Updating Cal-IBIS

# 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

## 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 1-4: 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.

In CCH2, you first must download all of the datasets, name the files in a specific way (I know, ugh), and then run the R code titled “Step3\_CCH2Cleanup”. For all of these datasets, it is absolutely critical that you follow the exact same file/folder naming – capitalization and spelling really do matter here. These are the searches you should run, and the format for the file names:

STEP 1 - CCH2:

* Download all files to “SBBG-Cal-IBIS/Data/[DATE]/CCH2”
* 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
* 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:

* Download all files to “SBBG-Cal-IBIS/Data/[DATE]/SEINet”
* 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
* 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

### 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:

* Download all files to “SBBG-Cal-IBIS/Data/[DATE]/SCAN”
* 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.
* 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
* Run the “Step3\_SCANCleanup.Rmd” code.

STEP 4 - Ecdysis:

* Download all files to “SBBG-Cal-IBIS/Data/[DATE]/Ecdysis”
* 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.
  + Save the resulting files as “Name.csv” within the folder specified, according to the same rules that apply for the WKTs in the other portals.
* Download locality-based 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 “SearchTerm.csv” in the “Terms” folder
  + Do not worry about unzipping the folders

After you have downloaded all of the data, run Steps 1-4 of the R code (each named “Step#\_PortalCleanup”). These chunks of code will generate clean data and save it to the master “clean data” folder (SBBG-Cal-IBIS/Data/[DATE]/Clean).

## Steps 5+6: 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). To upload the cleaned data to Symbiota, follow the instructions in the powerpoint saved in the GBIF folder.