

# TALLER GBIF.ES

## CALIDAD EN BASES DE DATOS DE BIODIVERSIDAD

Real Jardín Botánico Madrid, 2-4 Diciembre 2015



Santiago Martínez de la Riva Vivanco  
Unidad de Coordinación – GBIF.ES  
Nodo Nacional de Información en  
Biodiversidad en España  
[sama@gbif.es](mailto:sama@gbif.es)

### **Portal de Datos GBIF España:**

- El Portal de Datos es un sistema de información y gestión de instituciones, colecciones, recursos (juegos de datos) publicados por los proveedores nacionales a través de la red de GBIF España. Este sistema permite a los usuarios filtrar la información de los más de 11 millones de registros indexados a diferentes niveles y realizar la descarga de las búsquedas hechas.

[HTTP://DATOS.GBIF.ES](http://datos.gbif.es)

Información accesible a  
través del Portal de Datos de  
GBIF.ES.



## Jerarquía de Trabajo:

### Caso de estudio:

Institución -> (1 o varios) Colección -> (1 o varios) Recursos -> (1 o varios) Registros

- [Real Jardín Botánico \(CSIC\)](#)



- [Colección de Plantas vasculares \(MA\), Real Jardín Botánico \(CSIC\)](#)

- [Real Jardín Botánico, Madrid: Vascular Plant Herbarium \(MA\)](#)

- <http://datos.gbif.es/generic-hub/occurrences/ef0c2e69-aa92-437a-964c-9332d473c466>



### Test de Calidad a nivel de registro:

- Todos los registros contenidos en un juego de datos pasan un conjunto de test de calidad a diferentes niveles: Geoespacial (55), Taxonómico (14), Temporal (10) y Misceláneo (15).

### Original vs Procesado:

- Testeo de cada uno de los registros en tiempo de procesamiento.
- Carga -> **Procesamiento** -> Indexación.



### Test de Calidad a nivel de Colección:

- El informe que nos permite, hacer un seguimiento de la calidad de la información que ha sido cumplimentado por los proveedores a través de los cuestionarios, pudiendo identificar, qué es lo que falta para hacer cada colección en particular y todo el conjunto en global, un bloque de información de mayor calidad.

## [INFORME DE COLECCIONES](#)

## Calidad a nivel de Registros

Nombre de la prueba	Resultado
assumedPresentOccurrenceStatus ?	✖ Falló
Name not in national checklists ?	⚠ Advertencia
Basis of record not supplied ?	✔ Aprobada
Basis of record badly formed ?	✔ Aprobada
Invalid collection date ?	✔ Aprobada
incompleteCollectionDate ?	✔ Aprobada
First of the month ?	✔ Aprobada
Collector name unparseable ?	✔ Aprobada
Missing catalogue number ?	✔ Aprobada
Data are generalised ?	✔ Aprobada
Collection code not recognised ?	✔ Aprobada
Institution code not recognised ?	✔ Aprobada
Missing taxonomic rank ?	✔ Aprobada
Name not supplied ?	✔ Aprobada
Kingdom not recognised ?	✔ Aprobada
Name not recognised ?	✔ Aprobada
Invalid scientific name ?	✔ Aprobada
Location not supplied ?	✔ Aprobada
Country inferred from coordinates ?	✔ Aprobada
Supplied coordinates centre of state ?	✔ Aprobada
Coordinates centre of country ?	✔ Aprobada

**Calidad de Datos a nivel de REGISTRO:**

**Filtrando la información por las alertas encontradas:**

**Filtra tu búsqueda** x

	<b>Alertas</b>	<b>Contar</b>
<input type="checkbox"/>	assumedPresentOccurrenceStatus	11949148

  

**Filtra tu búsqueda** x

	<b>Incidencias taxonómicas</b>	<b>Contar</b>
<input type="checkbox"/>	No issues	11777987
<input type="checkbox"/>	Matched to homonym	99443
<input type="checkbox"/>	Parent concept is synonym of child	19859
<input type="checkbox"/>	Match to misapplied name	126
<input type="checkbox"/>	Species uncertain	72

  

INCLUIR los datos seleccionados ▼

EXCLUIR los datos seleccionados ▼

⬇
Cerrar

  

INCLUIR los datos seleccionados ▼

EXCLUIR los datos seleccionados ▼

⬇
Cerrar

## **Calidad de Datos a nivel de REGISTRO:**

### **Feedback:**

- Proceso continuo de mejora, con la ayuda de los usuarios del portal de datos a través de los reportes encontrados por los mimos en cada registro.
  - [http://datos.gbif.es/generic-hub/occurrences/search?q=%3A\\*&fq=assertions:%22incompleteCollectionDate%22#tab\\_map](http://datos.gbif.es/generic-hub/occurrences/search?q=%3A*&fq=assertions:%22incompleteCollectionDate%22#tab_map)  
[View](#)



### **Pruebas de calidad de datos**

con errores: 1

con advertencias: 1

aprobadas: 19

con falta de información: 11

no comprobadas: 49

[Ver informe de calidad de datos completo](#)

[Avisar de un error](#) (email > [feedback@gbif.es](mailto:feedback@gbif.es))

## Informe de datos

## Calidad a nivel de Colección

Totales		
Colecciones	160	
Instituciones	81	
proveedores de datos	1	
Recurso de datos	184	
Centros de datos	4	
Contactos	288	

## Calidad de datos de la colección

Collections with no collection type	1	0.6 %
Collections with no focus	63	39.4 %
Collections with no description	85	53.1 %
Collections with no keywords	1	0.6 %
Collections with no provider codes	1	0.6 %
Collections with no geo. description	11	6.9 %
Collections with no size	2	1.2 %
Collections with no digitised size	2	1.2 %

## Resumen del contacto

Collections with no contacts	2	1.2 %
Collections with no email contacts	23	14.4 %
Institutions with no contacts	2	2.5 %
Institutions with no email contacts	19	23.5 %



# TALLER GBIF.ES

## CALIDAD EN BASES DE DATOS DE BIODIVERSIDAD

Real Jardín Botánico Madrid, 2-4 Diciembre 2015

# Gracias



Santiago Martínez de la Riva Vivanco  
Unidad de Coordinación – GBIF.ES  
Nodo Nacional de Información en  
Biodiversidad en España  
[sama@gbif.es](mailto:sama@gbif.es)