Santa Barbara Botanic Garden Cal-IBIS Workflow: Adapted from an “Updating Cal-IBIS” document by Josie Lesage & Conservations with Ben Brandt

Created: March 2023 by Kylie Etter, Last updated January 2024

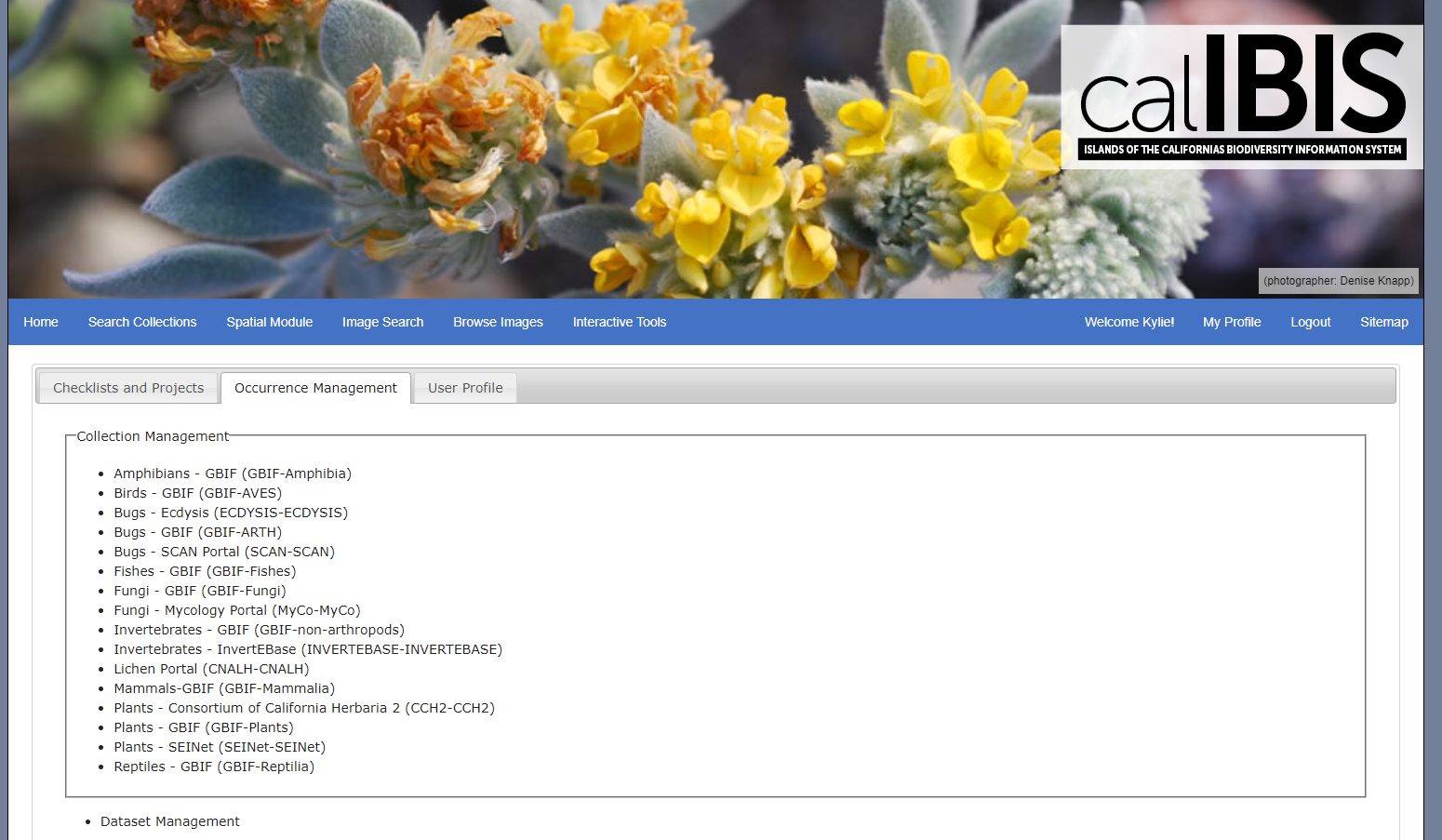
Background:

The California Islands Biodiversity Information System (Cal-IBIS, cal-ibis.org) is a database of biological specimen records meant to facilitate archipelago-wide data sharing, with the ultimate goal of protecting the archipelago’s biota and enhancing scientific understanding of it by better informing conservation management and research. This database includes snapshot records aggregated from other databases, including GBIF, SEINet, SCAN, and CCH2. Because there is some overlap between data sources, there’s a lot of cleaning that needs to happen “behind the scenes” before things get uploaded to Cal-IBIS. Cal-IBIS has snapshots of occurrences from most taxa on the Californias Channel Islands (Baja Islands included) and is a Symbiota Portal.

The snapshots have to be updated annually, manually. There is a different workflow for the GBIF datasets and all other databases, described below. Records are pulled from specified databases by both spatial location (their coordinates fit within the bounds of the island) and verbal locality information (the island name is mentioned within the locality information). There are many other places with names such as “Santa Rosa” and “Santa Cruz”, so some erroneous records might currently be pulled in via the non-GBIF databases. I, Kylie, have tried to adjust the R Code used for the GBIF data to remove most of the erroneous records while still trying to include all the records actually found on the islands and Ben Brandt helped do that for the non-GBIF snapshots. Ben(jamin) Brandt (Owner of Green Theory Studio) has been contracted to a lot of the initial set up and work for Cal-IBIS. He helped get the bat sound data on Cal-IBIS with bat and adapted the R code to download snapshots to an online process for databases other than GBIF), along with many other things. Symbiota portal functions are constantly being updated to make it better which means that the workflow below can become outdated quickly and things have adjusted with every update.

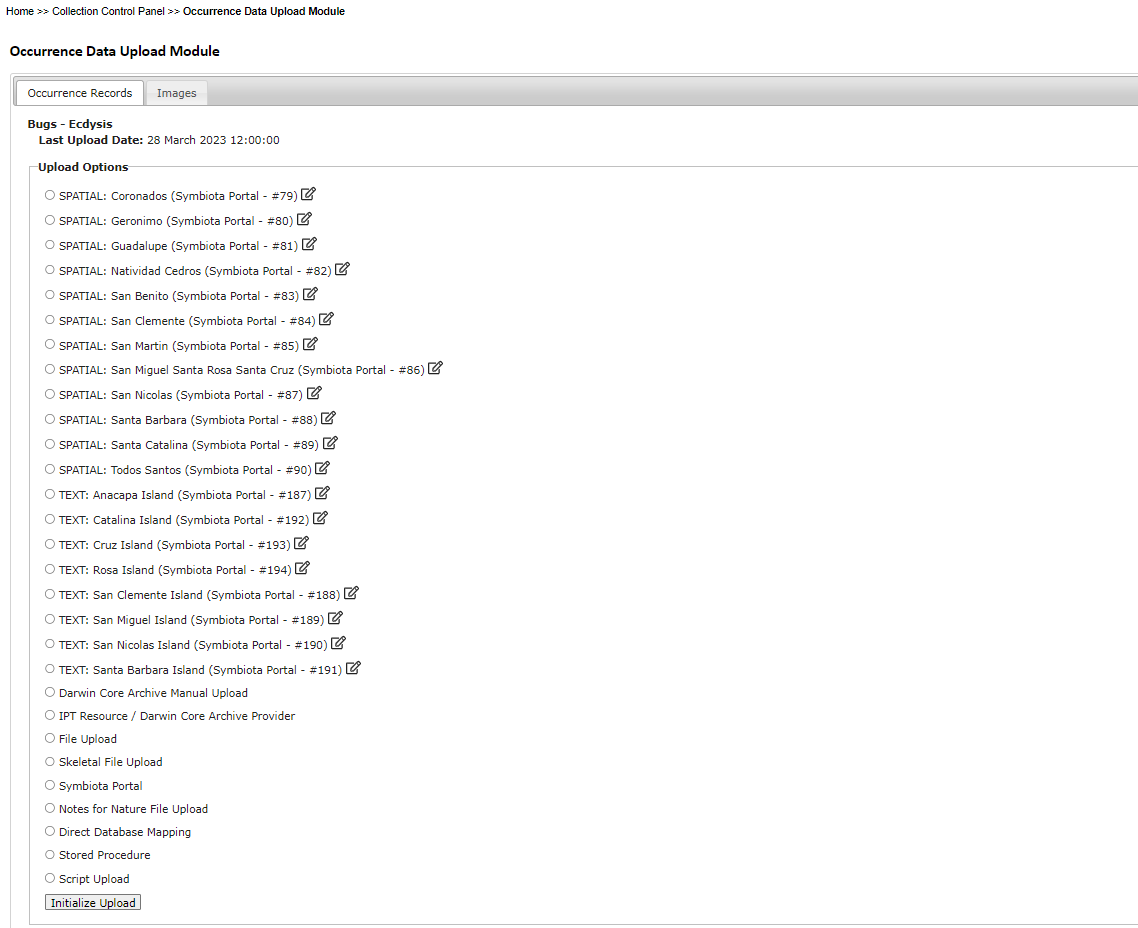
\*\*Either create a copy or print “Cal-IBIS\_Checklist.docx” to keep track of progress or else you can easily lose your place

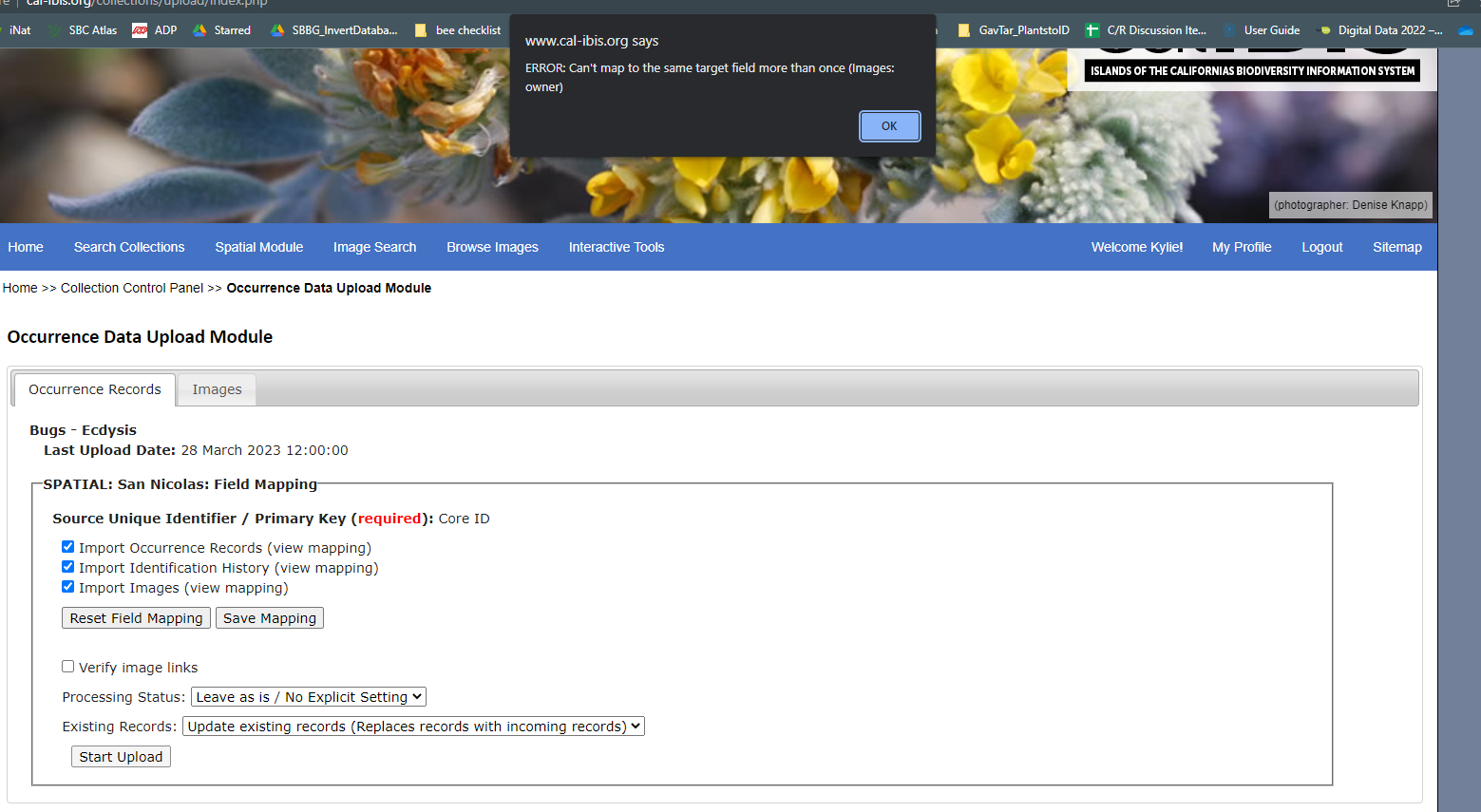
**Cal-IBIS Non-GBIF Collections Snapshot Updates:**

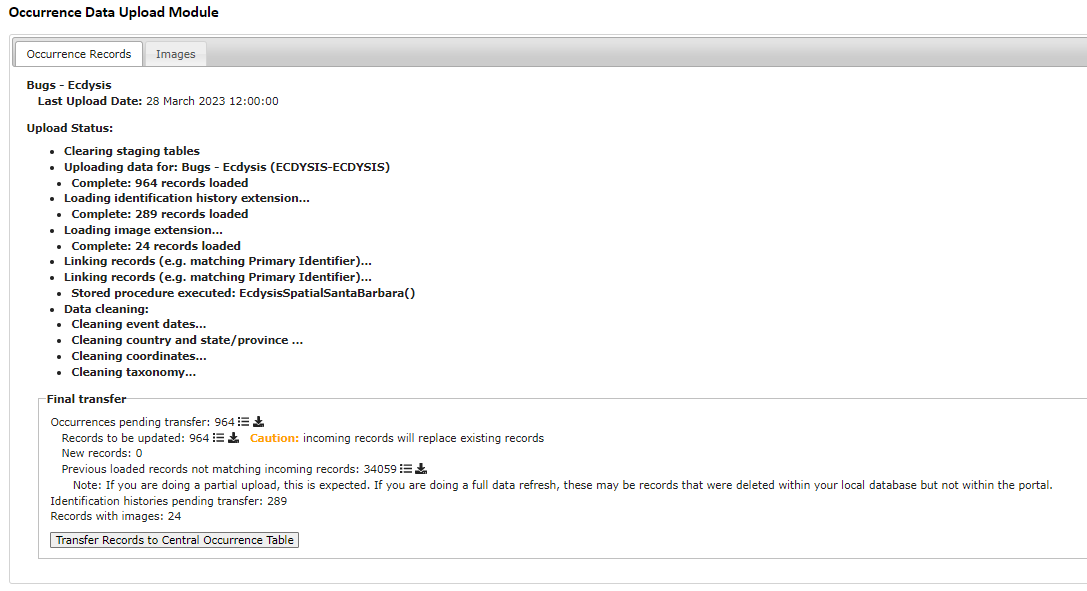
Step 1. Log into Cal-IBIS (cal-ibis.org), you will need to have permissions to manage collections. Click on “Search Collections” Here you will see a list of the collections you can manage, pictured below.

Step 2. Click on the first non-GBIF database you want to snapshot (if you are doing an annual update, it’s easiest to go in order). Once you open it you should see a drop down menu or a new page, click on “Occurrence Data Upload Module”. Repeat Step 2-3 for all non-GBIF databases. If the page does not look like below, click the gear icon in the upper right corner.

Step 3. Now you will see all the SPATIAL and TEXT upload profiles possible for this database. All of them should be updated, one by one. Click the first bubble, then click “Initialize Upload”. The next page might take a while to load. If it says “ERROR unpacking archive file: Upload file or target path does not lead to valid zip file” it means there are no occurrences that match those search criteria, so just move to the next one. If you get “ERROR There was a problem loading the resource at the path” go to Step 3.1 below.



Step 4. Click “Start Upload”. If you receive an error, go to step 4.1. The start upload may take a while (**up to 15 minutes**), if you don’t see evidence of the browser working (circle for loading) than something went wrong and you will want to refresh the page. If you continue to have issues, return to it another day. Once it loads, it will tell you how many records are being updated or are new. Hit “Transfer Records to Central Occurrence Table”, then “OK” on the Pop-up. This may take a while. Then click Collection Control Panel to take you back to the Collection Page to repeat Step 3-4 for all of the SPATIAL and TEXT profiles.

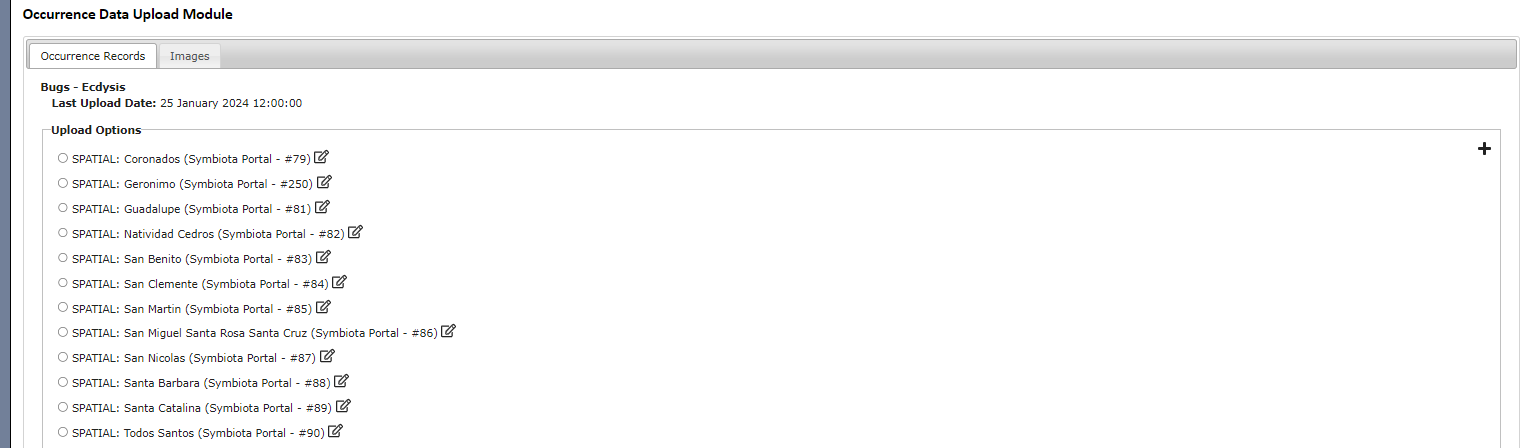


Step 3.1 If “ERROR There was a problem loading the resource at the path” pops up, that means the download url that used to work to download the data no longer works. The pathway is broken. This could be due to updates or changes in the structure of the symbiota portal we are trying to pull data from.

Go to the symbiota portal website and see if the url matches. The Mycology Portal changed from mycoportal.org to www.mycoportal.org which I only noticed by clicking on the url on the homepage because you can still get to the portal without www, so **check the url on the site itself**!

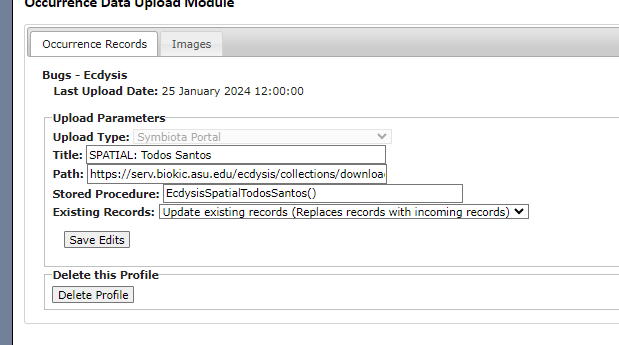
You can update the url link in the spatial or text profile (ex of profile, “SPATIAL: Coronados (Symbiota Portal - #79”) Click the edit button to the right of the profile name (Figure below, green arrow)

Example: This happened it 2024 with Ecdysis. Ecdysis changed its domain link, so I have to change the beginning of the link in “Path:” from https://serv.biokic.asu.edu/ecdysis/collections to https://ecdysis.org/collections It was finicky though, and sometimes that would not work, so I would copy over the link from another portal for the same spatial area (copy SEINet SPATIAL San Clemente link, replace the swbiodiversity.org/seinet with ecdysis.org and paste into the Ecdysis SPATIAL San Clemente ). Then save edits and try to upload in that profile again.

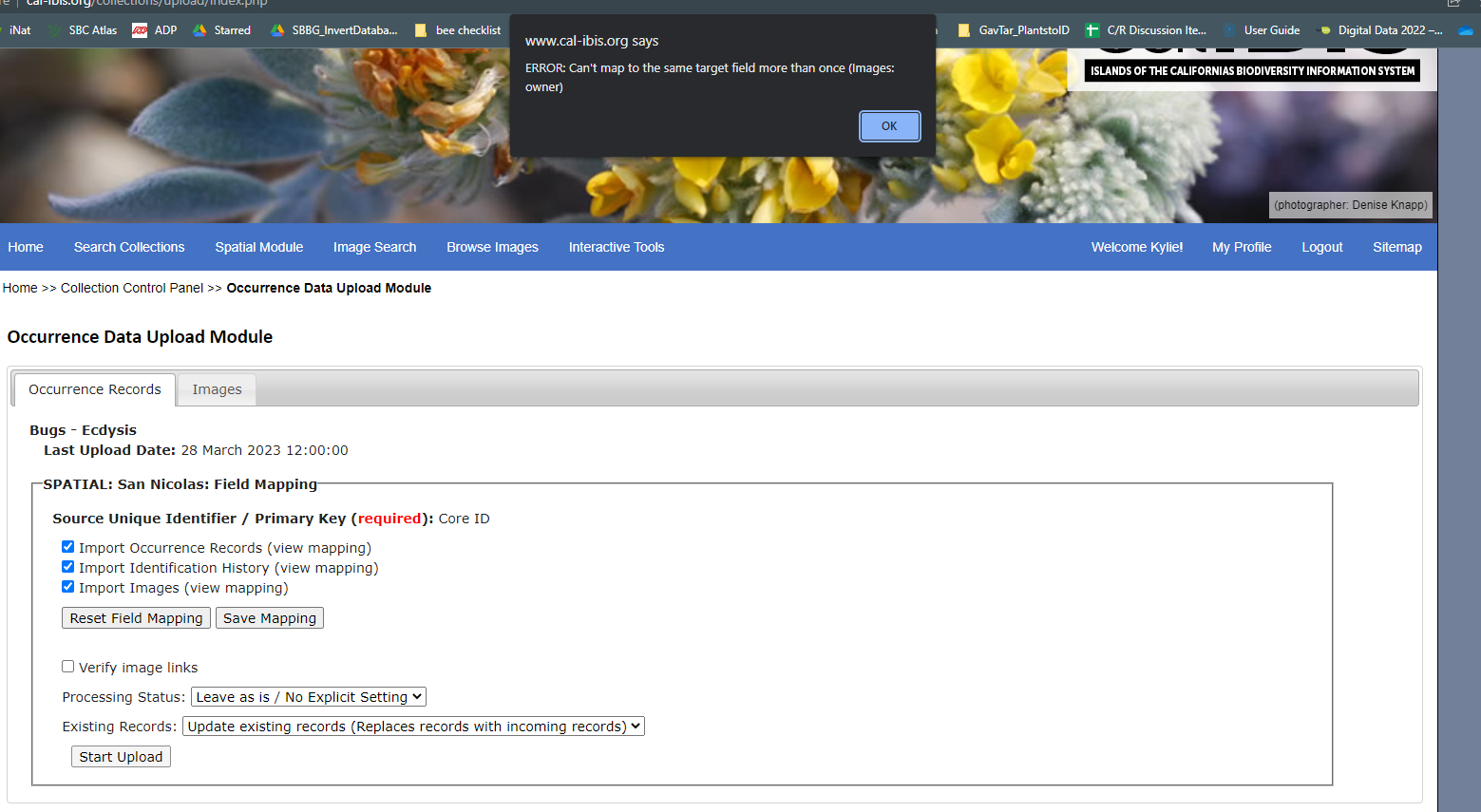
If that still does not work go to the portal you are trying to pull from (<http://ecdysis.org>) and search with a spatial bounding box around that island or text search and see if any recorded pop up, if there are records you have to make a new profile, click the plus sign in the corner (red circle). If there are no records, then the error is because there were no records to pull. Only time that happens to me was Ecydsis, SPATIAL: San Benito.

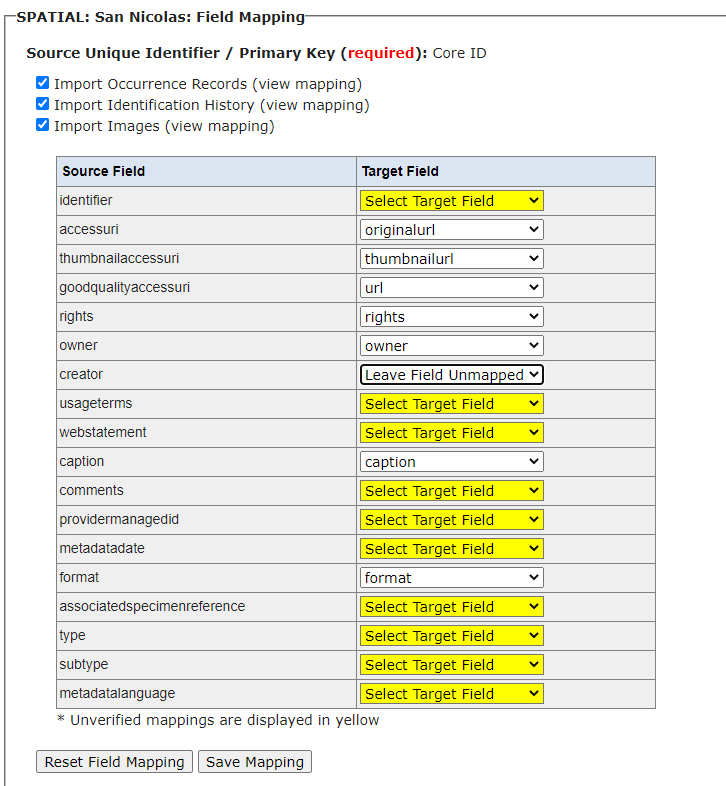
Click to create new profile

Click to edit existing profile

Making a new profile: Upload Type: Symbiota Portal, fill out Title: with the title of the profile you are trying to replace (i.e. SPATIAL: Todos Santos). Paste the URL into path (copy and paste from another portal as described above), fill out the Stored Procedure and make sure Existing Records is set to “Update existing records…”. Click create profile

Step 4.1 If “ERROR: Can’t map to the same target field more than once (Images:owner) pops up, this is how to fix it. Click on “(view mapping)” next to “Import Images”. For the source field “creator”, change the Target Field to “Leave Field Unmapped”. Then click “Save Mapping”. Go back to Step 4.



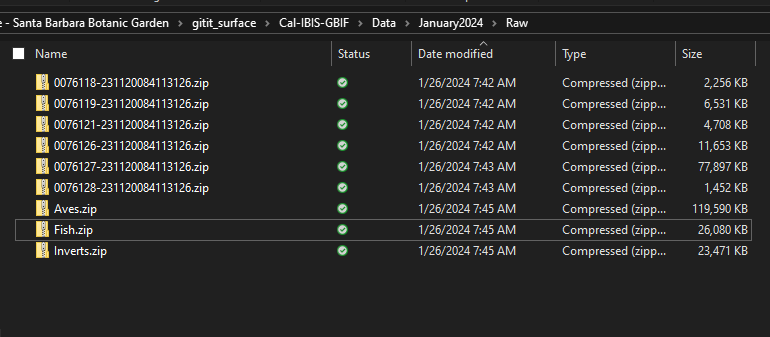


**Cal-IBIS GBIF Collections Snapshot Updates:**

Make sure you do this AFTER you update all the non-GBIF snapshots. A lot of symbiota portals put their data onto GBIF so we want to compare and remove duplicates with the R code.

These steps use 2 files with R code that batch download the data from GBIF, removes dupes found in other Cal-IBIS collections, and saves the files.

1. Run the GBIFStep1\_DownloadGBIF.Rmd R code. Follow the instructions in the code. Look at the size of your files, they should be more than 11KB, that is the size of a folder with zero occurrences!

Troubleshooting: The code pulls by taxonKeys, a number assigned to a taxonomic group (i.e. Mammals =359). However, in 2023 Reptiles and Fish didn’t work for me because some taxonKeys no longer were associated with taxonomic groups. Looking up the Reptilia taxonCode had 0 occurrences. I changed the code to include the three North American taxa in Reptilia. The same thing happened with a clade of Fish, so I had to look up the taxon Codes for many orders of that clade. If this happens to you for some group, the best thing is to look into the taxonomy of that group (Wikipedia is always handy), and then look up the taxonKeys in GBIF. If you type in the group you are looking for into occurrences, it will ask if you want to search within the taxon group (if it’s a valid GBIF taxon), if you click yes, then the taxonKey will be in the URL and you can copy it from there.

Make sure the zips are named correctly, if you reverse pasted the results$key column into the cc\_download\_get as specified in the Rmd, they should be in this order  
1. Reptiles

2. Mammals

3. Fungi

4. Arthropods

5. Plants

6. Amphibia

7. Aves

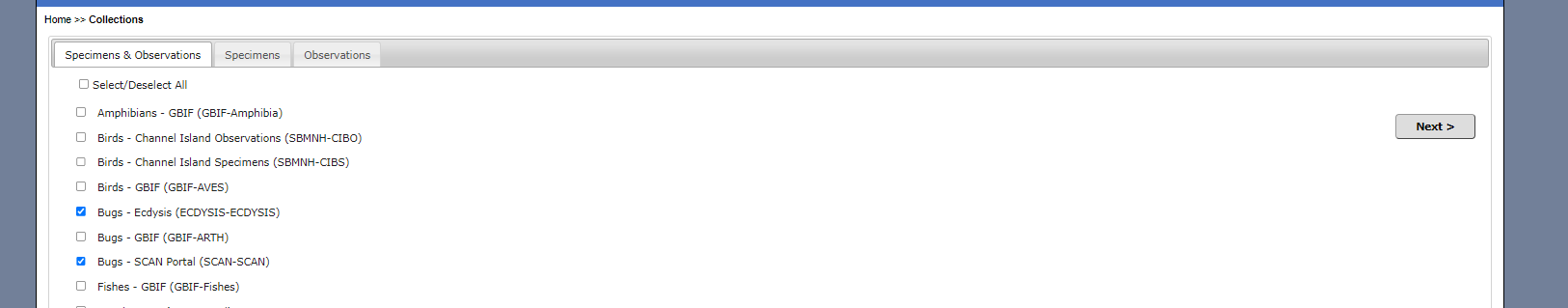
8. Fish

9. Inverts

You can check the download history by logging into GBIF with the credentials of the account used (https://www.gbif.org/user/download), if you click show in the bottom right corner of a occurrences query, the url link will have a number matching the zip file. Just check a few to make sure the zips are in the right order to name correctly.

2A. Open the GBIFStep2\_RemoveGBIFDupes.Rmd and follow the first step by downloading the non-GBIF data for taxa with GBIF and non-GBIF data: bugs, Fungi, Plants, and Invertebrates. Instructions also written below.

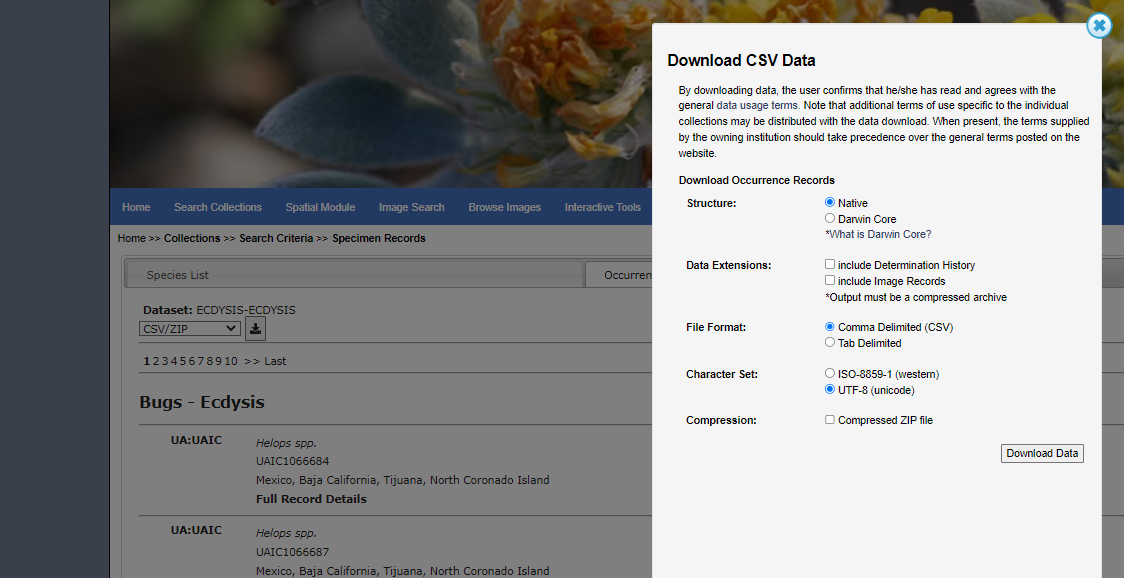
How to download data: Open Cal-IBIS, click “Search Collections”, uncheck the “Select/Deselect All” checkbox and check the collection(s) you want to download and then hit “Next”. (P.S. click the checkbox not on the words, the words will take you to a new page). The “Search Criteria” page will pop-up, click “Next”. Now you will be on a page with all the occurrences listed. Change the “Download Type” to “CSV/ZIP” and click the download button. On the pop-up you can unclick the Data Extensions and Compression, or just download as is. Once downloaded, move and rename the occurences.csv appropriately.



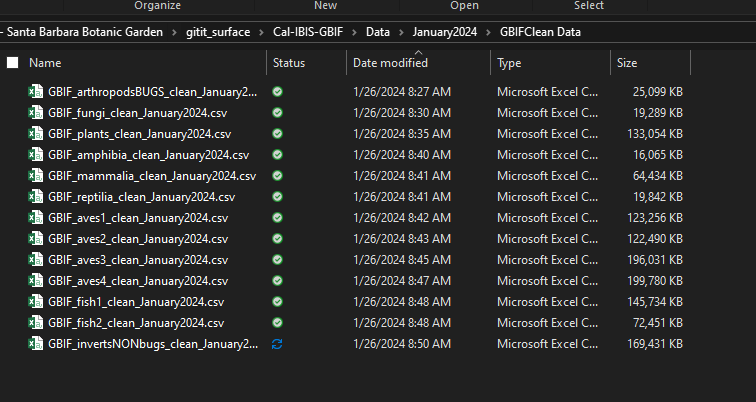
Finally, hit “Next”

Then select the collections you want

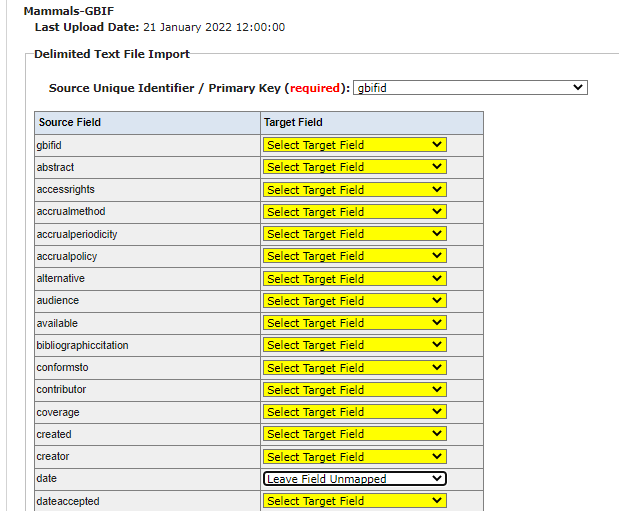
First click the Select/Deselect All checkbox



2B. Run the rest of the GBIFStep2\_RemoveGBIFDupes.Rmd code.

What you will have after you finish the Rmd code.

Now to upload the cleaned GBIF data to Cal-IBIS. Steps adapted from Josie’s instructions. For groups with multiple .CSVs (Birds and Fish) you will need to upload each one individually.

1. On Cal-IBIS, navigate to editor page (Collection Control Panel, see steps for non-GBIF collections if you forget how to get there) of the collection you want to upload (i.e. GBIF-Amphibia) and hit “Occurrence Data Upload Module” Click “File Upload” and the “Initialize Upload”.
2. Click “Choose File” and navigate to correct .csv. Click “Analyze file”. Keep automapping checked.
3. Choose “gbifid” from the drop-down menu for “Source Unique Identifier/Primary Key”. Change these fields: “scientificName” name to “sciname” and “species” to “Leave Field Unmapped”. Everything else can stay the same
4. Click “Save Mapping” and “Start Upload”. Wait for it to finish uploading in the Data Upload Module.
5. Click “Transfer Records to Central Occurrence Table”. Wait for the process to complete and then move on to the next collection to update.

Notes: Even if there are no New Records, it’s important to still update the records in case there were updates to any of the data within the record (Occurrences pending transfer). It is good to do all of one database within one sitting because you can’t differentiate when it was last updated between different SPATIAL or TEXT profiles. You will just see the last date the Database (i.e. Bugs – Ecdysis) was updated at all, even if it was just one SPATIAL or TEXT profile updated. You can easily forget which was the last SPATIAL/TEXT you did, so creating a list and checking it off is helpful, especially if you have to do other tasks in between. There is a checklist in the folder that you can print out and mark up during the upload process.