

# Conceptos básicos de Calidad de Datos

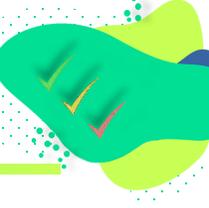
Publicación de datos de biodiversidad  
en GBIF y en revistas científicas

Paula Zermoglio, [pzermoglio@gmail.com](mailto:pzermoglio@gmail.com)  
Universidad de Buenos Aires - VertNet

**Gbif, Es**



# Estado de situación



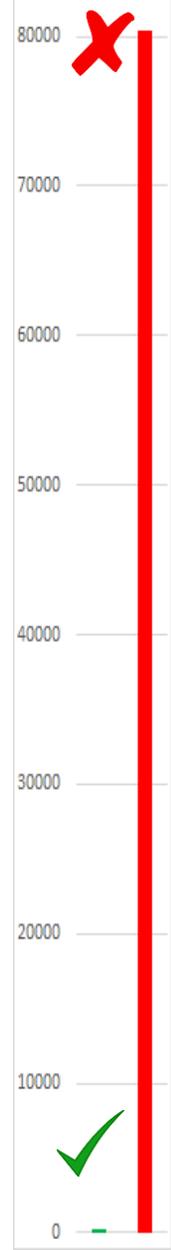
## Ejemplos:

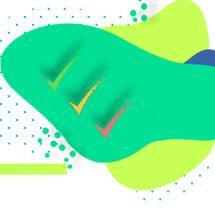
reproductiveCondition = **40.838** valores distintos



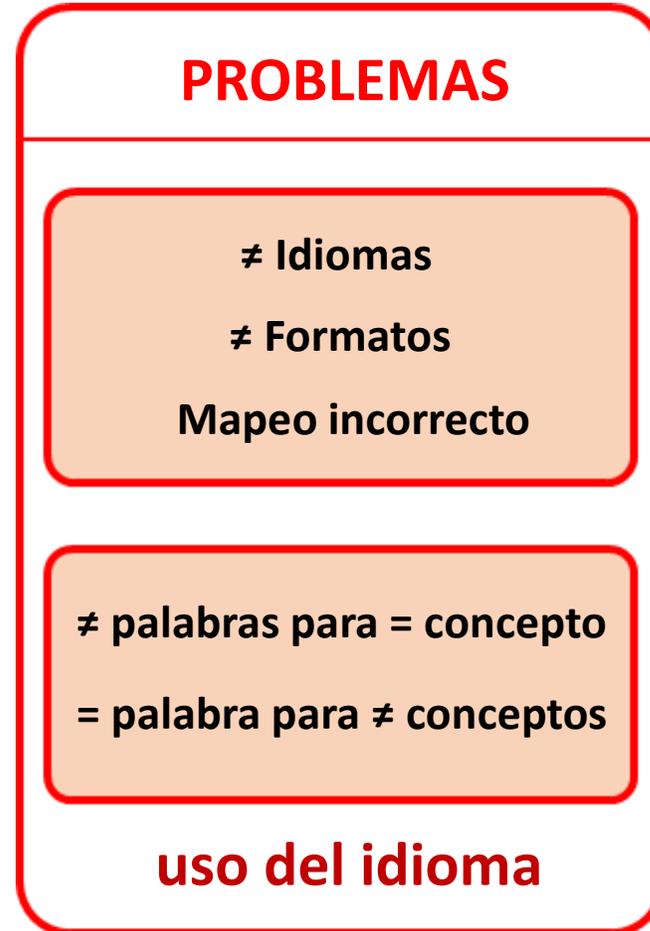
lifeStage = **33.402** valores distintos

country = **80.408** valores distintos  
(esperado = 250)





## Ejemplos:

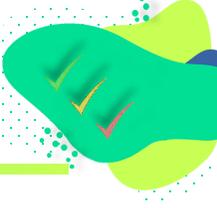


### Consecuencias

- Contenido excluído o ubicado incorrectamente
- Gran variabilidad

**Datos menos pasibles de descubrimiento y menos utilizables**

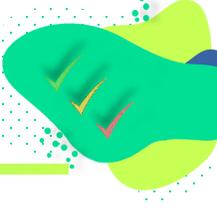
# Estado de situación



## Ejemplos:

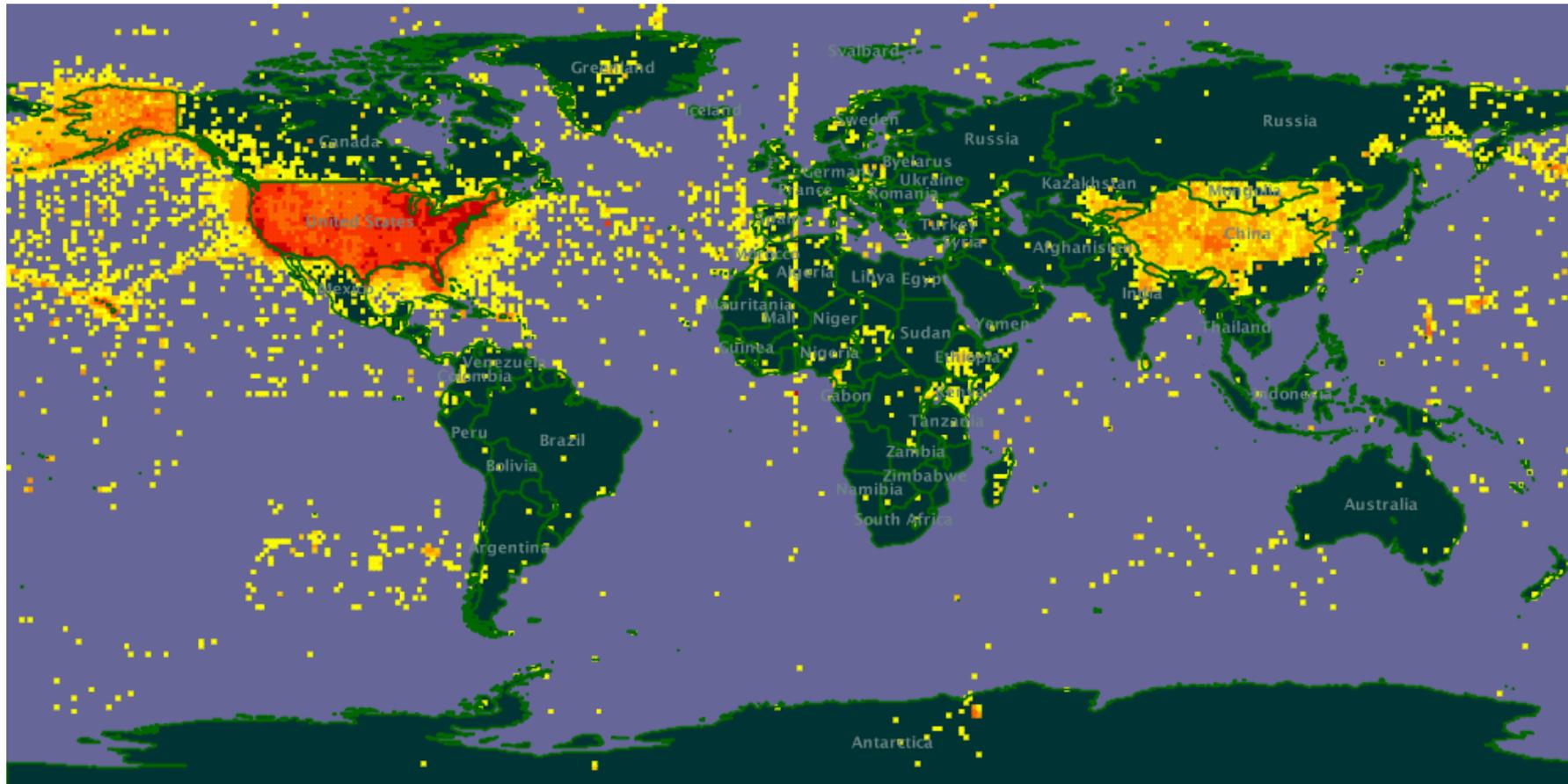


# Estado de situación

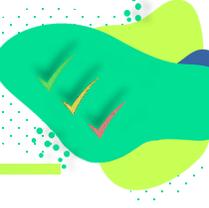


## Ejemplos:

Datos de ocurrencias de Estados Unidos en GBIF



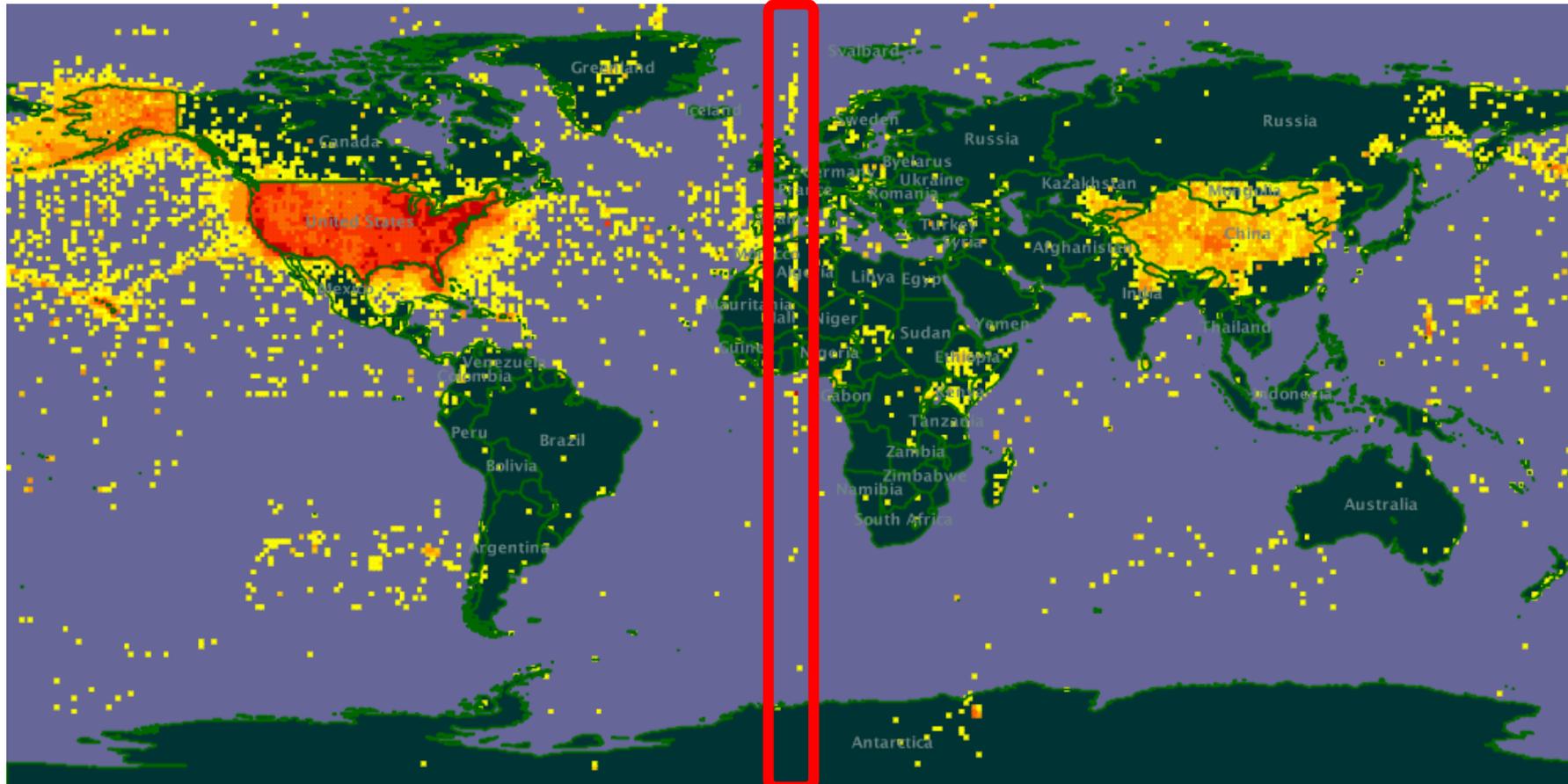
# Estado de situación



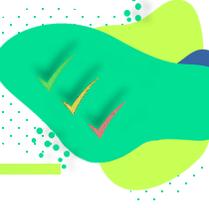
## Ejemplos:

Datos de ocurrencias de Estados Unidos en GBIF

Lon = 0



# Estado de situación

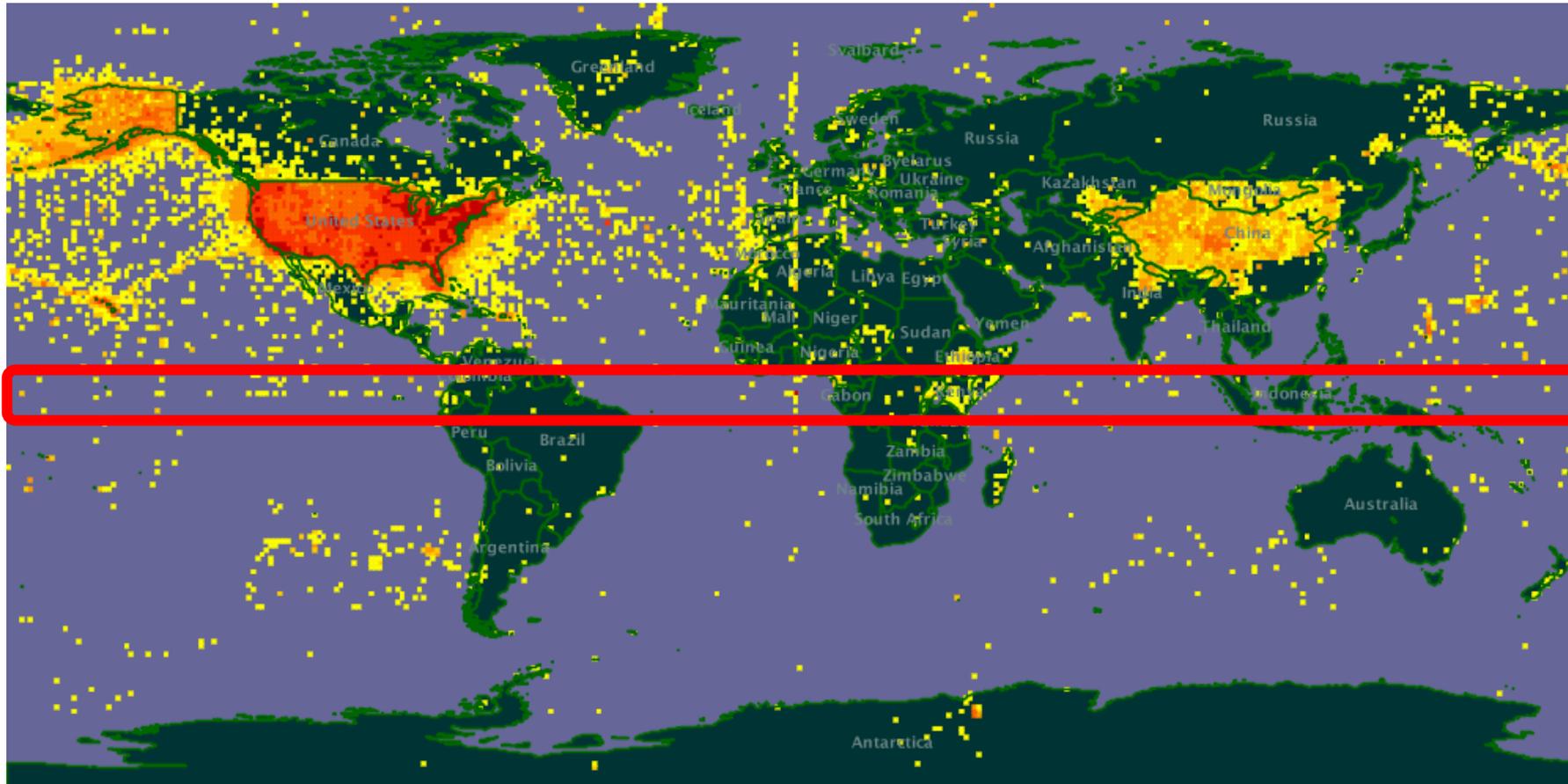


## Ejemplos:

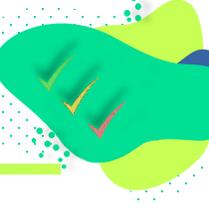
Datos de ocurrencias de Estados Unidos en GBIF

Lon = 0

Lat = 0



# Estado de situación



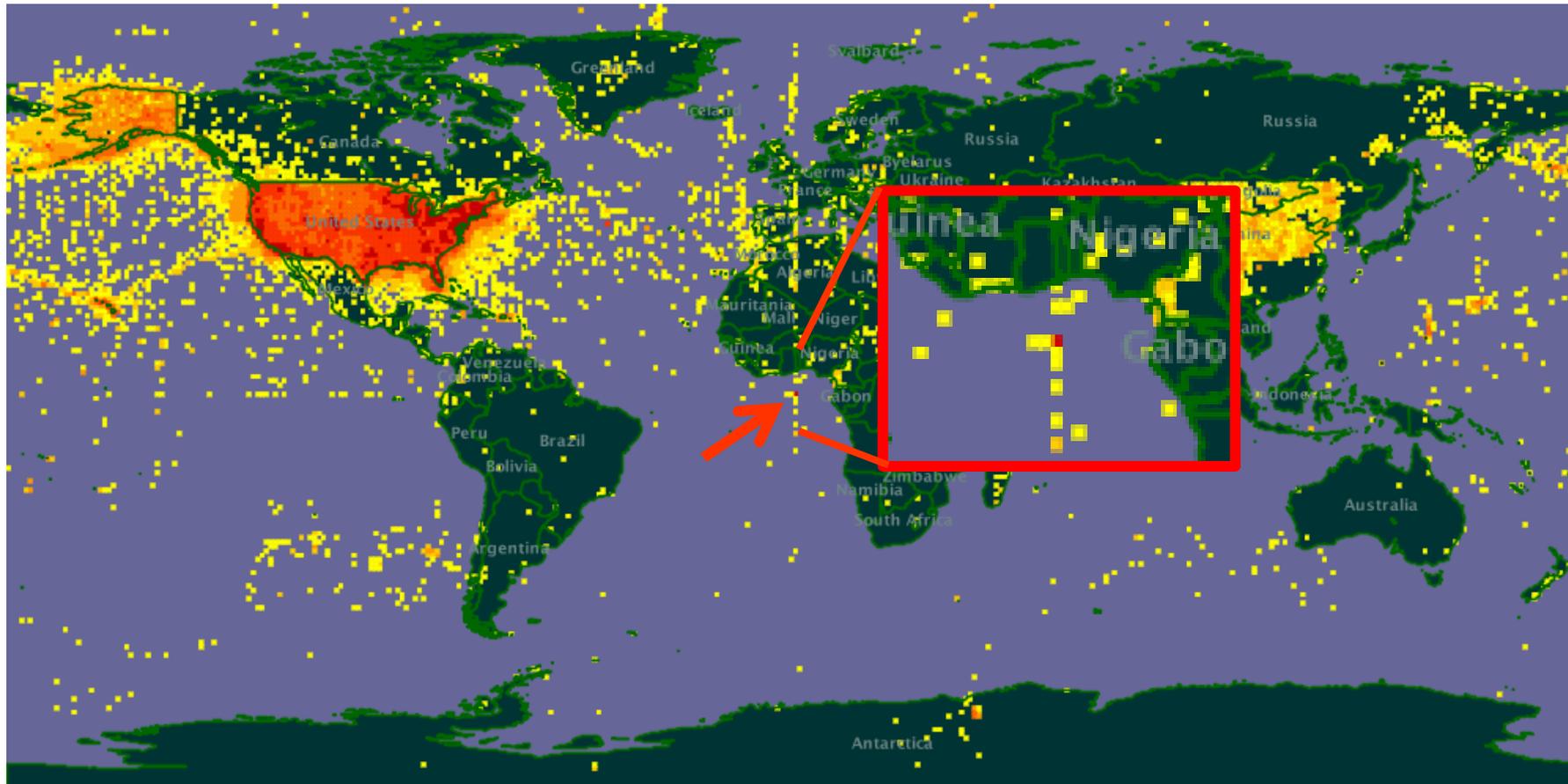
## Ejemplos:

Datos de ocurrencias de Estados Unidos en GBIF

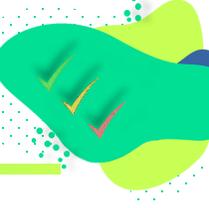
Lon = 0

Lat = 0

Lon y Lat = 0



# Estado de situación



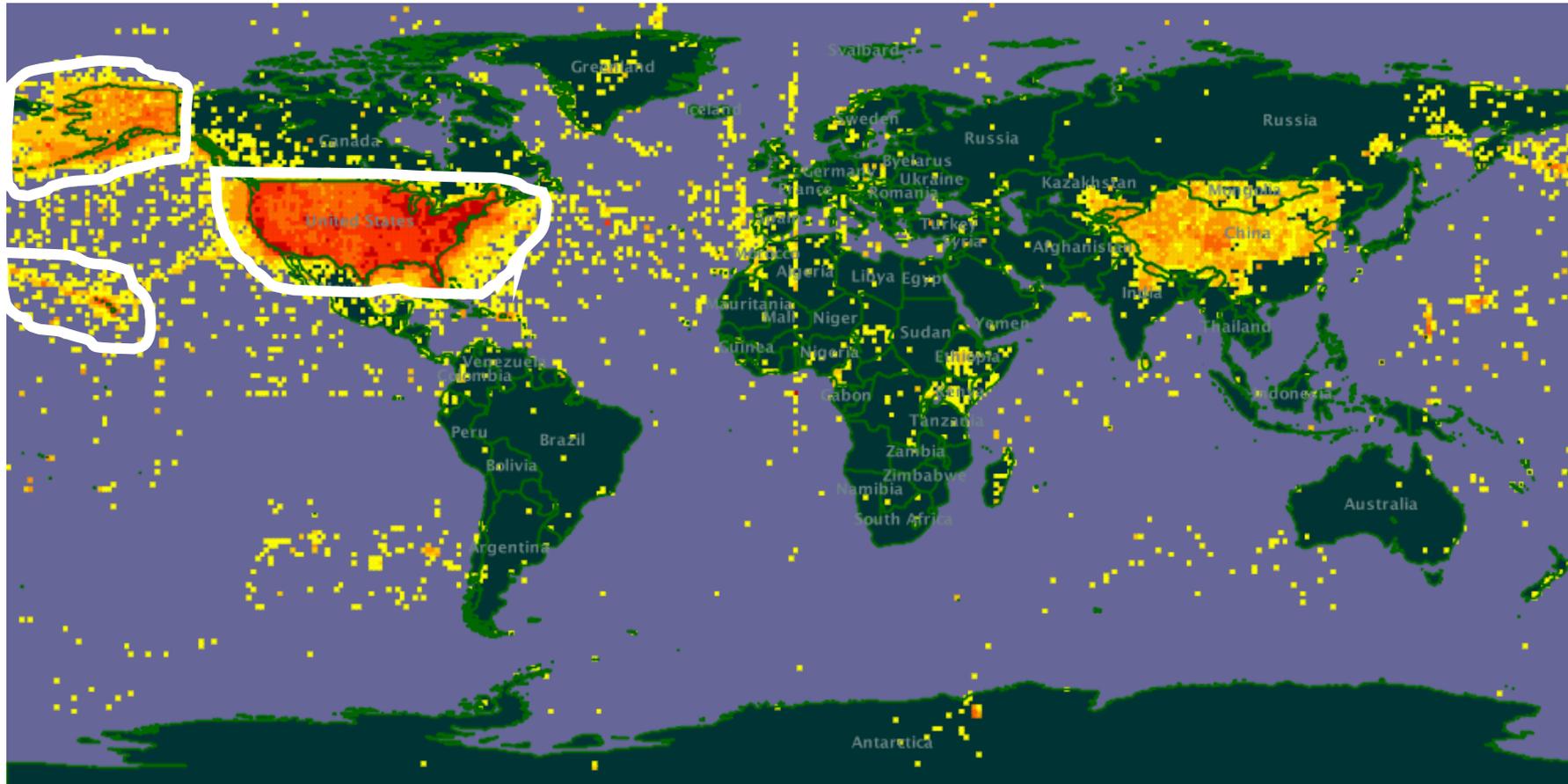
## Ejemplos:

Datos de ocurrencias de Estados Unidos en GBIF

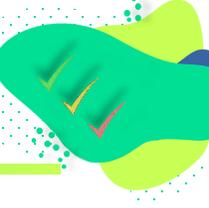
Lon = 0

Lat = 0

Lon y Lat = 0



# Estado de situación



## Ejemplos:

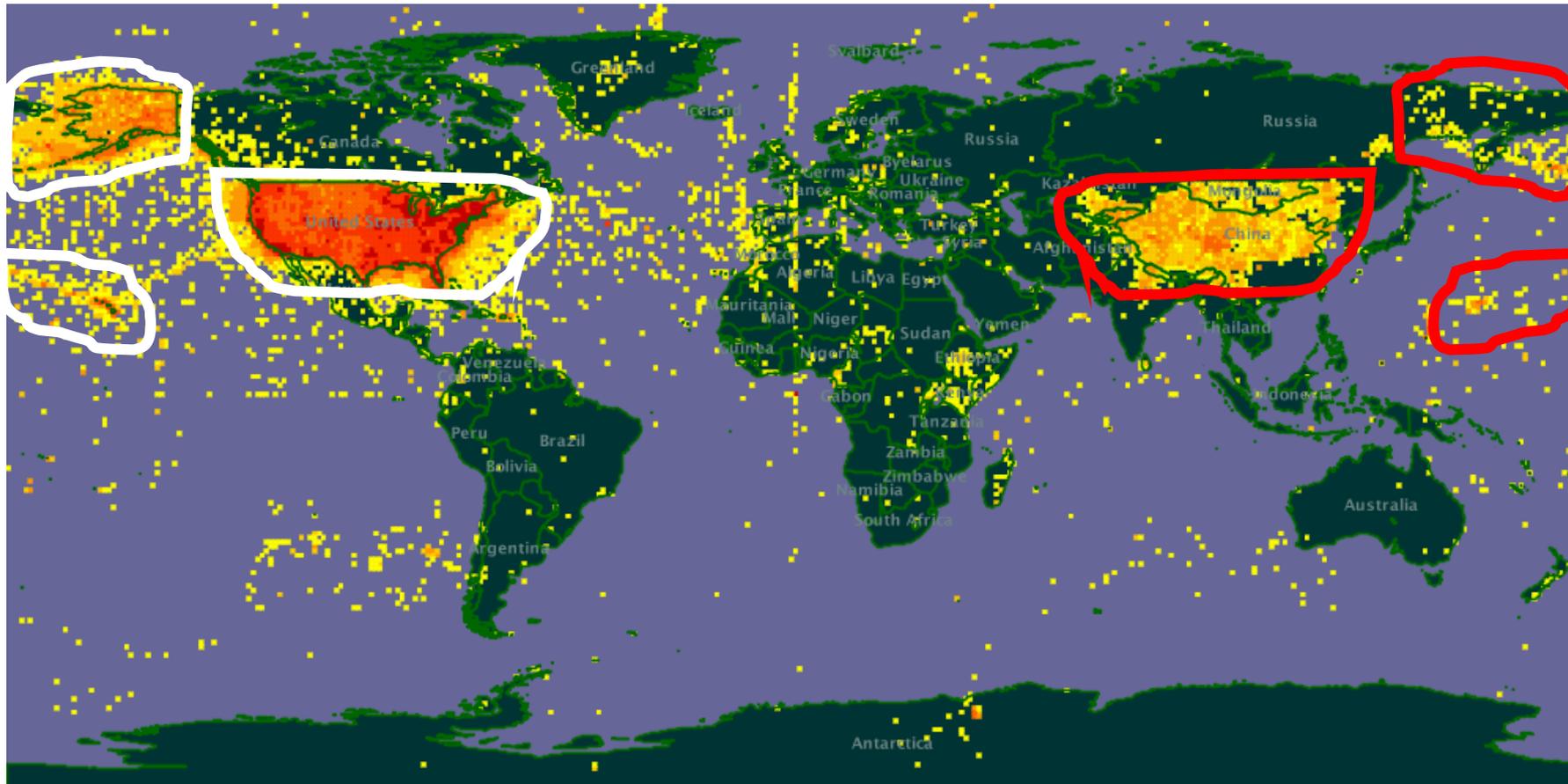
Datos de ocurrencias de Estados Unidos en GBIF

Lon = 0

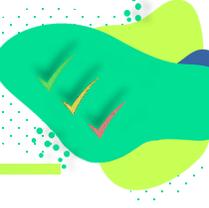
Lat = 0

Lon y Lat = 0

- Long



# Estado de situación



## Ejemplos:

Datos de ocurrencias de Estados Unidos en GBIF

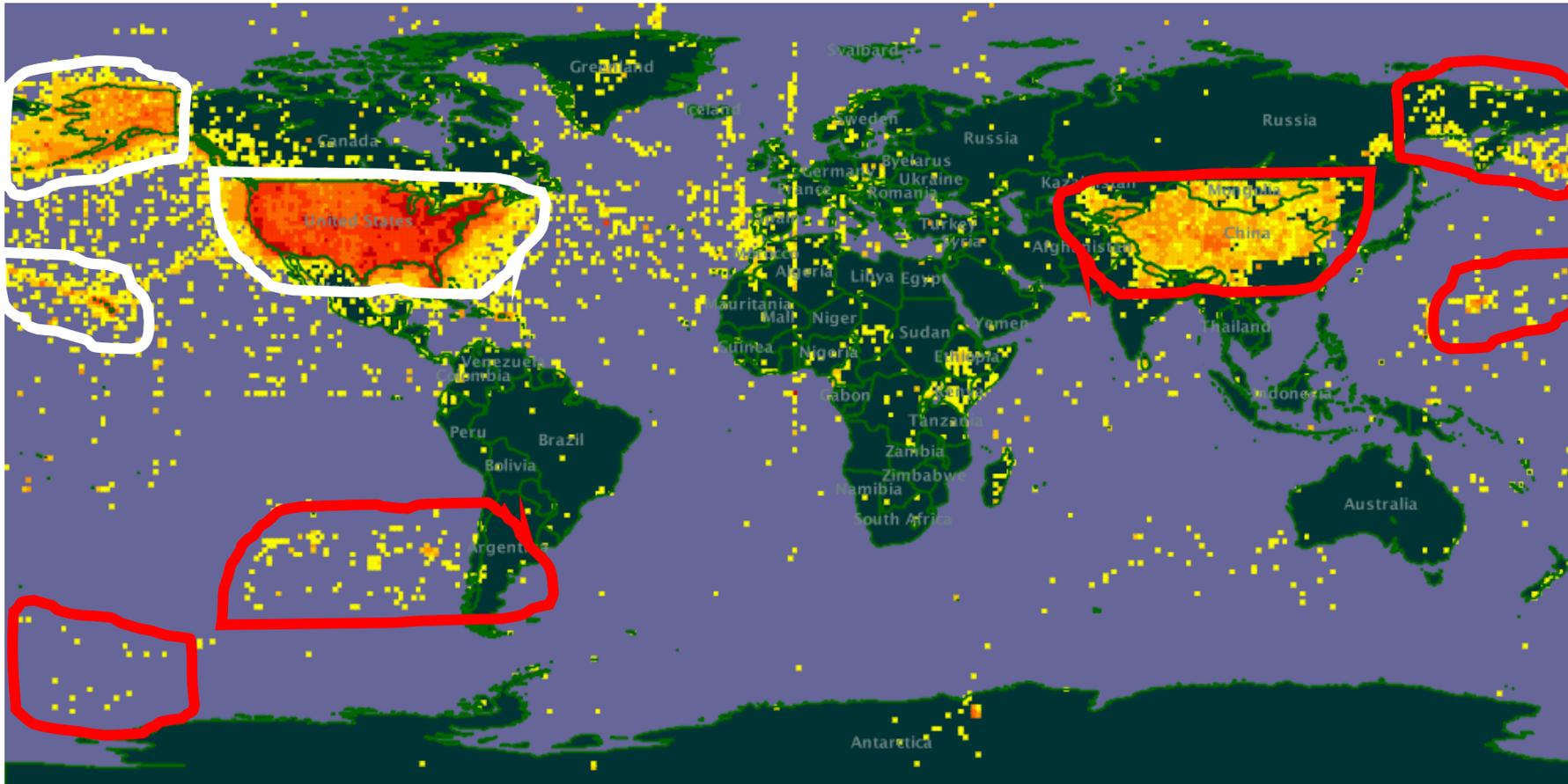
Lon = 0

Lat = 0

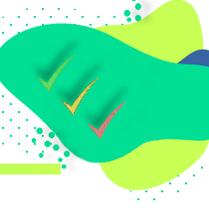
Lon y Lat = 0

- Long

- Lat



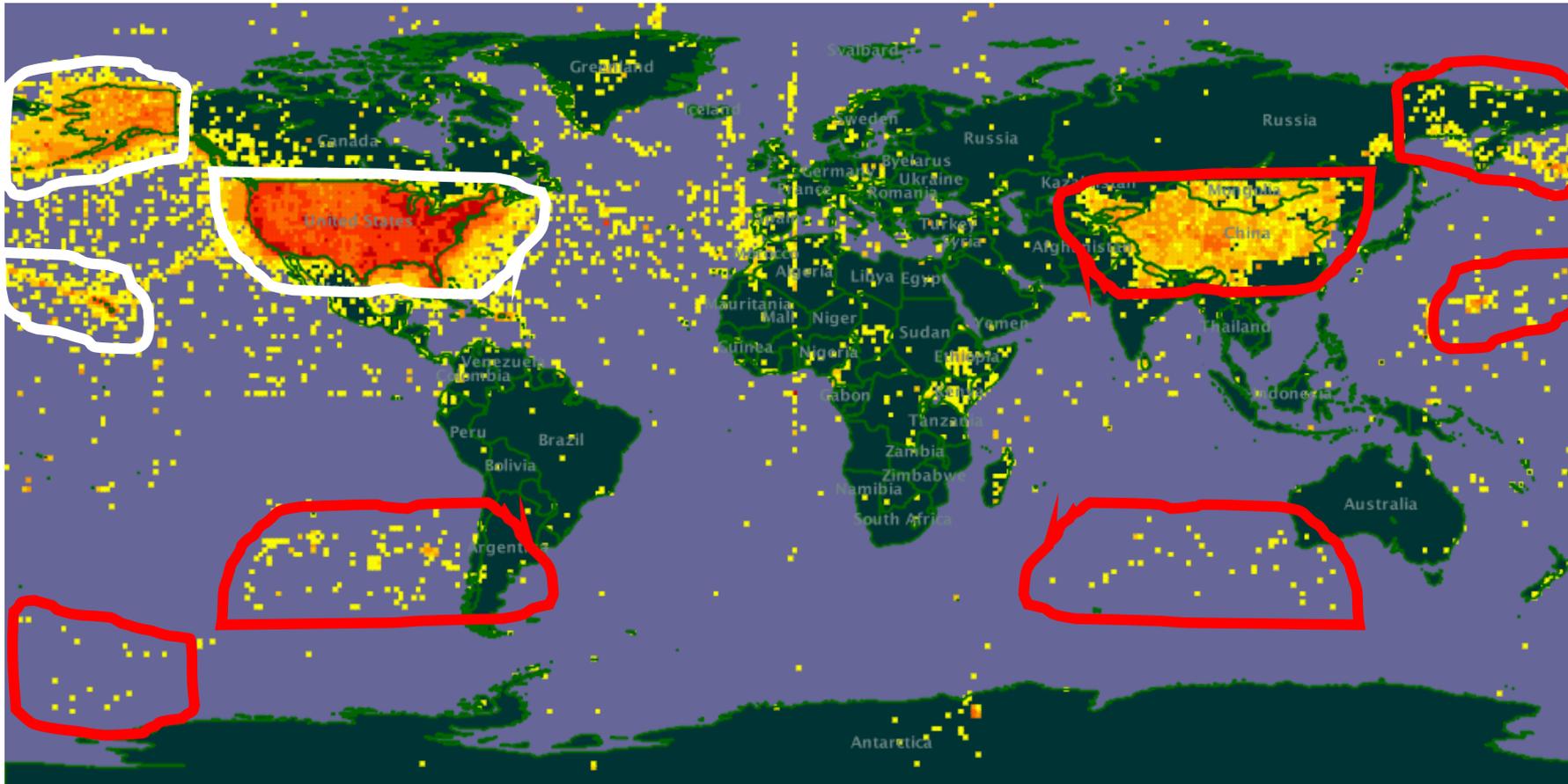
# Estado de situación

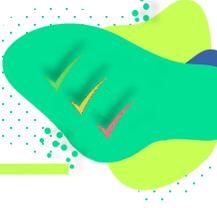


## Ejemplos:

Datos de ocurrencias de Estados Unidos en GBIF

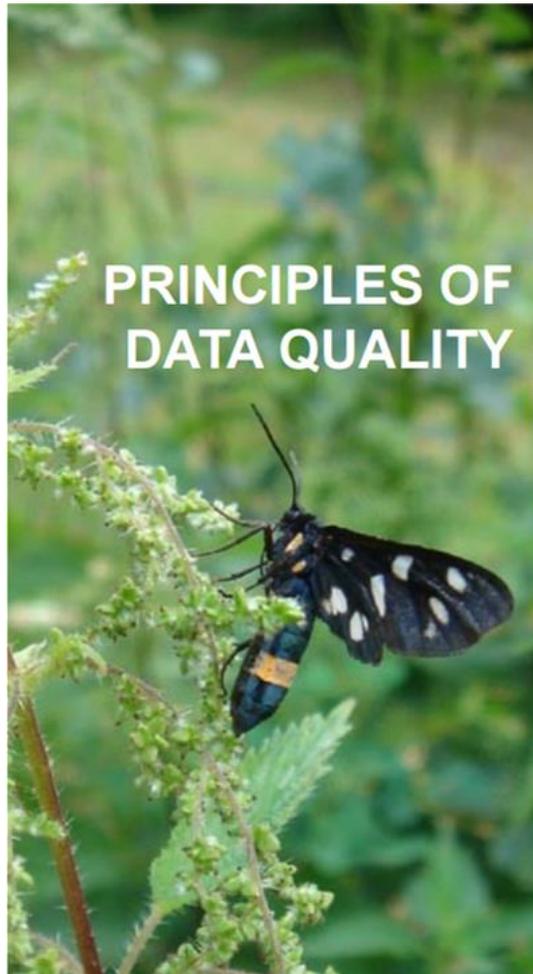
- Lon = 0
- Lat = 0
- Lon y Lat = 0
- Lon
- Lat
- Lon y Lat





## ¿Qué es la calidad de datos?

Lectura obligatoria



**Arthur D. Chapman'**

*Although most data gathering disciples treat error as an embarrassing issue to be expunged, the error inherent in [spatial] data deserves closer attention and public understanding ... because error provides a critical component in judging fitness for use.*  
(Chrisman 1991).

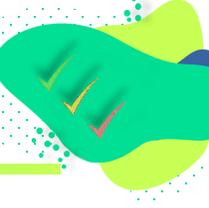


**PRIMARY SPECIES AND SPECIES-OCCURRENCE DATA**

**Arthur D. Chapman'**

*Error qui non resistitur, approbatur*  
An error not resisted is approved.  
(Ref. Doct. & Stud. c. 770).

**Keywords:**

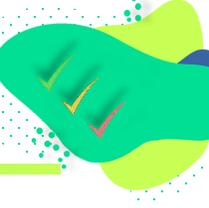


## ¿Qué es la calidad de datos?

**Los datos son de baja calidad si no pueden ser utilizados**

**y...**

**“La calidad de los datos no puede ser evaluada independientemente de los usuarios y los usos.”**



## ¿Qué es la calidad de datos?

**Los datos son de baja calidad si no pueden ser utilizados  
y...**

**“La calidad de los datos no puede ser evaluada  
independientemente de los usuarios y los usos.”**

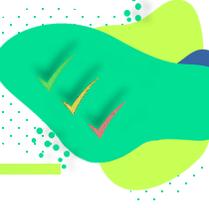
**Aptitud para el uso**  
***“Fitness for use”***



## ¿Qué es la calidad de datos?

**Aptitud depende de que los datos sean:**

- accesibles
- oportunos
- fáciles de leer
- fáciles de interpretar
- relevantes
- consistentes
- completos
- exactos
- detallados
- exhaustivos



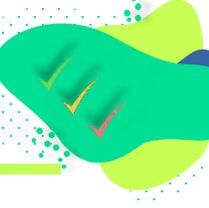
**¿Qué es la calidad de datos?**

**Dimensiones**

**Problemas**

**Soluciones**

**Principios**



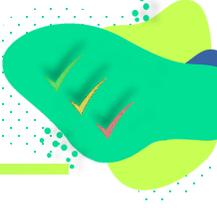
## Dimensiones

- ❖ **Compleitud**
- ❖ **Consistencia**
- ❖ **Exactitud**
- ❖ **Precisión**
- ❖ **Credibilidad**



## Problemas

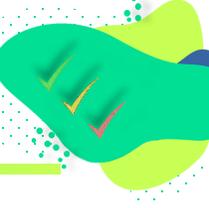
- ❖ **Redundancia:** sin estándares, sinónimos.
- ❖ **Valor faltante:** valor ausente.
- ❖ **Valor incorrecto:** valor no representa los hechos.
- ❖ **Valor no atomizado:** más de un valor en un campo atomizado.
- ❖ **Esquizofrenia:** campo utilizado con un fin incorrecto.
- ❖ **Registros duplicados:** más de un registro representando el mismo hecho.
- ❖ **Valores inconsistentes:** incongruencia en la representación de un hecho.
- ❖ **Contaminación:** creación de nuevos datos a partir de datos existentes incorrectos.



## Problemas

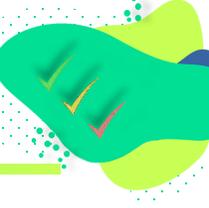
## Ejemplos

- ❖ **Redundancia:** “Capital Federal” / “Ciudad Autónoma de Buenos Aires”
- ❖ **Valor faltante:** ausencia de latitud o longitud
- ❖ **Valor incorrecto:** “Cyclocephalla” x “Cyclocephala”
- ❖ **Valor no atomizado:** “Diamante, Entre Ríos”
- ❖ **Esquizofrenia:** sex = “juvenil”
- ❖ **Registros duplicados:** dos registros idénticos
- ❖ **Valores inconsistentes:** country= “Chile” scientificName = “Daubentonia robusta”
- ❖ **Contaminación:** geografía superior completada a partir de municipio incorrecto



## Soluciones

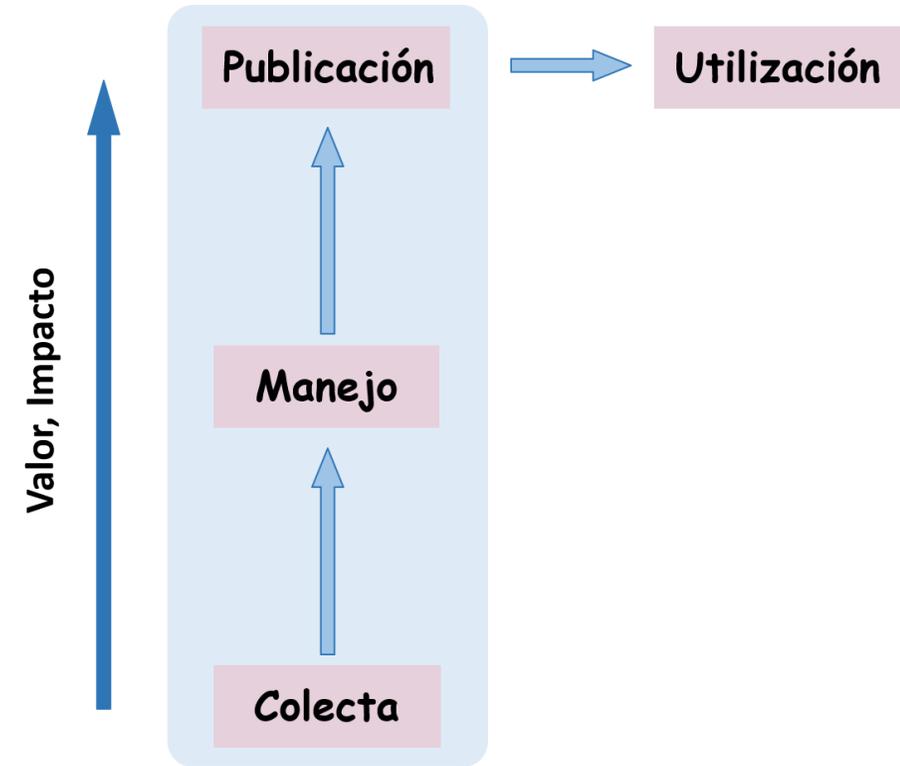
- ❖ **Prevención**
- ❖ **Detección y corrección**
- ❖ **Detección y recomendación**

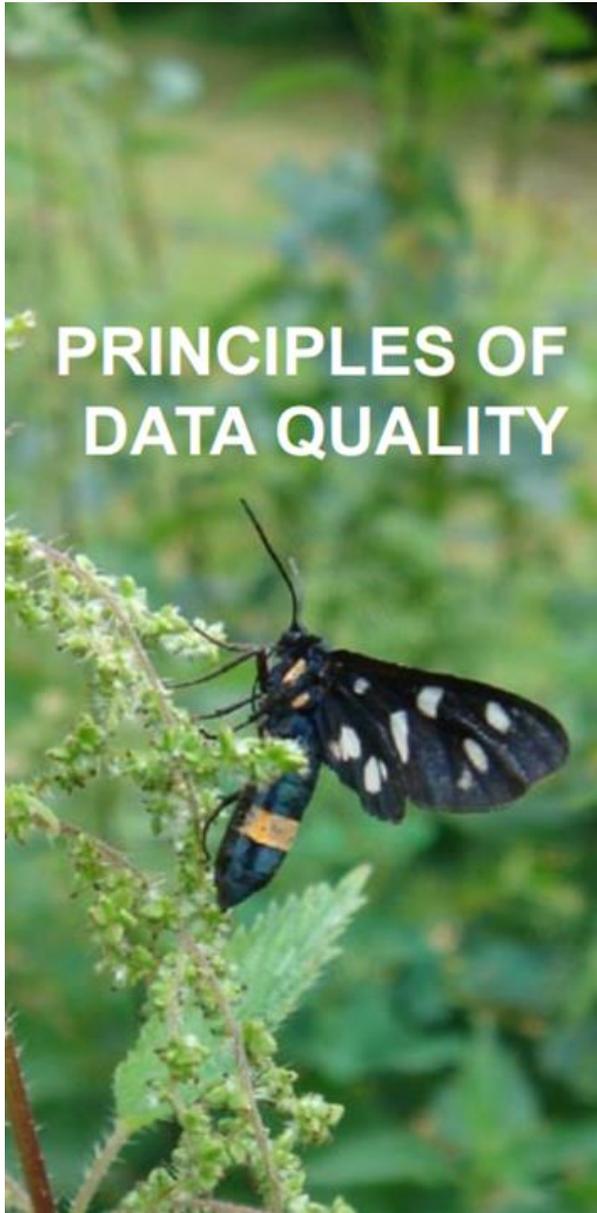
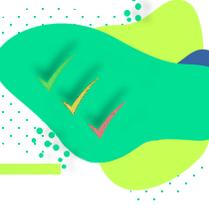


## Soluciones

- ❖ **Prevención**
- ❖ **Detección y corrección**
- ❖ **Detección y recomendación**

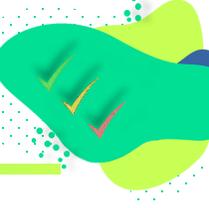
Pueden actuar a diferentes niveles





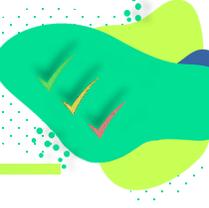
## Principios

- ❖ **Planeamiento**
- ❖ **Organización**
- ❖ **Prevención**
- ❖ **Responsabilidad**
- ❖ **Asociación**
- ❖ **Priorización**
- ❖ **Medidas de desempeño**
- ❖ **Optimización**
- ❖ **Retroalimentación**
- ❖ **Capacitación**
- ❖ **Transparencia**
- ❖ **Documentación**



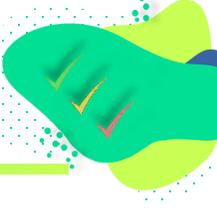
## Principios

- No podemos ir atrás en el tiempo...  
el potencial nunca será tan alto  
como en el punto de origen.



## Principios

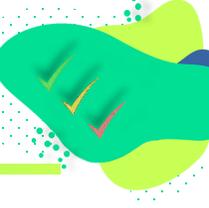
**Capturar tanto y tan inequívocamente  
como sea posible.**



## Principios

**Capturar tanto y tan inequívocamente  
como sea posible.**

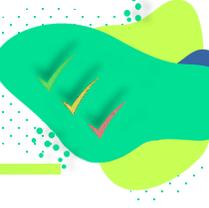
- La falta de conocimiento puede resultar en potenciales errores.



## Principios

**Capturar tanto y tan inequívocamente como sea posible.**

**Capacitar al personal que colecta y manipula datos.**



## Principios

**Capturar tanto y tan inequívocamente como sea posible.**

**Capacitar al personal que colecta y manipula datos.**

- Toda transformación de los datos puede potencialmente resultar en una pérdida neta de información.



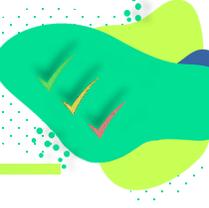
## Principios

**Capturar tanto y tan inequívocamente como sea posible.**

**Capacitar al personal que colecta y manipula datos.**

**Mantener todos los datos originales.**

**Documentar todos los cambios hechos.**



## Principios

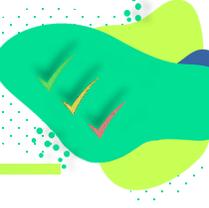
**Capturar tanto y tan inequívocamente como sea posible.**

**Capacitar al personal que colecta y manipula datos.**

**Mantener todos los datos originales.**

**Documentar todos los cambios hechos.**

- El uso de los datos aporta al control de calidad.



## Principios

**Capturar tanto y tan inequívocamente como sea posible.**

**Capacitar al personal que colecta y manipula datos.**

**Mantener todos los datos originales.**

**Documentar todos los cambios hechos.**

**Asegurar que los datos y la documentación sean accesibles.**

**Utilizar mecanismos de retroalimentación desde los usuarios.**



## Ejemplo final: ¿cómo usar este registro?

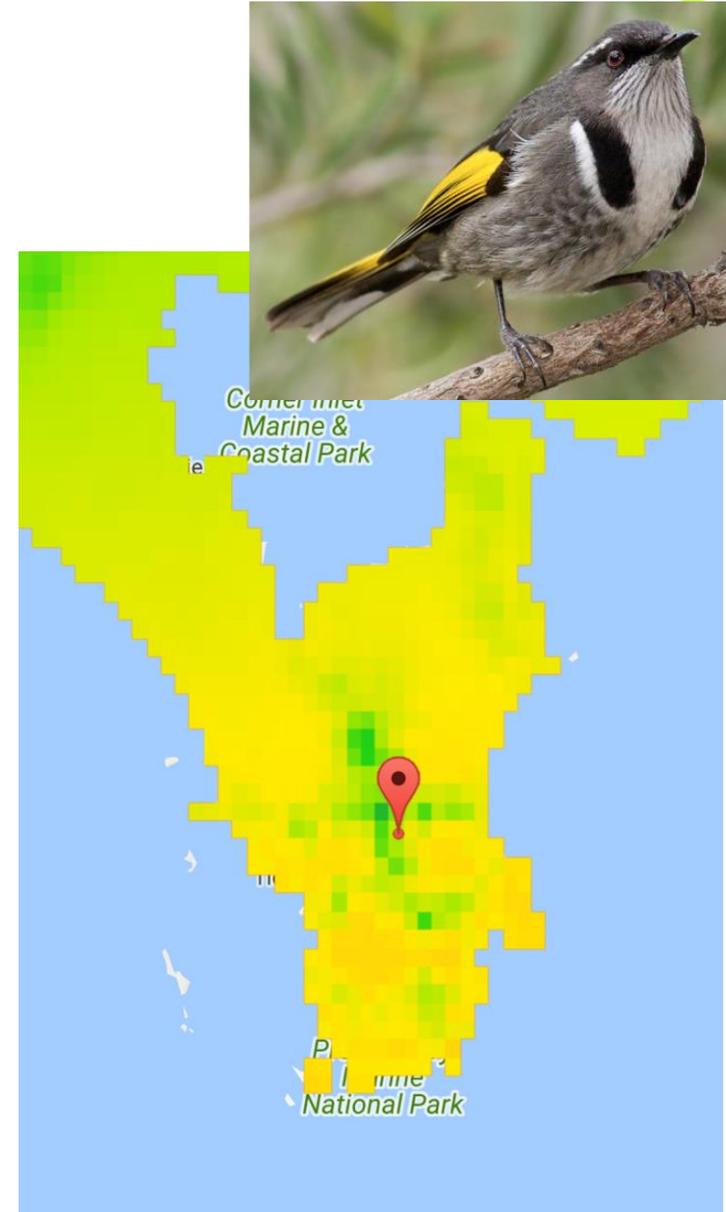


-38.99373, 146.37367

Elevación: 302m

Temp., mes más frío: 6.1°C

Precip., media anual: 107mm





## Ejemplo final: ¿cómo usar este registro?

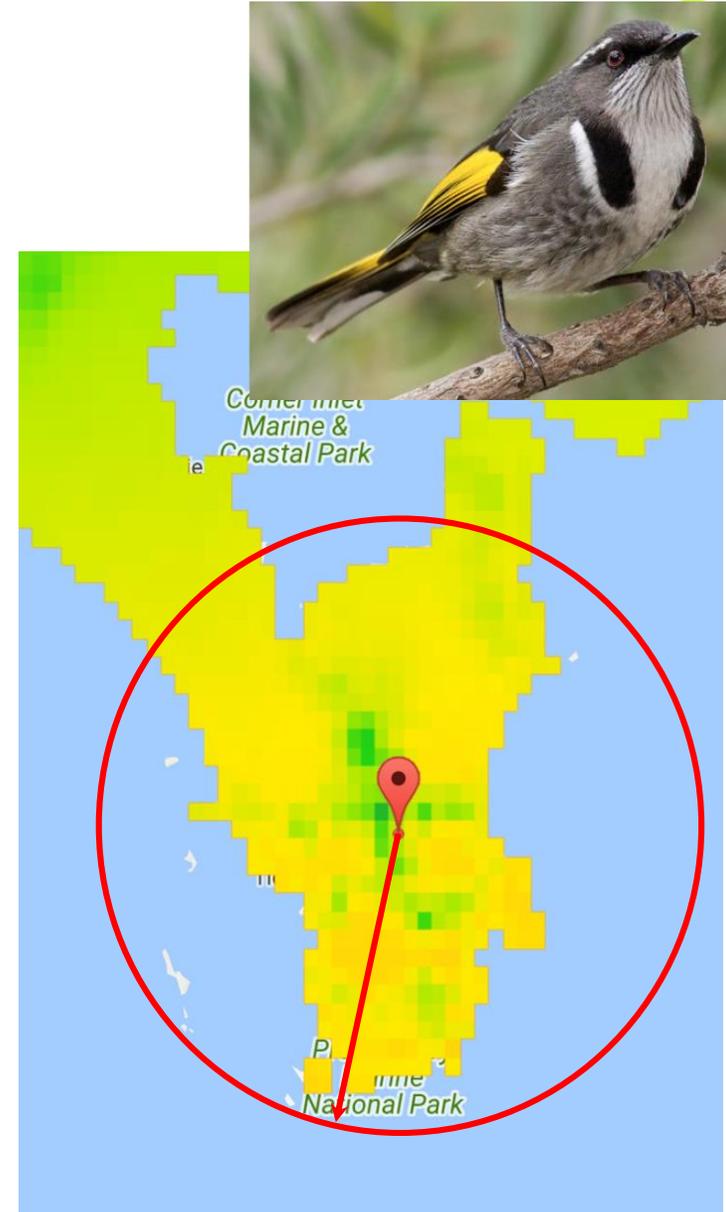


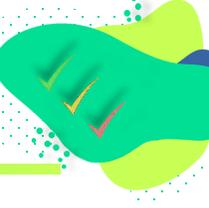
-38.99373, 146.37367  
(Wilson's Promontory)  
coordinateUncertainty: 14570m

Elevación: 0-544m (302m)

Temp., mes más frío: 6.1-9.1°C (6.1°C)

Precip., media anual: 65-107mm (107mm)





**La calidad de los datos determina su aptitud para usos particulares.**

**La calidad de los datos condiciona la calidad de la investigación y la toma de decisiones.**

# Gbif.es

¡Gracias!

pzermoglio@gmail.com

Gbif.es

VerNe



Universidad  
de Navarra