SCIENTIFIC NAME MANAGEMENT:

NORMALIZATION USING *SPECIES MATCHING* GBIF.ORG

CESP Project: Strengthening Zimbabwe’s GBIF node through collaboration with GBIF Spain

# 

# INTRODUCTION

In this exercise, we will check the validity of the scientific names in our dataset using the [*Species Matching*](https://www.gbif.org/tools/species-lookup) tool from GBIF. Where convenient, we will provide the correct name, its authors, taxonomic rank and the top taxonomic hierarchy in the work table. All this automatically and massively.

# PREVIOUS CONSIDERATIONS

[*Species Matching*](https://www.gbif.org/tools/species-lookup) It is a tool developed by the GBIF Secretariat (Global Biodiversity Information Facility) that allows verifying the nomenclatural validity of a given list of scientific names. To do this, it uses the GBIF taxonomic database (GBIF Backbone Taxonomy), which is the nomenclatural reference used by GBIF to integrate the information of scientific names from different resources, which provides consistency to the portal and facilitates taxonomic search in it.

Once a list of scientific names is provided in a .csv or .txt file, the Species Matching tool returns the following results that can be downloaded in .csv:

* Matching between the original names with those existing in the Backbone
* Correct name proposal
* Nomenclatural Status
* Taxonomic range
* Taxonomic hierarchy above the taxon provided
* Scientific names authors

# YOU WILL NEED

* The exercise data [Trees of food importance in biodiversity hotspots of Zimbabwe](https://drive.google.com/open?id=1bVdD38rpTx5fE_L-oaqPOumidgVT1D3w)previously adapted to the Darwin Core standard.
* [*Species Matching*](https://www.gbif.org/tools/species-lookup) tool.

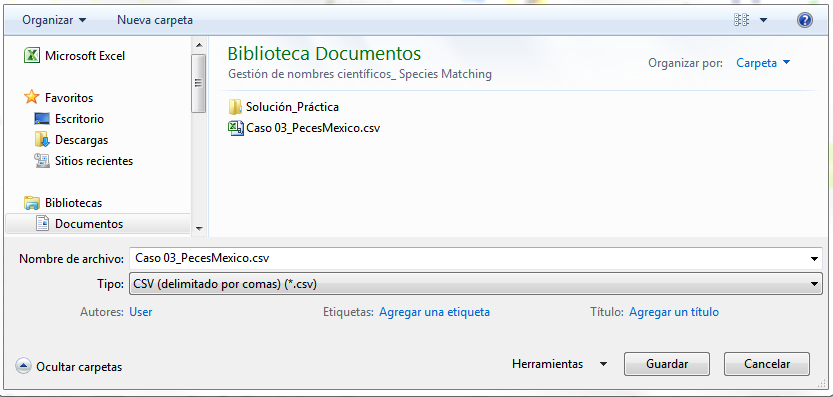
# INSTRUCTIONS

***Step 1.* Preparation of the work file.**

To enter the scientific names in *Species Matching* these must be provided as a .csv or .txt file that contains at least one column called "scientificName". Since our table is adapted to the standard (previous exercise) it contains this field and we simply have to save it in the appropriate format for Species Matching to interpret it. How to:

1. Open the work table that is in excel format. Check that you have the *ScientificName* field.
2. Save the excel file in .csv format by clicking on *File-*> *Save as* -> and in the drop-down "type" select *CSV (comma delimited) (\* .csv).*

*Note:* I have realized that sometimes it has problems interpreting tables with many fields so, as an alternative, you can simply copy the column *ScientificName* in a txt file.

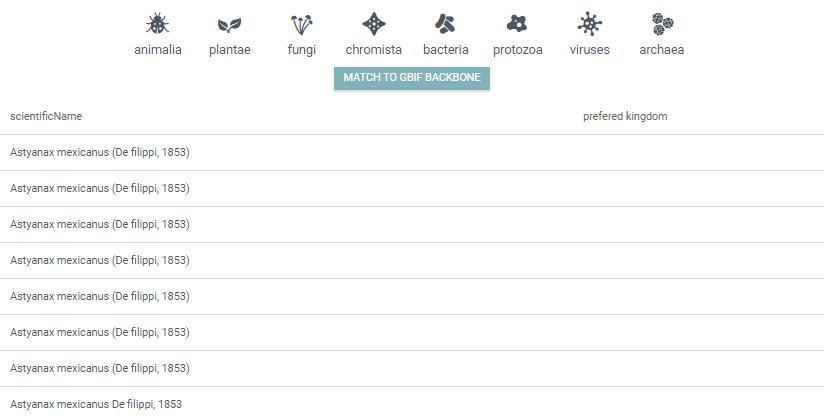


***Step 2****.*  **Names validation**

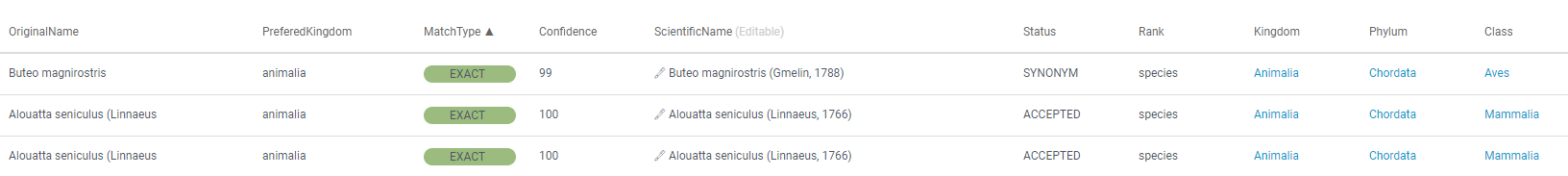
1. Open the [Species Matching](https://www.gbif.org/tools/species-lookup) tool at[*https://www.gbif.org/tools/species-lookup*](https://www.gbif.org/tools/species-lookup)*.*  There are two ways to upload the file to be normalized: clicking on the button **Select File** to search the file located in your documents or dragging the file to the **Drop here area.**

****

1. The list of names will automatically be previewed. If you didn't specify in the source, select the Kingdom **Plantae** according to the use case and click on **Match to GBIF backbone**.

****

1. A list with the following fields will be shown:

******

* **verbatimScientificName:** the original scientific name provided.
* **preferedKingdom:** the kingdom provided or selected.
* **matchType:** with text and in color code indicate the coincidence between the original name provided and the GBIF Backbone. FUZZY (doubtful), NONE or HIGHERRANK (It is assigned a name when no matches have been found at the same taxonomic level and a higher one is proposed or when homonyms are found).
* **confidence: match percentage**
* **scientificName** (editable): proposed matched scientific name.
* **status:** nomenclatural status.

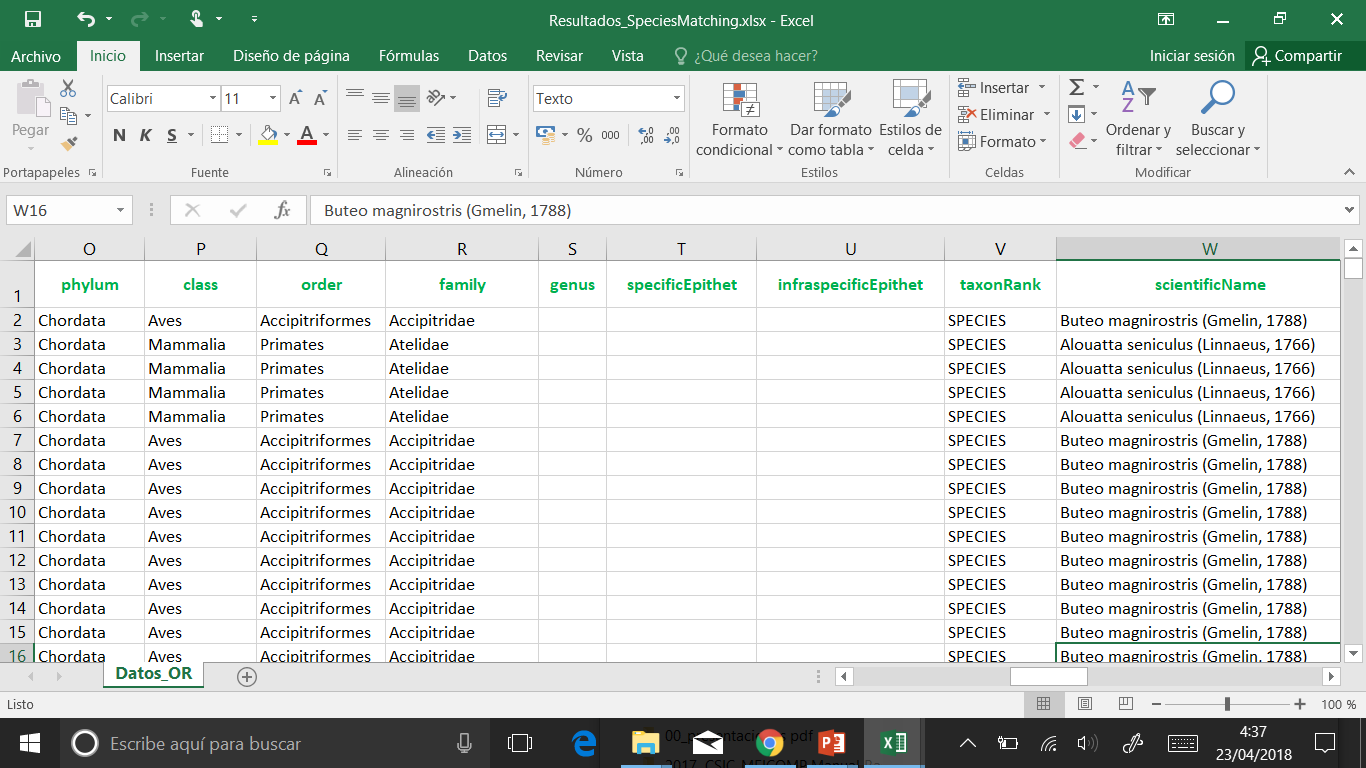
1. Sort matchType values (clicking on the column) to find the FUZZY, NONE and HIGHERRANK names.
2. From the **ScientificName** column, use the editing tool  to check manually the names matched as HIGHERRANKandFUZZY. Replace the name proposed by other more appropriate if you find a better option.

***Step 3.* Download and results display**

1. Download the results form option **Generate CSV** available at the bottom of the page.
2. Save the file **Normalized.csv** on your computer.
3. To display the above results, import the Normalized.csv data into an excel file following the instructions in the “[How to import text files into excel](https://drive.google.com/open?id=1rva4hsZtUQGV8Z5IM44chKhphH8x5Sc-)” guide. Rename the new file as **"Normalized\_Name.xlsx"**.

***Step 4.* Integration of the results in the original table**

1. Select, copy (Ctrl + C) and paste (Ctrl + V) in your table *Trees of food importance in biodiversity hotspots of Zimbabwe.xlsx* the values from the fields*ScientificName, kingdom, phylum, class, order, family* y *rank* obtained in the previous step.



1. Save the file.