Updating Cal-IBIS

Table of Contents

[Background and Rationale 1](#_Toc62200497)

[Methods 1](#_Toc62200498)

[Step 0. Prep before updating 1](#_Toc62200499)

[Steps 1-6: Symbiota Portals (SCAN Bugs, MyCo, Lichens CNALH, SEINet, CCH2) 1](#_Toc62200500)

[STEP 1 - CCH2: 2](#_Toc62200501)

[STEP 2 - SEINet: 2](#_Toc62200502)

[STEP 3 - SCAN: 2](#_Toc62200503)

[STEP 4 – Ecdysis: 3](#_Toc62200504)

[STEP 5 – MyCo Portal: 3](#_Toc62200505)

[STEP 6 – CNALH: 4](#_Toc62200506)

[Uploading Symbiota portal data to Cal-IBIS 4](#_Toc62200507)

[Steps 7+8: GBIF 4](#_Toc62200508)

[Still to do: 5](#_Toc62200509)

# Background and Rationale

The California Islands Biodiversity Information System (Cal-IBIS) 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.

# Methods

## Step 0. Prep before updating

Before conducting updates, you should download a backup of the collection(s) you’re updating. To do this, go to a collection page, click the “edit” button in the upper right, and then select “Download Backup Data File” under “General Maintenance Tasks”. Save these somewhere until you’re sure the new data has been imported and uploaded correctly.

## Steps 1-6: Symbiota Portals (SCAN Bugs, MyCo, Lichens CNALH, SEINet, CCH2)

CCH2 and SEINet

These two portals share many of the same collections, resulting in a large amount of duplicated data. In order to handle these duplicates, we download the CCH2 data first, and then the SEINet data.

These are the searches you should run, and where to save the documents:

### STEP 1 - CCH2:

1. Create folder “SBBG-Cal-IBIS/Data/[DATE]/CCH2”. You will download all files to “SBBG-Cal-IBIS/Data/[DATE]/CCH2”
2. Copy over the “Header.csv” file from the prior update into the new folder
3. Download spatial data:
   * Using the WKT codes in the “Island\_WKTs.csv” file in the “Spatial Files” folder, you’ll run and download 12 separate searches (one for each island polygon)
   * Save the resulting files as “IslandName\_WKT.csv” within the folder specified
4. Download locality-based data
   * Run the searches in the locality field for Vascular plants based on the appropriate tab in the Portal\_SearchTerms.xlsx document.
   * Save the resulting files as “SearchTerm.csv” in the “Terms” folder
   * Do not worry about unzipping the folders (you’ll do this in the code, after).

### STEP 2 - SEINet:

1. Create folder “SBBG-Cal-IBIS/Data/[DATE]/SEINet”. You will download all files to “SBBG-Cal-IBIS/Data/[DATE]/SEINet”
2. Copy over the “Header.csv” file from the prior update into the new folder
3. Download spatial data:
   * Using the WKT codes in the “Island\_WKTs.csv” file in the “Spatial Files” folder, you’ll run and download 12 separate searches (one for each island polygon) in the WKTs folder
   * Save the resulting files as “IslandName\_WKT.csv” within the folder specified
4. Download locality-based data
   * Run the searches in the locality field for Vascular plants based on the appropriate tab in the Portal\_SearchTerms.xlsx document.
   * Save the resulting files as “SearchTerm.csv” in the “Terms” folder
   * Do not worry about unzipping the folders

Then, run the two associated R Markdown files - “Step3\_CCH2Cleanup” and “Step4\_SEINetCleanup”.

SCAN-Bugs and Ecdysis

SCAN-Bugs works differently from other Symbiota portals, but Ecdysis is set up like CCH2 and SEINet. As before, you’ll want to save the files in specific folders, with specific names.

### STEP 3 - SCAN:

1. Create folder “SBBG-Cal-IBIS/Data/[DATE]/SCAN”. Download all files to “SBBG-Cal-IBIS/Data/[DATE]/SCAN”
2. Download spatial data:
   * In the “Spatial Files” folder, you’ll find 12 separate .kml files (one for each island polygon). Using the “Spatial module” of SCAN, drag one .kml over at a time and search within the polygon. Josie’s experience is that you’ll need to drag the polygon, click it until the border is bright blue and the search area says “Search within select polygon” and then run an otherwise empty (no terms) search.
   * Select to download the results as a CSV, and then save the resulting files as “Name.csv”, according to the same rules that apply for the WKTs in the other portals.
   * To ‘clear’ the map between searches, refresh the page.
3. Download locality-based data:
   * Run the searches in the locality field for Vascular plants based on the appropriate tab in the Portal\_SearchTerms.xlsx document. Be sure to add “California” as the State for the California records, and “Mexico” as the country for the Mexican island records; the SCAN locality search function works a little differently from other portals.
   * Save the resulting files as “SearchTerm.csv” in the “Terms” folder
   * Do not worry about unzipping the folders
4. Run the “Step3\_SCANCleanup.Rmd” code.

### STEP 4 – Ecdysis:

1. Create folder “SBBG-Cal-IBIS/Data/[DATE]/Ecydsis”. Download all files to “SBBG-Cal-IBIS/Data/[DATE]/Ecdysis”
2. Download spatial/wkt data:
   * In the “Spatial Files” folder, open “Island\_WKTS”. Paste WKTs, one at a time, into the Polygon box, and search.
   * Save the resulting files as “Name.csv” within the WKTs folder, according to the same rules that apply for the WKTs in the other portals.
3. Download locality-term data:
   * Run searches for each locality term, one at a time. Include “California” or “Mexico” when appropriate. The Portal\_SearchTerms.xlsx document has a list of search terms to run.
   * Save the resulting files as “SearchTerm.csv” in the “Terms” folder
4. Run “Step4\_EcdysisCleanup.Rmd”

### STEP 5 – MyCo Portal:

1. Create folder “SBBG-Cal-IBIS/Data/[DATE]/Ecydsis”. Download all files to “SBBG-Cal-IBIS/Data/[DATE]/Ecdysis”. Create a folder for “Terms” and a folder for “WKTs”. Copy over the header.csv file from the last time you did this.
2. There’s no good way to download spatial data, so just don’t worry about it for now. Check in with Josie in 2022 about it; they might have updated their portal.
3. Download locality-term data:
   * You can search for multiple terms in one search through SCAN, but otherwise follow the same instructions as above.
   * Run the searches in the locality field for Vascular plants based on the appropriate tab in the Portal\_SearchTerms.xlsx document.
   * Save the resulting files as “Search Term.csv” in the “Terms” folder
4. Run “Step5\_MyCoPortalCleanup.Rmd”
   * This one might have some snags, for some reason the quotes and backslashes can get messed up but you just need to go into the occurrence files and clean them up.

### STEP 6 – CNALH:

1. Create folder “SBBG-Cal-IBIS/Data/[DATE]/CNALH”. Download all files to “SBBG-Cal-IBIS/Data/[DATE]/CNALH”. Create a folder for “Terms” and a folder for “WKTs”. Copy over the header.csv file from the last time you did this.
2. Download spatial/wkt data:
   * In the “Spatial Files” folder, open “Island\_WKTS”. Paste WKTs, one at a time, into the Polygon box, and search. Save the resulting files as “Name” within the WKTs folder, according to the same rules that apply for the WKTs in the other portals.
3. Download locality-term data:
   * You can search for multiple terms in one search through SCAN, but otherwise follow the same instructions as above.
   * Run the searches in the locality field for Vascular plants based on the appropriate tab in the Portal\_SearchTerms.xlsx document.
   * Save the resulting files as “Search Term.csv” in the “Terms” folder
4. Run “Step6\_CNALHCleanup.Rmd”

### Uploading Symbiota portal data to Cal-IBIS

After you have downloaded all of the data and run Steps 1-6 of the R code (each named “Step#\_PortalCleanup”). These chunks of code will generate clean data and save it as a csv in the appropriate folder (eg. “SBBG-Cal-IBIS/Data/[DATE]/Portal/Ecdysis\_Clean.csv”).

Always add data as a “Full text file”. Set the unique source ID/primary identifier == “occurrence ID”. Let the rest of the fields automap.

## Steps 7+8: GBIF

To update the GBIF data, you should use 2 files with R code, titled “Step1\_DownloadGBIF” and “Step2\_RemoveGBIFDupes”. These files explain how to batch download the data for all 9 GBIF categories, and will do all of the downloading, duplicate removing, and saving of the files for you (with a few code-external steps of renaming files) – just follow along with the notes in those files. To upload the cleaned data to Symbiota, follow the instructions in the powerpoint saved in the GBIF folder.

### Uploading GBIF data to Cal-IBIS

# Still to do:

1. Update Steps 7, 8 to reflect new way to compare GBIF mycos and arthros against MycoPortal, SCAN, SEINet, etc. based on the way developed for the plants.
2. Add screenshots of upload to Cal-IBIS process.