

# Publicación de datos en línea: interés, contexto e implicaciones

Francisco Pando  
GBIF - España



# Qué vamos a ver

- El escenario actual: Colecciones, su relevancia y las tecnologías de la información y comunicación
- Metadatos, para dar contexto
- Licencias, para establecer las reglas del juego
- “Como citar”, para dar crédito y tener relevancia
- Que datos publicar
- Donde publicar
- El tema de la calidad

# El escenario actual: colecciones, su relevancia ... ¿colecciones para qué?

- Colecciones para taxónomos ya no basta
- Una colección puede y debe ser una infraestructura de información “para todo:”
  - Salud pública
  - Contaminación ambiental (DDT, Hg, ...)
  - Agricultura (plagas, OGM, ...)
  - Pérdida de habitat
  - Especies invasoras
  - Cambio climático
  - ...

Suarez, A.V. & Tsutsui, N.D. ( 2004). The Value of Museum Collections for Research and Society. *BioScience* 54 (1): 66-74

# El escenario actual: Colecciones, su relevancia ...y las tecnologías de la información y comunicación

- Las tecnologías informáticas están devolviendo las colecciones a la escena de la ciencia actual y de las preocupaciones sociales. Sin embargo, este procesado de conlleva una descontextualización de los datos.
- El contexto de la información es algo muy difícil de aprehender pero vital para un uso óptimo de los ejemplares y las colecciones.
- Los metadatos proporciona contexto.
- Esta situación hace que una buena documentación sea más importante que nunca.

Hay que luchar contra el síndrome de “todo el mundo sabe eso” (y no se documenta)

# Metadatos, para dar contexto

“Los metadatos son un elemento fundamental en la generación del conocimiento”

...y en su descubrimiento y recuperación

Y en su uso

- Generar, descubrir, recuperar, usar  
⇒ Intercambiar, integrar, reutilizar  
⇒ estándares

# Dublin Core

- [The Dublin Core metadata element set](#) se convirtió en norma [ISO 15836/2003](#) en febrero de 2003.
- Creada por una comunidad de individuos de diferentes procedencias y disciplinas, de organizaciones de todo el mundo que incluyen tanto al sector público como al privado.
- Define 15 elementos básicos y generales para describir un recurso (un programa, una página Web, un mapa,. ..).
- <http://dublincore.org/documents/dces/>

# Dublin Core Metadata Element Set

## Contenido:

1. **Título:** el nombre dado a un recurso, habitualmente por el autor.
2. **Claves:** (Palabras clave)
3. **Descripción:** una descripción textual del recurso. Puede ser un resumen en el caso de un documento o una descripción del contenido en el caso de un documento visual.
4. **Fuente:** secuencia de caracteres usados para identificar unívocamente un trabajo a partir del cual proviene el recurso actual.
5. **Lengua:** lengua/s del contenido intelectual del recurso.
6. **Relación:** es un identificador de un segundo recurso y su relación con el recurso actual. Este elemento permite enlazar los recursos relacionados y las descripciones de los recursos.
7. **Cobertura:** espacial y/o temporal del contenido intelectual del recurso.

## Propiedad Intelectual:

8. **Autor o Creador:** la persona o organización responsable de la creación del contenido intelectual del recurso.
9. **Editor:** la entidad responsable de hacer que el recurso se encuentre disponible.
10. **Otros Colaboradores:** una persona u organización que haya tenido una contribución intelectual significativa,
11. **Derechos:** son una referencia (p.e. URL) sobre términos y condiciones de acceso a un recurso.

## Instanciación:

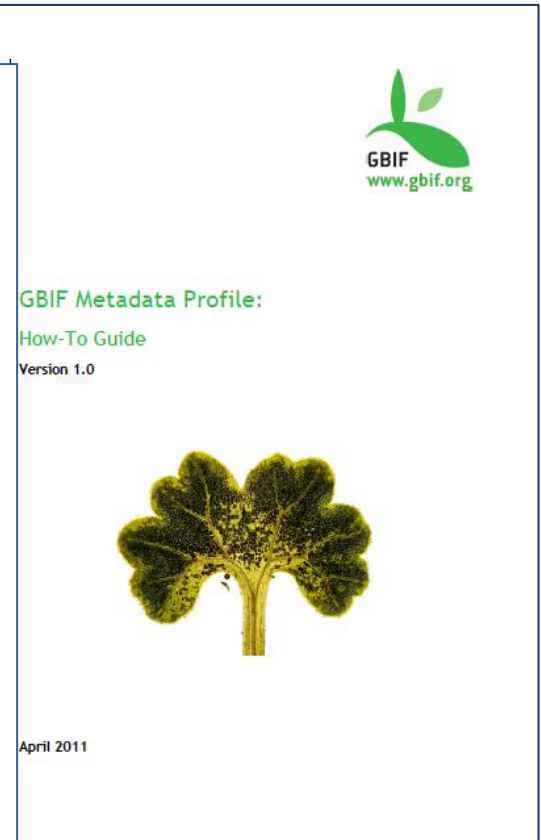
12. **Fecha:** una fecha de publicación del recurso en su forma actual.
13. **Tipo del Recurso:** la categoría del recurso.
14. **Formato:** es el formato usado para identificar el software y el hardware para mostrar el recurso.
15. **Identificador del Recurso:** secuencia de caracteres utilizados para identificar unívocamente un recurso. URL, URN, ISBN, DOI,...

# Metadatos de conjuntos de datos

GMP (GBIF Metadata profile):

EML (Ecological Metadata Language)

- Dataset (Resource)
- Project
- People and Organisations
- Keyword Set (General Keywords)
- Coverage
  - o Taxonomic Coverage
  - o Geographic Coverage
  - o Temporal Coverage
- Methods
- Intellectual Property Rights
- Additional Metadata
  - + NCD (Natural Collections Descriptions Data) Related



[http://links.gbif.org/gbif\\_metadata\\_profile\\_how-to\\_en\\_v1/](http://links.gbif.org/gbif_metadata_profile_how-to_en_v1/)  
[http://www.gbif.org/orc/?doc\\_id=2820](http://www.gbif.org/orc/?doc_id=2820)



# ejemplo

- <http://www.gbif.es:8080/catalogue/#>

# Licencias para publicar (compartir) datos

- “El valor de las colecciones (y de sus datos) está en su uso”
- La sociedad reclama poder usar la información que paga (investigación pública  $\Rightarrow$  acceso público)
- Los actores involucrados (científicos, conservadores) deben recibir reconocimiento profesional por el trabajo realizado

## ... ¿Qué está pasando?

- Open data
- Open access journals
- Open Government
- Ley 27/2006, y Directivas 2003/4/CE y 2003/35/CE
- [European Commission embraces open access](#)

Tolle, K. M., Tansley, D. S. W., & Hey, A. J. G. (2011). The Fourth Paradigm: Data-Intensive Scientific Discovery. (T. Hey, S. Tansley, & K. Tolle, Eds.) Proceedings of the IEEE. IEEE. Retrieved from <http://research.microsoft.com/en-us/collaboration/fourthparadigm/>

Chavan, V., and Penev, L. (2011). The data paper: a mechanism to incentivize data publishing in biodiversity science. BMC Bioinformatics 12, S2. Available at: <http://www.biomedcentral.com/1471-2105/12/S15/S2>.



# Licencias “Creative Commons”



## Atribución CC BY

Esta licencia permite a otros distribuir, remezclar, retocar, y crear a partir de tu obra, incluso con fines comerciales, siempre y cuando te den crédito por la creación original. Esta es la más flexible de las licencias ofrecidas. Se recomienda para la máxima difusión y utilización de los materiales licenciados.

[Ver Resumen de la licencia](#) | [Ver Código Legal](#)



## Atribución-CompartirIgual CC BY-SA

Esta licencia permite a otros remezclar, retocar, y crear a partir de tu obra, incluso con fines comerciales, siempre y cuando te den crédito y licencien sus nuevas creaciones bajo condiciones idénticas. Esta licencia suele ser comparada con las licencias “copyleft” de software libre y de código abierto. Todas las nuevas obras basadas en la tuya portarán la misma licencia, así que cualesquiera obras derivadas permitirán también uso comercial. Esa es la licencia que usa Wikipedia, y se recomienda para materiales que se beneficiarían de incorporar contenido de Wikipedia y proyectos con licencias similares.

[Ver Resumen de la licencia](#) | [Ver Código Legal](#)



## Atribución-SinDerivadas CC BY-ND

Esta licencia permite la redistribución, comercial o no comercial, siempre y cuando la obra circule íntegra y sin cambios, dándote crédito.

[Ver Resumen de la licencia](#) | [Ver Código Legal](#)



## Atribución-NoComercial CC BY-NC

Esta licencia permite a otros distribuir, remezclar, retocar, y crear a partir de tu obra de modo no comercial, y a pesar de que sus nuevas obras deben siempre mencionarte y mantenerse sin fines comerciales, no están obligados a licenciar sus obras derivadas bajo las mismas condiciones.

[Ver Resumen de la licencia](#) | [Ver Código Legal](#)



## Atribución-NoComercial-CompartirIgual CC BY-NC-SA

Esta licencia permite a otros distribuir, remezclar, retocar, y crear a partir de tu obra de modo no comercial, siempre y cuando te den crédito y licencien sus nuevas creaciones bajo condiciones idénticas.

[Ver Resumen de la licencia](#) | [Ver Código Legal](#)



## Atribución-NoComercial-SinDerivadas CC BY-NC-ND

Esta licencia es la más restrictiva de nuestras seis licencias principales, permitiendo a otros descargar tus obras y compartirlas con otros siempre y cuando te den crédito, pero no permiten cambiarlas de forma alguna ni usarlas comercialmente.

[Ver Resumen de la licencia](#) | [Ver Código Legal](#)

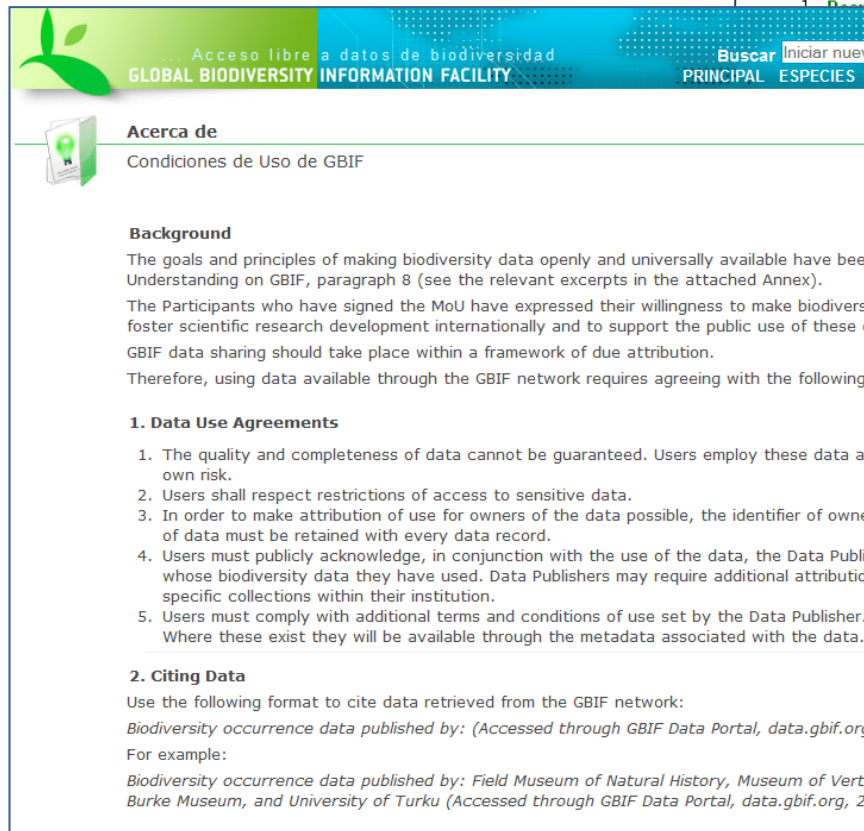
<http://creativecommons.org/licenses/?lang=es>

# Como citar

## Information

Name:	CIBIO, Alicante:ABH- GBIF
Description:	CIBIO, Universidad de Alicante, Herbario ABH
How to cite this dataset:	Herbario de la Universidad de Alicante, ABH (2005). (accessed through GBIF data portal, <a href="http://data.gbif.org/datasets/resource/251">http://data.gbif.org/datasets/resource/251</a> , 2011-11-29)
Basis of record:	Specimen
Access point URL:	<a href="http://taray.csic.es:6000/digir/DiGIR.php">http://taray.csic.es:6000/digir/DiGIR.php</a>

<http://data.gbif.org/datasets/resource/240>



Acceso libre a datos de biodiversidad  
GLOBAL BIODIVERSITY INFORMATION FACILITY

Buscar Iniciar nueva sesión  
PRINCIPAL ESPECIES

### Acerca de

Condiciones de Uso de GBIF

#### Background

The goals and principles of making biodiversity data openly and universally available have been outlined in the GBIF Data Use Agreement (see the relevant excerpts in the attached Annex).

The Participants who have signed the MoU have expressed their willingness to make biodiversity data openly and universally available to foster scientific research development internationally and to support the public use of these data. Data sharing should take place within a framework of due attribution.

Therefore, using data available through the GBIF network requires agreeing with the following conditions:

#### 1. Data Use Agreements

1. The quality and completeness of data cannot be guaranteed. Users employ these data at their own risk.
2. Users shall respect restrictions of access to sensitive data.
3. In order to make attribution of use for owners of the data possible, the identifier of owners of data must be retained with every data record.
4. Users must publicly acknowledge, in conjunction with the use of the data, the Data Publisher whose biodiversity data they have used. Data Publishers may require additional attribution for specific collections within their institution.
5. Users must comply with additional terms and conditions of use set by the Data Publisher. Where these exist they will be available through the metadata associated with the data.

#### 2. Citing Data

Use the following format to cite data retrieved from the GBIF network:

*Biodiversity occurrence data published by: (Accessed through GBIF Data Portal, data.gbif.org)*

For example:

*Biodiversity occurrence data published by: Field Museum of Natural History, Museum of Vertebrate Zoology, and University of Turku (Accessed through GBIF Data Portal, data.gbif.org, 2011-11-29)*

## Herbario de Criptogamia

RJB > Colecciones > Herbario > Criptogamia > Bases de datos > Condiciones

Databases

CONDICIONES Y RESERVAS DE USO DE LOS DATOS

CONDITIONS AND PROVISOS FOR USE OF CONTENT

**Responsabilidad:** Ni el Real Jardín Botánico ni sus conservadores se responsabilizan de la exactitud, fiabilidad y de cobertura de los datos, ni del uso que de ellos se les dé.

**Uso comercial de los datos:** Se prohíbe el uso comercial de los datos obtenidos, o que genere beneficio económico, sin el consentimiento explícito del Real Jardín Botánico.

**Atribución:** Queda estrictamente prohibida la edición, reproducción, distribución, adición o reducción total o parcial de la información, incluida explícitamente su incorporación a otras bases de datos de acceso libre o restringido.

**Reconocimientos:** El uso de los datos implica la obligación de reconocer y citar su origen explícitamente (ver "[como citar este recurso](#)"). Esta atribución se debe mantener en cualquier uso subsecuente de los datos.

**Información obtenida:** el usuario se compromete a que la información obtenida no sea utilizada en detrimento de la conservación de la flora y fauna, ni en procesos que alteren irreversiblemente el medio ecológico (extracción, explotación, criminada o introducción de especies).

**Información negativa:** los datos de biodiversidad constatan generalmente

1. **Data are provided on an "as is" basis-** no responsibility by the Real Jardín Botánico and their curators of the accuracy, reliability, or completeness of the data or information provided data is accepted for uses or misuses.
2. **Data are not expected to be used on a for-profit basis-** use of the data for commercial or for-profit applications are permitted only via written permission from the Real Jardín Botánico
3. **Data are provided on an end-user basis-** that is, data are provided to users, but should not be passed on to third parties or redistributed. It is explicitly forbidden to incorporate these data into other databases of free or restricted access.
4. **Proper acknowledgment of the data source institutions is expected;** The data use implies explicit acknowledgment (see "[how to cite this resource](#)") which should be kept with any subsequent use of the data.
5. **Data obtained:** Data are provided under the understanding that their use will be generally consistent with the philosophy of not causing detriment to wild species and ecosystem balance.
6. **Use of negative information -** Biodiversity data generally constitute positive information only; that is, absence of a record does not always mean absence of a particular species from particular point in space and time. Because biodiversity

<http://data.gbif.org/tutorial/datauseagreement>

# Que datos publicar

## Criterios:

- Compatibilizar rendimiento científico y acceso
- Compatibilizar conservación y acceso

## Herramientas:

- Retrasar la publicación
- Difuminar la precisión
- Metadatos

(los metadatos del “recurso publicado” no tienen porqué ser los de la colección)

# Otros aspectos a cuidar

- Usar identificadores únicos, estables
  - Considerar dar URLs a cada registro
- Gestionar los datos de uso de los distintos portales por los que publiquemos: todos tienen que sumar, no diluir
- “Feedback”

# El tema de la calidad

- La perfección no existe
- El dato malo tampoco; un dato no es bueno o malo intrínsecamente, sino que su bondad depende del uso

La clave es proporcionar una buena documentación sobre que estamos publicando (metadatos)

Hay herramientas, que ayudan, p.e.:

**Darwin Test**

Una aplicación para la validación y chequeo de los datos en formato Darwin Core 1.2 o Darwin Core 1.4, para asegurar una calidad en las bases de datos que se ven a través de la red de GBIF

[http://www.gbif.es/darwin\\_test/Darwin\\_test.php](http://www.gbif.es/darwin_test/Darwin_test.php)

# Donde (y como) publicar

## disponible.

(De *disponer*).

1. adj. Dicho de una cosa: Que se puede disponer libremente de ella o que está lista para usarse o utilizarse.

## acceso.

(Del lat. *accessus*).

1. m. Acción de llegar o acercarse.



Se trataría de hacer disponible, dar acceso no basta (ir más allá de los PDFs, usar estándares)

El potencial de los datos de las colecciones se explota si se publican por varias vías



# Donde publicar

El potencial de los datos de las colecciones se explota adecuadamente solo si se publican por varias vías, que en su conjunto deben responder a una estrategia de publicación. p.e.:

- GBIF: Portal global, generalista: da datos básicos
  - <http://data.gbif.org>
- OBIF: Portal global, temático: vista parcial con capas relevantes para el ambiente marino
  - <http://www.iobis.org/mapper/>
- Portales nacionales: dan contexto específico (áreas protegidas, listas rojas, legislación, nombres vernáculos...)
  - [www.gbif.es/datos](http://www.gbif.es/datos)
- Portal de la propia colección: información completa, de conjuntos de datos específicos (enlaces a imágenes, publicaciones, Genbank, historial de identificaciones,... )
  - <http://161.111.170.202/herb/asp/>

# A modo de conclusión: lista de comprobación para publicar (1/2)

- Ser parte del paisaje global o desaparecer
  - Publicar “en línea” y por varias vías
  - De manera estratégica y consistente
  - Hacerse con los datos de uso
- Hacer metadatos; vale la pena cuidarlo
  - Usar formatos estándar
  - Los metadatos de la colección y del recurso publicado no tienen porqué coincidir

# A modo de conclusión: lista de comprobación para publicar (2/2)

- Identificadores únicos para ejemplares
- Proporcionar mecanismos de “feedback”
- Indicar:
  - “Citar como”
  - Licencias de uso
  - Condiciones de uso de los datos

Francisco Pando

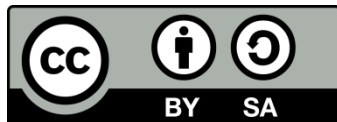
Unidad de coordinación, GBIF España

Real Jardín Botánico - CSIC

Claudio Moyano 1, 28014 Madrid

[pando@gbif.es](mailto:pando@gbif.es)

[www.gbif.es](http://www.gbif.es)



<http://creativecommons.org/licenses/by-sa/3.0/es/>

