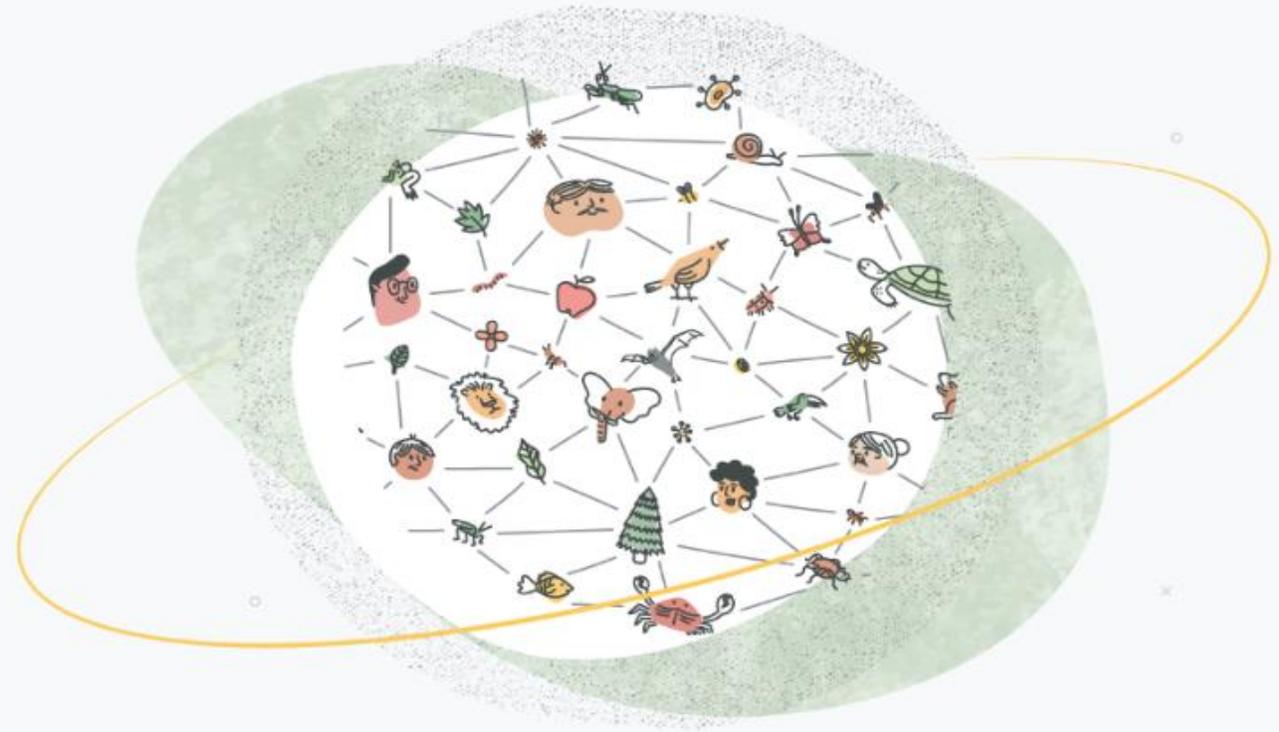


Mejora de la calidad de datos
de biodiversidad:
herramientas para el control
de calidad de los datos,
desde la captura hasta su
publicación y su uso en el
contexto de GBIF

Octubre 2024

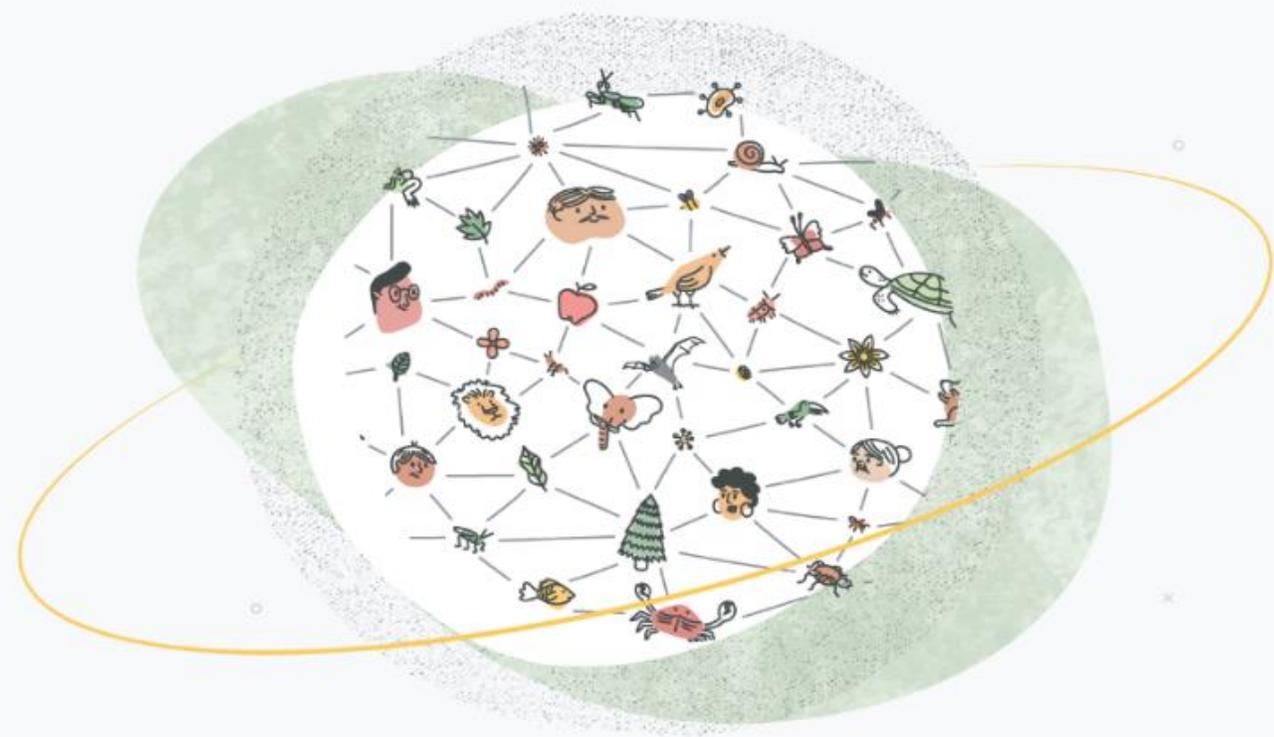


Introducción a la calidad de datos de biodiversidad

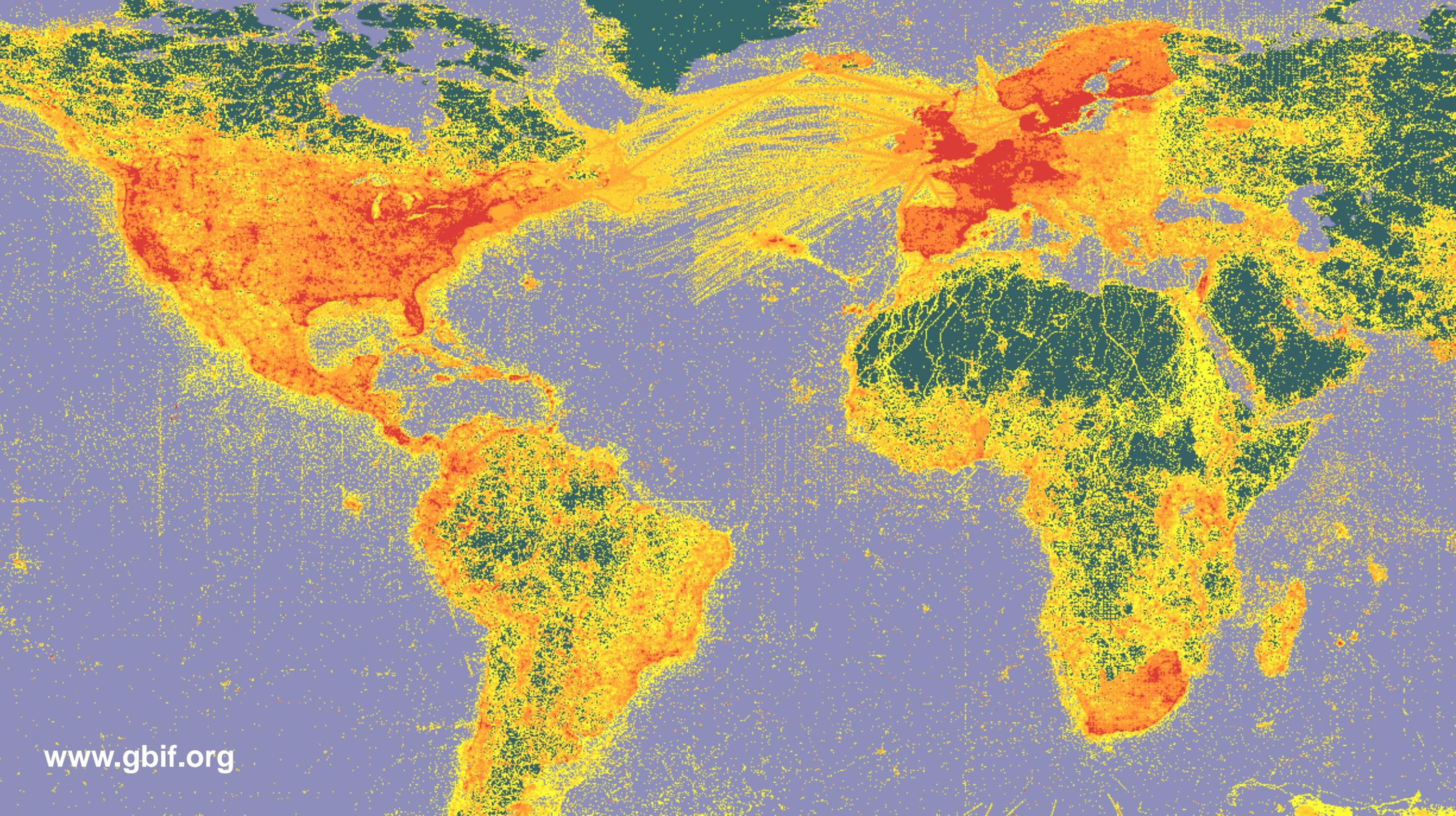
Conceptos y principios generales

I Taller GBIF.ES: Mejora de la calidad de datos de biodiversidad: herramientas para el control de calidad de los datos, desde la captura hasta su publicación y su uso en el contexto de GBIF

Octubre 2024

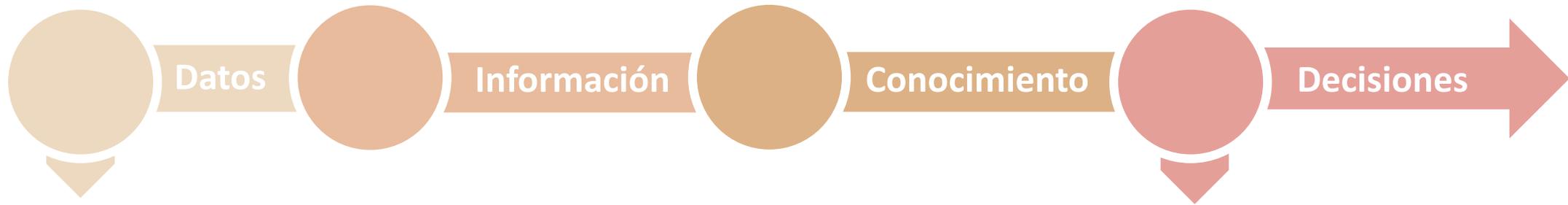


Katia Cezón
katia@gbif.es



Datos primarios de biodiversidad

Fuente de datos como base para el conocimiento

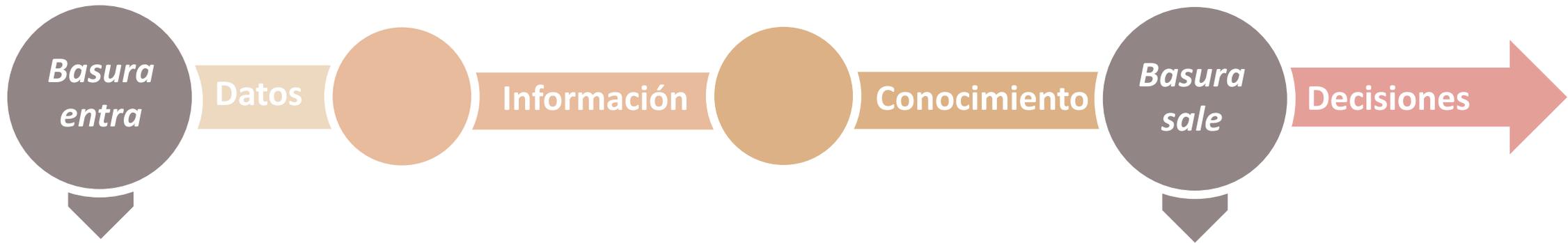


- ✓ Datos digitalizados
- ✓ Localizables
- ✓ Formato accesible
- ? Datos de calidad

Necesitamos un universo de datos lo suficientemente representativo, accesible y de calidad para generar conocimiento y tomar decisiones bien fundamentadas

Datos primarios de biodiversidad

Premisa: Basura entra – Basura sale



Los *problemas de calidad*

- Dominio taxonómico
- Dominio geoespacial
- Dominio temporal

Impactos

- Investigación científica
- Decisiones de conservación
- Políticas públicas

Impacto de la calidad de los datos

SCIENTIFIC REPORTS

Article | [Open access](#) | Published: 14 April 2020

Identification errors in camera-trap studies result in systematic population overestimation

[Örjan Johansson](#) , [Gustaf Samelius](#), [Ewa Wikberg](#), [Guillaume Chapron](#), [Charudutt Mishra](#) & [Matthew Low](#)

[10.1038/s41598-020-63367-z](https://doi.org/10.1038/s41598-020-63367-z)

Population Ecology

ORIGINAL ARTICLE |  Full Access

Sampling bias in snow leopard population estimation studies

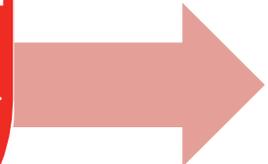
Kulbhushansingh R. Suryawanshi , Munib Khanyari, Koustubh Sharma, Purevjav Lkhagvajav, Charudutt Mishra

First published: 10 February 2019 | <https://doi.org/10.1002/1438-390X.1027> | Citations: 36

[10.1002/1438-390x.1027](https://doi.org/10.1002/1438-390x.1027)

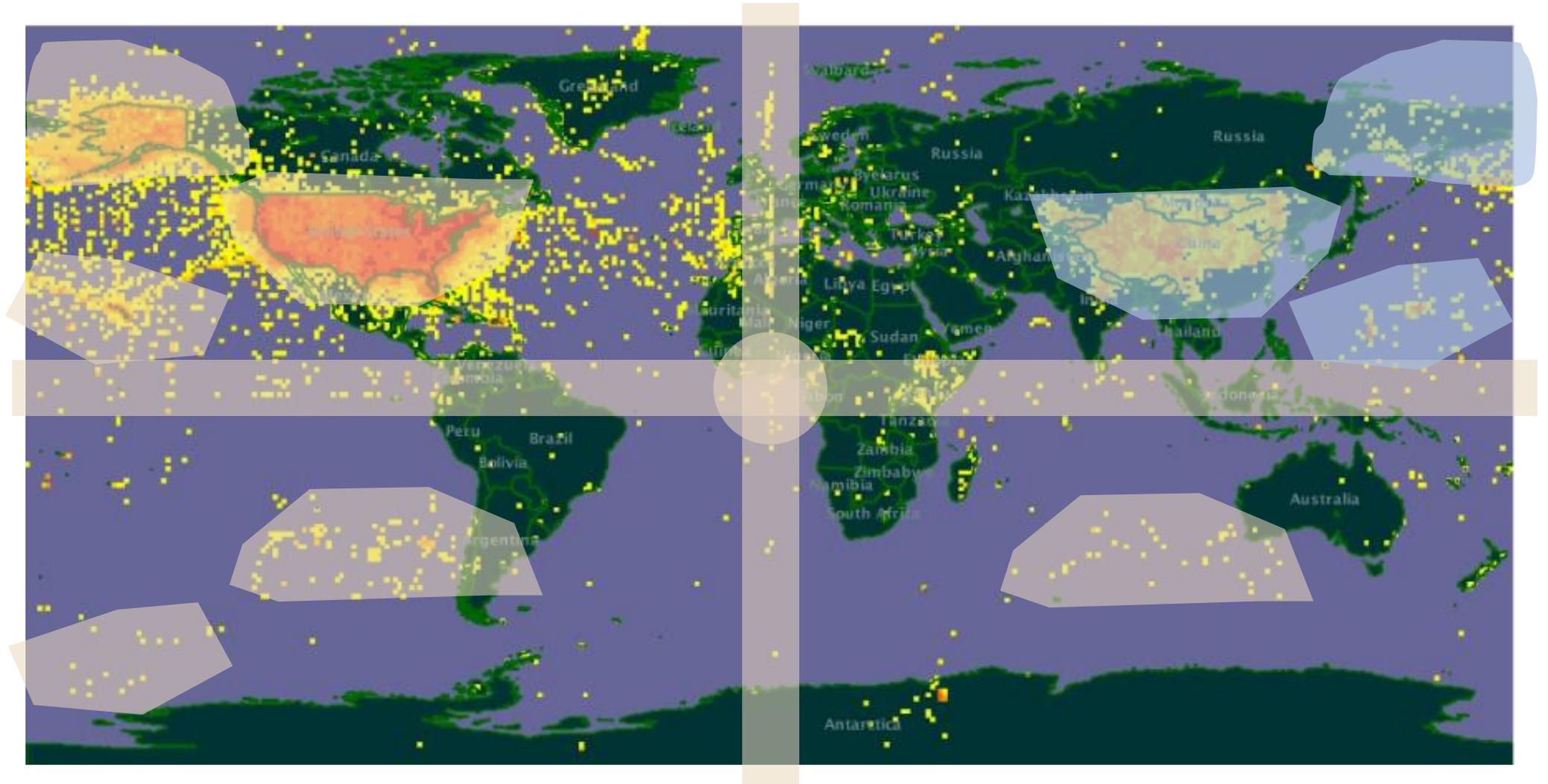


¡Dramatización!



Impacto de la calidad de los datos

- Long = 0
- Lat = 0
- Long y Lat = 0
- Long
- Lat
- Long y Lat



Datos de presencia Estados Unidos GBIF (2013)

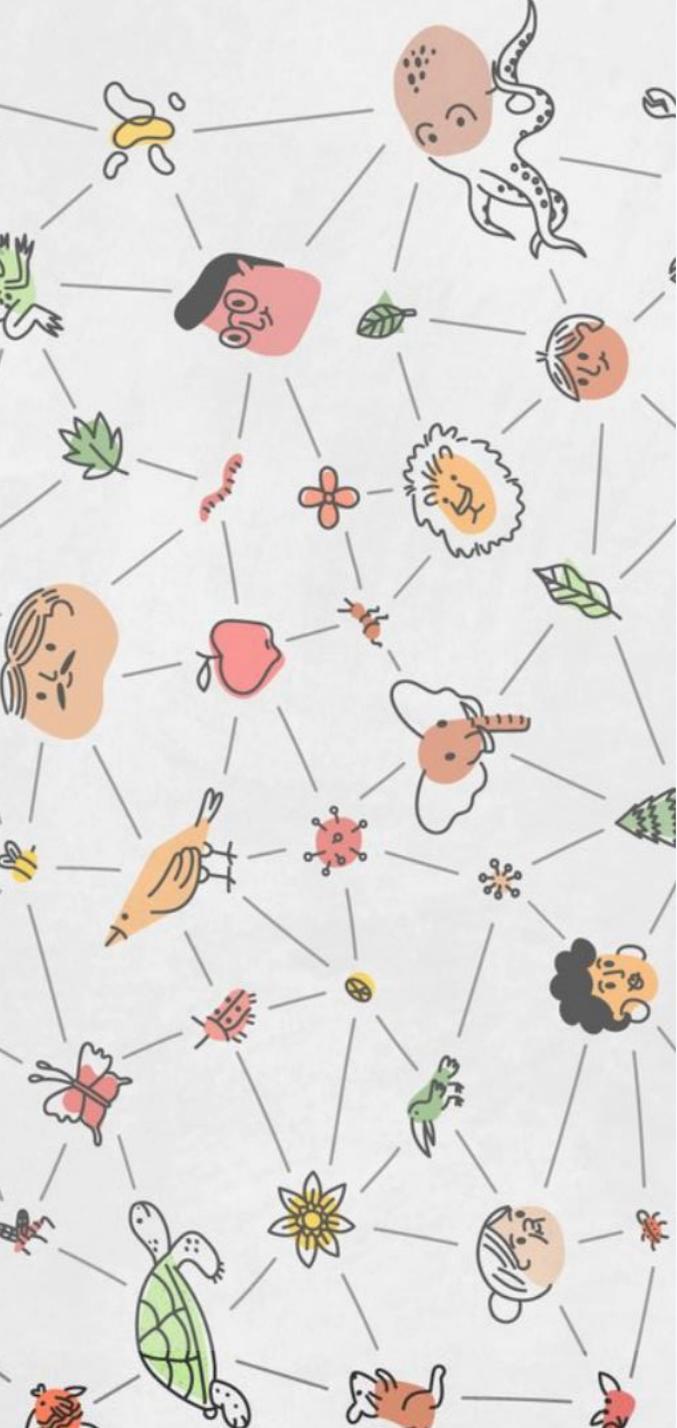
Tim Hirsch

Impacto de la calidad de los datos



Datos de presencia Estados Unidos GBIF (2013)

Tim Hirsch

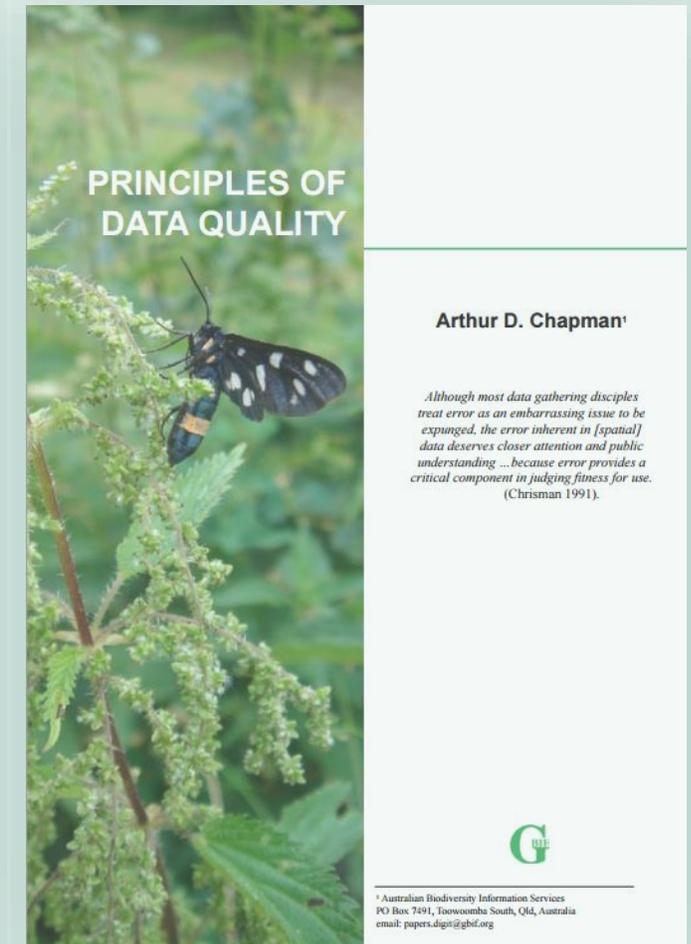
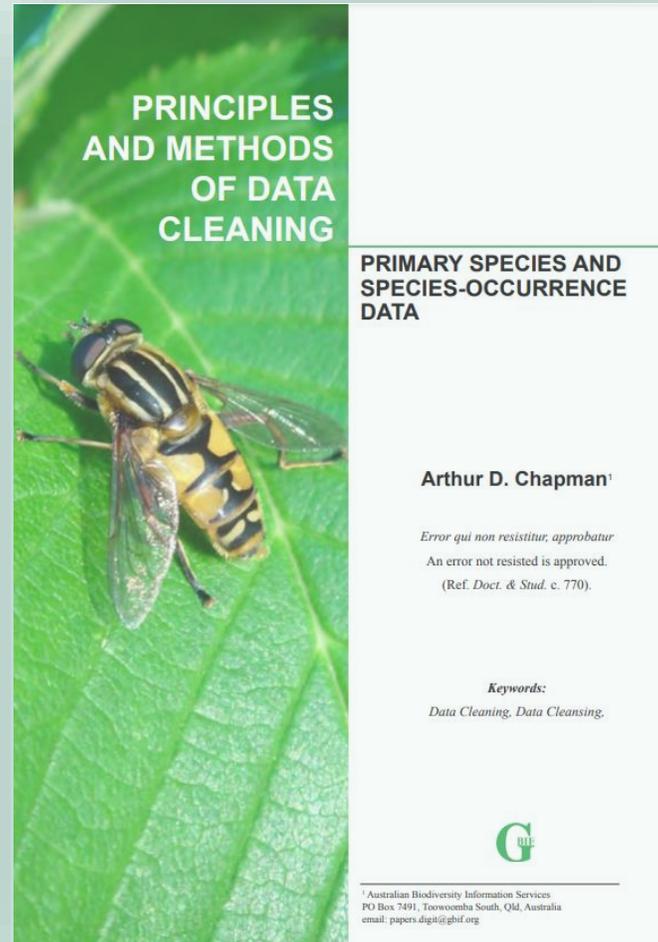


Calidad de datos
Algunos conceptos

Calidad de los datos

Obras de referencia

- Chapman, A. D. 2005. **Principles of Data Quality**. Report for the GBIF, Copenhagen. Disponible [aquí](#).
- Chapman, A. D. 2005. **Principles and Methods of Data Cleaning – Primary Species and Species Occurrence Data**. Report for GBIF, Copenhagen. Disponible [aquí](#).



Calidad de los datos

¿Qué es la calidad de datos?

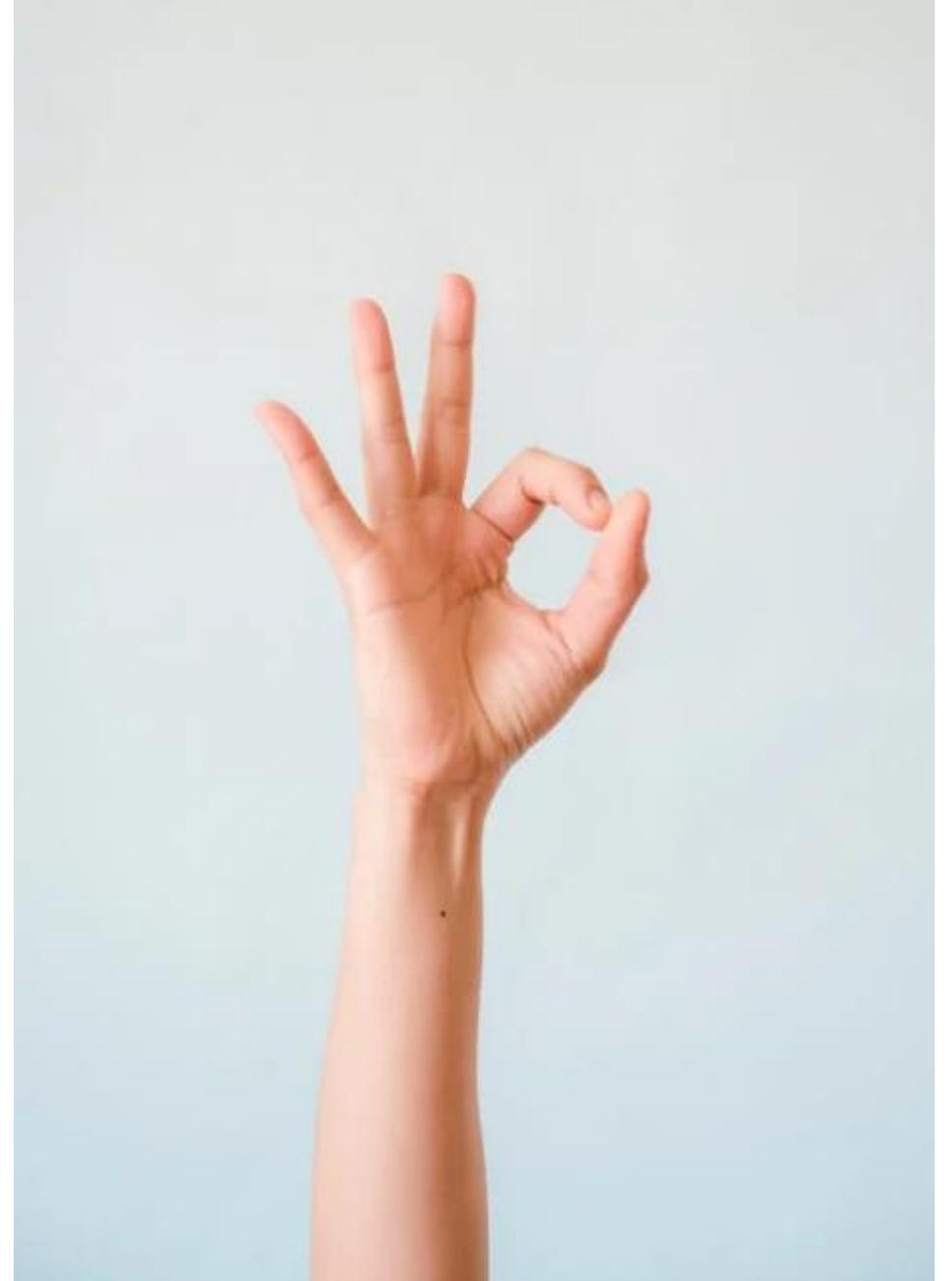
Fitness for use = aptitud para el uso

La calidad de los datos está relacionada con el **uso** y no puede evaluarse independientemente del **usuario**.

La calidad está dada por el **propósito**.

Un dato tiene calidad si es adecuado para ser usado.

Si el dato no sirve para un propósito del que lo usa, puede ser útil para otros.



Calidad de los datos

¿Qué es la calidad de datos?

Dato: mapa de la **presencia** de *Acacia acuminata* en el continente australiano.



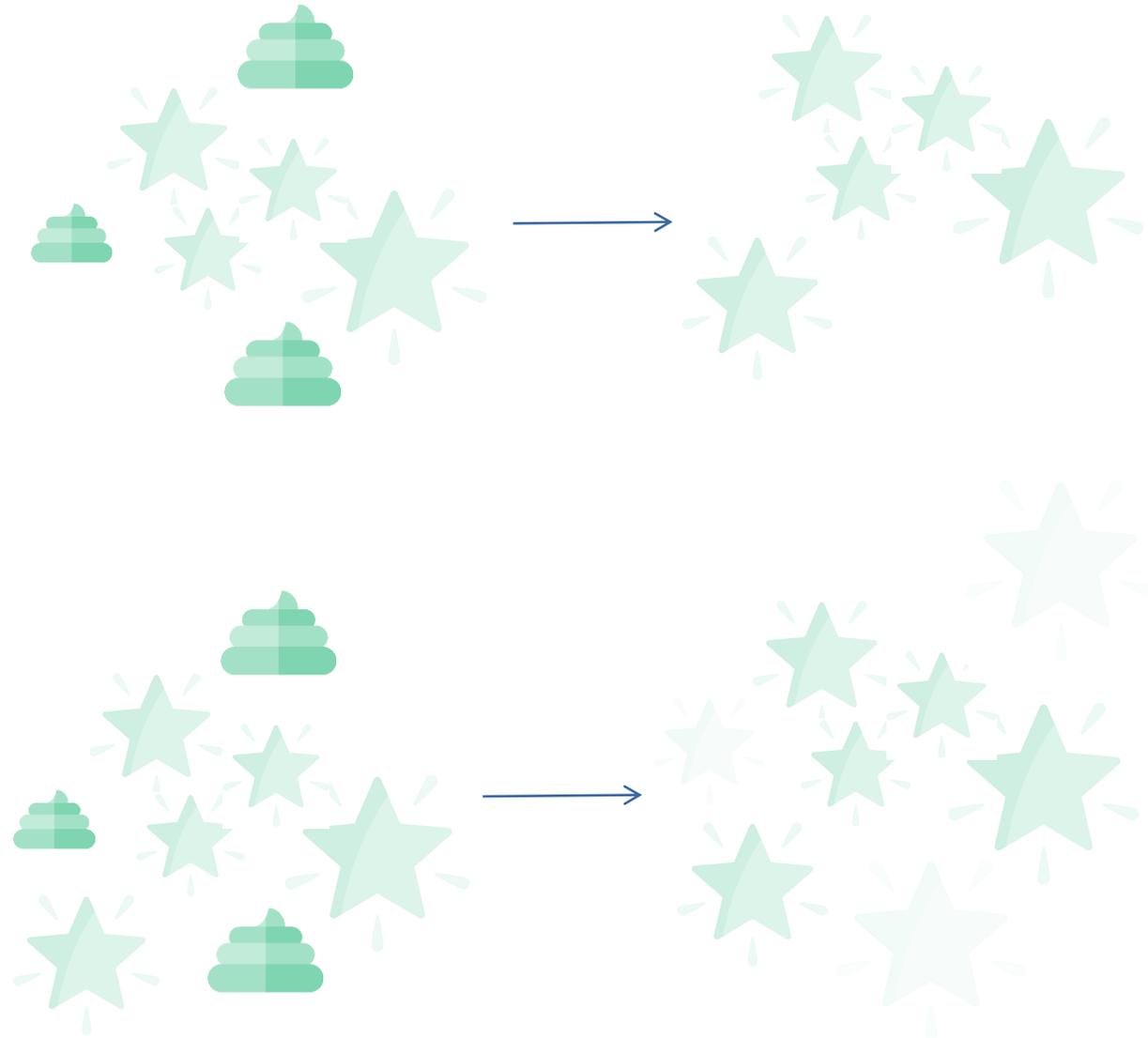
¿Es este un dato de calidad?

- ¿*Acacia acuminata* se encuentra en Australia?
- ¿*Acacia acuminata* se encuentra en el área protegida Ngaanyatjara?

Calidad de los datos

¿Quién decide si un dato es de calidad?

El **usuario** es quien decide si un dato es de calidad o no



Calidad de los datos

¿Qué es la calidad de datos?

La aptitud depende de que los datos sean:

- Exactos
- Precisos
- Consistentes
- Completos
- Exhaustivos
- Accesibles
- Oportunos
- Fáciles de leer
- Fáciles de interpretar

Tipos de errores

Errores técnicos

1

Omisión o completitud

Información está ausente o incompleta

“September” “0” “_”

4

Redundancia

Mismo datos capturado más de una vez

“17 ? 1963” “17 00 1963”

7

Congruencia

Datos imposibles

“04/10/2030”

2

Tipográficos

Errores tipográficos y ortográficos

“19o3”

5

Convención

Datos sin utilizar convenciones

“17 ? 1963” “17 00 1963”

3

Contexto

Información que no corresponde al campo

“2050 m”

6

Uniformidad

Misma información escrita de forma diferente

“10-7-1992” “12/10/1993”

Tipos de errores

Errores de consistencia

1

Taxonómicos

¿Si se realizó una identificación correcta?

2

Continuidad

¿Existe una línea temporal clara de las fechas de recolección, identificación, digitalización y actualización consistentes?

3

Geográficos

¿Están las coordenadas dentro de la localidad o región identificada?

4

Tendencias de recolecta

¿El detalle del registro coincide con los patrones de colecta de la organización o del colector?

5

Precisión y exactitud

¿Son los datos preciosos y exactos?

6

Métodos de colecta

¿Son los registros consistentes con el método de colecta?

Tipos de errores

Errores técnicos

Son relativamente simples, usualmente su corrección se puede automatizar.

Validación

Incluye procesos para controlar el formato de los datos, el chequeo de su integridad, la verificación de su consistencia y del cumplimiento de las normas del estándar.

Errores de consistencia

Se detectan al aplicar reglas del mundo real a los datos.

Verificación

Evalúa la exactitud de los datos, cómo se ajustan estos a la realidad. Usualmente requieren un especialista para identificarlos, y en ocasiones, es necesario comparar con datos de referencia.

¿Dónde surgen los problemas?

Cadena de la información



“Ya que el error es algo ineludible, debería ser reconocido como una dimensión fundamental del dato. Y necesita ser detectado, registrado y documentado.” [Chrisman, 1991](#).

Mecanismos de mejora

1

Prevención

Prevenir es mejor que curar (y es mucho más barato).

2

Detección y limpieza

Identificar errores en el conjunto de datos y corregirlos.

3

Documentación y recomendaciones

Detectar errores en el conjunto de datos y generar recomendaciones de limpieza.

4

Mecanismos de retroalimentación

Informar de los errores a los actores implicados en los pasos previos.

5

Capacitación

Del personal que colecta y manipula datos



¡Gracias!

katia@gbif.es

www.gbif.es

