Connecting Students to Citizen Science and Curated Collections

Protocol: Exploring Data with BISON

Through completing your own collections, you have been able to experience first-hand what goes into collecting, identifying, and archiving biodiversity data. Now that you know where the data comes from, let's explore where the data goes. Research-grade observations on iNaturalist are shared with other data archives, such as the **Global Biodiversity Information Facility** (GBIF, https://gbif.org), a database of biological collections and occurrence data. This is important because aggregated data allows researchers, land managers, policy-makers, and others all over the world access to your observation information in the broad context of many other sources, both observational and collections-based. **Observational data** is species occurrence data, and represents where someone saw, and often imaged, a living specimen. **Collections-based data** takes this a step further by **vouchering** specimens. Vouchering ensures that specimens have been sampled and archived indefinitely. If we want to explore some aggregated observation and collection information, we can do this through BISON, or **Biodiversity Information Serving Our Nation** (https://bison.usgs.gov), a tool for visualizing GBIF data.

Below are a series of exercises to help acquaint you with the capabilities of GBIF and BISON. This protocol has two sections:

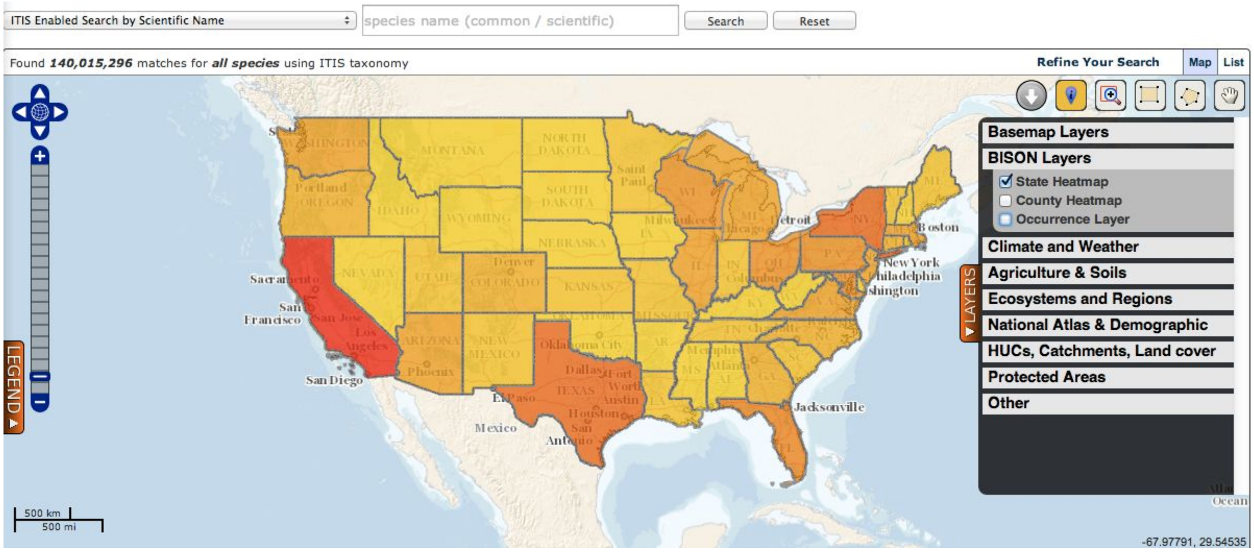
I. Browse Data
II. Query Data

Students completing this exercise will be able to do the following:

I. Recognize the importance of herbaria in plant biology research.
II. Evaluate the quality of information and data obtained from a variety of sources.
III. Discuss the value of big data for investigating large spatial- or temporal-scale phenomena.
IV. Evaluate the importance of citizen scientists to big data.

Exercises

Let’s begin by going to the BISON website: https://bison.usgs.gov. The homepage (shown below) displays the map interface we will be using.



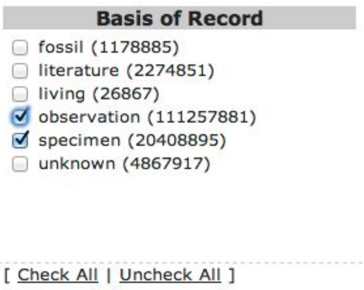
Found **140,015,296** matches for ***all species*** using ITIS taxonomy

☑ State Heatmap
☐ County Heatmap
☐ Occurrence Layer

I. Browse Data

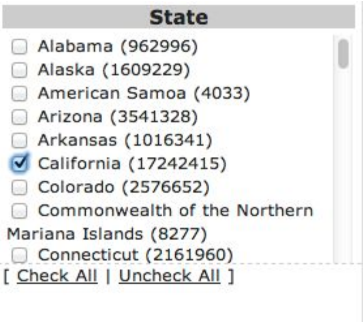
We want to explore plant data available for our local area; to do so click *“Refine Your Search”* in the upper right of the map area. This brings us to a series of checkboxes: Basis of Record, Provider, State/Province, Marine EEZ, Higher Taxa, Year Range, and Centroid. We are going to modify the following:

Basis of Record - uncheck everything except “observation,” and “specimen”
Provider - leave as is (all checked)
State/Province - select your own
Marine EEZ - uncheck everything
Higher Taxa - uncheck everything except “Plants”
Year Range - leave as is (blank)
Centroid - leave as is (“All Data” selected)



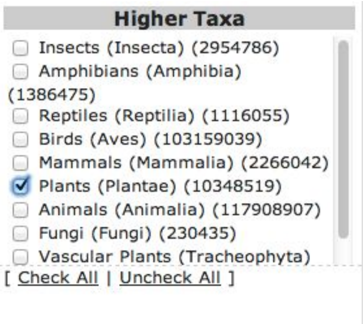
Basis of Record
☐ fossil (1178885)
☐ literature (2274851)
☐ living (26867)
☑ observation (111257881)
☑ specimen (20408895)
☐ unknown (4867917)

[Check All | Uncheck All]



State
☐ Alabama (962996)
☐ Alaska (1609229)
☐ American Samoa (4033)
☐ Arizona (3541328)
☐ Arkansas (1016341)
☑ California (17242415)
☐ Colorado (2576652)
☐ Commonwealth of the Northern Mariana Islands (8277)
☐ Connecticut (2161960)

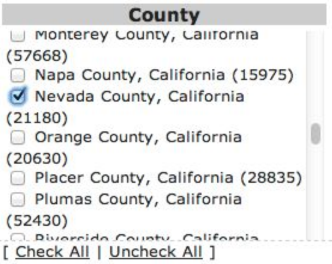
[Check All | Uncheck All]



Higher Taxa
☐ Insects (Insecta) (2954786)
☐ Amphibians (Amphibia) (1386475)
☐ Reptiles (Reptilia) (1116055)
☐ Birds (Aves) (103159039)
☐ Mammals (Mammalia) (2266042)
☑ Plants (Plantae) (10348519)
☐ Animals (Animalia) (117908907)
☐ Fungi (Fungi) (230435)
☐ Vascular Plants (Tracheophyta)

[Check All | Uncheck All]

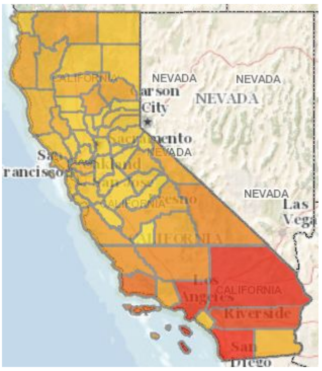
Return to the map by selecting *"Update Search Results"* in the small text at the bottom of the refine search pop-out. Now the map is centered on your state of interest–the example here is from California–and displaying a heat map of the number of records for each county. From here, you can go back to *"Refine Your Search"* and select your specific county–the example is from Nevada County.



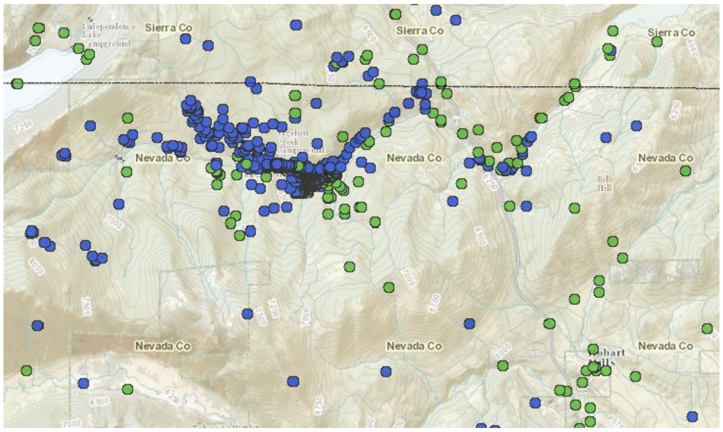
County

☐ Monterey County, California (57668)
☐ Napa County, California (15975)
☑ Nevada County, California (21180)
☐ Orange County, California (20630)
☐ Placer County, California (28835)
☐ Plumas County, California (52430)
☐ Riverside County, California

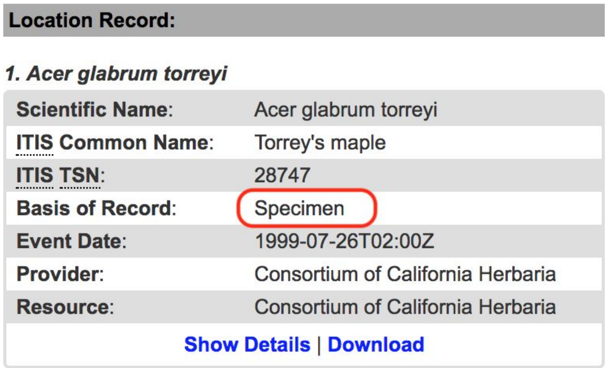
[Check All | Uncheck All]



Next, we want to be able to see individual data points, so you will need to check the box that says *"Points Layer"* under *"BISON Layers"* on the Layers tab (you may have to open this tab by clicking the orange "*Layers*" button on the middle right of the map screen). After checking this box, BISON will display blue and green dots all over the county map we already selected.



Clicking on a **blue dot will bring up any observations** for this geographical location, while clicking on a **green dot will bring up any specimen data** for this geographical location. Notice how the two types of records look different in the images below.

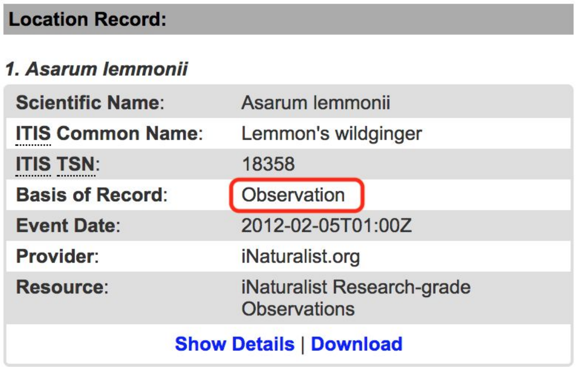


Location Record:

1. Acer glabrum torreyi

Scientific Name:	Acer glabrum torreyi
ITIS Common Name:	Torrey's maple
ITIS TSN:	28747
Basis of Record:	Specimen
Event Date:	1999-07-26T02:00Z
Provider:	Consortium of California Herbaria
Resource:	Consortium of California Herbaria

Show Details | **Download**



Location Record:

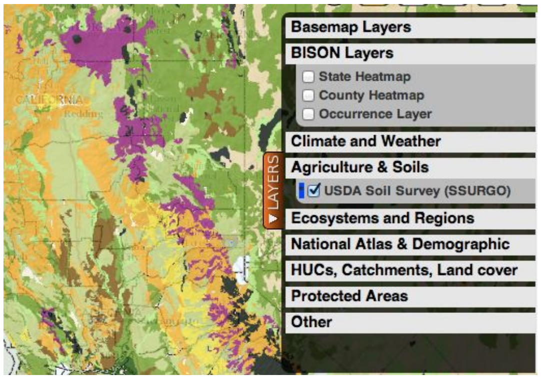
1. Asarum lemmonii

Scientific Name:	Asarum lemmonii
ITIS Common Name:	Lemmon's wildginger
ITIS TSN:	18358
Basis of Record:	Observation
Event Date:	2012-02-05T01:00Z
Provider:	iNaturalist.org
Resource:	iNaturalist Research-grade Observations

Show Details | **Download**

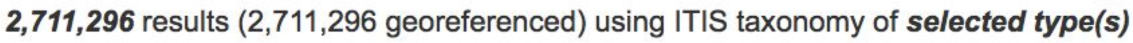
There may be multiple records for one dot, if they were collected at the same geographic location. Scroll down to view other records. At the bottom of the popup, you have the option to show details, and download data, for the records you are viewing.

You can also explore visualizations with different map layers. In the example to the right, the goal was to look at a specific geographic area in the context of soil types. To do this we clicked on the *"USDA Soil Survey"* layer from under *"Agriculture & Soils"* in the Layers tab. Explore some of the other layers you can view on your own.



II. Query Data

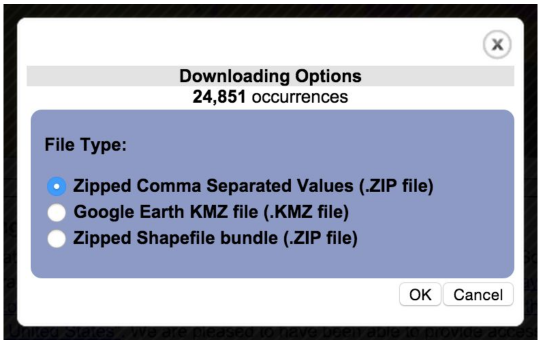
You can download data from BISON by clicking the ⬇ button at any time, but to do so, you'll want to be more specific in your search area and criteria than we were while browsing. Notice the line at the top of the BISON map that says something like this:



2,711,296 results (2,711,296 georeferenced) using ITIS taxonomy of ***selected type(s)***

If you were to click on the green download button with 2,711,296 matches showing, you would be downloading a *lot* of data. Instead, let's narrow down our search before downloading.

You have already learned how to filter your search results by basis of record, provider, state, county, higher taxa, year range, and centroid. Now, let's use the geographic area selectors to find records for an even more specific region than your county. Think of a region that you'd like to explore and get data. In this example we will look at all of the records within a watershed. Choose either the ▢ or ⬠ geographic selector tool, depending on whether the region you chose is a rectangle or needs to be drawn as a polygon. With the tool selected, draw a box or polygon around your region of interest. BISON should update the map and you will notice that a) no green or blue data points occur outside your selected region, and b) the number of matches in the top left of the map should be less than before.



Downloading Options
24,851 occurrences

File Type:

● **Zipped Comma Separated Values (.ZIP file)**
○ **Google Earth KMZ file (.KMZ file)**
○ **Zipped Shapefile bundle (.ZIP file)**

Feel free to redraw your boundaries, or modify your search terms. When the map displays the data you are interested in, click the green download button. A popup will appear describing the use terms for this data; please read this, especially the part where it briefly explains how data can be cited in a paper. If you agree to the terms you will then have the option of downloading data in one of three ways, as shown in the image to the left.

Choose "Comma Separated Values" if you want to further explore this data in a spreadsheet application such as Excel or with a programming language like R. A KMZ file can only be used in Google Earth. The last option, a zipped Shapefile bundle, is appropriate if you want to import this data into a GIS application. Save this file for use in the results and reflection.

Connecting Students to Citizen Science and Curated Collections

Results: Exploring Data with BISON

Below are a series of questions related to your experience browsing and querying BISON. You can record your answers below as you move through the assignment. Make sure you save your downloaded data from BISON!

I. Browse Data

What state and county did you select for browsing the data in BISON?

Did you find observation and occurrence data in the county you selected? If your answer is no, pick a new county and or state. Keep this up until you find a locality with observation *and* occurrence data.

Name three map layers available for you to visualize alongside the observation and occurrence data (aside from USDA Soil Survey).

II. Query Data

What region did you select within your state and county?

How many occurrences/observations were found in your selected region?

When you download the data, what information is made available? Answer this question by opening your downloaded data and looking at the fields that are included.

Connecting Students to Citizen Science and Curated Collections

Reflection: Exploring Data with BISON

Being able to visualize all this data may be obviously neat, but how is it useful to research? Answer the following questions on a separate sheet of paper.

1. Examine the data you queried; do you note anything of interest? Is there an abundance of data? For which species? Do you think this accurately represents the plant biodiversity of this region?

2. Think of a research question related to occurrence and collections-based data. What is the question and how would the data help you address this question?

3. How could you modify your search parameters in BISON to download data valuable to the question you posed above?

4. Are you confident in the quality of the data you downloaded from BISON? Does it matter whether it is occurrence data or collections-based data? Does it matter who collected the data? Why or why not?

5. If you were conducting a field based inventory of the region you selected, how could iNaturalist and/or BISON help your survey or project design?

6. Do a search on Google Scholar for research conducted using herbarium data–this could be anything from a paper describing a new species to a paper predicting range expansions due to climate change. Select one manuscript and cite it below. State the research question addressed and describe the data used (e.g. physical specimens viewed by visiting herbaria in person or via loan; digital occurrence data, such as that we just explored on BISON; fieldwork that resulted in new herbaria specimens; etc.). Finally, comment on how this study would be affected if all herbaria specimen data were available through a common database, associated with high resolution images and searchable from your computer?