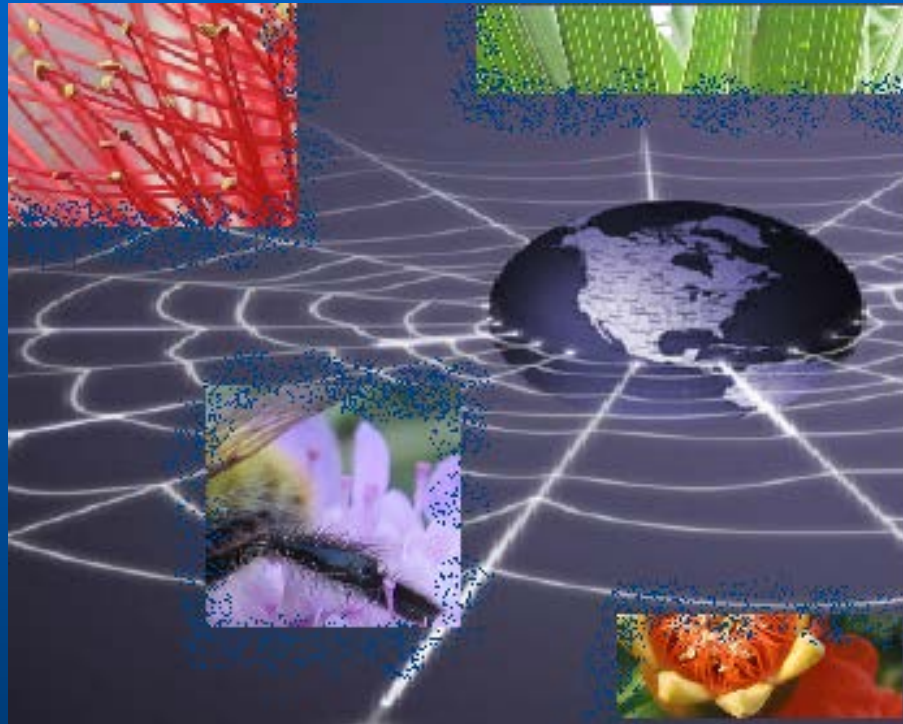


# I TALLER SOBRE GESTIÓN DE DATOS SENSIBLES



Isabel Ortega Maqueda  
ortega@gbif.es

**I TALLER SOBRE GESTIÓN DE DATOS SENSIBLES**  
Madrid, 15-16 de Octubre 2008

# Lo que vamos a tratar

1. El punto de partida de GBIF
2. Qué consideramos datos sensibles
3. Principios de compartición de datos
4. Nacional/regional frente a Global
5. En relación a los duplicados
6. Análisis correlacionales
7. Temas sociales y políticos

# El punto de partida

- En marzo de 2006, el **Secretariado de GBIF** llevó a cabo **una encuesta** entre los proveedores de datos, cuya finalidad era:
  - **Definir** qué se consideran **datos sensibles**
  - Conocer qué se está haciendo **actualmente** para gestionar la **distribución** de la información sensible.
- En mayo de 2006 se publica el informe con **los resultados** de las encuestas:

## Questionnaire on Dealing with Sensitive Data

- El 1 de Agosto de 2006, Arthur D. Chapman publica un borrador con el **análisis de los resultados**. El 6 de abril de 2007 publica **el informe definitivo**:

## Report - Dealing with Sensitive data

# El enfoque de la encuesta ...

- El foco se pone en la **distribución sin protección** de datos primarios de especies o de especímenes (por ejemplo, la localidad exacta de taxones raros, en peligro de extinción o con valor comercial)
- La protección de los datos sensibles necesita ser tratado en relación con los datos publicados para ser compartidos a través de la **red de GBIF**.
- La **generalización** de datos, aunque restringe la disponibilidad o precisión de ciertos datos, no restringe demasiado el uso que se puede dar a estos datos.
- La generalización permite que la **información fundamental** esté todavía disponible, mientras se oculta la información sensible al público general en el caso de que la publicación de esta información tenga un riesgo contra el medioambiente, especies concretas o personas vivas.

# Qué consideramos datos sensibles

## ■ Definición de dato sensible:

- Cualquier **característica o atributo** de la colección, que por su naturaleza, un proveedor de datos no desea hacer accesible en su estado original (por ejemplo: nombre del colector)
- El **taxon** en sí mismo o **una parte** de él (por ejemplo: una población en particular)
- Aspectos del **medioambiente** en el que el taxon es citado (terrenos particulares, determinados hábitats)

# Categorías de datos sensibles

La encuesta identifica dos categorías de datos sensibles:

- Datos que son “**permanentemente**” sensibles (como especies en peligro de extinción).
- Datos que son “**temporalmente**” sensibles (como los que están en espera de una publicación científica).

Las categorías de **datos sensibles temporales** incluyen:

- Datos pendientes de su publicación
- Datos pertenecientes a una investigación en curso
- Datos incompletos o sin chequear

# Datos sensibles permanentemente

- Taxones raros o amenazados, protegidos bajo legislación
- Taxones con valor comercial
- Taxones vistosos o frágiles
  - Orquídeas, cactus, cicadáceas y algunos lugares de nidificación/pernoctación de algunos animales.
- Datos sometidos a no revelación por parte de los propietarios de terrenos
- Datos usados para la obtención de ingresos
- Otros:
  - Usos culturales y tradicionales
  - Localizaciones en áreas protegidas
  - Datos bajo acuerdos con terceras partes
  - Interceptación de productos en cuarentena y registros de productos en cuarentena
  - Privacidad y derechos de Propiedad Intelectual de otros

# Categorías de sensibilidad

Los custodios de los datos necesitan **identificar** qué dato puede ser considerado “**sensible**”, es decir, todo aquel que su publicación en la **Web** pueda conllevar resultados **adversos** para el taxon o el medioambiente.

Según los **niveles de amenaza** del taxon, su **vulnerabilidad** y el **tipo** de información **asociada**, cada **institución** adoptará unas **categorías** de sensibilidad que aplicará a los registros que se van a publicar en **Internet**.

Basadas en las categorías usadas por el New South Wales Departament of Environment and Conservation (2007):

CATEGORÍA	RAZONES PARA SU INCLUSIÓN
Categoría 1 - Especies o registros cuyos datos no deben ser proporcionados en absoluto, o pueden ser georreferenciados a una gran región o cuenca de agua.	<ol style="list-style-type: none"><li>1. Una especie de <u>alto valor biológico</u> que está bajo <u>alto peligro</u> de sobre-explotación/enfermedad u otra amenaza identificable, de forma que la publicación de su localidad pueda amenazar al taxon.</li><li>2. La información del registro es de tal naturaleza que su publicación puede causar <u>daños irreparables</u> al medio ambiente ó al taxon.</li></ol>
Categoría 2 – Especies o registros cuyas coordenadas se harán públicas tras su pérdida de precisión a 0.1 grado.	<ol style="list-style-type: none"><li>1. La especie es clasificada como <u>altamente sensible</u>, y la provisión de la localidad precisa podría llevar a la especie a sobre-explotación y perturbación de su hábitat.</li><li>2. Los registros información <u>altamente sensible</u> y su publicación puede causar <u>daños</u> al medio ambiente, al taxon.</li></ol>



# Categorías de sensibilidad (cont.)

Continuación de las categorías usadas por el New South Wales Department of Environment and Conservation (2007):

<p>Categoría 3 – Especies o registros cuyas coordenadas se harán públicas tras su pérdida de precisión a 0.01 grado.</p>	<ol style="list-style-type: none"><li>1. La especie es clasificada de <u>sensibilidad media a alta</u>, y la provisión de la localidad podría llevar a la especie a explotación y perturbación.</li><li>2. Los registros incluyen información <u>sensible</u>, cuya publicación podría causar daños al individuo o al <u>medioambiente</u>.</li></ol>
<p>Categoría 4 – Especies o registros cuyas coordenadas se harán públicas tras su pérdida de precisión a 0.001 grado.</p>	<ol style="list-style-type: none"><li>1. La especie es clasificada de <u>sensibilidad baja a media</u>, y la provisión de la localidad podría llevar a la especie a recolección y daño deliberado.</li><li>2. Los registros incluyen información <u>sensible</u>, cuya publicación podría causar daños al individuo o al medioambiente</li></ol>
<p>No sensible</p>	<p>Georreferenciación sin restricciones</p>

# Principios de compartición de datos sensibles

1. Cuando sea posible, la información medioambiental deberá estar **disponible para todos**, en interés de la ciencia, del medioambiente y de la biodiversidad.
2. Todos los datos considerados sensibles deberán tener una **fecha de revisión** de dicho estatus de sensibilidad, así como las **razones por las que se incluyen** en dicho estatus.
3. Si los datos van a ser **restringidos para su distribución**, deberá hacerse **solo en la copia de los datos a distribuir**. Los datos **originales nunca** deben ser **alterados, falsificados ni borrados** de los registros de la colección.
4. La **documentación es esencial** cuando los datos han sufrido **restricciones o generalización**. Es importante que la información se almacene como un metadato dentro del propio registro.

# Principios (continuación)

5. Cuando los datos son **restingidos para su distribución** (ej.: colector, localidad, etc) deberá ser **documentado** reemplazando el campo con el **texto** correspondiente que **informe** de este hecho. El dato **no debe dejarse en blanco** ni con valor nulo.
6. Los **usuarios** de datos sensibles deberán **respetar** cualquier **restricción de acceso a los datos** que el proveedor haya establecido. Si se les concede mayor acceso a la información, los usuarios no deberán infringir la **confidencialidad** de dicha información.
7. Los **proveedores** de datos deberán **respetar** las necesidades de los **usuarios** de datos para acceder a los datos y a la documentación, de manera que los **usuarios** pueden determinar la “**adecuación para el uso**” de los datos y asegurar que sus análisis estén ajustados a la realidad.

# Nacional/regional frente a Global

- Un dato es sensible según el **área** en que se halle: un taxon puede ser sensible en una región y no en otra.
- Si una institución alberga datos de **otra jurisdicción** (otra autonomía, otro país, etc.) puede no saber qué táxones están protegidos **dentro** de esa otra jurisdicción.
- Esta dificultad se agrava cuando un taxon sensible en un área puede ser una mala hierba o una plaga en otra área.
- Se ha sugerido la conveniencia de tener **una lista global de táxones sensibles**. Las dificultades en desarrollar dicha lista (taxonomicamente y geográficamente), intercambiando y manteniendo esta información no es un tema simple.

# En relación a los duplicados

- A menudo los datos se distribuyen alrededor del mundo a través de especímenes **duplicados**. Aunque la institución de origen restrinja los datos sensibles, otras instituciones que albergan duplicados pueden publicar la misma información sin restricciones.
- Esto puede ser debido a **desconocimiento** de lo que puede ser considerado como sensible **en el área de origen** del taxon, ya que no existe actualmente una lista mundial de datos considerados sensibles asociados a un área.
- El uso de los **LSIDs (Life Science Identifiers)** para la identificación de duplicados/registros relacionados y el intercambio automático de información (**Push Technology**) servirá para hacer un seguimiento de los duplicados.
- Así, la institución que alberga los datos originales podrá **automáticamente notificar** a las instituciones que albergan **duplicados**, cualquier cambio en el estatus del taxon.

# Lista global de datos sensibles

- Está en estudio por parte de GBIF, el desarrollo de una **lista mundial de táxones (o sus atributos) potencialmente sensibles**, que esté ligada con **ECat (Electronic Catalogue of Names of Known Organisms)**, y que serviría para alertar a los proveedores de que el taxon o algún atributo ha sido identificado como sensible.
- La lista debe ser concebida **exclusivamente** como un “**disparador**” (trigger) que sirva para señalar al proveedor de datos de la necesidad de **tomar una decisión** sobre la actual situación del dato sensible en la **Web**, y no como una lista para generar restricciones generales.
- Tendría la ventaja de **alertar** a a los proveedores de datos de **otras jurisdicciones** sobre la conveniencia de generalizar o restringir el dato.

# Qué deberá incluir la lista global

- Esta lista deberá incluir la siguiente información:
  - Nombre del Taxon
  - Criterio y razones para su inclusión
  - Nombre de la persona u organización responsable de su inclusión
  - Cobertura geográfica en la que es sensible (especialmente si sólo es sensible dentro de una jurisdicción)
  - Categoría de sensibilidad
  - Fecha para su revisión.
- Con cada taxon incluido en esta lista se evaluará que los **beneficios de su inclusión** son mucho mayores que los **beneficios de su exclusión**.

# Presente y futuro de la lista global

- No es fácil llegar a un **consenso**, por lo que ese tema necesita ser **discutido** en profundidad
- Previamente sería necesario adoptar una **criterios de sensibilidad** según los cuales un taxon sería o no incluido en la lista.
- Se tratará de **evitar** que la lista se convierta en una razón para **bloquear** todos los datos de los táxones de la lista.



# Análisis correlacionales

- Muchos encuestados señalaron la idea de **proteger** la **información adjunta** para restringir las posibilidades de realizar **análisis correlacionales** y el análisis lógico y estadístico de los datos (data mining).
- Esta información adjunta incluye el
  - colector
  - número de colector
  - fecha de recolección,
  - localidad
  - hábitat
- Esta información puede ser usada **para identificar secuencias** en la recolección, para el mismo colector y fecha de recolección y por tanto ser usada por deducción para identificar las coordenadas geográficas de registros que carecen de ella.

# Lo que opinan los encuestados sobre los análisis correlacionales

- Muchas instituciones harían accesibles sus datos si tuvieran la **seguridad** de que la información oculta **no puede ser deducida** a través de análisis correlacionales, de manera que los táxones no se vean amenazados a través de la publicación de esta información.
- Algunas instituciones **continuarán restringiendo** los datos de la localidad que se muestran en Internet.
- Otras los tendrán disponibles a **usuarios “fiables”** (bona-fide users)
- Gran número de instituciones desearían hacer públicos sus datos tras una **generalización de la información sensible**.

# Temas políticos y sociales

- Varios encuestados plantearon el tema del **balance** entre la protección de especies a través del **conocimiento de su localización** opuesto a la protección a través de la **restricción del conocimiento de su localización**.
- Por un lado tenemos muchos ejemplos en donde las especies han salido **perjudicadas** por la divulgación de su situación geográfica. Por esta razón, la localización de muchas especies **no ha sido publicada**.
- Por otro lado, muchas especies han sido **puestas en peligro por falta de conocimiento de su existencia** en un lugar concreto.
- El punto de unión de estas dos situaciones sería que la información de datos sensibles fuera suministrada a **agencias y organismos gubernamentales**, que **gestionaran** adecuadamente esta información.

# Más sobre temas políticos y sociales

- A nivel social, sería deseable la implicación de **colectivos o asociaciones** que **colaboren** para la conservación de especies sensibles y otras especies.
- El **compromiso** de estos colectivos en la conservación es mejor que mantener al margen al público interesado y crear así **confrontaciones**.
- Esto está en relación con la necesidad de fomentar una **cultura de la conservación** por parte de los países.
- Hay un deseo por parte de GBIF de **profundizar en la discusión** de temas socio-políticos relacionados con los datos sensibles

¿Preguntas, ideas, sugerencias???