USE CASES: DATA ACCESS AND USE WORKSHOP

DAY 1: Data access through the GBIF Data Portal

Lantana camara L. Foto via The New York Botanical Garden (CC BY 4.0)

05-06 March 2020 at FSE Complex Board Room; Bindura University of Science Education

As part of the GBIF's CESP Project: **Strengthening Zimbabwe's GBIF node through collaboration with GBIF Spain (CESP2019-001)**

INTRODUCTION

The purpose of the following exercises is to get familiar with the process of searching and downloading data and other information from the GBIF Data Portal (https://www.gbif.org/). In each section you will find some questions that you will be able to answer following the instructions explained below.

CONTENTS

USE CASE

Mudiwa Onai is at the beginning of his Ph.D. at the Bindura University of Science Education and is conducting a study about the species "lantana" (*Lantana camara*), a small perennial shrub from tropical America commonly introduced to other countries around the world as an ornamental plant. It is considered a major invasive shrub globally, impacting biodiversity, economies, ecosystem services, and driving socio-ecological changes. One of his goals is to run ecological niche models for the species to know its potential spatial distribution in southern Africa. His

supervisor adviseshim to visit consult www.gbif.org where he can find open species- occurrence data for his models, freely available for download.

1. GBIF.org homepage

When Mudiwa visits the GBIF.org website he finds a homepage with a lot of information to be explored. Before starting with his research he wants to explore this homepage to have a better idea on what information he is going to find in this biodiversity data portal and how that information is structured. In his exploration of the homepage, he would like to answer the following questions:

1. What are the available languages for this website?
2. What are the dates for the 3 next events promoted by GBIF? Are there any of them hosted in Africa?
3. What are the topics of the last peer-reviewed paper using GBIF mediated-data shared in the GBIF website?
4. In what island are the majority of data collected between 1600 and 1700 and published through GBIF?
5. What are the social media channels used by GBIF?

Follow the instructions below to answer the questions.

Instructions

1. Open the GBIF Data Portal https://www.gbif.org/
2. Click on the button of the language selector and see how many languages there are available.

3. Click on the "About" tab and go to "Events". See what are the 3 next events promoted by GBIF and take notes on their dates. If you see in detail the provided information, you will see which event is going to be hosted in Africa.
4. Come back to the homepage by clicking on the GBIF logo in the top left side of the website.
5. Click on the number of peer-reviewed papers using data and check what are the topics of the last paper shared by GBIF.
6. Come back to the homepage by clicking on the GBIF logo in the top left side of the website.

7. Scroll down on the homepage to see the map. Go to the time selector and move the right control to set approximately the year 1700. Zoom in the map to see what island has more occurrences records for that time period.

8. Now, go to the footer of the website and have a look at the logos of social media channels. If you do not know any of them you can click on them and see what each social media is.

2. Account creation on GBIF.org

If you are not already logged into GBIF.org (or have never created an account), you are prompted to do so. Having an account in the system is free and allows you to download data, keep track of all your downloads, which will make it easier to repeat the same queries later as more records become available. To create a new account follow these steps:

1. Open the GBIF Data Portal https://www.gbif.org, and click on **Login -> Register** in the upper right side of the page**.** Create the new account by using your email and choosing a username, or by using your Google, Facebook or GitHub account (recommended) .
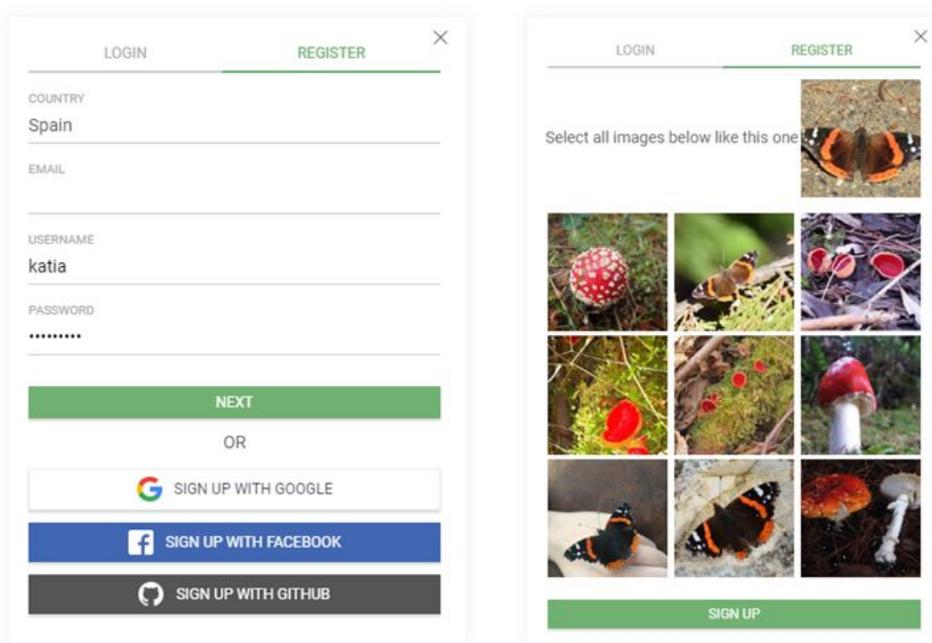
2. A small test will come out to prove that you are not a robot. Follow the instructions and click on **sign up.**
3. You will receive an email in your inbox to confirm your account, please check your spam folder if you do not find it! Click on the provided link to activate it.

Account creation on GBIF.org

ⓘ Este mensaje ha sido identificado como un correo no deseado. Se eliminará después de 10 días. No es un correo no deseado

H helpdesk@gbif.org
Mié 26/02/2020 10:50
Usted ⌄

Hello Katia,

Thanks for registering at GBIF.org. Please confirm your GBIF account by clicking the following link:
https://www.gbif.org/user/confirm?username=akiesta&code=18a6bb64-53e4-475e-a29e-01486f1a2a4c

The GBIF Secretariat

3. Taxonomic search

Before downloading any records, Mudiwa would like to look at the data available for *Lantana camara* L. on the GBIF.org portal to answer the following questions:

2. Is *Lantana camara* L. an accepted name? How many synonyms does it have?
3. What family does this species belong to?
4. How many infraspecies does it have?
5. How many records are available for this species?
6. Which is the country that has registered more records of this species?

Follow the instructions below to answer the questions.

Instructions

1. Open the Biodiversity Data Portal https://www.gbif.org/
2. Type the species **Lantana camara** in the search box and then click on the search button
   🔍 .
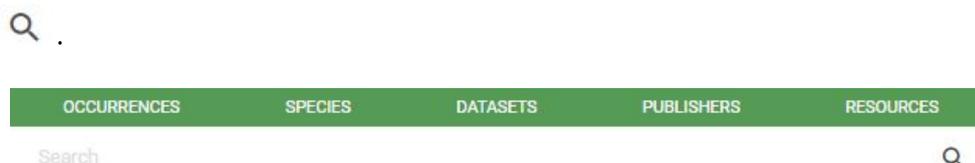
3. A new page will appear with the result of your search. This result includes not only the *Lantana camara* species but also other taxa, resources, and datasets that contain the words that you typed.

*Note: February 2020, keep in mind that if you do the search in the future the result may change due to updates in the data portal .

4. Click on the **SPECIES** tab and select the entry of *Lantana camara* L. which is the accepted name to see the information about this species.

5. Explore the sections **CLASSIFICATION** on the left side menu , **OVERVIEW**, and **METRICS** to answer the exercise's questions.

* Note: the results that appear in the attached images do not have to correspond to the current result.

4. Occurrence search, visualization, and data download

After exploring taxonomic information of *Lantana camara* L., Mudiwa would like to know the current distribution and records available of this species in the study area (**Zambia**, **Zimbabwe** and **Mozambique**). He only wants to use the **georeferenced records** (records that have coordinates), so he decides to use the Global Biodiversity Portal to perform a search with filters and download data.

1. How many records correspond to Mudiwa's criteria after applying filters?
2. How many of them have images?
3. Explore and download the records that correspond to Mudiwa's criteria using the following guidelines.
4. What is the DOI (Digital Object Identifier) for this search? Cite the data using the DOI.

Instructions

1. Open the Biodiversity Data Portal https://www.gbif.org/.
2. Click on OCCURRENCES using any of the accesses indicated in the following image.
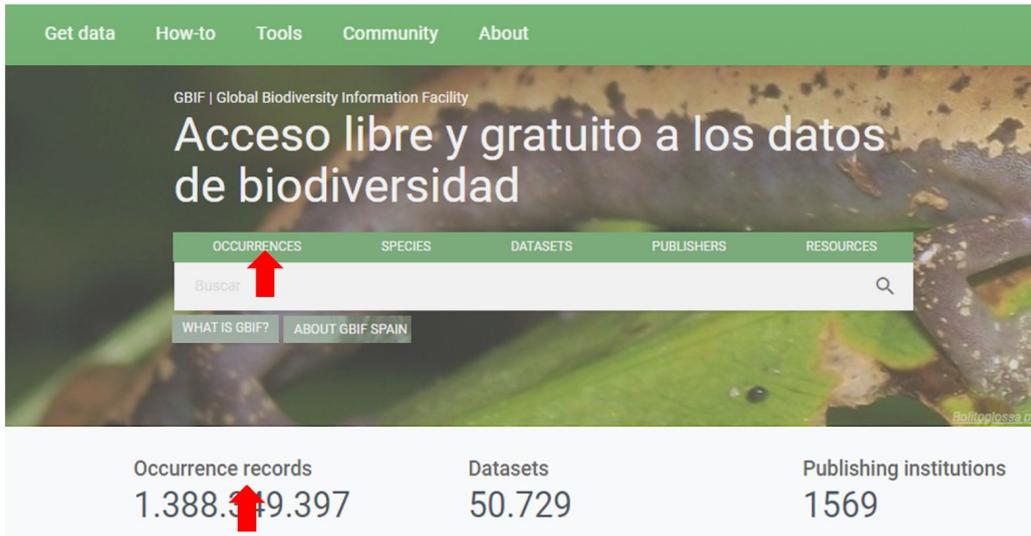
2. A new page will come up with all occurrences published in gbif.org.

   a) All records can be visualized in 5 ways:
      - List of records with a summary of information
      - Gallery with the imagen linked to the records selected
      - Map with the georeferenced records
   - Taxonomy
      - Metrics
   b) Download button
   c) Bar with filters
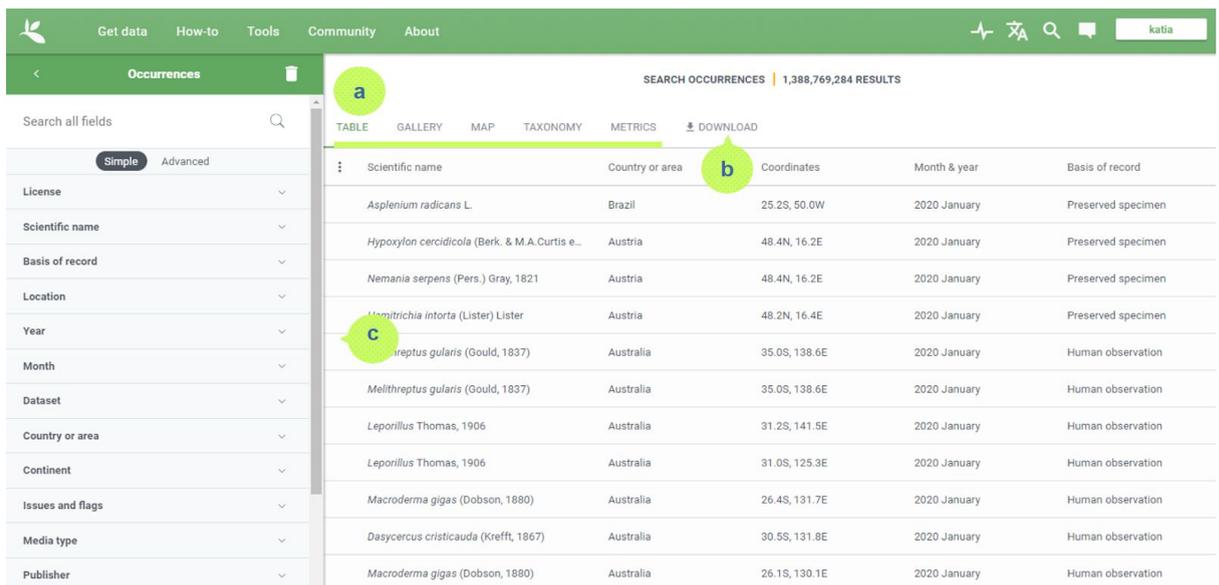
3. Refine your search through filters:

3.1. In the filtering area, unfold the **Scientific Name** filter and type the species *Lantana camara* **L.** in the search box. Click enter to apply the filter so you will have ONLY occurrences corresponding to species Lantana camara L.

3.2. Unfold the **Location** filter and select the option **Including coordinates.**

3.3. From the **Country or area** field, type your study countries (**Zambia, Mozambique** and **Zimbabwe**) to obtain only the records from these areas. After applying these filters you can answer the first question.

4. Explore the **gallery** to answer the second question.

5. Click on the **Download** button to download the results (remember that you have to be logged in to the platform). A new dialogue window will open to select the download option (explained below). Click on **SIMPLE** option.

- **Simple:** a simple, tab-delimited format which includes only the GBIF-interpreted version of the data, as a result of the indexing process. Good for making quick tests and importing directly to spreadsheets.
- **Darwin Core Archive:** richer format that includes all data - interpreted as well as the original version provided by the publisher (prior to indexing and interpretation by GBIF).

It includes all the metadata and issue flags provides a richer view of the downloaded dataset.

- **Species List**: a simple table format that includes an interpreted list of species contained within the dataset.

6. After a few minutes you will receive an email once your download is ready providing a link for the download. Clicking in that URL you will be able to download that dataset any time you want. Past downloads are stored and can be accessed anytime via your downloads page.

7. GBIF.org assigns a DOI (Digital Object Identifier) to all downloads.

DOIs are unique codes that can be resolved using standard mechanisms. They allow other people to access exactly the same dataset that you downloaded. This has proven to be a very useful element to use in citations, and to allow others to verify and reuse your source data.

Citing data downloads from GBIF.org is easy. This citation appears again **in the confirmation email sent to the registered user** and in the download/DOI page . Keep this reference close so you can cite it.

Visit "Citing GBIF.org" for more information on correct citation of the use of GBIF data.

5. Import downloaded data into an Excel file

In this exercise we will explain how to import the CSV results file into an Excel sheet so they are more easily readable.

We will move from this situation:

To this one:

	C	D	
1	Match Taxon Concept GUID	Scientific Name	Ver
2	urn:lsid:catalogueoflife.org:taxon:2dd15317-4661-11e1-9b0d-e752e483e0da:col20120124	Quercus suber L.	
3	urn:lsid:catalogueoflife.org:taxon:2dd15317-4661-11e1-9b0d-e752e483e0da:col20120124	Quercus suber L.	
4	urn:lsid:catalogueoflife.org:taxon:2dd15317-4661-11e1-9b0d-e752e483e0da:col20120124	Quercus suber L.	
5	urn:lsid:catalogueoflife.org:taxon:2dd15317-4661-11e1-9b0d-e752e483e0da:col20120124	Quercus suber L.	
6	urn:lsid:catalogueoflife.org:taxon:2dd15317-4661-11e1-9b0d-e752e483e0da:col20120124	Quercus suber L.	
7	urn:lsid:catalogueoflife.org:taxon:2dd15317-4661-11e1-9b0d-e752e483e0da:col20120124	Quercus suber L.	
8	urn:lsid:catalogueoflife.org:taxon:2dd15317-4661-11e1-9b0d-e752e483e0da:col20120124	Quercus suber L.	
9	urn:lsid:catalogueoflife.org:taxon:2dd15317-4661-11e1-9b0d-e752e483e0da:col20120124	Quercus suber L.	
10	urn:lsid:catalogueoflife.org:taxon:2dd15317-4661-11e1-9b0d-e752e483e0da:col20120124	Quercus suber L.	
11	urn:lsid:catalogueoflife.org:taxon:2dd15317-4661-11e1-9b0d-e752e483e0da:col20120124	Quercus suber L.	
12	urn:lsid:catalogueoflife.org:taxon:2dd15317-4661-11e1-9b0d-e752e483e0da:col20120124	Quercus suber L.	
13	urn:lsid:catalogueoflife.org:taxon:2dd15317-4661-11e1-9b0d-e752e483e0da:col20120124	Quercus suber L.	
14	urn:lsid:catalogueoflife.org:taxon:2dd15317-4661-11e1-9b0d-e752e483e0da:col20120124	Quercus suber L.	
15	urn:lsid:catalogueoflife.org:taxon:2dd15317-4661-11e1-9b0d-e752e483e0da:col20120124	Quercus suber L.	
16	urn:lsid:catalogueoflife.org:taxon:2dd15317-4661-11e1-9b0d-e752e483e0da:col20120124	Quercus suber L.	
17	urn:lsid:catalogueoflife.org:taxon:2dd15317-4661-11e1-9b0d-e752e483e0da:col20120124	Quercus suber L.	
18	urn:lsid:catalogueoflife.org:taxon:2dd15317-4661-11e1-9b0d-e752e483e0da:col20120124	Quercus suber L.	
19	urn:lsid:catalogueoflife.org:taxon:2dd15317-4661-11e1-9b0d-e752e483e0da:col20120124	Quercus suber L.	
20	urn:lsid:catalogueoflife.org:taxon:2dd15317-4661-11e1-9b0d-e752e483e0da:col20120124	Quercus suber L.	

Instructions

1. Open a blank Excel file and name it "Lantana".
2. From the top menu, click on **Data** -> **Get external data** and choose the type of the source file TEXT (csv).

3. A new *Open dialog box* will appear, select **Text Files (*.prn,*.txt,*.csv)** from the drop-down list in the lower right-hand corner. Browse for the CSV file and open it as usual by double-clicking.
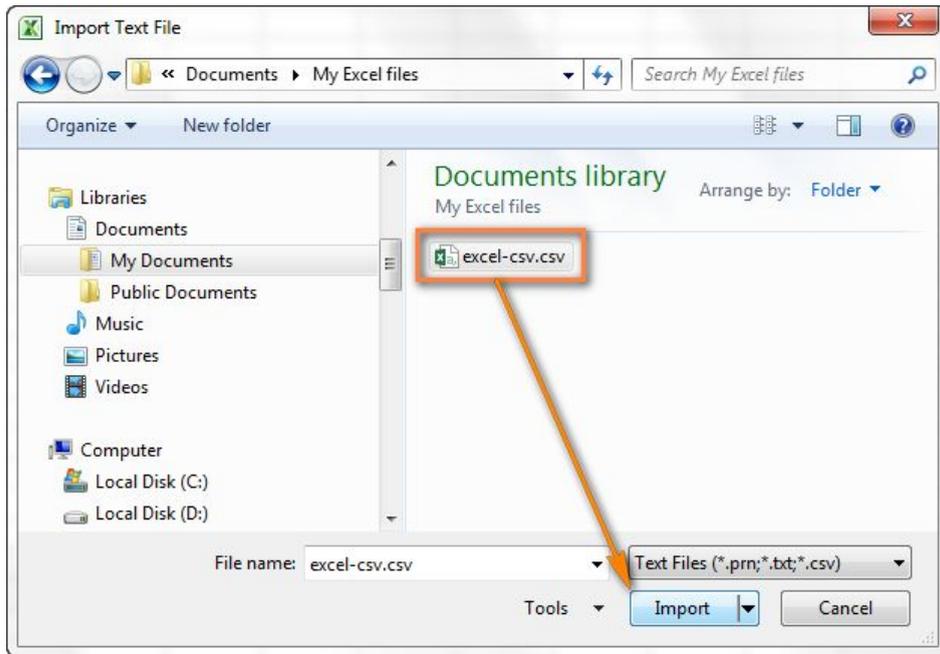
4. Then the Excel wizard will be opened to import text, which will have to be configured so that the adaptation of our file is correct. Be sure to follow these steps and leave the options as shown in the picture:

Step 1. Choose the **file type** and the **row number** to start importing the data. Typically, you choose *Delimited* and *row 1*, respectively. The *file origin* must be *Unicode UTF 8.* The preview window in the lower part of the wizard shows the first few entries of your CSV file. Click on **Next**.
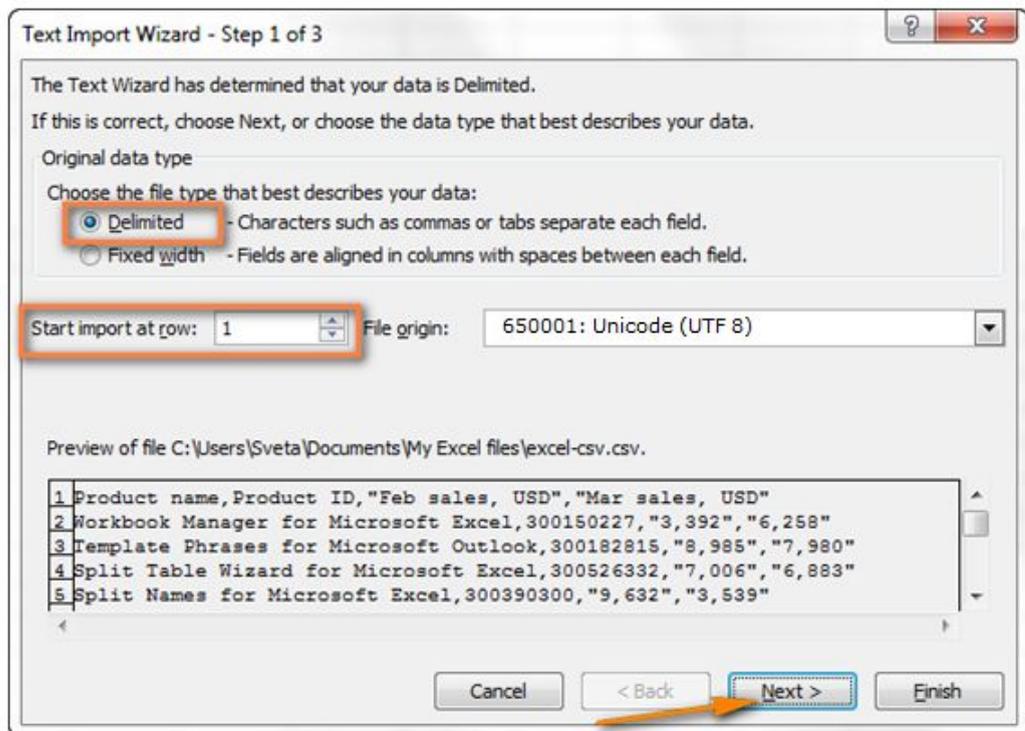
Step 2. In this step, please select a delimiter(s) and text qualifier.**Delimiter** is the character that separates values in your .csv file. Specify *Tab* and *Comma* since we want to have each product (separated by tabs) to start with a new row and the product details such as IDs and sales numbers (separated by commas) to be imported in separate cells. Click on **Next**.
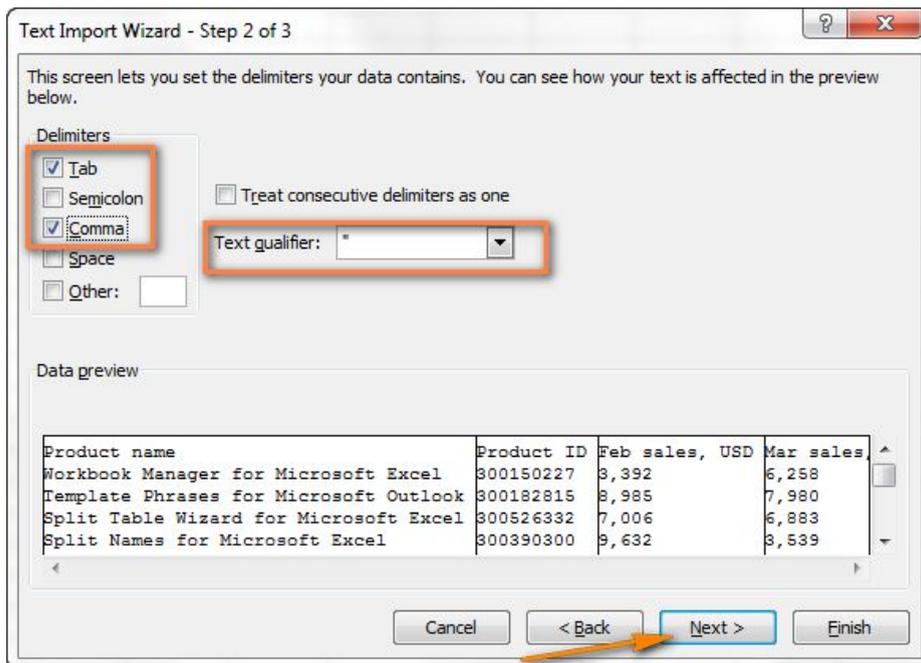
Step 3. If you are happy with the Data preview, then click the **Finish** button.
Step 4. Choose the destination for the imported data, either an existing or a new worksheet, and click *OK* to finish importing your CSV file to Excel.
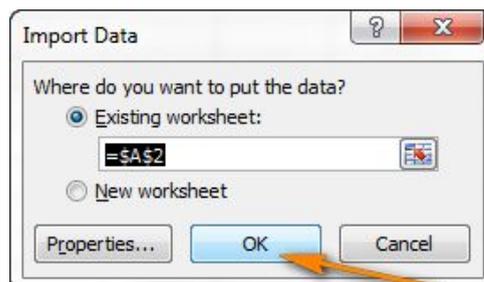
5. Finally you will have your results in a readable way, with each field easily identifiable.

|	C	D	
1	Match Taxon Concept GUID	Scientific Name	Ver
2	urn:lsid:catalogueoflife.org:taxon:2dd15317-4661-11e1-9b0d-e752e483e0da:col20120124	Quercus suber L.	
3	urn:lsid:catalogueoflife.org:taxon:2dd15317-4661-11e1-9b0d-e752e483e0da:col20120124	Quercus suber L.	
4	urn:lsid:catalogueoflife.org:taxon:2dd15317-4661-11e1-9b0d-e752e483e0da:col20120124	Quercus suber L.	
5	urn:lsid:catalogueoflife.org:taxon:2dd15317-4661-11e1-9b0d-e752e483e0da:col20120124	Quercus suber L.	
6	urn:lsid:catalogueoflife.org:taxon:2dd15317-4661-11e1-9b0d-e752e483e0da:col20120124	Quercus suber L.	
7	urn:lsid:catalogueoflife.org:taxon:2dd15317-4661-11e1-9b0d-e752e483e0da:col20120124	Quercus suber L.	
8	urn:lsid:catalogueoflife.org:taxon:2dd15317-4661-11e1-9b0d-e752e483e0da:col20120124	Quercus suber L.	
9	urn:lsid:catalogueoflife.org:taxon:2dd15317-4661-11e1-9b0d-e752e483e0da:col20120124	Quercus suber L.	
10	urn:lsid:catalogueoflife.org:taxon:2dd15317-4661-11e1-9b0d-e752e483e0da:col20120124	Quercus suber L.	
11	urn:lsid:catalogueoflife.org:taxon:2dd15317-4661-11e1-9b0d-e752e483e0da:col20120124	Quercus suber L.	
12	urn:lsid:catalogueoflife.org:taxon:2dd15317-4661-11e1-9b0d-e752e483e0da:col20120124	Quercus suber L.	
13	urn:lsid:catalogueoflife.org:taxon:2dd15317-4661-11e1-9b0d-e752e483e0da:col20120124	Quercus suber L.	
14	urn:lsid:catalogueoflife.org:taxon:2dd15317-4661-11e1-9b0d-e752e483e0da:col20120124	Quercus suber L.	
15	urn:lsid:catalogueoflife.org:taxon:2dd15317-4661-11e1-9b0d-e752e483e0da:col20120124	Quercus suber L.	
16	urn:lsid:catalogueoflife.org:taxon:2dd15317-4661-11e1-9b0d-e752e483e0da:col20120124	Quercus suber L.	
17	urn:lsid:catalogueoflife.org:taxon:2dd15317-4661-11e1-9b0d-e752e483e0da:col20120124	Quercus suber L.	
18	urn:lsid:catalogueoflife.org:taxon:2dd15317-4661-11e1-9b0d-e752e483e0da:col20120124	Quercus suber L.	
19	urn:lsid:catalogueoflife.org:taxon:2dd15317-4661-11e1-9b0d-e752e483e0da:col20120124	Quercus suber L.	
20	urn:lsid:catalogueoflife.org:taxon:2dd15317-4661-11e1-9b0d-e752e483e0da:col20120124	Quercus suber L.	

Related tutorial

https://www.youtube.com/watch?v=xyWI4ER2cB4&feature=player_embedded

6. Spatial search

Mudiwa's work has not gone unnoticed in his laboratory as his work has been published in a high impact journal. His lab is currently working on updating the species **plant list of the Hwange National Park** and they would be interested to know how GBIF data could be used to help in the process. They would like to use the data published in GBIF as an starting point to update that list. With Mudiwa's help they manage to get the following:

1. A polygon to delimit the study area
2. A query that includes all the occurrences
3. A list with the taxa included in the Hwange National Park.
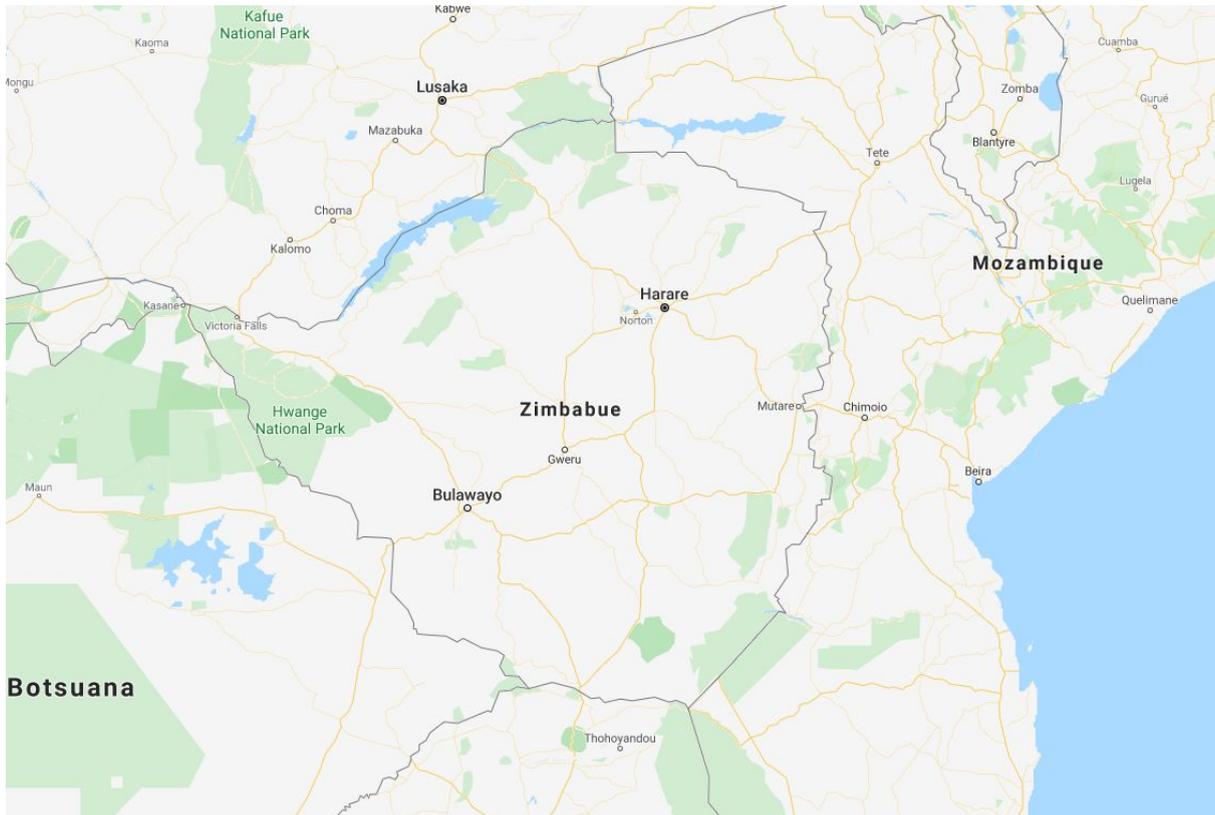
Instructions

1. Open the GBIF Data Portal https://www.gbif.org/
2. Click on OCCURRENCES using any of the accesses indicated in the following image.
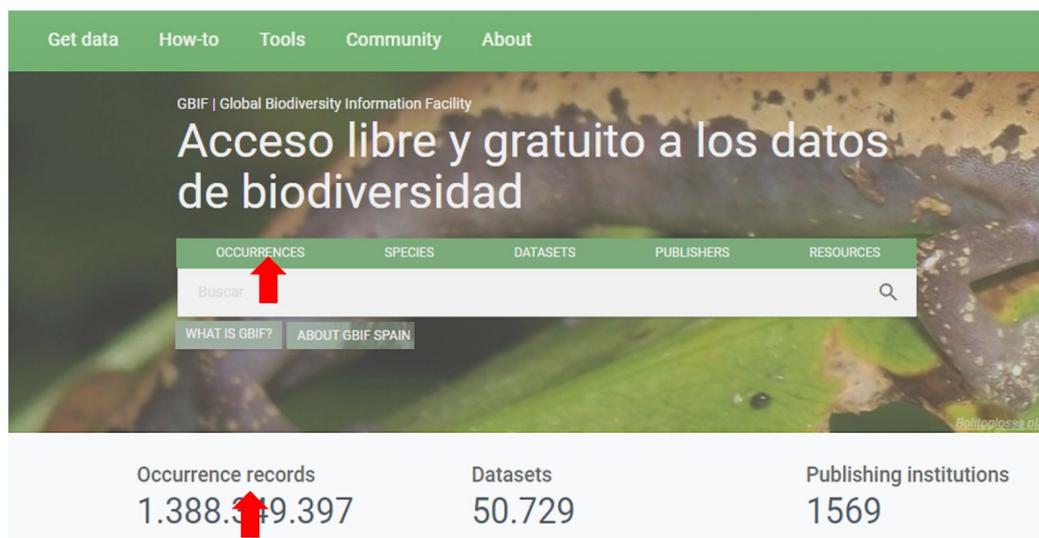
3. Go to the **Map** tab and zoom **+** over the National Park area.

4. Draw a **polygon** ⬠ to delimit the National Park (don't need to be precise :)).. Click the first point of the polygon to close the shape. The portal automatically will re-calculate the occurrences count.
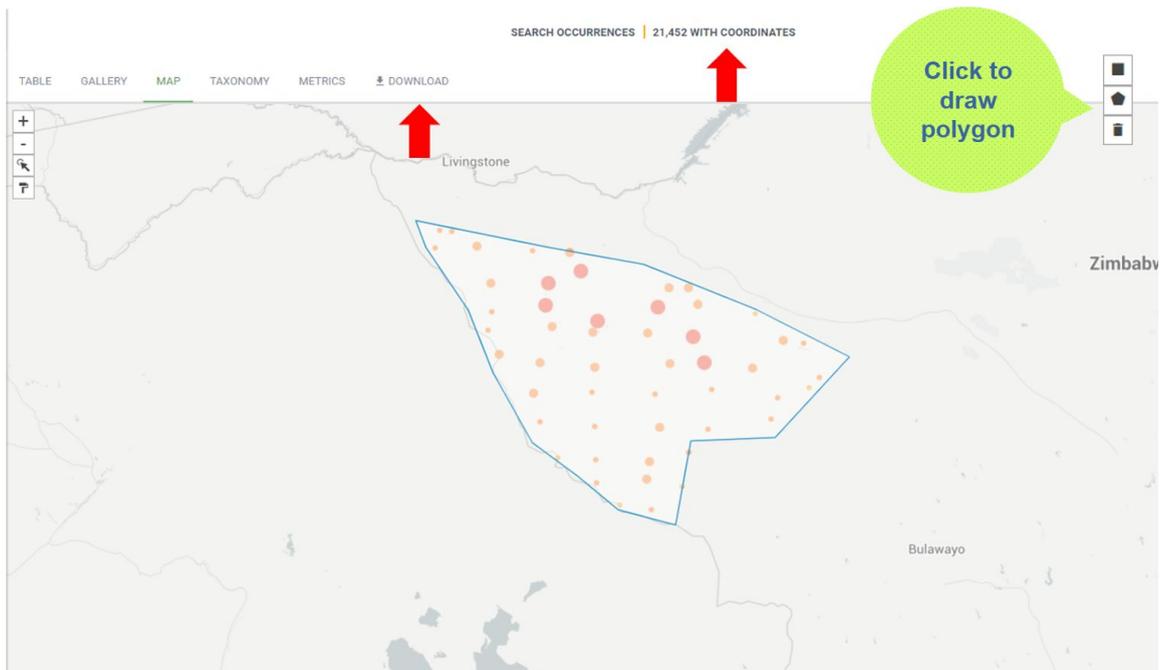
5. Explore the **Taxonomy** tab to know the taxonomic distribution of occurrences.
6. Click on the **Download** button and select the **Species List** option.

7. Country page

At this moment, Mudiwa already knows how to explore the biodiversity information available through GBIF.org. Now he would like to get some information about the countries (Zimbabwe, Zambia and Mozambique) where he conducts his study and make a comparison between them. He would like to answer the following questions:

1. Have any of these countries established a GBIF node? When were they established?
2. What is the country with more occurrence records published in his territory?
3. What is the kingdom most represented in each country?
4. Generate an activity report for each country.

Follow the instructions below to answer the questions.

Instructions

1. Open the GBIF Data Portal https://www.gbif.org/
2. Type the name of each country in the search bar and look for them in the returned results.

3. Once you have opened the three country pages, look if these pages include the "PARTICIPATION" tab, which will be activated if the country has established a node and it will contain information about the node.

4. Next, select the "DATA ABOUT" tab in each country page to see how many occurrence records have been published in those countries.
5. In the same tab for each country, scroll down and explore the chart for occurrences per kingdom to see the most represented kingdom in each country.
6. Lastly, go to the top part of each country page and click in the "ACTIVITY REPORT" button to download that document.
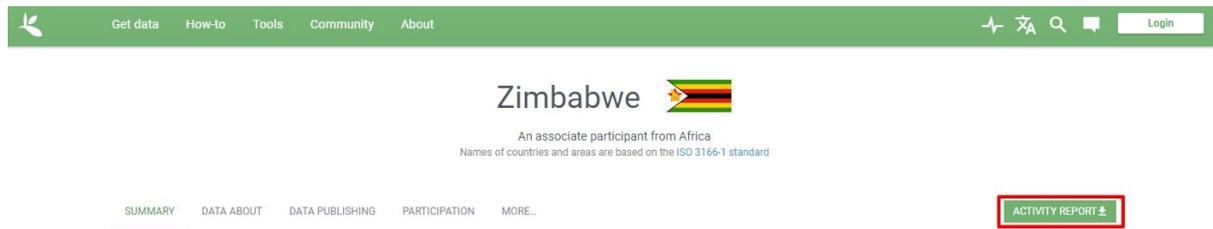
## 8. Institution and dataset pages

Now, Mudiwa would like to know more about how active is the Natural History Museum of Zimbabwe and the kind of information this institution is mobilising through GBIF. He would like to answer the following questions:

1. When did this institution register in GBIF?
2. How many occurrence records and datasets have already been published by the Museum?
3. How many papers have been published using data mobilised by the Museum?
4. What are the main topics covered by the first paper?
5. How many checklist datasets have been published by the Museum?
6. What dataset published by the Museum is mobilising more data through GBIF?
7. Who is the principal investigator and administrative point of contact for this dataset?
8. In this dataset, what is the most represented order? And the most represented family within that order?

Follow the instructions below to answer the questions.

Instructions

1. Open the GBIF Data Portal https://www.gbif.org/
2. Type the name of the institution (Natural History Museum of Zimbabwe) in the search bar and look for it in the returned results. To make your search easier, in the results generated by GBIF, select the "PUBLISHERS" tab and click on the Natural History Museum of Zimbabwe institution. There you can find the requested information.
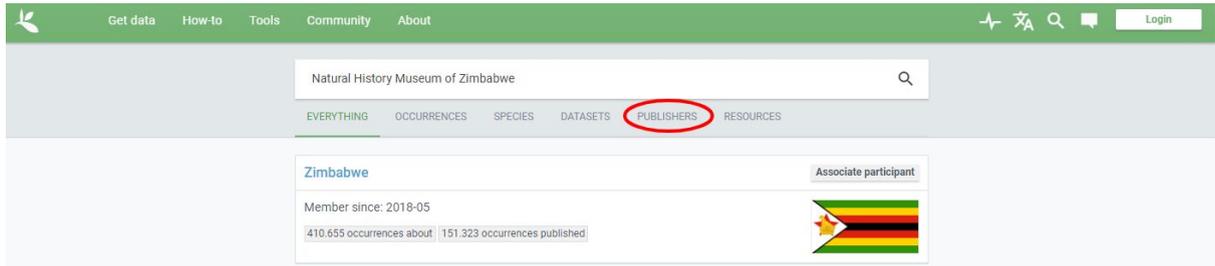
3. Click on the citation button to see all the papers using data from this institution and take note on what are the topics of the first paper in the list.
4. Come back to the institution page and click on the dataset button.
5. Select the "CHECKLIST" tab and you will see how many datasets are checklist datasets.
6. Now, select the "ALL" tab and see which of them is mobilising more occurrence records.
7. Select that dataset to see the dataset page. In the "DATASET" tab, go to the "Contacts" section and explore the different contact people associated with this dataset.
8. Finally, select the "METRICS" tab and go to the phylogenetic tree to see the most represented order.

9. Click in the most represented order and see what is the most represented family.