VALIDACIÓN DE DATOS: DATA VALIDATOR GBIF

Taller GBIF.ES: Publicación de datos de biodiversidad en GBIF y en revistas científicas

# 

# INTRODUCCIÓN

Durante esta práctica detectaremos posibles incidencias en estructura y contenido de nuestro juego de datos, y comprobar su validez para ser publicado en el portal de GBIF.

ANTES DE EMPEZAR

El [validador de datos GBIF](https://www.gbif.org/tools/data-validator) es un servicio que permite que cualquier persona con un conjunto de datos relevante para GBIF reciba un informe sobre la corrección sintáctica y la validez del contenido de dicho conjunto de datos. Al enviar un conjunto de datos al validador, puede determinar rápidamente los posibles problemas en los datos, sin tener que publicarlos.

Los archivos admitidos por esta herramienta son:

* Archivos Core de Darwin comprimidos en ZIP (DwC-A) que contienen términos de Darwin Core en la primera fila.
* Plantillas de Excel de Integrated Processing Toolkit (IPT) que cque contienen términos de Darwin Core en la primera fila.
* Archivos CSV simples que contienen términos de Darwin Core en la primera fila.

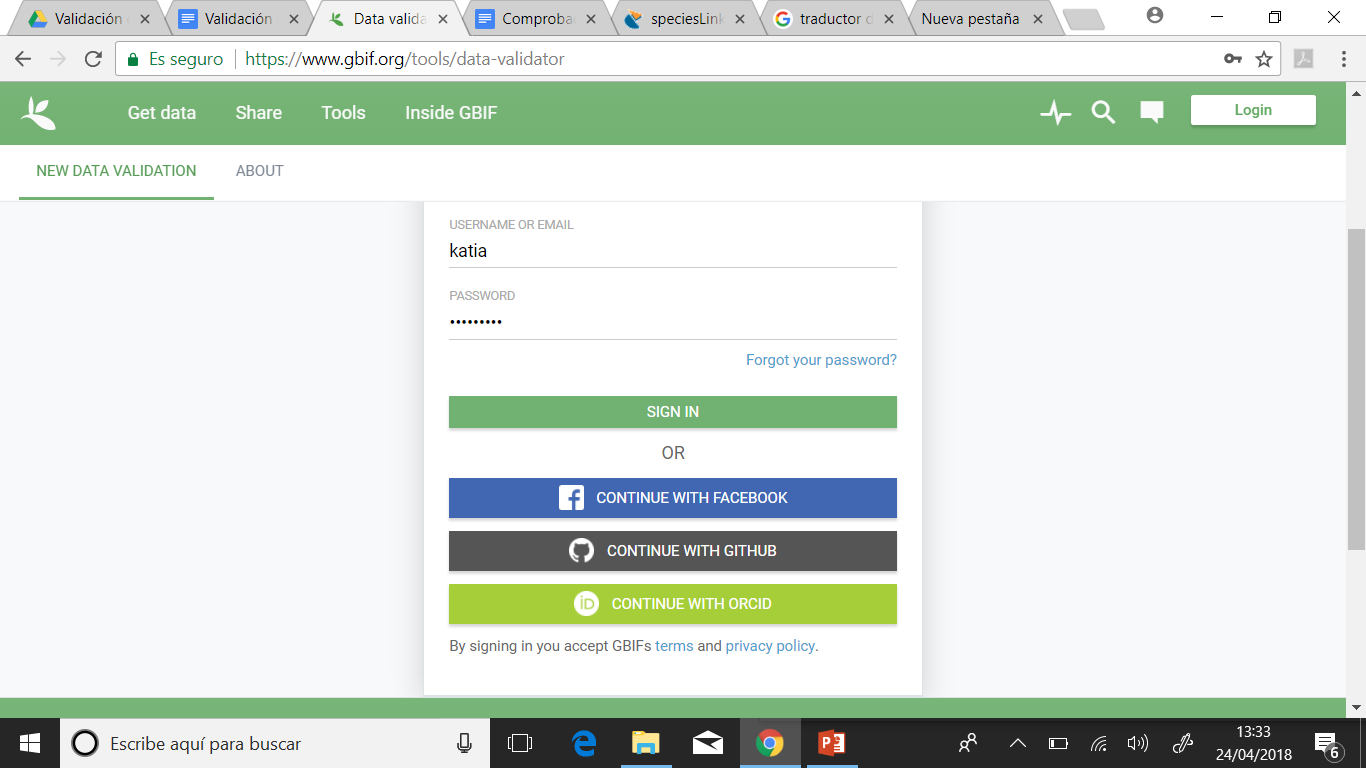
# NECESITARÁS

Enlace web a la herramienta: <https://www.gbif.org/tools/data-validator>

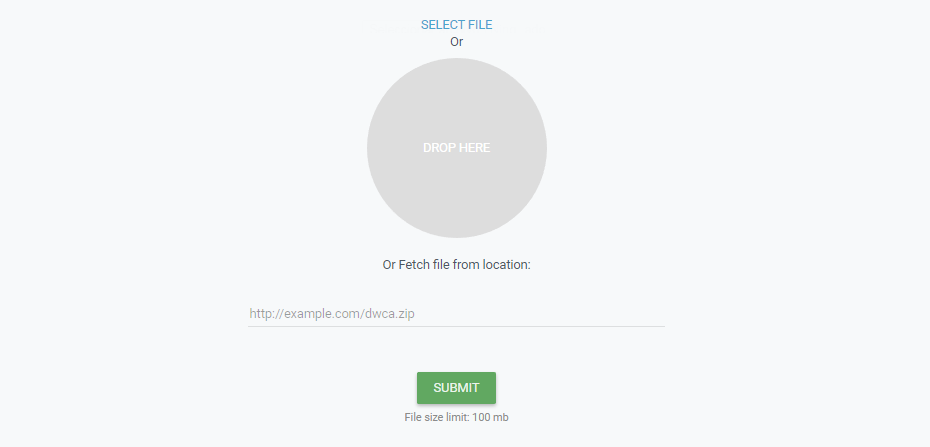
Tabla de trabajo: [Caso 07\_Validacion\_PecesMexico.xlsx](https://drive.google.com/open?id=1Zy7Ms61BhvhDD0tnHOI1JGrR6uNliyCJ)

INSTRUCCIONES

1. Abra la herramienta Data Validator.Si no la tiene ya, **cree una cuenta de usuario** en gbif.org.

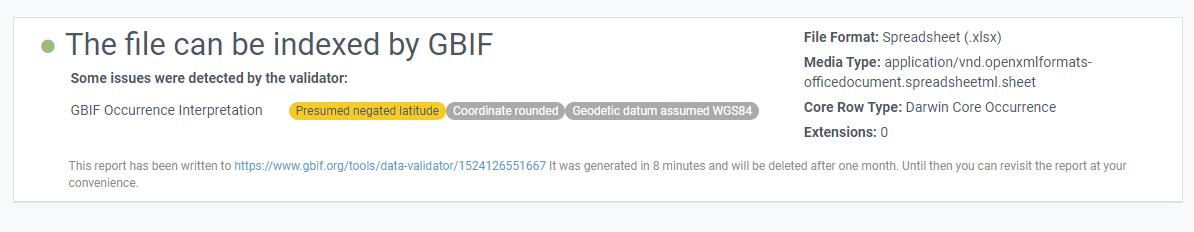


1. Comience cargando el archivo de datos de *Caso 07\_Validacion\_PecesMexico.xlsx* al validador, ya sea 1) haciendo clic en **SELECT FILE** y seleccionando en su computadora local o 2) arrastrando el archivo desde una carpeta local y soltandolo en el ícono **Drop here.**

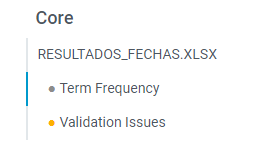
****

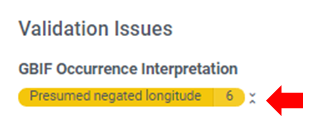
Rápidamente le ofrecerá el resultado de la validación indicando si el juego de datos es apto o no para la publicación en GBIF y podrá ver el informe de validación que contiene la siguiente información:

* Un resumen del tipo de conjunto de datos y un indicador simple de si puede ser indexado por GBIF o no
* Un resumen de los problemas encontrados durante la interpretación de GBIF del conjunto de datos
* Número de registros interpretados con éxito
* Frecuencia de los términos utilizados en el conjunto de datos



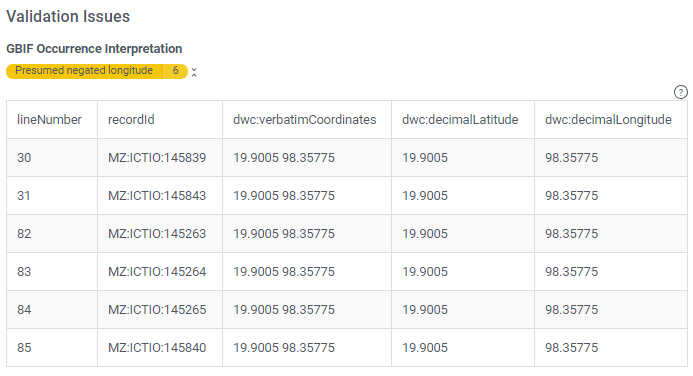


1. Analice el cabecero del resultado. Según el análisis realizado ¿es su juego de datos apto para la publicación en GBIF? Si el resultado es que sí, enhorabuena : )
2. Haga clic sobre **Validation isues** y analice los resultados. Se espera que se haya obtenido el siguiente aviso sobre un problema con la latitud de las coordenadas.



Este aviso informa de que se han encontrado registros con longitud positiva que aparentemente debería ser negativa.

1. Haga clic sobre las flechas del aviso (indicado arriba con flecha roja) para obtener un desglose de los registros con incidencias. Aquí se muestran todos los registros y los campos que se han interpretado como problemáticos, pudiendo identificarlos en el origen a través de su recordId.



**Tengo el informe de validación, ¿ahora qué?**

En una situación real, si el validador determina que su conjunto de datos no puede ser indexado por GBIF, debe abordar los problemas planteados por el informe de validación antes de considerar publicarlo en GBIF. Por otro lado, si obtiene luz verde y su conjunto de datos es indexable por GBIF, debe revisar cuidadosamente cualquier problema que pueda ser el resultado de, por ej. errores de conversión, etc. que podrían afectar la calidad de los datos. Si encuentra y corrige cualquier error, desde un solo error tipográfico hasta grandes problemas sistemáticos, puede volver a enviar el conjunto de datos tantas veces como desee.