

TALLER GBIF.ES

CALIDAD EN BASES DE DATOS DE BIODIVERSIDAD

Madrid, 2-4 Diciembre 2015



Katia Cezón
Unidad de Coordinación – GBIF.ES
Nodo Nacional de Información en
Biodiversidad en España

PRINCIPIOS GENERALES SOBRE LA CALIDAD DE DATOS

Madrid, 2-4 Diciembre 2015



Katia Cezón
Unidad de Coordinación – GBIF.ES
Nodo Nacional de Información en
Biodiversidad en España

PRINCIPIOS SOBRE LA CALIDAD DE DATOS

OBRA DE REFERENCIA



Chapman, A. D. 2005. Principles of Data Quality, version 1.0. Report for the Global Biodiversity Information Facility, Copenhagen. ISBN 87-92020-03-8. Available online at http://www.gbif.org/orc/?doc_id=1229.



Arthur D. Chapman¹

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).



¹ Australian Biodiversity Information Services
PO Box 7491, Toowoomba South, Qld, Australia
email: papers.digit@gbif.org

PRINCIPIOS SOBRE LA CALIDAD DE DATOS

PREMISAS



Prevenir es mejor que curar y es mucho más barato

La **prevención de errores** nada tiene que hacer con los datos que ya existen en la base de datos. En estos casos, la **validación y la corrección** serán muy importantes en el proceso hacia la calidad.



Detectar las causas del error nos ayudará a prevenirlas

Corregir los datos y no hacer nada para **prevenir los errores** significa que los errores seguirán apareciendo sistemáticamente y no los reduciremos nunca.

PRINCIPIOS SOBRE LA CALIDAD DE DATOS

- 1 **Exactitud.** Debemos tender a que el dato esté lo más próximo posible al valor real.
- 2 **Consistencia.** Debemos establecerse **reglas** y **convenciones** para garantizar que los datos sean lo más consistentes posible: presentados siempre de la misma manera y que se mantienen en el tiempo y sin ambigüedad.

GENUS	SPECIES	INFRASPECIES
Eucalyptus	globulus	subsp. bicostata

GENUS	SPECIES	INFRASP_RANK	INFRASPECIES
Eucalyptus	globulus	subsp.	bicostata

PRINCIPIOS SOBRE LA CALIDAD DE DATOS

- 3 Conservación de datos originales.** Es importante no perder los valores originales durante el proceso de limpieza o depuración.
- 4 Depuración y validación de datos.** Detección de los datos que no sean correctos ni consistentes.
- 5 Accesibilidad.** Con mecanismos de acceso claros para que cualquier persona pueda realizar sus análisis.
- 6 Documentación y transparencia.** Hacer públicos los procedimientos y la documentación para el manejo de la colección, los análisis realizados, los informes y las actualizaciones.
- 7 Actualización.** Se refiere a la frecuencia de actualización del conjunto de los datos de la colección.

PRINCIPIOS SOBRE LA CALIDAD DE DATOS

CONSIDERACIONES RELEVANTES

- **Formación y entrenamiento** del personal - Seleccionar al personal adecuado para cada una de las labores en la conformación, incremento y depuración de la base de datos.
- Crear **criterios de captura** (convenciones, estándares)
- **Verificar y corregir** los datos, control interno (quincenal, mensual)
- **Almacenamiento seguro** de archivos (backup)
- Producir máximos resultados mínimos recursos (**eficiencia**) alcanzar objetivos planteados (**eficacia**)
- **Feedback** - Mecanismo de retroalimentación a través del que los usuarios informan a cerca de errores, y hacen que esta información se refleje en la calidad de los datos.



Gracias