DATA VALIDATION: GBIF DATA VALIDATOR

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

# 

# INTRODUCTION

The [GBIF data validator](https://www.gbif.org/tools/data-validator) is a service that allows anyone with a GBIF-relevant dataset to receive a report on the syntactical correctness and the validity of the content contained within the dataset. By submitting a dataset to the validator, you can go through the validation and interpretation procedures usually associated with publishing in GBIF and quickly determine potential issues in data - without having to publish it.

The files accepted by this tool are:

* ZIP-compressed Darwin Core Archives (DwC-A) (containing cores Occurrence, Taxon, or Event).
* Integrated Processing Toolkit (IPT) Excel templates containing [Checklist](https://github.com/gbif/ipt/wiki/checklistData#templates), [Occurrence](https://github.com/gbif/ipt/wiki/occurrenceData#templates), or [Sampling-event](https://github.com/gbif/ipt/wiki/samplingEventData#templates) data.
* Simple CSV and excel files containing Darwin Core terms in the first row.

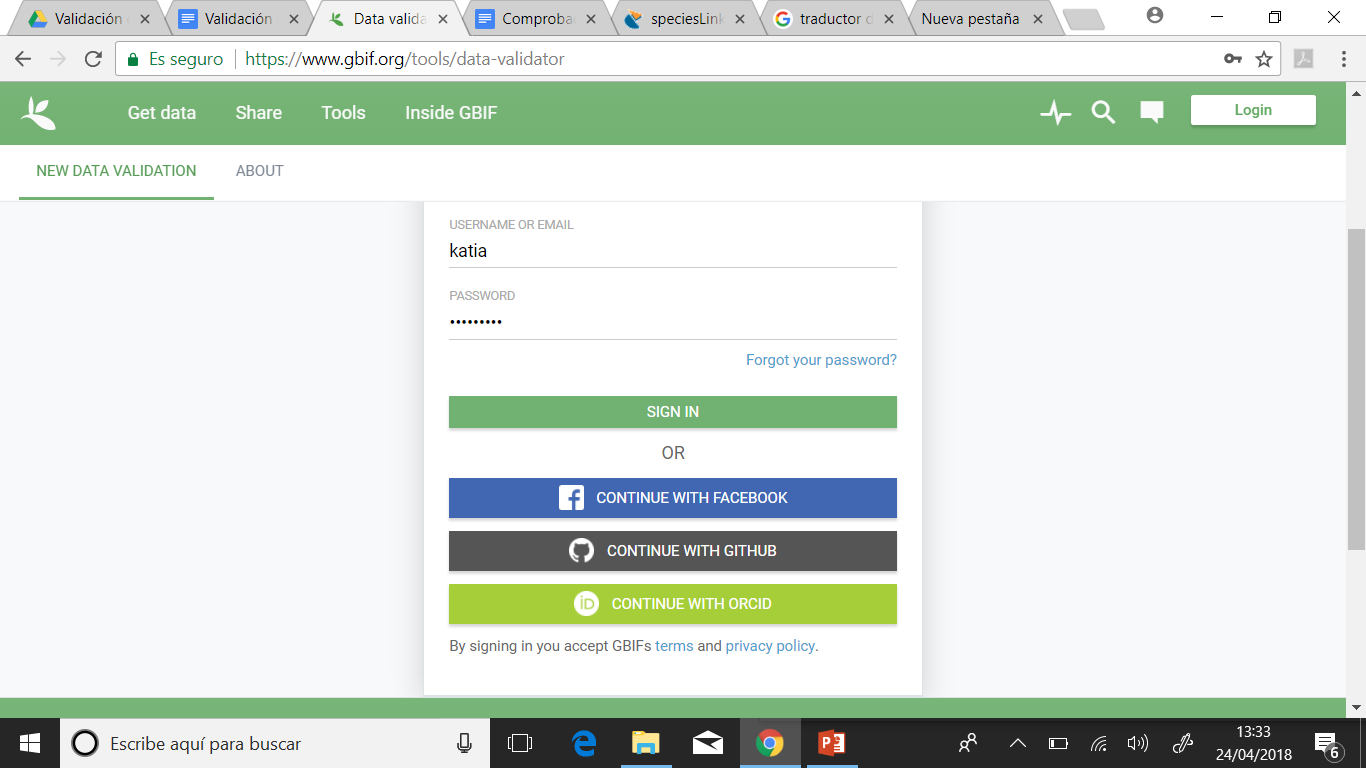
# MATERIALS

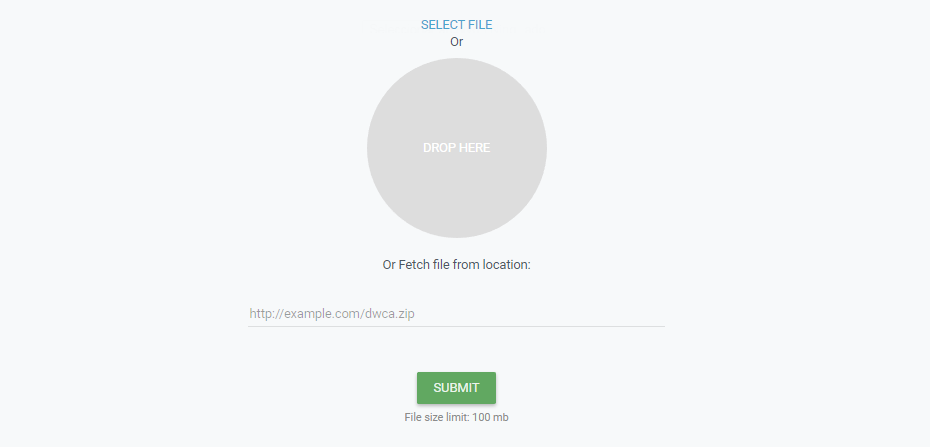
Tool URL: <https://www.gbif.org/tools/data-validator>

Work dataset: [Caso 07\_Validacion\_PecesMexico.xlsx](https://drive.google.com/open?id=1Zy7Ms61BhvhDD0tnHOI1JGrR6uNliyCJ)

INSTRUCTIONS

1. Open the Data Validator tool.If you don't already have one, **create a user account at** gbif.org.

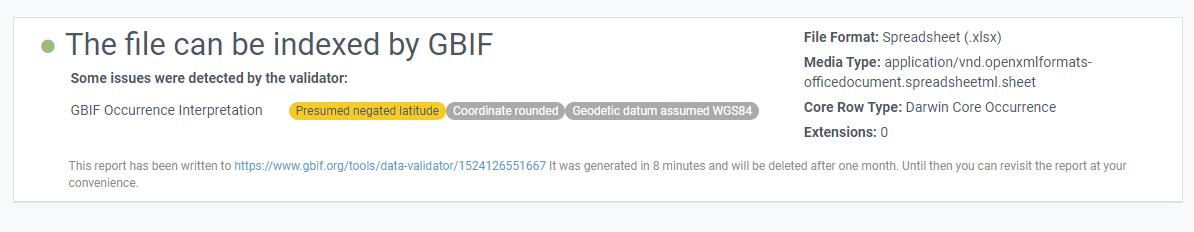


1. Upload the dataset file to the validator, either by 1) clicking **SELECT FILE** and selecting it on your local computer or 2) dragging the file from a local folder and dropping it on the **Drop here** icon. Once you hit the **Submit** button, the validator starts processing your dataset file. You will be taken straight to a page showing the status of the validation.****

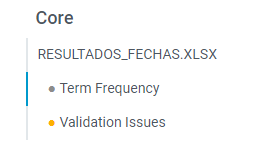
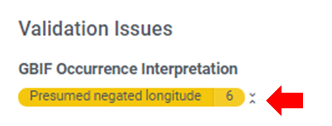
Depending on the size of your dataset, processing might take a while. Once processing is done, you will be able to see the validation report containing the following information:

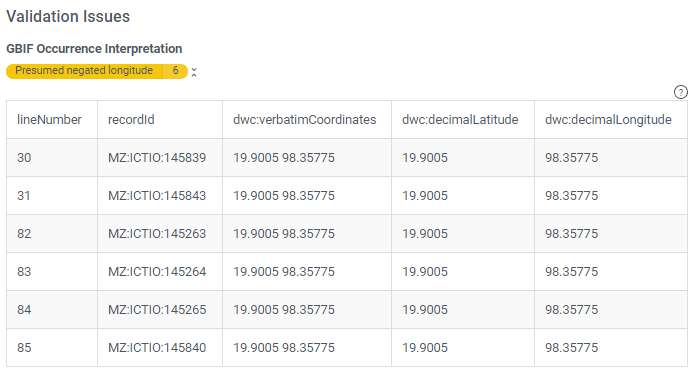
* A summary of the dataset type and a simple indicator of whether it can be indexed by GBIF or not.
* A summary of issues found during the GBIF interpretation of the dataset.
* Detailed break-down of issues found in metadata, dataset core, and extensions (if any), respectively.
* The number of records successfully interpreted.
* Frequency of terms used in the dataset.

You will also be able to view the metadata as a draft version of the dataset page as it would appear when the dataset is published and registered with GBIF.





1. Analyze the header displayed to know if the dataset can be indexed by GBIF or not.
2. Click on **Validation isues** and analyze the results. 
3. Click on the warning arrows (indicated above with red arrow) to obtain a breakdown of the records with incidents. Here all the records and fields that have been interpreted as problematic are shown, being able to identify them at the source through their recordId.



6. Identify all the errors detected by the system.

### **I’ve got the validation report - now what?**

If the validator finds that your dataset cannot be indexed by GBIF, you should address the issues raised by the validation report before you consider publishing it to GBIF. On the other hand, if you get the green light and your dataset is indexable by GBIF, you should still carefully review any issues that may be the result of e.g. conversion errors, etc. which could affect the quality of the data. If you find and correct any error - from a single typo to large systematic problems - feel free to resubmit your dataset as many times you like.

23/09/2019

This report has been written to <https://www.gbif.org/tools/data-validator/1568816668926> It was generated 2 hours ago And will be deleted after one month. Until then you can revisit the report at your convenience.