

# PRINCIPIOS GENERALES SOBRE LA CALIDAD DE DATOS

Cristina Villaverde – GBIF España

Katia Cezón - GBIF España

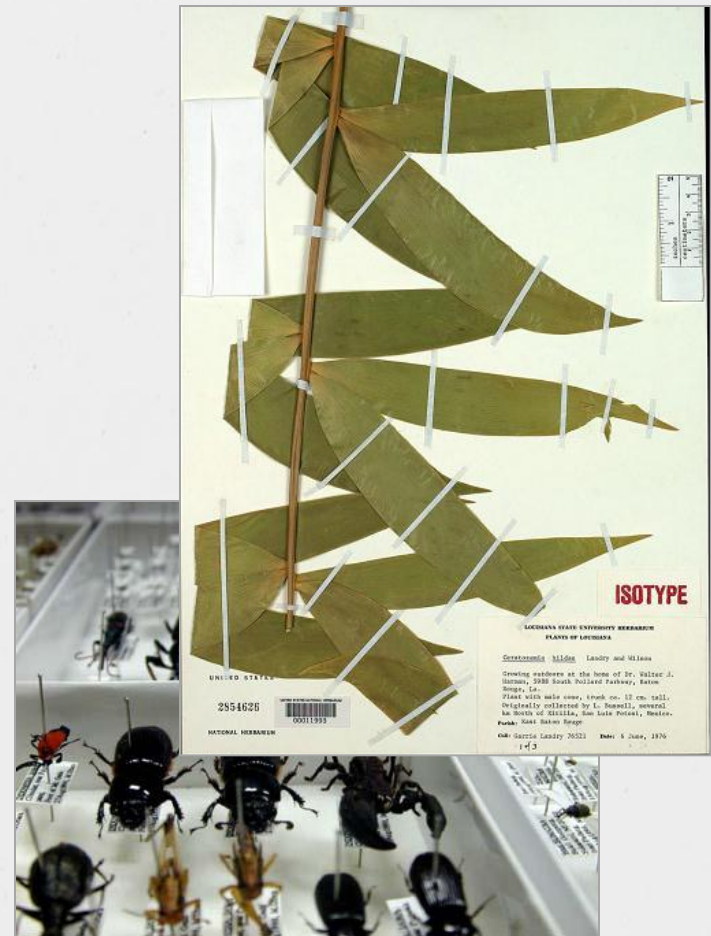
V Jornadas sobre Información de Biodiversidad y Administraciones Ambientales  
Cáceres, Noviembre 2013



CHAPMAN, A.D. 2005. Principles of Data Quality, version 1.0.  
Report for the GBIF, Copenhagen.

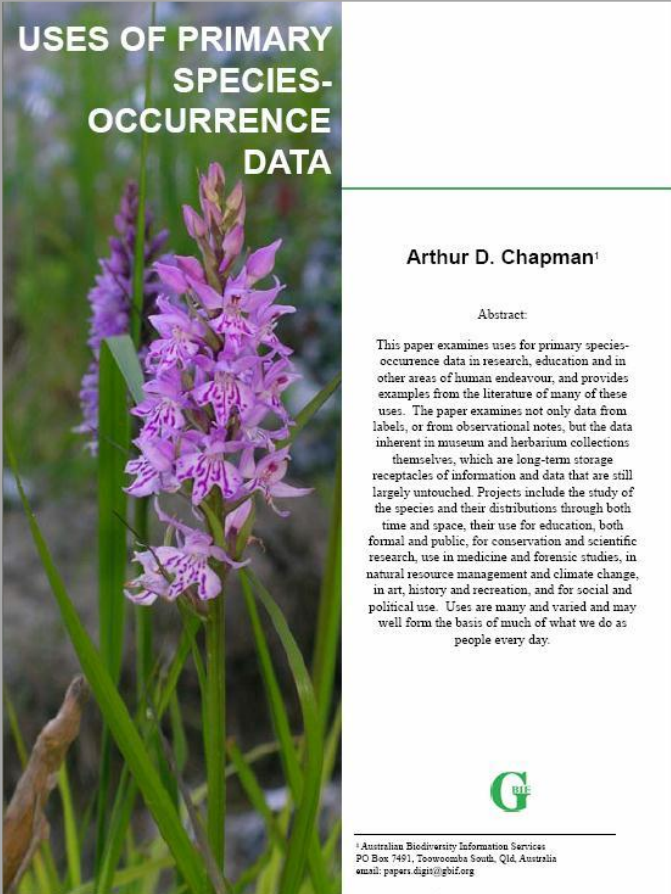
# Usando los datos sobre ejemplares

- Estudios taxonómicos, ecológicos, biogeográficos, filogenéticos.
- Estudios de población y distribución de especies.
- Estudios sobre especies amenazadas.
- Sobre migración de especies.
- Planificación sobre la conservación de espacios protegidos.
- Gestión de recursos naturales.
- Modelado de datos de especies.
- Impacto del cambio climático.



# Usando los datos sobre ejemplares

- Agricultura, Montes y Pesca
- Perspectivas basadas en productos biológicos.
- Salud y seguridad públicas.
- Medicina forense.
- Ecoturismo.
- Arte e Historia, Ciencias y política
- Planificación de infraestructuras humanas.
- Etc.




**USES OF PRIMARY SPECIES-OCCURRENCE DATA**

Arthur D. Chapman<sup>1</sup>

Abstract

This paper examines uses for primary species-occurrence data in research, education and in other areas of human endeavour, and provides examples from the literature of many of these uses. The paper examines not only data from labels, or from observational notes, but the data inherent in museum and herbarium collections themselves, which are long-term storage receptacles of information and data that are still largely untouched. Projects include the study of the species and their distributions through both time and space, their use for education, both formal and public, for conservation and scientific research, use in medicine and forensic studies, in natural resource management and climate change, in art, history and recreation, and for social and political use. Uses are many and varied and may well form the basis of much of what we do as people every day.



<sup>1</sup> Australian Biodiversity Information Services  
PO Box 7491, Toowoomba South, Qld, Australia  
email: papers.digit@gibf.org

# ¿Qué es la calidad de datos?

*Una característica esencial y necesaria para que los datos sean "adecuados para el uso".*

El propósito general al describir la calidad de los datos de un registro concreto es describir la adecuación del registro para un uso particular que el usuario tenga en mente para dichos datos.

Chrisman, 1991

# ¿Qué es la calidad de datos?

¿Este dato es de buena calidad?:



- ¿La especie 'A' se encuentra en Tasmania?
- ¿La especie 'A' se encuentra en el Área de conservación del patrimonio de Tasmania'?

# CADENA DE INFORMACIÓN Y PÉRDIDA DE LA CALIDAD

# Cadena de información de la calidad de datos

COSTE DE LA CORRECCIÓN DE ERRORES





# PRINCIPIOS DE LA CALIDAD

# Principios de la calidad

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 de la calidad

**EXACTITUD** – Debemos tender a que el dato esté lo más próximo posible al valor real.

## CONSISTENCIA

Datos presentados siempre de la misma manera y se mantienen en el tiempo de forma clara, consistente y sin ambigüedad:

- **Consistencia semántica:** la información que se almacena
- **Consistencia estructural:** la forma en que se almacena

| Genus      | Species  | Infraspecies     |
|------------|----------|------------------|
| Eucalyptus | globulus | subsp. bicostata |
| Eucalyptus | globulus | bicostata        |

**Table 1.** Showing semantic inconsistency in the *Infraspecies* field.

| Genus      | Species  | Infrasp_rank | Infraspecies |
|------------|----------|--------------|--------------|
| Eucalyptus | globulus | subsp.       | bicostata    |
| Eucalyptus | globulus |              | bicostata    |

**Table 2.** Showing semantic consistency in the *Infraspecies* field by addition of a second (“*Infrasp\_rank*”) field.

# Principios de la calidad

**DEPURACIÓN** – Detección y exclusión de los datos que no sean correctos ni consistentes.

**EFFECTIVIDAD** – La probabilidad de que una tarea alcance los objetivos deseados.

- Ej: el porcentaje de registros para los cuales la latitud y la longitud pueden ser determinados exactamente.

**EFICIENCIA** – Producir los máximos resultados con los mínimos recursos.

- Ej: optimizar los procesos de georreferenciación ordenando por localidad y georreferenciándolos utilizando los mismos mapas para este conjunto de registros.

# Principios de la calidad

**ACCESIBILIDAD** – cómo de accesibles son los resultados para los usuarios/el público.

- Ej: la facilidad con la que los usuarios acceden a la georreferenciación de una localidad particular que acaba de ser georreferenciada.

**TRANSPARENCIA** – hacer públicos los procedimientos y la documentación para el manejo de la colección, los análisis realizados, los informes y las actualizaciones.

- Ej.: conocer los métodos con que han sido georreferenciados un grupo de registros y disponer de la documentación asociada a esta tarea.

# Principios de la calidad

ACTUALIDAD DE LOS DATOS – Se refiere a la frecuencia de actualización del conjunto de los datos de la colección.

- ¿Cuándo fueron los datos actualizados por última vez?
- ¿Con qué frecuencia se actualizan y son puestos a disposición de los usuarios?

La frecuencia de actualización deben ser concretada y documentada.

# También debemos tener en cuenta...

Otras características que deben ser observadas son:

- **Documentación:** es un principio clave. Permite a los usuarios verificar si los datos se ajustan al uso que necesitan en ese momento.
- **Feedback:** mecanismo de retroalimentación a través del cual los usuarios informan a cerca de errores, y hacen que esta información se refleje en la calidad de los datos.
- **Formación y entrenamiento del personal:** debe incluir desde los colectores, hasta los operadores de digitalización de los datos y los gestores de las bases de datos.
- Crear **protocolos de actuación** que sirvan de base para la formación del personal, y para las tareas del día-a-día.

# Errores en los datos

*En general, un buen entendimiento de los errores y su propagación conduce a un control activo de la calidad.*

Burrough and McDonnell, 1998

*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