



Cutting-Edge Mapping Technology at UC Berkeley

Find and Format Species Data from GBIF: How to format GBIF species data for Maxent

There are many sources of Species' data available through web services. Below is an example of how you would download and format data for a species through the GBIF portal.

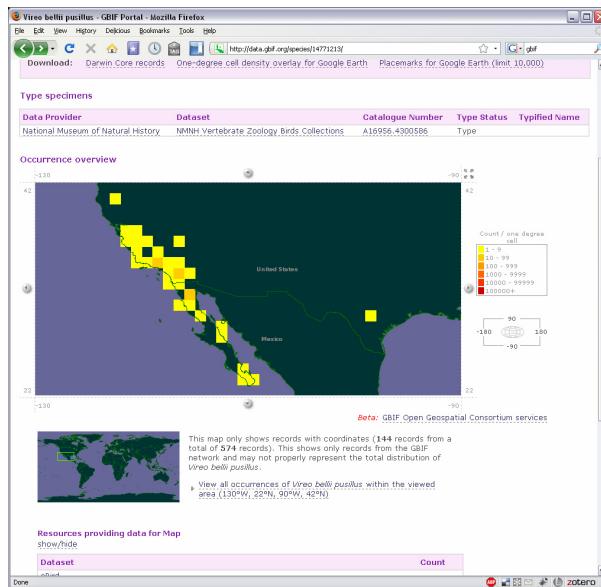
1. To access GBIF, go to <http://www.gbif.org>.

The screenshot shows the GBIF portal homepage. On the left, there's a sidebar with links for GBIF DATA (Browse, Search, How to search, Providers, Data policy), ABOUT GBIF (Press, GBIF Q&A, GBIF Data Sharing, GBIF Symposia, etc., Ebbe Nielsen Prize, GBF Posters, GBF Publications, GBF Documents, GBF Membership, GBF Nodes, GBF Directory), TOOLS AND SERVICES (Newsletters, Mailing lists, Web, LDOI registry, Standards, CIRCA, GBF tools download), and SUPPORT (Become a data provider, GB documents [login], GB15). The main content area has a "HOME" link and a "Latest news" section with several items. A "Calendar of events" for November 2008 is shown, followed by a "GBIF LDOI Registry" table with columns for registration, update information, Data Providers (266), Datasets (7621), and Searchable Records (151,222,333). A "Personalized news channels" section follows, listing GBIF CIRCA - MOST POPULAR ITEM, Helpdesk procedures, Guide to Best Practices for Georeferencing, and DLOR Provider Package for Windows - User Guide.

2. Choose **Data > Search** and for this example, we will search for: **Vireo bellii pusillus**, the Least Bell's Vireo, currently listed on California's endangered species list.

The screenshot shows the GBIF Data Portal. At the top, it says "151,222,333 occurrence records from 266 data providers". The main navigation bar includes SPECIES, COUNTRIES, DATASETS, OCCURRENCES, SETTINGS, and ABOUT. Below the navigation is a banner with the text "... free and open access to biodiversity data". The central search results page displays a search bar with "Vireo bellii pusillus" and a "Search" button. It also features three main sections: "Explore Species" (Find data for a species or other group of organisms), "Explore Countries" (Find data on the species recorded in a particular country), and "Explore Datasets" (Find data from a data provider, dataset or data network). The "Datasets" section includes a table with columns for Name, Description, and Details. A "See data for:" dropdown menu is visible at the bottom.

3. Click on the species name at the top of the screen. You will see a map that highlights the distribution of data available for the species.



4. Click “View all occurrences of Vireo bellii pusillus within the viewed area” and you will be brought to a download page. Here you can see a sample of the results listed on the page.

5. Click on the Download: **Spreadsheet of Results** link to download the data as a table.

Occurrence search - Table view - Classification includes Subspecies: Vireo bellii pusillus , Bounding box is 130.0°W, 22.0°N, 90.0°W, 42.0°N - ...																
free and open access to biodiversity data																
GLOBAL BIODIVERSITY INFORMATION FACILITY HOME SPECIES COUNTRIES DATASETS OCCURRENCES SETTINGS ABOUT																
Occurrence search - Table view									Please add filters and click Search to perform a search for occurrences records. Specify more filters to narrow your search to get more accurate results for the species, area or time period that is of interest.							
Actions			Your current search													
View:			Classification includes Subspecies: Vireo bellii pusillus													
Specify:			Bounding box is 130.0°W, 22.0°N, 90.0°W, 42.0°N													
Download:			Change your current search													
Create:			New! Niche Model													
Table of results																
Displaying 1 to 20. [First / Prev] 1 , 2 , 3 , 4 , 5 , 6 [Next]																
Scientific Name	Dataset	Institution Code	Collection Code	Catalogue Number	Basis of Record	Date	Coordinates	Country								
Vireo bellii pusillus	Terrestrial vertebrate specimens	MVZ	Bird	3601	Specimen	27/04/1906	32.64°N, 116.7799°W		View							
Vireo bellii pusillus	Terrestrial vertebrate specimens	MVZ	Bird	3602	Specimen	28/04/1906	32.64°N, 116.7799°W		View							
Vireo bellii pusillus	Terrestrial vertebrate specimens	MVZ	Bird	3603	Specimen	29/04/1906	32.64°N, 116.7799°W		View							
Vireo bellii pusillus	Terrestrial vertebrate specimens	MVZ	Bird	3604	Specimen	20/05/1906	32.64°N, 116.7799°W		View							
Vireo bellii pusillus	Terrestrial vertebrate specimens	MVZ	Bird	4249	Specimen	23/04/1862	32.7324°N, 117.1119°W		View							
Vireo bellii pusillus	Terrestrial vertebrate specimens	MVZ	Bird	8294	Specimen	24/05/1909	32.97°N, 116.36°W		View							
Vireo bellii pusillus	Terrestrial vertebrate specimens	MVZ	Bird	8295	Specimen	29/05/1909	32.97°N, 116.36°W		View							
Vireo bellii pusillus	Terrestrial vertebrate specimens	MVZ	Egg	602	Specimen	28/05/1909	32.97°N, 116.36°W		View							

6. In the following page, you can select the format that you would like to receive the data in, for maxent we will use **Comma separated values**.

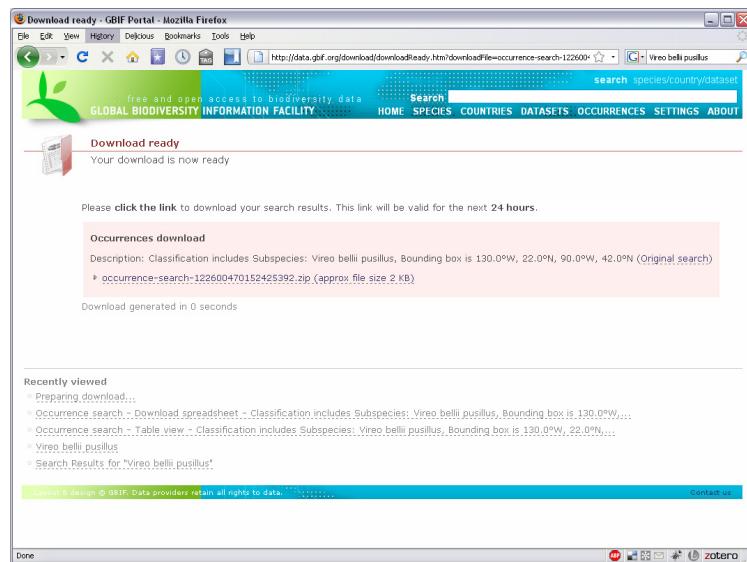
7. Next, you can choose from a variety of different fields that are available for the data. You can use all of these fields to have more information, but for this exercise, we will only choose **Scientific name**, **Latitude**, and **Longitude** as these are the three fields necessary to run maxent. We can also choose **State/Province** so that we can be sure that the data covers California.

The screenshot shows a web browser window titled "Occurrence search - Download spreadsheet - Classification includes Subspecies: Vireo bellii pusillus , Bounding box is 130.0°W, 22.0°N, 90.0°...". The URL in the address bar is [http://data.gbif.org/occurrences/downloadSpreadsheet.htm?c\[0\].s=20&c\[0\].p=0&c\[0\].t=gbif](http://data.gbif.org/occurrences/downloadSpreadsheet.htm?c[0].s=20&c[0].p=0&c[0].t=gbif). The main content area displays the "Download spreadsheet of results" page. It includes a summary of the query: "Classification includes Subspecies: Vireo bellii pusillus , Bounding box is 130.0°W, 22.0°N, 90.0°W, 42.0°N". It states that the download will return a maximum of 250,000 records and will be in a zipped format with a citation. The page has several sections for configuration:

- Format options:** Radio buttons for "Tab-delimited" (unchecked), "Comma separated values" (checked), and "Excel (maximum of 65,535 records)" (unchecked).
- Required fields:** Checkboxes for "Data provider" (checked), "Dataset" (checked), "Scientific name" (checked), and "GBIF portal url" (checked).
- Dataset:** Buttons for "Select all" and "Deselect all". Below are checkboxes for: Institution code, Collection code, Catalogue No, Basis of record, Collector name, Date collected, Collector number, Field number, GUID, Image url, Last indexed, Identifier, Identification date, and GBIF webservice url.
- Taxonomy:** Buttons for "Select all" and "Deselect all". Below are checkboxes for: Scientific name (interpreted) (checked), Author, Kingdom, Phylum, Class, Order, Family, and Genus.
- Geospatial:** Buttons for "Select all" and "Deselect all". Below are checkboxes for: Country, Country (interpreted), Locality, County, Continent or Ocean, State/Province, Region, Provider country, Latitude, Longitude, Coordinate precision, Cell id, Centi cell id, Min depth, Max depth, Max depth, Max depth, Min altitude, Max altitude, Altitude precision.

A large "Download now!" button is located at the bottom of the configuration area. At the very bottom, there is a "Recently viewed" section with links to previous searches and a toolbar with various icons.

8. Once the download is complete, you will have to unzip the file. Using the free utility WinRAR, you can right click on the zip file and choose “Extract Here” and the csv file will be added to the current folder.



9. Next, we will open the csv file in excel for some minor formatting. In excel, choose **File > Open**. Be sure to set the Files of Type to csv to read the file.

10. To prepare the data for maxent, we can first sort by state and remove the point that do not fall in California. Then delete the columns for all but **Scientific Name**, **Latitude**, and **Longitude**.

The screenshot shows a Microsoft Excel window titled "Microsoft Excel - occurrence-search-1226004701524.csv". The spreadsheet has a single sheet named "Data provider". The columns are labeled A through L. The data starts with a header row (row 1) and includes the following rows:

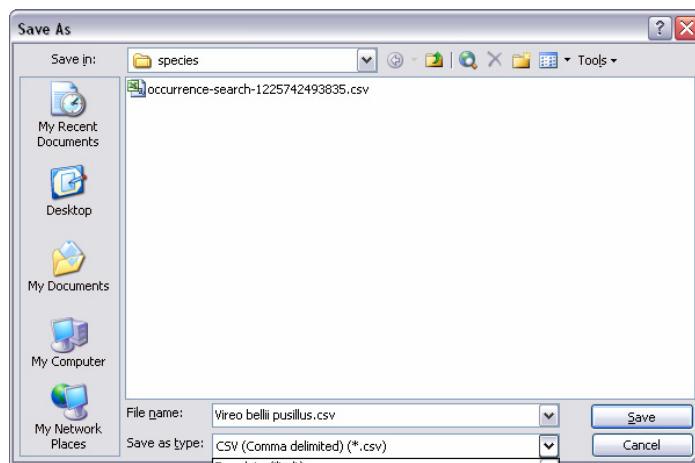
A	B	C	D	E	F	G	H	I	J	K	L
1	Data provider	Dataset	Scientific name	State/Prov	Latitude	Longitude	GBIF portal url				
2	Museum o	Terrestrial	Vireo bellii	Vireo bellii	California	32.64	-116.78	http://data.gbif.org/occurrences/38501102			
3	Museum o	Terrestrial	Vireo bellii	Vireo bellii	California	32.64	-116.78	http://data.gbif.org/occurrences/38501104			
4	Museum o	Terrestrial	Vireo bellii	Vireo bellii	California	32.64	-116.78	http://data.gbif.org/occurrences/38501108			
5	Museum o	Terrestrial	Vireo bellii	Vireo bellii	California	32.64	-116.78	http://data.gbif.org/occurrences/38501110			
6	Museum o	Terrestrial	Vireo bellii	Vireo bellii	California	32.7324	-117.112	http://data.gbif.org/occurrences/38501119			
7	Museum o	Terrestrial	Vireo bellii	Vireo bellii	California	32.97	-116.36	http://data.gbif.org/occurrences/38501121			
8	Museum o	Terrestrial	Vireo bellii	Vireo bellii	California	32.97	-116.36	http://data.gbif.org/occurrences/38501123			
9	Museum o	Terrestrial	Vireo bellii	Vireo bellii	California	32.97	-116.36	http://data.gbif.org/occurrences/38506844			
10	Museum o	Terrestrial	Vireo bellii	Vireo bellii	Baja Calif.	23.45	-110.217	http://data.gbif.org/occurrences/38501221			
11	Museum o	Terrestrial	Vireo bellii	Vireo bellii	Baja Calif.	30.7197	-116.018	http://data.gbif.org/occurrences/38501266			
12	Museum o	Terrestrial	Vireo bellii	Vireo bellii	Baja Calif.	30.7197	-116.018	http://data.gbif.org/occurrences/38501270			

11. Lastly, we must swap the Longitude and Latitude columns, as maxent requires the data format specifically ordered to species name, x (long), y (lat).

12. Save the file as a .csv, and it will be ready for use in Maxent!

The screenshot shows a Microsoft Excel window titled "Microsoft Excel - occurrence-search-1226004701524.csv". The spreadsheet contains a single sheet with data starting from row 1. Column A is labeled "Scientific name" and contains the value "Vireo bellii pusillus" repeated 21 times. Columns B and C are labeled "Longitude" and "Latitude" respectively, with values ranging from -116.78 to -116.112 for longitude and 32.64 to 32.7324 for latitude. The bottom status bar shows the path "\occurrence-search-1226004701524 /".

	A	B	C	D	E	F	G
1	Scientific name	Longitude	Latitude				
2	Vireo bellii pusillus	-116.78	32.64				
3	Vireo bellii pusillus	-116.78	32.64				
4	Vireo bellii pusillus	-116.78	32.64				
5	Vireo bellii pusillus	-116.78	32.64				
6	Vireo bellii pusillus	-117.112	32.7324				
7	Vireo bellii pusillus	-116.36	32.97				
8	Vireo bellii pusillus	-116.36	32.97				
9	Vireo bellii pusillus	-116.36	32.97				
10	Vireo bellii pusillus	-120	34.65				
11	Vireo bellii pusillus	-116.933	33.083				
12	Vireo bellii pusillus	-120	34.65				
13	Vireo bellii pusillus	-116.933	33.083				
14	Vireo bellii pusillus	-118.25	34.05				
15	Vireo bellii pusillus	-118.469	34.3231				
16	Vireo bellii pusillus	-116.687	33.8795				
17	Vireo bellii pusillus	-116.687	33.8795				
18	Vireo bellii pusillus	-116.687	33.8795				
19	Vireo bellii pusillus	-116.519	33.7882				
20	Vireo bellii pusillus	-116.519	33.7882				
21	Vireo bellii pusillus	-116.519	33.7882				



RS/GIS Quick Start Guides

These materials are freely available for use and modification through a Creative Commons Attribution-Share Alike 3.0 License.

License details: <http://creativecommons.org/licenses/by-sa/3.0/us/>
Version: Original document was created by GfG (2008)

