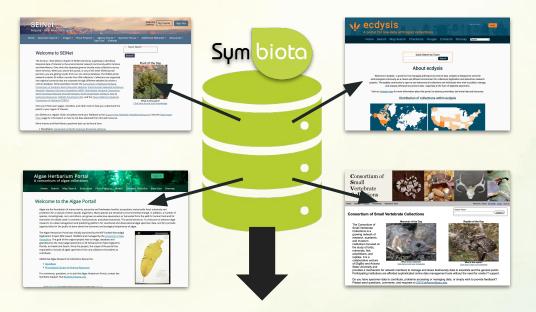






# Symbiota + Data Publishing

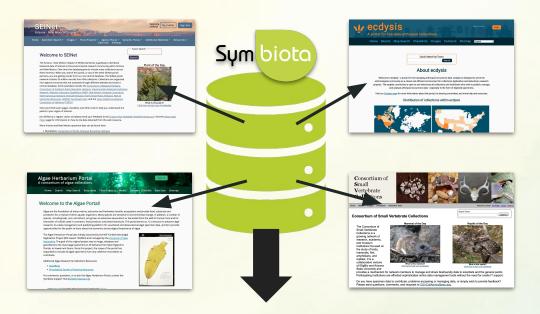
A Brief Introduction



# Symbiota + Data Publishing

Symbiota has **built-in** data publishing tools to help you share data from Symbiota to other aggregators





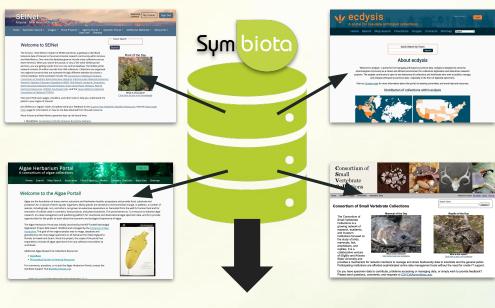
# Symbiota + Data Publishing

Symbiota has **built-in** data publishing tools to help you share data from Symbiota to other aggregators





GBIF
iDigBio
+ other Symbiota portals

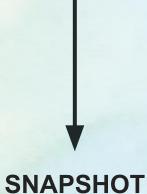






(The portal is your database)

**LIVE-MANAGED** 



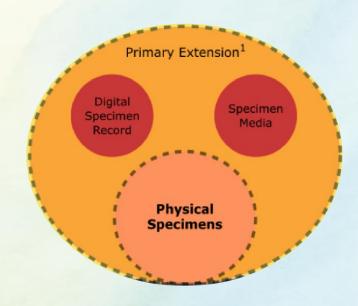
(You actively manage your data elsewhere)

# Prepare for Data Publishing: What kinds of data can Symbiota publish?

# **Primary Occurrence Data**

Digitized metadata describing physical specimens

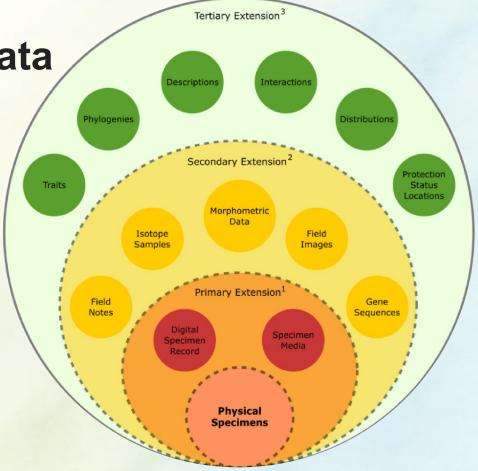
- Catalog Number
- Taxonomic identification
- Collecting event
- Collecting location
- Geological context



**Extended Specimen Data** 

A single specimen extends "beyond the singular physical object to potentially limitless additional physical preparations and digital resources"

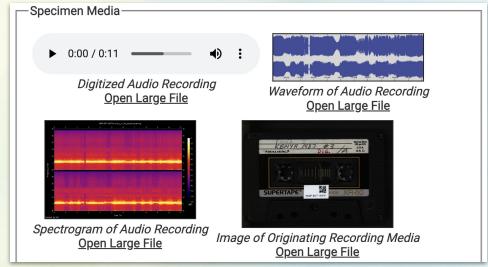
–Lendemer et al. 2020



## **Media Records**

- = Links to images and audio files
  - Files can be hosted externally or by the SSH
  - Now includes image tag values,
     e.g. "HabitusLateral"
  - Published to multimedia.csv following the Audiovisual Core standard

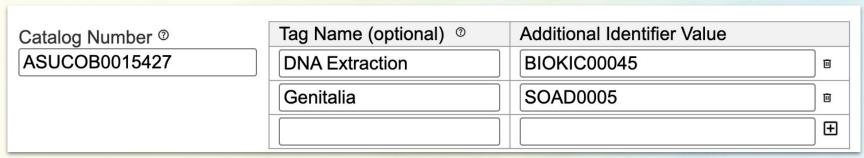




ANSP-ENT-143711-side\_a\_clip4 Academy of Natural Sciences (CC-BY-NC)
UCSB-IZC00045302 University of California Santa Barbara (CC0)
https://ecdysis.org/collections/individual/index.php?occid=5702178
https://library.big-bee.net/portal/collections/individual/index.php?occid=2128034

## **Alternative Identifiers**

- = Deprecated catalog numbers, subsample numbers, accession numbers, etc.
  - Can be directly entered or imported into Symbiota as key:value pairs
  - Published as parsed values in identifiers.csv
    - Also concatenated in dwc:otherCatalogNumbers



# **Determination History**

- Can be added to occurrences one-by-one or imported in bulk
- "Current determination" exported via dwc:scientificName
- All determinations are parsed and published to identifications.csv

# Toxonotus championi (Jordan, 1906: 342) CURRENT DETERMINATION Determiner: K.S. Orellana (0000-0002-4098-5823) Date: 2022-01-28 Reference: Rheinheimer, J. (2004) Illustrierter Katalog und Bibliographie der Anthribidae der Welt (Insecta: Coleoptera). Mitteilungen des Entomologischen Vereins Stuttgart, 39(1/2), 1-243. Notes: Nomenclatural change. Neanthribus championi Jordan, 1906 Determiner: C.W. O'Brien Date: 2012

## **Traits**

- = Phenological, reproductive,substrate-related, etc.
  - Configured on a portal-by-portal basis in the Traits module
- Entered record-by-record or using Attribute Mining Tool
- Published to measurementOrFact.csv

Trait: Angiosperm Pher Reproductive Unopen Flower: Present Absent	nological Traits————————————————————————————————————
Open Flower:  Present Absent  Senesced Flower: Present Absent  Fruit: Present	Lactation:  Lactating  Not Lactating  Cannot be determined  Pregnancy:  Pregnant  Not pregnant  Cannot be determined
☐ Absent ☐ Sterile ☐ Not scorable	

https://www.cch2.org | https://biorepo.neonscience.org

# **Material Sample**

#### E.g., subsample data

- Configured on a portal-by-portal basis
- Developed for the NEON Biorepository
- Can be associated with occurrence records one-by-one
- Published to materialSample.csv



#### Material Sample

Sample Type: skull

Catalog Number / Barcode: D00000013639

Condition: unchecked
Disposition: in collection
Preservation type: dry
Preparation date: 2021-04-09

Prepared by: Engasser, Emmy

#### Material Sample

Sample Type: whole organism

Catalog Number / Barcode: D00000013639

Condition: fair

Disposition: in collection

Preservation type: ethanol-fixed, stored in 75% ethanol

<u>Preparation details</u>: skull removed <u>Preparation date</u>: 2021-04-09 <u>Prepared by</u>: Engasser, Emmy

Remarks: abrasions on face, wounds on back anterior to tail, notch in right ear

### **Genetic Data**

- = Links to records in external repositories
  - Manually entered or imported/exported using dwc:associatedSequences



UVGA007015 Universidad del Valle de Guatemala (CC-BY-NC-SA)

VRT 27-APR-2021



Bolitoglossa qeqom voucher UVG A-007015 large subunit ribosomal RNA gene, partial sequence; mitochondrial

GenBank: MW991409.1

FASTA Graphics

Go to: ♥

LOCUS MW991409 525 bp DNA

DEFINITION Bolitoglossa qeqom voucher UVG A-007015 large subunit ribosomal RNA gene, partial sequence; mitochondrial.

ACCESSION MW991409
VERSION MW991409 1

https://www.ncbi.nlm.nih.gov/nuccore/MW991409

# Associations (Linked Resources)

- Associations between cataloged specimens and other resources
  - Internal or external to your portal
  - Includes occurrences, uncataloged taxon observations, and more
  - Coming soon: associations.csv



#### **Related Occurrences**

isReferencedBy:

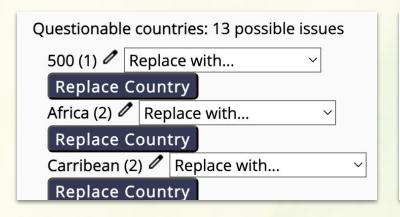
https://www.biodiversitylibrary.org/page/7764350 isReferencedBy:

https://hdl.handle.net/2027/mdp.39015018645831 ecologicallyOccursWith: [Fagopsis longifolia (Lesq.) Hollick]

https://paleo.symbiota.org/portal/collections/individual/index.php?occid=763783 https://www.biodiversitylibrary.org/page/7764350 (CC0) | www.morphosource.org

# Prepare for Data Publishing: How do I publish data using Symbiota?

- 1) Ingest and prepare your data in a Symbiota portal
- ★ Use your portal's built-in data cleaning tools to correct typos in scientific names and geographic data (if "live managed")



Resolving #11: Haemorhouse purpureus (Fringillidae) [9 specimens
Checking Catalog of Life...
Taxon not found
Checking close matches in thesaurus...
Haemorhous purpureus => remap to this taxon
Haemorhous purpureus californicus => remap to this taxon
Haemorhous purpureus purpureus => remap to this taxon
Manual search:

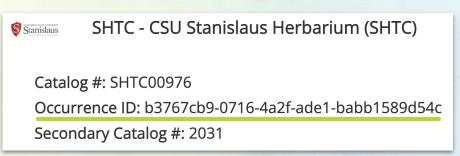
Remap

- 1) Ingest and prepare your data in a Symbiota portal
- **★** Review the **Symbiota Data Quality Toolkit**



- 1) Ingest and prepare your data in a Symbiota portal
- ★ Avoid duplication of records by maintaining consistent occurrenceid values

Please ask for help if you're not sure!



https://www.cch2.org/portal/collections/individual/index.php?occid=3613371

#### 2) Make sure your collection's metadata are accurate

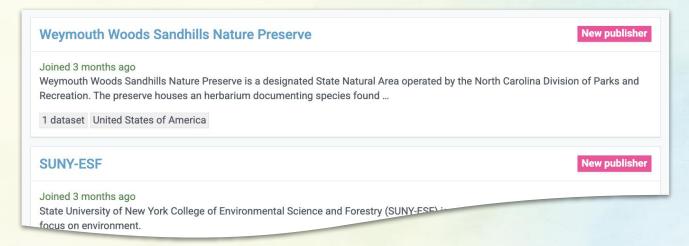
Collection title, description, contacts, etc.





3) Request to become a GBIF Publisher (if needed)

As a GBIF Associate Participant Node, the Symbiota Support Hub can help with this step for US-based collections.



- 4) Notify SSH once you're endorsed
- ★ Often, we can do the rest for you!





# Summary

Symbiota portals are platforms for publishing primary and extended specimen data.



The Symbiota Support Hub can facilitate many steps in the data publishing process!









## Thank you!

NSF Award # 2027654 SPNHC Local Organizing Committee Relevant feature development:

Academy of Natural Sciences
Big-Bee TCN
Botanical Research Institute of Texas
California Botanic Garden
George Mason University
National Ecological Observatory Network

Contact: <a href="mailto:help@symbiota.org">help@symbiota.org</a>
or come to our SPNHC Help Desk!



Portal Directory
(Find a portal & publish your data!)