



Como publicar datos en Internet (aspectos a tener en cuenta)

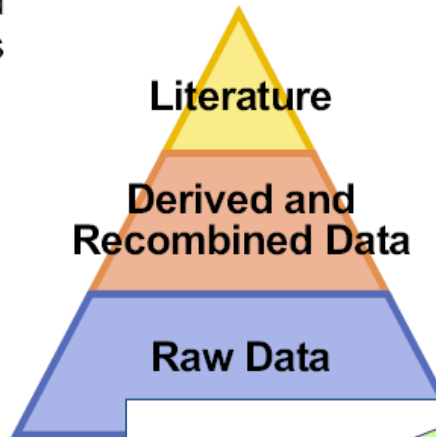
Francisco Pando
GBIF - España



Los datos se reusan, se integran

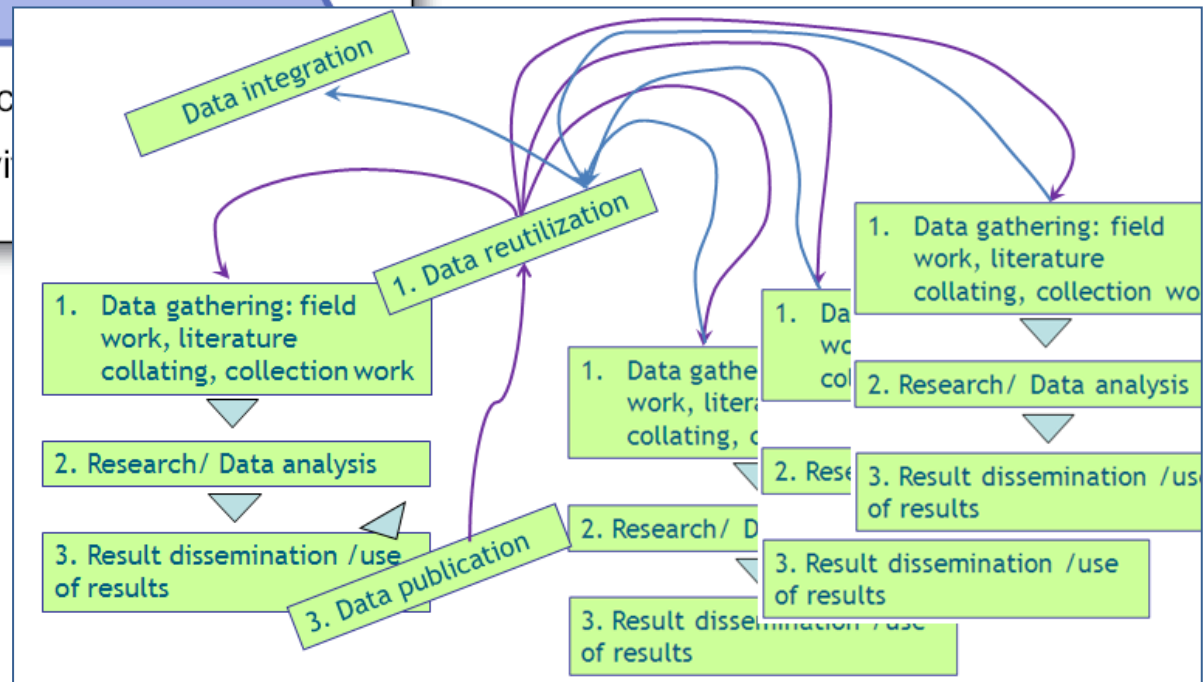
All Scientific Data Online

- 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



The Fourth Paradigm: Data-Intensive Scientific Discovery

<http://research.microsoft.com/en-us/collaboration/fourthparadigm/content.aspx>

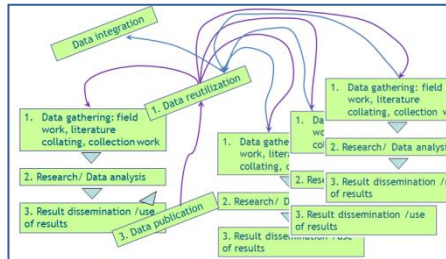


¿Que hace falta para hacer real esta visión?

- 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,...

XML

Los perfiles de metadatos se tienden a codificar usando un metalenguaje llamado XML (Extensible markup Language).

La definiciones de los metadatos se expresan usando una implementación de XML llamada XSD (*XML Schema Definition*)

Validación: Esquemas XML

```
<?xml version="1.0" encoding="UTF-8"?>
<correo xmlns="http://www.idee.es" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.idee.es/correo.xsd">
  <origen>alice@mfom.es</origen>
  <destino>bob@mfom.es</destino>
  <asunto>Avances IDEE</asunto>
  <texto>La IDE de Andalucía se incorporará a la IDEE en breve.</texto>
</correo>
```

❖ Documento XML

```
<?xml version="1.0"?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema"
targetNamespace="http://www.idee.es"
xmlns=http://www.idee.es" elementFormDefault="qualified">
  <xs:element name="correo">
    <xs:complexType>
      <xs:sequence>
        <xs:element name="origen" type="xs:string"/>
        <xs:element name="destino" type="xs:string"/>
        <xs:element name="asunto" type="xs:string"/>
        <xs:element name="cuerpo" type="xs:string"/>
      </xs:sequence>
    </xs:complexType>
  </xs:element></xs:schema>
```

❖ Esquema XML



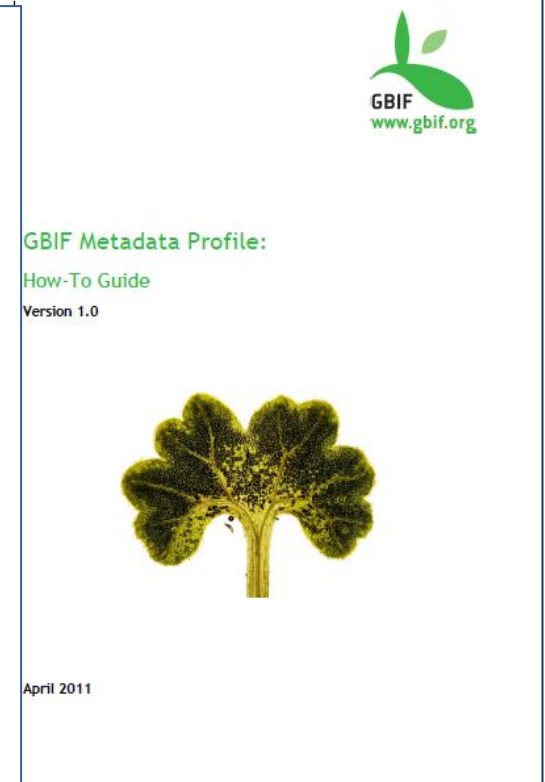
BY

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://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:basisOfRecord](#) | [dcterms:institutionID](#) | [dcterms:collectionID](#) | [dcterms:datasetID](#) | [dcterms:institutionCode](#) | [dcterms:collectionCode](#) | [dcterms:datasetName](#) | [dcterms:ownerInstitutionCode](#) | [dcterms: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](#) | [verticalElevation](#)

dcterms:Location

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

GeologicalContext

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

Identification

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

Taxon

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

Auxiliary Terms

ResourceRelationship

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

MeasurementOrFact

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

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 Wiecezorek (MVZ), Markus Döring (GBIF), Renato De Giovanni (CRIA), Tim Robertson (GBIF), Dave Vieglais (KUNHM)

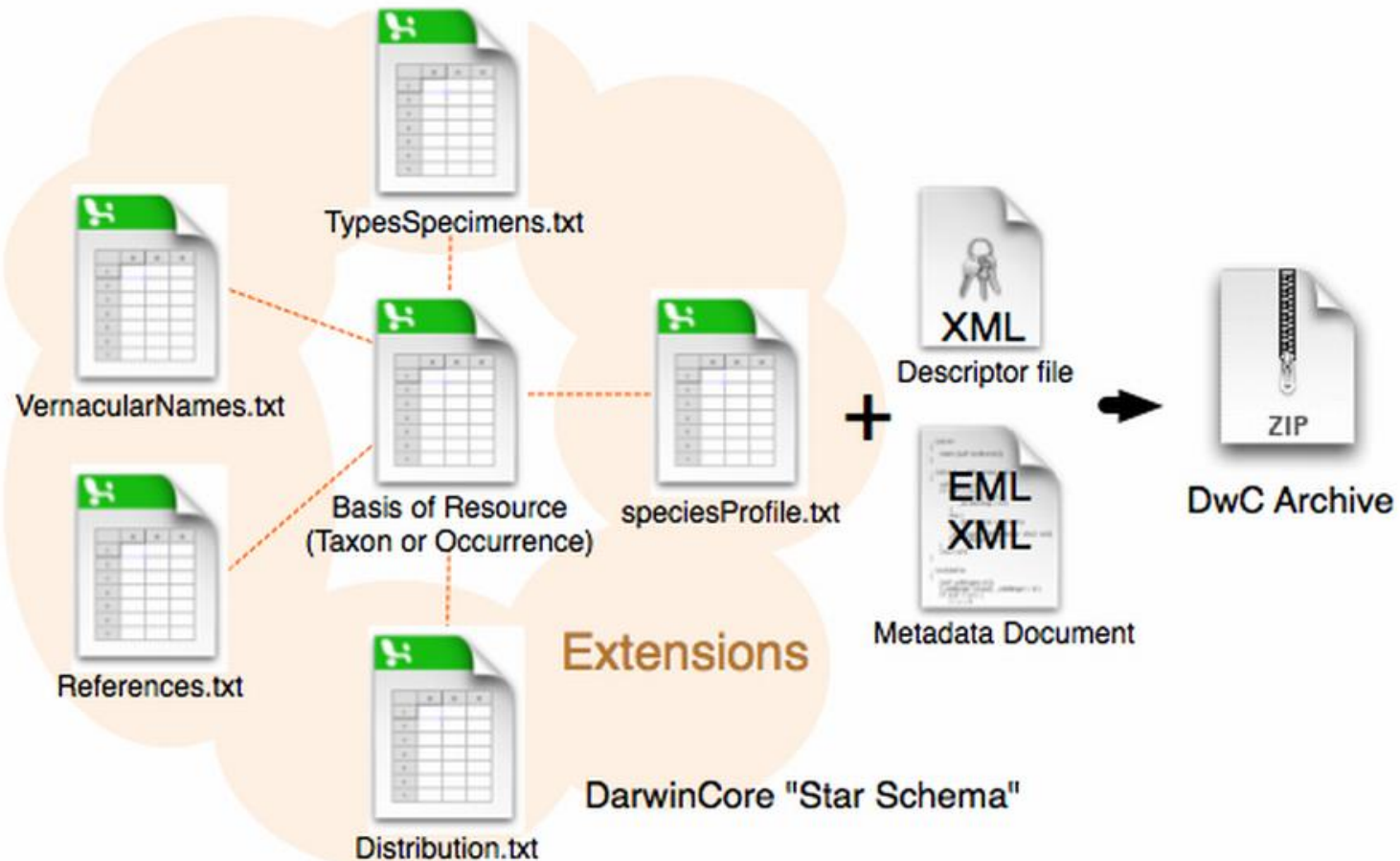
Legal: This document is governed by the standard legal, copyright, licensing provisions and disclaimers issued by the Taxonomic Databases Working Group.

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)

“Darwin Core Archive”

Darwin Core Archive Components



En portales, indexadores



gbif.es Nodo Nacional de Información en Biodiversidad

Nodo Español de GBIF / Spanish GBIF Node | Inicio | Vistazo a Datos | Consulta de Datos | Financiación | Participantes | Software | Portales |

Inicio
GBIF.ORG
Sobre GBIF
Actualidad
Participación en GBIF.ES
Consultar Datos
Servicios
Metadatos
(Centros, colecciones, bdd)
Formación y Divulgación

Inicio > Registro de colecciones, proyectos y bases de datos de biodiversidad en España > Búsquedas

 **REGISTRO DE COLECCIONES, PROYECTOS Y BASES DE DATOS DE BIODIVERSIDAD EN ESPAÑA**

Búsqueda de Centros, Colecciones y Personas.

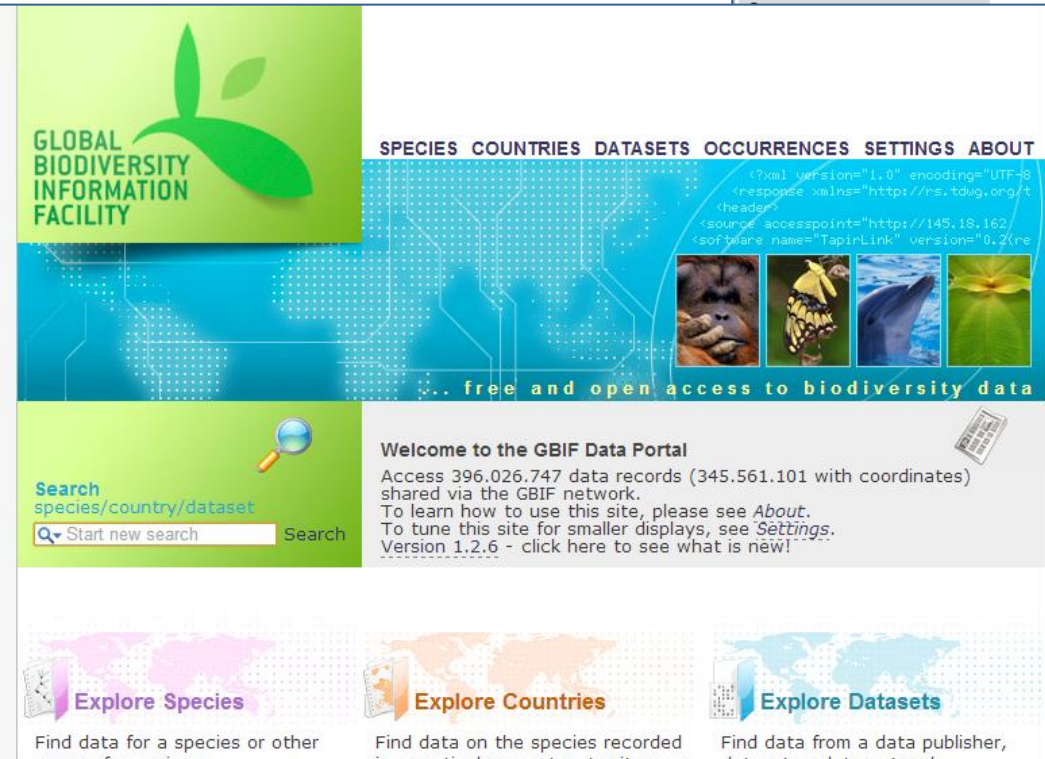
Desde esta página puede acceder a un sistema de búsqueda de **colecciones y proyectos de biodiversidad** de los **centros** que las albergan, y de **personas** relacionadas con las mismas.

datos mostrados aquí provienen del trabajo conjunto de dos proyectos internacionales:

- Por una lado, los datos provienen en su mayor parte de la base de datos **NoDIT**, proporcionada por el proyecto europeo **BioCASE** (*Biological Collection Access Service for Europe*): un proyecto de investigación científica financiado por la Unión Europea dentro del programa temático "Energy, Environment and Sustainable Development" del 5º Programa Marco (1998-2002).
- Los formularios de búsqueda y de acceso a los datos y los informes han sido desarrollados por la **Unidad de Coordinación del Nodo Nacional de Información sobre Biodiversidad GBIF España**.

Acceso a la base de datos

- Búsqueda de **CENTROS, INSTITUCIONES** y **PROYECTOS** de



GLOBAL BIODIVERSITY INFORMATION FACILITY

SPECIES COUNTRIES DATASETS OCCURRENCES SETTINGS ABOUT

... free and open access to biodiversity data

Welcome to the GBIF Data Portal
Access 396.026.747 data records (345.561.101 with coordinates) shared via the GBIF network.
To learn how to use this site, please see *About*.
To tune this site for smaller displays, see *Settings*.
Version 1.2.6 - click here to see what is new!

Search
species/country/dataset
Start new search Search

Explore Species
Find data for a species or other

Explore Countries
Find data on the species recorded

Explore Datasets
Find data from a data publisher,

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 iniciativas ciudadanas

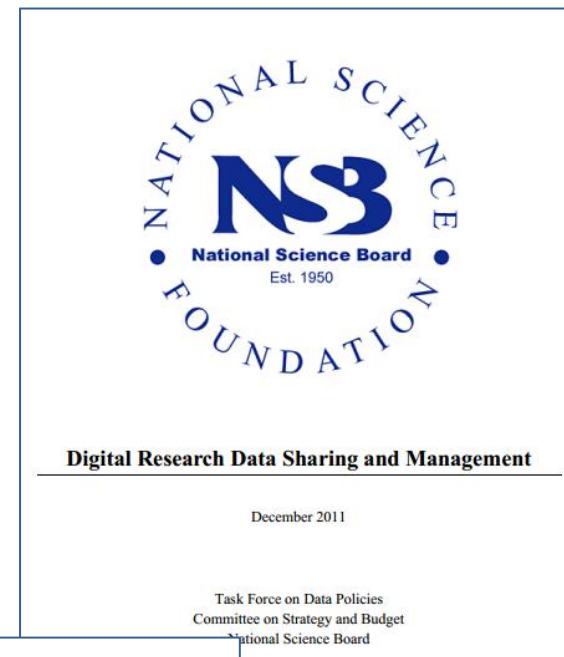
- Que tenga una base de socialización
- Que haya visibilidad y reconocimiento por el esfuerzo
- Que sea sencillo
- Que haya aprendizaje
- Que sea divertido

http://www.conabio.gob.mx/otros/nabci/doctos/a_ver_aves.html



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/pdf_06/recommendation-access-and-preservation-scientific-information_en.pdf



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>.



Que los datos se publiquen: requerimientos e incentivos para gestores

- 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 información relacionada 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: elementos técnicos

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



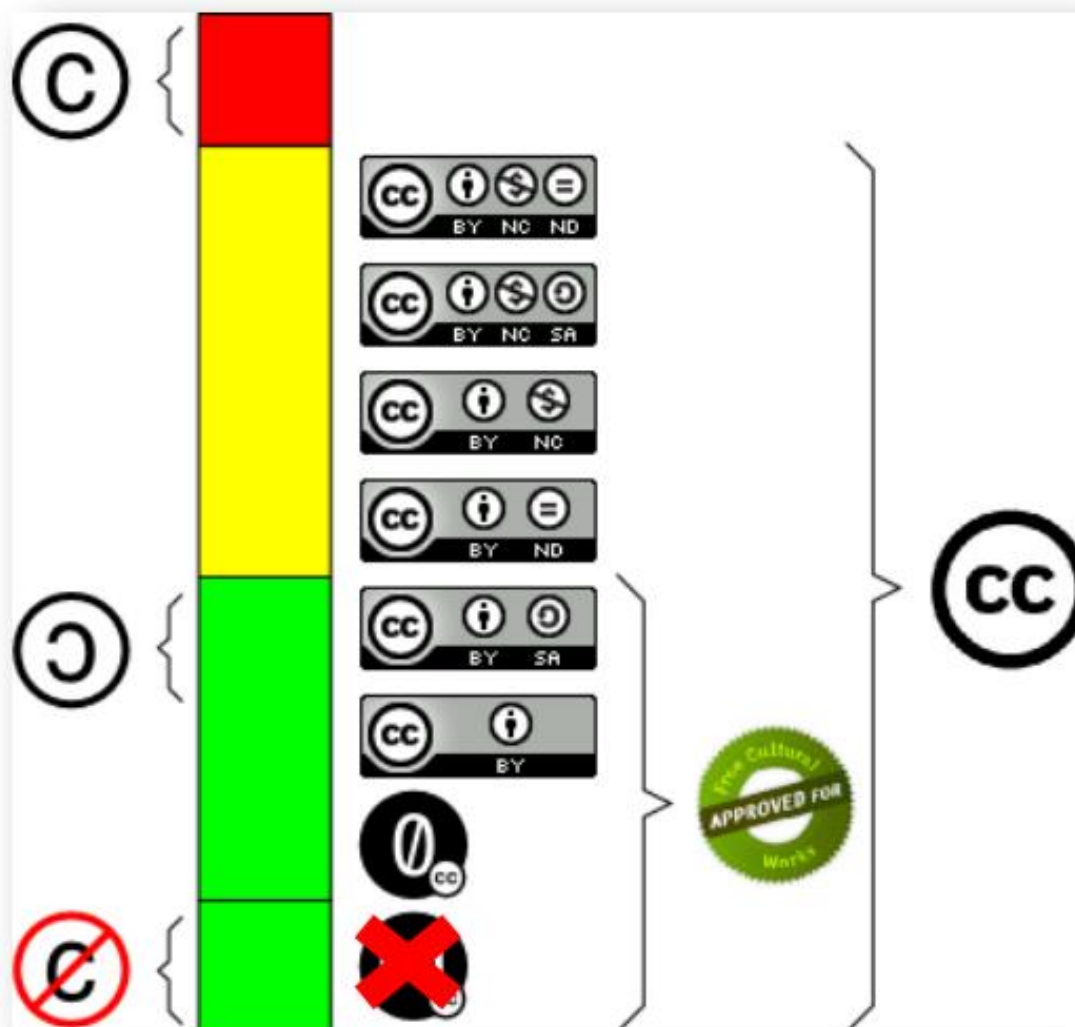
*"A piece of data or content is open if **anyone** is **free to use, reuse, and redistribute** it — subject only, at most, to the requirement to attribute and/or share-alike."*

The Open Definition



Aquél que hace sus datos públicos debe definir claramente la licencia bajo la cual lo hace, de manera que los derechos de reutilización queden claros para los potenciales usuarios.

Creative Commons



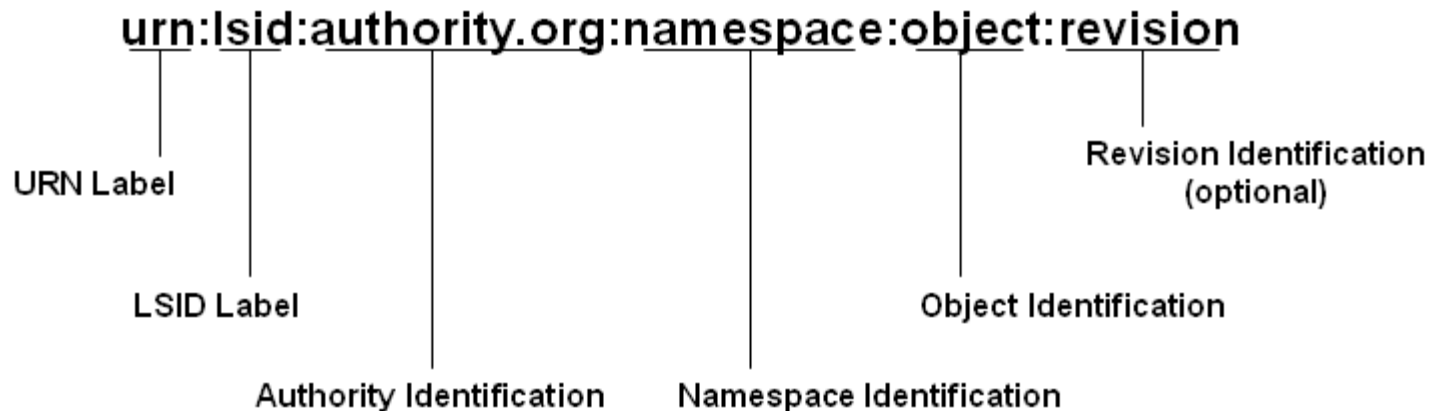
Datos. Licencias. Tabla resumen

Licencia	Dominio	BY	SA	Comentarios
CC BY	Contenido	SÍ	NO	
CC BY-SA	Contenido	SÍ	SÍ	
CC Zero	Contenido y datos	NO	NO	Dedicado al dominio público. Renuncia a todos los derechos
ODC – BY	Datos	SÍ	NO	
ODC – PDDL	Datos	NO	NO	Dedicado al dominio público. Renuncia a todos los derechos
ODC – ODbL	Datos	SÍ	SÍ	Restringe la compartición a los mismos términos que el original
CC BY-ND-*	Contenido	SÍ	S/N	Restringe la creación de trabajos derivados
CC BY-NC-*	Contenido	SÍ	S/N	Restringe el uso a actividades no comerciales

Identificadores

... únicos, persistentes, resolvibles

LSIDs

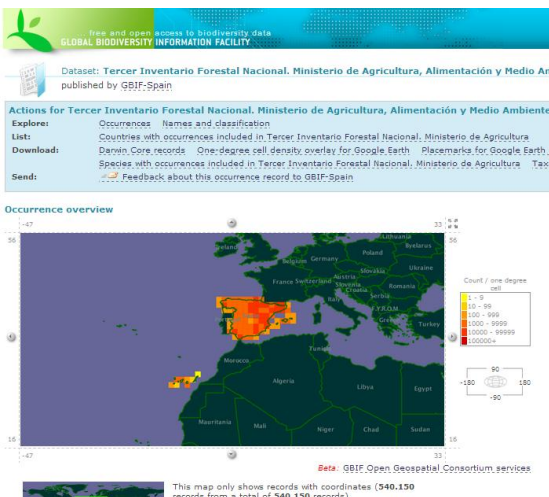


DOIs

<http://dx.doi.org/10.1080/0022293031000156312>

Un sistema de citación

- En GMP / EML
 - **Citation Identifier** - the URL or other unique identifier to be used to cite the resource.
 - **Resource Citation** - the traditional textual citation for the resource with author, date, and publisher information.
 - **Bibliography** - this header separates the single citation for the resource from the additional citations used to produce or as a result of the production of the resource.



	names spelling and authorship. More info about IFN3 at http://www.magrama.gob.es/es/biodiversidad/servicios
Rights:	by-nc-sa Attribution non-commercial share-alike
How to cite this dataset:	GBIF-Spain: Tercer Inventario Forestal Nacional. Ministerio de Agricultura, Alimentación y Medio Ambiente (España/Spain) (IFN3) (accessed through GBIF data portal 04-14)
Basis of record:	Unknown
Access point URL:	http://www.gbif.es/ficheros/Data_Resource/IFN3/DwC

Lista de chequeo para publicar datos

- Documentados
- Estandarizados
- “Mecanizable”
- Identificables
- Fáciles de encontrar
- Fáciles de reusar
- Metadatos
- Según estándar
- estructurados
- Con identificadores
- En portales, indexados
- Con licencias, con “como citar”

Francisco Pando

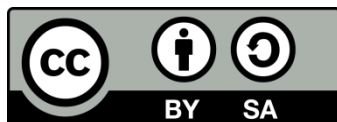
Unidad de coordinación, GBIF España

Real Jardín Botánico - CSIC

Claudio Moyano 1, 28014 Madrid

pando@gbif.es

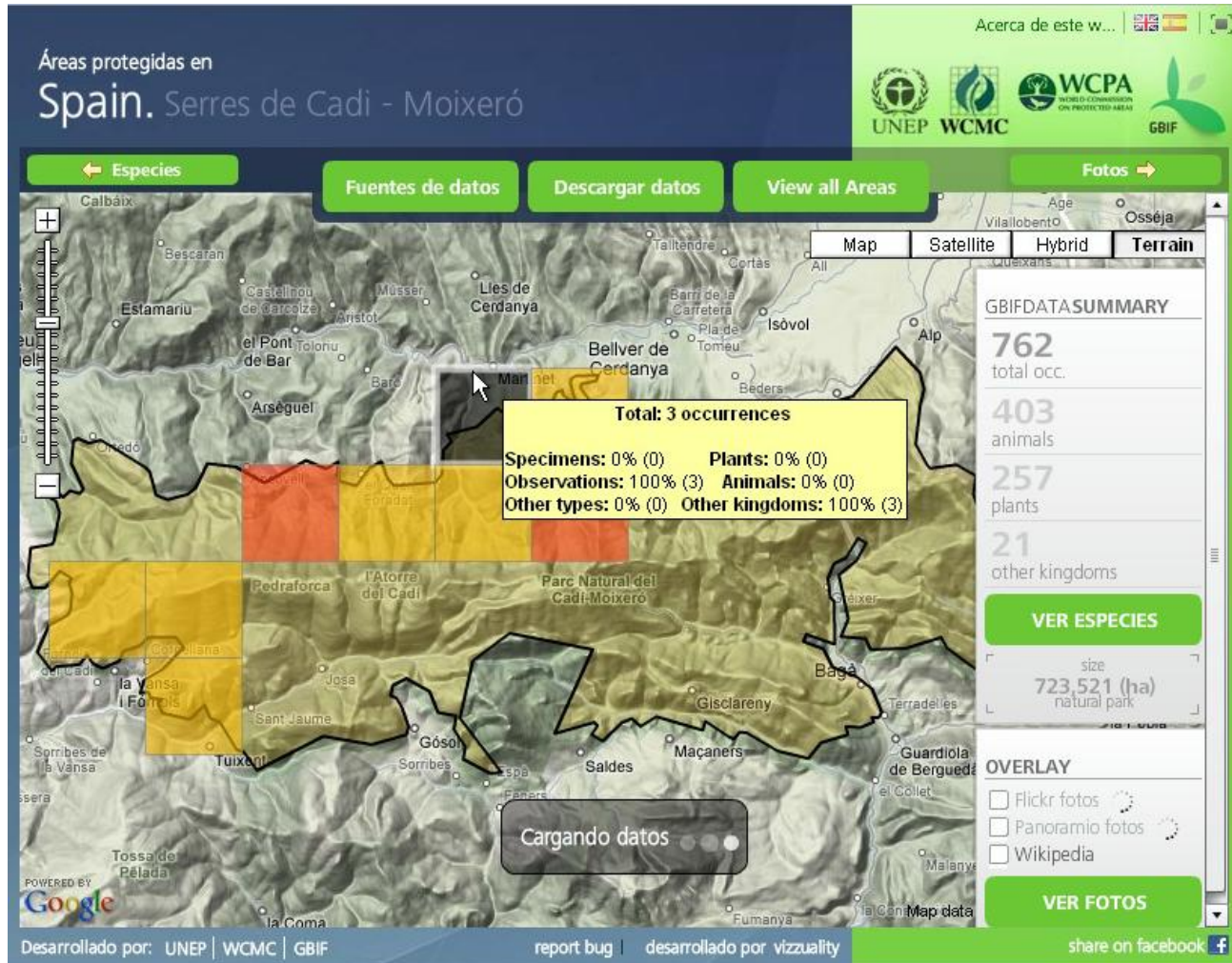
www.gbif.es



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



Integración con otras fuentes de datos



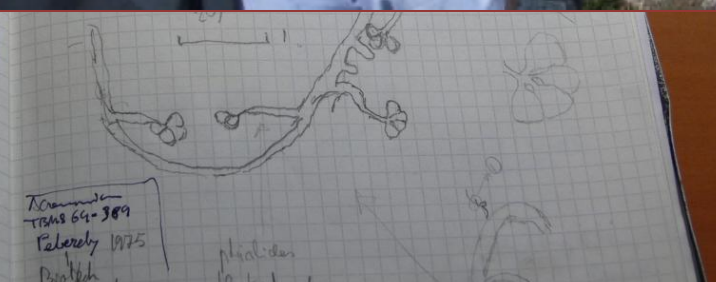
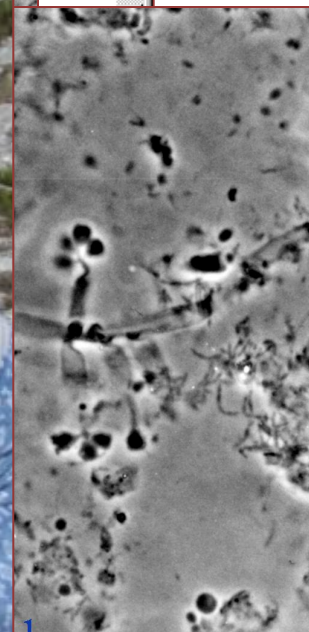
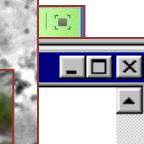
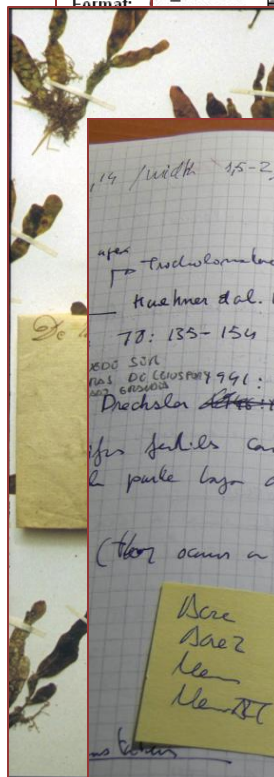
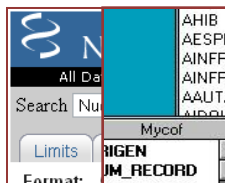
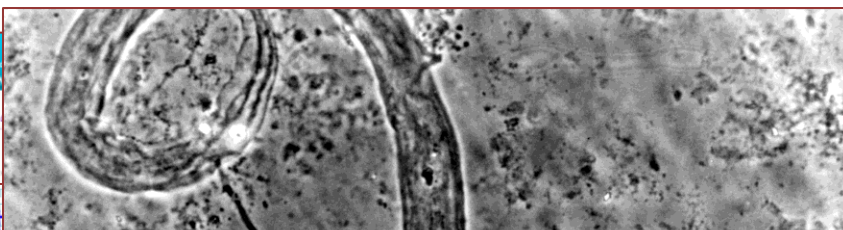
“Widget”
WCMC-GBIF:
Combina
información de
polígonos
(áreas
protegidas) con
datos de puntos
(GBIF)



... free and open access
GLOBAL BIODIVERSITY INFORMATION FINDER

Species: *Ulex europaeus*
Gorse

Kingdom: Unknown xGenus: *Ulex* xSpecies: *europaeus*



- ☐ 1. smooth
☐ 2. with spiculae
☐ 3. with small protuberances

A modo de conclusión...

- Lo que sabemos de la biodiversidad está extremadamente disperso
- Un acceso unificado y público a esta información es esencial no solo para hacer mejor ciencia, sino también mejor gestión y mejor conservación
- Los datos realmente se hacen útiles cuando estos se pueden enlazar y combinar, y para esto hay que publicar datos (no solo resultados) y usar estándares