



Native Bee Occurrence Data Standardization and Sharing Guidelines

Background

The following guidelines are the culmination of several collaborative sessions held by participating members of the Texas Native Bee Research Coordination Network from March to July of 2024.

Members of this RCN work with native bees and entomological data in various capacities and represent a wide array of universities, non-profits, government agencies, and natural history collections across the state of Texas (and surrounding region) and included professors, graduate students, professional entomologists, collections curators, conservation professionals, and federal and state wildlife biologists.

Collaboratively the group discussed and synthesized best practices for standardizing occurrence data for sharing and uploading to publicly accessible online databases, with the goal of updating occurrence records for native bees across the state of Texas to inform future conservation priorities for native bees in the state.

Table of Contents

<u>Introduction</u>	<u>2</u>
<u>Overview</u>	<u>2</u>
<u>Data Standardization Guidelines</u>	<u>3</u>
<u> Occurrence Data Templates</u>	<u>3</u>
<u>RCN Preferred Online Databases</u>	<u>4</u>
<u> Decision Tree: Where should I upload my data?</u>	<u>5</u>
<u>Data Sharing and Upload Guidelines</u>	<u>6</u>
<u>Using Symbiota Portals for Research</u>	<u>8</u>
<u>Resources and FAQ's.....</u>	<u>9</u>



Introduction

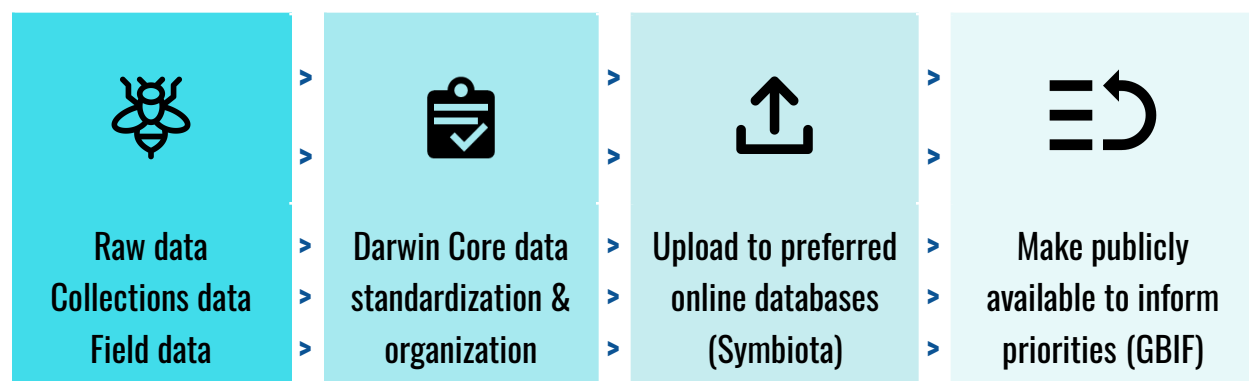
As a state, Texas has a vibrant community of environmental professionals who work with native bees in a variety of capacities, yet the region still suffers from a lack of publicly available occurrence records for native bees. These records, shared via platforms such as [Symbiota](#) and [GBIF](#) (Global Biodiversity Information Facility), are used by the research and conservation communities at large to inform work priorities across the region. Therefore, it is important that these records be as current and comprehensive as possible.

As members of the Texas Native Bee RCN and the entomological and ecological research/conservation communities in Texas, we are committed to making relevant occurrence data from our work publicly available when possible. The following guidelines provide a step by step protocol that will help members of the native bee research community standardize their data and streamline the uploading process to online hosting sites and databases.

Overview

As data contributors can come from a variety of fields and are collecting native bee occurrence data for a variety of purposes, we have provided a general overview of the data standardization and sharing/upload workflow we will be discussing in the following guidelines. In general, we hope to help those collecting native bee occurrence data to transform their field or collections-based records into a standardized Darwin Core format that will be optimal for uploading onto the [preferred online databases of the Texas Native Bee RCN](#) and streamline the data sharing process.

General Workflow:





Data Standardization Guidelines

In adherence with [data fields in Symbiota portals](#), the Texas Native Bee Research Coordination Network strongly aligns its data standards with the [Darwin Core Standards](#) established by Biodiversity Information Standards (formerly known as the Taxonomic Databases Working Group aka “TDWG”). These standards provide a set of universal terms that can be used to label your data so it is easier to upload and share, and these terms are to be used as column headers in your data sheets. The Texas Native Bee RCN collaboratively decided on a subset of Darwin Core terms that are likely to be relevant to native bee researchers and developed a data template for the community to use in organizing and standardizing their data. We encourage all data contributors to consider adopting this Darwin Core terminology when developing their data sheets into the future as it will streamline the data sharing process and encourage reciprocity throughout the community.

Occurrence Data Templates:

We have developed the following occurrence data templates to help our community streamline the process of utilizing Darwin Core headers when labeling their data. These templates have been collaboratively curated by members of the Texas Native Bee RCN so that other members of the native bee research community in Texas will not need to search through the immense Darwin Core glossary of terms to find ones that are relevant to their research.

[Texas Native Bee RCN Darwin Core Data Template - Annotated](#)

[Texas Native Bee RCN Darwin Core Data Template - Blank](#)

Using these templates:

We have provided two data templates to help you in your data standardization process:

- The first is **annotated** and will help you identify terms that are relevant to your data. Please read the notes about the colors of the column headers. We have provided basic definitions and potential associated terms you may be using in your data sets. We have also provided links to the Darwin Core glossary if you want to explore these terms in more depth.
- The second template, which is **blank**, is so you can easily transfer column headers to your data set. You can either create an editable copy of this template and build your data set in the template, or you can copy/paste individual headers onto existing data sets.



RCN Preferred Online Databases

There are a number of online databases and hosting sites available for scientists working with biodiversity to share their data. Currently, the native bee researchers community across the state are not using any one of these sites universally. This means that digitized native bee occurrence records are dispersed across a variety of platforms, making it difficult to find comprehensive data on a given taxa.

As such, the RCN thought it was important to determine a **preferred online database** that, as a community, we agree to prioritize for sharing our occurrence data. From our collaborative sessions and discussions we determined the following:

For data that will be “live-managed” (data are continuously/regularly updated and maintained within the portal’s user interface; any records you add to this portal will represent your canonical specimen records)

- **Ecdysis** (a [Symbiota](#)-based portal): Symbiota is an open-source software platform for managing and sharing biodiversity data through a network of institution, project, or taxa-specific “research portals”. **Ecdysis is most relevant for vouchered specimen data that is being regularly updated (adding new records/editing existing records) within the portal**, such as collections data, or collecting efforts/research that will be ongoing.
 - ◆ Within Symbiota, we use the [Ecdysis portal](#) for our live-managed data. The Ecdysis portal is specifically focused on terrestrial arthropod occurrence data.
 - Ecdysis is not limited to bee data. Data associated with any terrestrial arthropod taxa can be shared in this portal.
 - ◆ Data from Ecdysis can be published to GBIF, as long as you are affiliated with an institution that has been “endorsed” by Symbiota ([See below](#)).

For “snapshot” data, or a static copy of your data (not actively managed data)

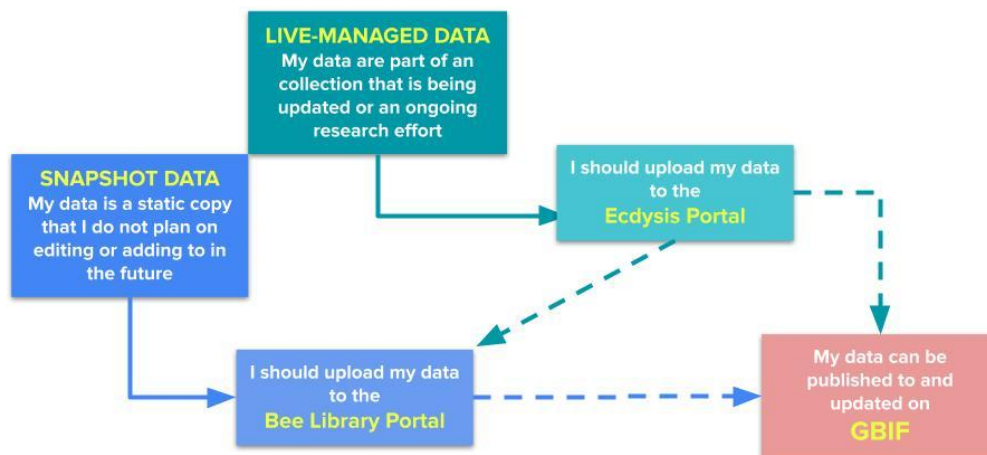
- **The Bee Library** (another [Symbiota](#)-based portal) is an online repository of bee image, trait, and specimen data. **The Bee Library is most useful for “snapshot data”** or occurrence records that come from datasets that are a static copy of your data, meaning **you do not intend to actively manage these records within the portal**, though you may occasionally share refreshed copies of your records by uploading a new spreadsheet.
 - ◆ For example, we encourage members of the RCN to upload data from discrete or completed projects and past publications on [The Bee Library](#).
 - ◆ Data from the Bee Library can also be published to GBIF ([see below](#))
 - ◆ Note: Additionally, data hosted on Ecdysis will be automatically pushed to the Bee Library, making it a fantastic search engine for our group.

- **Note:** If you are new to Ecdysis, the resources linked at the bottom of [this page](#) will help you learn more about how to use this portal for managing your data. [Similar resources](#) have been created for the Bee Library.
- **Note:** A copy of bee data published to Ecdysis will automatically be sent to the Bee Library each time it is updated, but data from the Bee Library will not be shared on Ecdysis.

Data access for the conservation community and the general public

- **GBIF** (Global Biodiversity Information Facility) is the gold standard in the conservation community. The U.S. Fish and Wildlife Service and the Texas Parks and Wildlife Department use records on GBIF to make their species status/listing determinations. Therefore, we want our data to ultimately end up on GBIF.
 - ◆ Unfortunately, GBIF is not the most user-friendly for data uploading, so we will be using Ecdysis (for live-managed collections) and the Bee Library (for snapshot datasets) to send data to GBIF.
 - ◆ Conveniently, data uploaded/published on Symbiota portals can be published to the GBIF site if your collection is affiliated with an academic/government/research institution or organization.

Decision Tree: Where should I upload my data?



Broadly, data will be uploaded to the two preferred RCN Portals (Ecdysis or the Bee Library), then can be published directly to GBIF from these sites.

- A copy of data from Ecdysis will also automatically be published to the Bee Library each time it is updated. Data from the Bee Library will not appear on Ecdysis.

If you have any questions about which hosting site is most appropriate for your data, please reach out to the Symbiota help desk at help@symbiota.org.



Data Sharing and Upload Guidelines

Now that you have organized and standardized your data, then selected the appropriate hosting site for your data, it is time to begin the upload process! Please watch [this video](#) for a more detailed explanation of the Symbiota upload process.

→ Getting data uploaded to Ecdysis and Bee Library Portals

- ◆ Step 1: You will fill out [this form](#), then **you must email help@symbiota.org to identify yourself as a member of the Texas Native Bee RCN.**
 - Select “Bee Library: Texas Bee RCN” as the portal you’d like to join.
 - Your response to the form will alert the Symbiota Support Hub to forward data shared with Ecdysis to the Bee Library, as applicable.
- ◆ Step 2: If you do not already have a user account in the portal(s) that you’ll be joining, create one in each portal. The Symbiota Support Hub will link your user account to your dataset:
 - [Create an account in Ecdysis](#)
 - [Create an account in Bee Library](#)
- ◆ Step 4: Collect your Darwin Core formatted data files
 - If you have questions about your uploads, the Symbiota Support Hub can help upload your files, reach out to help@symbiota.org
 - **Important:** Each specimen **must** be associated with a unique identifier (“Catalog Number”) in order for its record to be uploaded to the portal, e.g. [“ASUCOB0002331”](#).
- ◆ Step 5: Login to the portal and navigate to the *Administration Control Panel* for your dataset (*My Profile > Occurrence Management > Admin Control Panel*). elect the **Import/Update Specimen Records** option
 - If you have a CSV to import, you will use the Full Text File Import option
 - Video tutorial: <https://youtu.be/zrwHlDtae-c?si=DpfuzcolzKwvULgo>
 - Instructions: https://biokic.github.io/symbiota-docs/coll_manager/upload/
- ◆ Step 6: Map your data headers to Symbiota fields, verify, save your import profile
- ◆ Step 7: Double check your processed data in the staging table
 - To ensure there are no errors you can check your data before submitting your data.
 - **Tip:** The first time you add data to the portal, upload a small test batch of representative records (<10) to ensure your data will be imported into the portal properly *before* uploading a larger number of records. You will be able to manually delete erroneous records [one-by-one](#), but not in bulk.
 - **If you have any doubts** about your data upload, reach out to help@symbiota.org **before hitting publish** (next step).
- ◆ Step 6: Hit **Publish** and your data will be live on Ecdysis portal shortly!

→ Note for adding Snapshot Data to the Bee Library

- ◆ For those of us who may be uploading discrete datasets from projects and/or past publications, **datasets that will not continue to be updated within the portal (i.e. the portal is not the canonical source of your specimen data)**, the Symbiota Help Desk gave us a recommendation:
 - **If possible, try to upload your discrete datasets as one large, notated dataset from you or your lab.**
 - *For example, it makes the process a lot easier for the Symbiota folks if all the data from past publications from the Jha Lab is uploaded as one large dataset all at once.*
 - You can add notes/unique identifiers to indicate which publication this data came from, but uploading this data together as one unit will help Symbiota process this data faster.
 - *If desired, once your data are in either Ecdysis or the Bee Library portal, you can use the [Dataset](#) feature to group your records by publication/project.*

→ Publishing data in Ecdysis to GBIF for the first time

- ◆ Step 1: Make sure your institution/organization is endorsed as a GBIF publisher.
 - To do so, create an account/login to GBIF
 - Then, click this [Become a Publisher Link](#)
 - Follow these [written instructions](#) and watch [this recording](#) for more guidance
 - If you would like to know more about the GBIF data publishing process in general, additional recordings are available: [YouTube Playlist](#)
- ◆ Step 2: Once you are registered/endorsed, email Symbiota at help@symbiota.org to let them know you would like to publish your Ecdysis dataset to GBIF.
- ◆ Note: After you send your data to GBIF for the first time, you do not need to repeat Steps 1-2. You can then refresh your data in GBIF by pushing this button:
Administration Control Panel > Darwin Core Archive Publishing > Create/Refresh Darwin Core Archive

→ Publishing data in the Bee Library to GBIF for the first time

- ◆ The general workflow for publishing data from the Bee Library to GBIF will be the same as with Ecdysis.
 - Step 1: A collection is established and data are added to The Bee Library (outlined above).
 - Step 2: The data provider needs to become an endorsed GBIF publisher (or can already be affiliated with one, i.e. their university) (see Step 1



outlined for publishing data from Ecdysis). GBIF will not accept data from collections that are not affiliated with an academic or government organization.

- Step 3: Contact the Symbiota Support Hub for assistance (help@symbiota.org). **Snapshot data requires that each specimen occurrence record has an associated Global Unique Identifier (GUID) across all platforms where data may be hosted.**
 - This way duplicates cannot be published to GBIF. The Symbiota portal can generate these identifiers for you, but be aware that you must always keep a copy of these identifiers in your home database with your canonical specimen records.
- Step 4: Data is then pushed from the portal to GBIF.

Using Symbiota Portals for Research

If you are interested in accessing data through Ecdysis and the Bee Library for your research, we wanted to provide some guidance to help you streamline this process.

Remember:

- **Ecdysis** will be where live managed data from our RCN collaborators will be maintained and updated.
- **Bee Library** will be where snapshot data (copies of datasets that will not be updated) from our RCN collaborators will be stored. A category dedicated to the RCN will help identify and aggregate datasets from our research group on [this page](#).
- Data from both of these portals will ideally be pushed and published to **GBIF**.
- Copies of data from Ecdysis **will be** eventually available on the Bee Library, snapshot data on Bee Library **will not be** available on Ecdysis.

Important Note: Depending on whether collaborators go through the process of getting registered with an endorsed institution and take the step to publish to GBIF, not all data present on each portal may be on GBIF at the time of your search.

Where to Begin: Ideally, all data across these platforms would be instantaneously up to date, but you may find this is not the case. Therefore, we recommend the following search order for your endeavors:

1. Start with the Bee Library.
 - a. The Bee Library will become the most comprehensive place to search for both snapshot records from past publications and research efforts, as well as occurrence records from live managed collections hosted on Ecdysis.



-
- b. The Bee Library team will be creating a subcategory for data on [this search form](#) that is uploaded by members of the Texas Native Bee RCN so we can track our uploads as a group.
 - c. If you have questions about the data you find on the Bee Library, the contact information of the person managing this data can be found with the data.
 2. Check out Ecdysis to find collections data and data that is being continuously updated. Unlike the Bee Library, this portal is not exclusive to bee collections.
 - a. If you have questions about the data you find on Ecdysis, the contact information of the person managing this data can be found with the data.
 3. Crosscheck with GBIF.
 - a. Most of the data you will find on GBIF from our RCN will have been uploaded through Ecdysis and the Bee Library.

Important Note: Additional relevant data from efforts outside the RCN will also be on all three of these sites, please check all three sites to cover your bases.

Resources and FAQ's

What if I have data on other sites that I would like to add to a Symbiota portal?

- If you have data uploaded on VertNet, iDigBio, or GBIF and you want it to also be uploaded on these Symbiota portals so it can be easily accessed with the RCN's data you can email Symbiota and they will help - help@symbiota.org.

What if I am an individual collector, and my records are not part of a collection at an endorsed institution?

- Data from personal collections can be uploaded to the Bee Library and Ecdysis, but to be published to GBIF (our ultimate goal), personal collections will need to be affiliated with an endorsed institution. The Texas Native Bee RCN is happy to connect individual collectors with appropriate endorsed institutions if they would like to make their data publicly available.

Can I edit data in two different portals at once?

- Please watch [this video](#) from Symbiota addressing data differences across portals and online hosting sites.



Resources

→ [Symbiota Docs \(written tutorials\)](#)

→ [Symbiota YouTube Tutorials](#)

→ **Bee Library:**

◆ symbiota.org/portal-advancement-campaigns/bee-library-portal-campaign

- Recordings on data ingestion, data publishing, and other topics

→ **Ecdysis:**

◆ symbiota.org/portal-advancement-campaigns/ecdysis-portal-campaign

- Recordings on data entry, data publishing, and other topics