

# Datos (primarios) de biodiversidad: ¿Por qué publicarlos?

**Francisco Pando**  
**GBIF - España**



# Qué vamos a ver

- Datos, ciencia y biología
- ¿Qué hace falta para que los datos se publiquen?
  - Que los datos se usen
  - Que se publiquen de manera que se puedan usar
  - Que los datos se publiquen
- Requerimientos e incentivos

# El acceso a la información sobre la biodiversidad es complejo

La Biodiversidad tiene:

- implicaciones múltiples, más allá de la ciencia
- dispersión de la información
- complejidad intrínseca
- que solo se publican los resultados

# Data-Intensive Scientific Discovery

- Data intensive. Basic science is data intensive in its own right, but data sources that support basic science are often insufficient to support applications. Localized impacts with global extent, such as intrusion of invasive species, are often difficult for centralized projects with small numbers of researchers to ascertain. **New applications-appropriate sources must be identified**, and new ways of observing (including the use of communities as data gatherers) must be developed.

Hey, T., Tansley, S., & Tolle, K. (Eds.) 2009. The fourth Paradigm. Data-Intensive Scientific Discovery. Microsoft Research. Redmon. E.E.U.U.

# Open access & data

Los datos pagados con dinero público deben ser accesibles para todos; dos ejemplos de cómo se está poniendo en práctica:

Congress passed a law that recommended that if you take NIH (National Institutes of Health) funding for your research, you should deposit your research reports with the National Library of Medicine (NLM) so that the full text of your papers should be in the public domain. Voluntary compliance with this law has been only 3 percent, so things are about to change. We are now likely to see all of the publicly funded science literature forced online by the funding agencies. There is currently a bill



## ¿Qué es Open Data Euskadi?

Es el portal de acceso a los datos públicos del Gobierno Vasco en formato reutilizable. Entre sus objetivos destacan:



### Generar valor y riqueza

Obteniendo productos derivados de los datos por parte de terceros.



### Generar transparencia

Reutilizando los datos para analizar y evaluar la gestión pública.

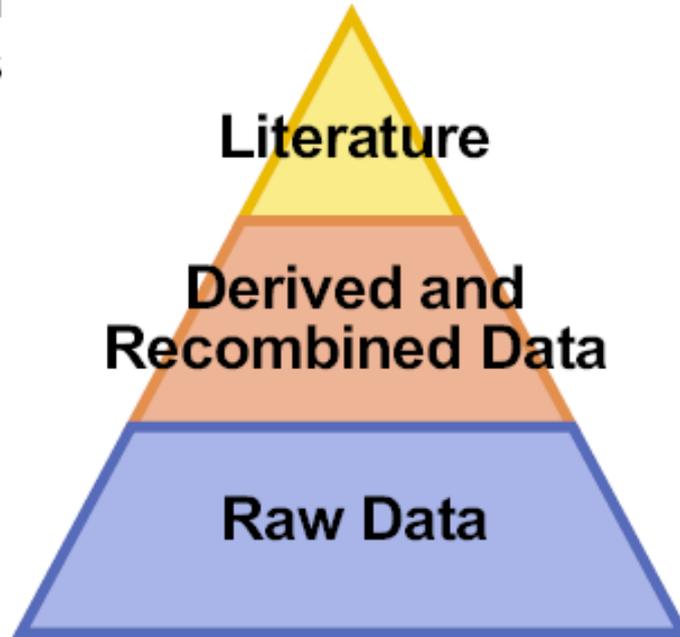


### Facilitar la interoperabilidad entre administraciones

Creando servicios que utilicen datos de diferentes

# All Scientific Data Online

- Many disciplines overlap and use data from other sciences
- Internet can unify all literature and data
- Go from literature to computation to data back to literature
- Information at your fingertips for everyone-everywhere
- Increase Scientific Information Velocity
- Huge increase in Science Productivity



# 2020 SCIENCE

En 2005 Microsoft Research reunió a 30 distinguidos científicos para dilucidar el camino de la ciencia para los siguientes 14 años y determinar el impacto de los avances informáticos en la misma. El informe se centra en la biología (molecular, de organismos, de sistemas) por dos razones:

- Es donde va a ser más grande el impacto de la informática
- Es donde están los retos científicos, sociales y globales más grandes

<http://research.microsoft.com/towards2020science/>

# GBIF como respuesta

Los ministros de ciencia de un grupo de países subscriben en 1999 la creación de un **mecanismo de coordinación científica** para que los usuarios de todo el mundo puedan **descubrir y utilizar** las enormes cantidades de **datos sobre la biodiversidad** global, **impulsar la investigación** científica en muchas disciplinas, fomentar el **desarrollo tecnológico** continuado, facilitar la **participación equitativa** en los beneficios de la biodiversidad y **eleva la calidad de vida** de los miembros de la sociedad...

# ¿Que hace falta para hacer real esta visión?

- Que los datos se publiquen
- Que se publiquen de manera que se puedan usar
- Que los datos se usen

# Que los datos se usen

- Capacidad para usarlos
  - Formación
- Cambio de cultura:
- Reusar los datos
  - Cambiar procedimientos
  - <http://www.nsf.gov/nsb/publications/2011/nsb1124.pdf>

# Que se publiquen de manera que se puedan usar

- Documentados
  - Los contenidos
  - Las condiciones de uso
- Estandarizados
- “As application appropriate sources”
  - Que puedan ser procesados por programas > estructurados
- Fáciles de descubrir

\* “en digital”, se supone

# Datos documentados:

## 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:

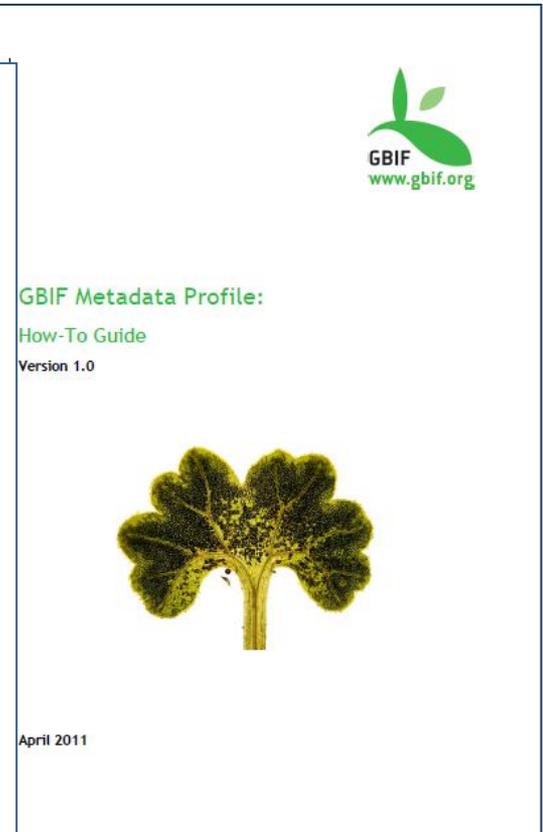
12. **Fecha:** una fecha de publicación del recurso en su forma actual.
13. **Tipo del Recurso:** la categoría del recurso.
14. **Formato:** es el formato usado para identificar el software y el hardware para mostrar el recurso.
15. **Identificador del Recurso:** secuencia de caracteres utilizados para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

# Datos estandarizados: estándar

# Sobre Darwin Core

Datos estandarizados

Biodiversity  
Information  
Standards  
TDWG

## Introduction

References

## Quick Reference Guide

Term Index

Record-level Terms

Occurrence

Event

Location

GeologicalContext

Identification

Taxon

ResourceRelationship

MeasurementOrFact

Term Definitions

## Simple Darwin Core

## Type Vocabulary

## Namespace Policy

## XML Guide

## Text Guide

## Darwin Core Terms: A quick reference guide

**Title:** Darwin Core Terms: A quick reference guide

**Date Issued:** 2009-02-12

**Date** 2011-10-26

**Modified:**

**Abstract:** This document is a quick reference for all recommended Darwin Core terms. For complete historical term information, including version changes and pre-standard terms, see [[HISTORY](#)]. For a comparative table of elements from pre-standard versions of Darwin Core to the current terms in the standard, see [[VERSIONS](#)].

**Contributors:** John Wieczorek (MVZ), Markus Döring (GBIF), Renato De Giovanni (CRIA), Tim Robertson (GBIF), Dave Vieglais (KUNHM)

**Legal:** This document is governed by the standard legal, copyright, licensing provisions and disclaimers issued by the Taxonomic Databases Working Group.

**Part of TDWG Standard:** <http://www.tdwg.org/standards/450/>

**Creator:** Darwin Core Task Group

**Identifier:** <http://rs.tdwg.org/dwc/2011-10-26/terms/>

**Latest Version:** <http://rs.tdwg.org/dwc/terms/>

**Replaces:** <http://rs.tdwg.org/dwc/2009-12-07/terms/>

**Document Status:** Current Standard

# Darwin Core: Términos

<http://rs.tdwg.org/dwc/terms/>

## **Record-level Terms**

[dcterms:type](#) | [dcterms:modified](#) | [dcterms:language](#) | [dcterms:rights](#) | [dcterms:rightsHolder](#) | [dcterms:accessRights](#) | [dcterms:b  
institutionID](#) | [collectionID](#) | [datasetID](#) | [institutionCode](#) | [collectionCode](#) | [datasetName](#) | [ownerInstitutionCode](#) | [basisOfRecord](#) |

## **Occurrence**

[occurrenceID](#) | [catalogNumber](#) | [occurrenceRemarks](#) | [recordNumber](#) | [recordedBy](#) | [individualID](#) | [individualCount](#) | [sex](#) | [lifeStage](#)  
[associatedReferences](#) | [associatedOccurrences](#) | [associatedSequences](#) | [associatedTaxa](#)

## **Event**

[eventID](#) | [samplingProtocol](#) | [samplingEffort](#) | [eventDate](#) | [eventTime](#) | [startDayOfYear](#) | [endDayOfYear](#) | [year](#) | [month](#) | [day](#) | [vert](#)

## **dcterms:Location**

[locationID](#) | [higherGeographyID](#) | [higherGeography](#) | [continent](#) | [waterBody](#) | [islandGroup](#) | [island](#) | [country](#) | [countryCode](#) | [state](#)  
[maximumDepthInMeters](#) | [minimumDistanceAboveSurfaceInMeters](#) | [maximumDistanceAboveSurfaceInMeters](#) | [locationAccordingT  
coordinateUncertaintyInMeters](#) | [coordinatePrecision](#) | [pointRadiusSpatialFit](#) | [footprintWKT](#) | [footprintSRS](#) | [footprintSpatialFit](#) | [g](#)

## **GeologicalContext**

[geologicalContextID](#) | [earliestEonOrLowestEonothem](#) | [latestEonOrHighestEonothem](#) | [earliestEraOrLowestErathem](#) | [latestEraOrH  
latestAgeOrHighestStage](#) | [lowestBiostratigraphicZone](#) | [highestBiostratigraphicZone](#) | [lithostratigraphicTerms](#) | [group](#) | [formation](#)

## **Identification**

[identificationID](#) | [identifiedBy](#) | [dateIdentified](#) | [identificationReferences](#) | [identificationVerificationStatus](#) | [identificationRemarks](#) | [id](#)

## **Taxon**

[taxonID](#) | [scientificNameID](#) | [acceptedNameUsageID](#) | [parentNameUsageID](#) | [originalNameUsageID](#) | [nameAccordingToID](#) | [namePu  
higherClassification](#) | [kingdom](#) | [phylum](#) | [class](#) | [order](#) | [family](#) | [genus](#) | [subgenus](#) | [specificEpithet](#) | [infraspecificEpithet](#) | [taxonR](#)

## **Auxiliary Terms**

### **ResourceRelationship**

[resourceRelationshipID](#) | [resourceID](#) | [relatedResourceID](#) | [relationshipOfResource](#) | [relationshipAccordingTo](#) | [relationshipEstablis](#)

### **MeasurementOrFact**

[measurementID](#) | [measurementType](#) | [measurementValue](#) | [measurementAccuracy](#) | [measurementUnit](#) | [measurementDetermined](#)

# DwC: múltiples versiones

## Darwin Core Versions

The following versions of the Darwin Core are of historical significance:

- **DwC 1.2 schema** -- first deployed version (used in GBIF, also known as [DarwinCoreV2?](#))
- **DwC 1.21 schema** -- revised version (used in GBIF, MaNIS, HerpNet, OrNIS, and FishNet2)
- **DwC 1.3 schema** -- draft standard of the Darwin Core accepted at TDWG Meeting 2004, Christchurch, New Zealand. (unused)
- **DwC 1.4 schema** -- draft standard under discussion (not for use)(but used in GBIF, and recommended to use in GBIF see: <http://www.gbif.org/DataProviders/HowTo>)
- **OBIS** -- based on DwC 1.2 used in GBIF, Ocean Biogeographic Information System)
- **PaleoPortal** -- based on DwC 1.2 (used in [The Paleontology Portal](#))

Biodiversity  
Information  
Standards  
TDWG

Introduction

References

Quick Reference Guide

Simple Darwin Core

Type Vocabulary

Namespace Policy

XML Guide

Text Guide

Complete History

Decision History

Mapping to ABCD

Mapping to Old Versions

## Darwin Core

**Title:** Darwin Core

**Date Issued:** 2009-02-12

**Date** 2009-10-08

**Modified:**

**Abstract:** This document is a cover page, an entry-level document to the Darwin Core standard. It describes the purpose of the standard and orients the reader to the documents that cover specific topics within the standard, such as the quick guide to the list of terms.

**Contributors:** John Wieczorek (MVZ), Markus Döring (GBIF), Renato De Giovanni (CRIA), Tim Robertson (GBIF), Dave Vieglais (KUNHM)

**Legal:** This document is governed by the standard legal, copyright, licensing provisions and disclaimers issued by the Taxonomic Databases Working Group.

**Part of TDWG** <http://www.tdwg.org/standards/450/>

- As application appropriate sources”
  - Que puedan ser procesados por programas > estructurados
    - Tabulados
    - Formato (semántico, sintáctico), xml
    - Bases de datos
    - RDF
- Fáciles de descubrir
  - Indexados, vinculados a portales (p.e., GBIF)

# Que los datos se publiquen

## Cambio cultural

Pasar del “yo acopio mis datos, los analizo y publico,” a “yo uso los datos de manera parecida a como usamos la electricidad o las carreteras,” como infraestructura

# Que los datos se publiquen

- Los datos hasta ahora, se han publicado poco y mal
- Cambiar los requerimientos, cambiar los incentivos

# Que los datos se publiquen: requerimientos, e incentivos para administraciones públicas

- Que los datos sean parte del "paquete"

## Ejemplos de petición de datos primarios en pliegos de condiciones

Uno de los problemas ambientales es la falta de datos relacionados con

Ejemplo. PLIEGO DE BASES TÉCNICAS. CONTRATO DE SERVICIOS PARA EL PROCESO DE DESIGNACIÓN DE 13 ZONAS ESPECIALES DE CONSERVACIÓN (ZEC) DE LA RED NATURA 2000 Y LA ELABORACIÓN DE LOS INSTRUMENTOS DE CONSERVACIÓN Y GESTIÓN DE LAS 13 ZEC Y DE UNA ZEPA (Descarga del documento aquí).

<http://www.gbif.es/AdministracionesAmbientales.php?tab=0>

- Que las secciones de "recopilación de la información" de los informes y consultorías citen las fuentes
- Que se haga ley, p.e.:
  - Open Government
  - Ley 27/2006, y Directivas 2003/4/CE y 2003/35/CE
  - [European Commission embraces open access](#)

# Que los datos se publiquen: requerimientos, e incentivos para administraciones públicas

- Que tenga una base de socialización
- Que haya visibilidad y reconocimiento por el esfuerzo
- Que sea sencillo
- Que haya aprendizaje
- Que sea divertido

[http://www.conabio.gob.mx/otros/nabci/doctos/a\\_ver\\_aves.html](http://www.conabio.gob.mx/otros/nabci/doctos/a_ver_aves.html)



CONABIO

Iniciativa para la Conservación de las Aves de América

*Todas las aves, todos los hábitat, i  
Hacemos más tra*

nabci México

Aves de México A ver aves Extinciones/paleontología Especies en riesgo Aves en la cultura Conteos colectivos Exóticas, introducidas y plagas

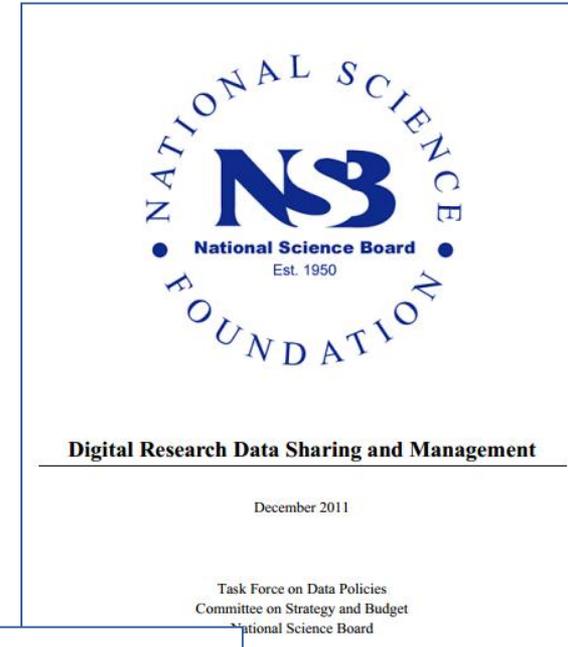
Usted esta aquí: Nabci » Aves » A ver aves

**¡Conviértete en observador profesional!**

Necesitas un par de binoculares, una guía de campo, una libreta para tomar apuntes, una cámara.

# Que los datos se publiquen: requerimientos, e incentivos para científicos

- Cambio cultural:
  - a) Que publicar datos se promueva  
**=> reconocimiento, mérito científico**
  - b) Que sea una exigencia de las políticas científicas:
    - <http://www.nsf.gov/nsb/publications/2011/nsb1124.pdf>
    - [http://ec.europa.eu/research/society/document\\_library/pdf06/recommendation-access-and-preservation-scientific-information\\_en.pdf](http://ec.europa.eu/research/society/document_library/pdf06/recommendation-access-and-preservation-scientific-information_en.pdf)



# Que los datos se publiquen: requerimientos, e incentivos para científicos

- Mientras llega, Hacen falta ciertos **elementos técnicos** que facilite el reusar y citar
  - Un sistema de citación
  - Un marco de licencias
  - Un sistema de referencia (identificadores únicos, persistentes resolvíbles)
- Hace falta **más cambio de cultura**
  - Reusar
  - Citar, “metadatar”
  - ... reconocer

# Que los datos se publiquen: requerimientos, e incentivos para científicos

- Hoy por hoy, lo que cuenta para el científico como mérito profesional son las publicaciones en revistas con factor de impacto, y de ahí la utilidad del concepto de “data paper” como incentivo

## Data Paper:

“A specific, stand-alone publication describing a particular dataset or a collection of datasets”

-----

Una publicación independiente que describe un juego de datos o una colección de ellos (publicados en línea)

Chavan, V., and Penev, L. (2011). The data paper: a mechanism to incentivize data publishing in biodiversity science. *BMC Bioinformatics* 12, S2.

<http://www.biomedcentral.com/1471-2105/12/S15/S2>.

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/>

