

## VIII Taller GBIF.ES: Imágenes digitales para estudios de biodiversidad

4 – 6 de febrero de 2025, Jardí Botànic, Valencia

# Metadatos y estándares para imágenes digitales en biodiversidad

Francisco Pando  
GBIF España



VNIVERSITAT ID VALÈNCIA  
Jardí Botànic



M<sup>[UV]</sup>  
HN

*museus de la universitat  
de valència:  
història natural*

Museo de Ciencias  
Universidad de Navarra





# Contenido

- Metadatos: qué, para qué, cuales
- Dublin Core
- Exif
- IPTC
- XMP
- TDWG
- Audiovisual Core
- MIDS
- IPR & licencias

HOW I  
LEARNED  
TO STOP  
WORRYING  
AND  
LOVE  
METADATA

# Metadatos: qué

- Son datos sobre otros datos (el “recurso”)
- Recogen el contexto del recurso
- Pueden ser un subrogado del recurso; permitir conocerlo sin acceder al recurso

# Metadatos: para qué

- Descubrimiento
- Determinar la utilidad del recurso para fines concretos (fitness for use)
- Especificar las condiciones de uso del recurso (licencia, “como citar”)
- Ordenar, clasificar
- Herramienta de gestión

Y además:

Intercambiar, integrar, reutilizar

⇒ estándares

# Imágenes, biodiversidad y metadatos

**Darwin Core:** Es un estándar de metadatos utilizado específicamente en el ámbito de la **biodiversidad** - TDWG

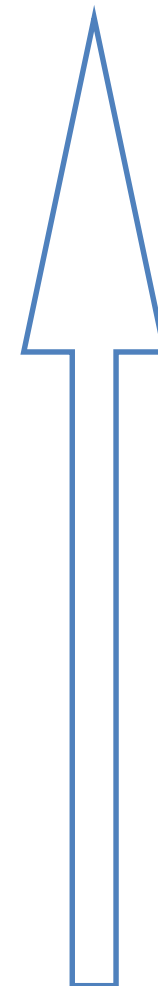
**Audiovisual Core:** Es un estándar de metadatos utilizado específicamente en el ámbito de la **biodiversidad** para recursos y colecciones **multimedia** - TDWG

**Dublin Core:** Es un estándar de metadatos **general**

**IPTC:** Estándar para archivos **multimedia** enfocado en la industria de noticias y medios

**EXIF:** Información técnica sobre la captura de **imágenes**

**XMP (Extensible Metadata Platform)**  
Es un estándar de metadatos para archivos **multimedia** - Adobe / MWG



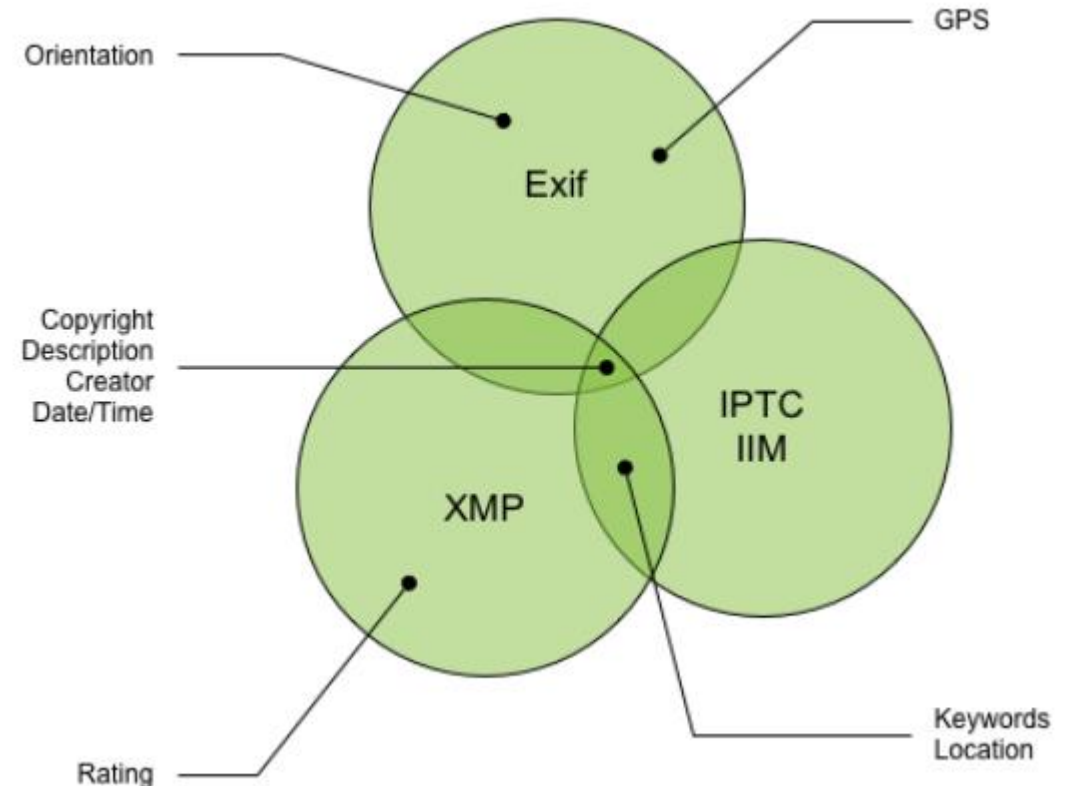
# Imágenes y estándares para metadatos

**Dublin Core:** Es un estándar de metadatos general

**EXIF:** Información técnica sobre la captura de imágenes

**IPTC:** Estándar para archivos multimedia enfocado en la industria de noticias y medios

**XMP (Extensible Metadata Platform)**  
Es un estándar de metadatos para archivos multimedia - Adobe / MWG



# 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, ..).
- Dos espacios de nombres:
  - **dc**: Este espacio de nombres contiene los 15 elementos originales del Conjunto de Elementos de Metadatos de Dublin Core, como title (título), creator (creador), subject (tema), etc.
  - **dcterms**: Este espacio de nombres incluye términos adicionales y refinamientos desarrollados para extender el conjunto original, proporcionando más granularidad y especificidad
- <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,...



# Dublin Core (un estándar en metadatos)

1. Title
2. Creator
3. Subject
4. Description
5. Publisher
6. Contributor
7. Date
8. Type
9. Format
10. Identifier
11. Source
12. Language
13. Relation
14. Coverage
15. Rights

Title: Lección de anatomía del Dr. Nicolaes Tulp

Creator : Rembrandt Harmenszoon van Rijn

Description : Cuadro del pintor neerlandés Rembrandt. Fue pintado en 1632. Se trata de una pintura al óleo sobre lienzo, que mide 169,5 centímetros de alto y 216,5 cm de ancho. El cuadro muestra una lección de anatomía con un grupo cirujanos impartida por el doctor Nicolaes Tulp. El doctor Nicolaes Tulp está representado explicando la musculatura del brazo a profesionales de la medicina. El cadáver es del criminal Aris Kindt (Adriaan Adriaanszoon), de 41 años, ahorcado ese mismo día por robo a mano armada. Algunos de los espectadores son varios patrones que pagaban comisiones por ser incluidos en la pintura

Type: Pintura al óleo

Coverage: 1632

Rights: Mauritshuis, La Haya

...



# Exif

***Exchangeable image file format*** (en español, *Formato de archivo de imagen intercambiable*; abreviatura oficial **Exif**, no EXIF)

Especificación de datos técnicos para cámaras digitales, creada desde el sector industrial

Las etiquetas (*tags*) de [metadatos](#) definidas en el estándar Exif cubren un amplio espectro incluido:

- Información de fecha y hora.
- Configuración de la cámara: estática (modelo de cámara y el fabricante y dependiente de la imagen (apertura, velocidad del obturador, distancia focal, medidor de exposición y sensibilidad ISO).
- Información sobre localización (GPS).
- Descripción e información sobre [copyright](#).

Datos embebidos en la imagen (JPEG, JPEG2000, PNG, TIFF pero no GIF)

IMG20241025141310 (002).jpg - IrfanView (Zoom: 683 x 910)

File Edit Image Options View Help

22.8

PERA VESSAMENTS

IMG20241025141310 (002).jpg - EXIF Info

EXIF Tag	Value
Filename	IMG20241025141310 (002).jpg
ImageWidth	3000
ImageLength	4000
ImageDescription	
Make	OPPO
Model	OPPO Reno4 Z 5G
Orientation	Top left
XResolution	72
YResolution	72
ResolutionUnit	Inch
Software	MediaTek Camera Application
Date Time	2024:10:25 14:13:10
YCbCrPositioning	Co-Sited
Exif-IFD-Offset	438
ExposureTime	1/50.0 seconds
FNumber	1.70
ExposureProgram	Not defined
ISO Speed Ratings	768
Recommended Exposure I...	0
ExifVersion	0220
Date Time Original	2024:10:25 14:13:10
Date Time Digitized	2024:10:25 14:13:10
Offset Time Original	+02:00
Components Configuration	YCbCr
ShutterSpeedValue	1/49 seconds
ApertureValue	F 1.70
BrightnessValue	6.20
ExposureBiasValue	0.00
MaxApertureValue	F 1.00
MeteringMode	Center weighted average
LightSource	Other
Flash	Flash not fired, compulsory flash mod...
FocalLength	4.71 mm
UserComment	
SubsecTime	216
SubsecTimeOriginal	216
SubsecTimeDigitized	216
FlashPixVersion	0100
ColorSpace	sRGB
ExifImageWidth	3000
ExifImageHeight	4000
InteroperabilityOffset	1650
SensingMethod	None
ExposureMode	Auto
White Balance	Auto
DigitalZoomRatio	100.00 x

Copy to clipboard

More options... Exit

3000 x 4000 x 24 BPP 54/83 23% 1.84 MB / 34.33 MB 25/10/2024 / 14:38:00



IMG20241025141310.jpg

Descripción	IPTC	IPTC Extension	Datos de cámara	Datos de vídeo
Marca:	OPPO		Dimensión X en píxeles:	3000
Modelo:	OPPO Reno4 Z 5G		Y:	4000
Fecha y hora:	25/10/2024 – 14:13:10		Orientación:	Normal
Velocidad del obturador:	1/50 seg		Resolución X:	72
Programa de exposición:	No definido		Y:	72
F-Stop:	f/1.7		Unidad de resolución:	Pulgada
Valor de apertura:	f/1.7		Bits comprimidos por píxel:	
Valor máximo de apertura:	f/1.0		Espacio de color:	sRGB
Índices de velocidad ISO:	768		Fuente de luz:	Otra fuente de lu...
Distancia focal:	4.7 mm		Origen de archivo:	
Lente:				
Flash:	No disparado			
	Sin detección de luz estrobosc...			
	Supresión obligatoria de flash (2)			
	Función de flash presente			
	Sin reducción de ojos rojos			
Modo de medición:	Centrar media ponderada			

Powered By xmp™

Preferencias Importar... Aceptar



# IPTC

En rigor: IPTC-IIM: “International Press Telecommunications Council”- “Information Interchange Model”

Creado para apoyar un flujo de trabajo concreto: “del fotógrafo a la noticia”, y un fin específico: “proteger los intereses de la industria”

Primera versión en 2004; actualmente integrado en XMP

# Metadata Working Group

- Metadata Working Group
- Based on a 2006 proposal by Microsoft, the Metadata Working Group (MWG) organization was
- created in 2007 by 5 founding members: Apple, Adobe, Canon, Microsoft and Nokia. Sony joined this
- initiative in 2008

# XMP

- Adobe's Extensible Metadata Platform
- XMP es una tecnología de etiquetado que permite incrustar metadatos directamente en un archivo.
- Permite automatización de flujos de trabajo, gestión de derechos y muchas otras posibles mejoras.
- estandariza la definición, creación y procesamiento de metadatos extensibles
- XMP se convierte en contenedor para los términos de Exif e IPTC-IIM

```
Xml Copiar
<?xml version="1.0" encoding="UTF-8"?>
<x:xmpmeta xmlns:x="adobe:ns:meta/">
  <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:dc="http://purl.org/dc/elements/1.1/"
    xmlns:xmp="http://ns.adobe.com/xap/1.0/"
    xmlns:xmpRights="http://ns.adobe.com/xap/1.0/rights/"
    xmlns:photoshop="http://ns.adobe.com/photoshop/1.0/">
    <dc:title>
      <rdf:Alt>
        <rdf:li xml:lang="x-default">Ejemplo de título</rdf:li>
      </rdf:Alt>
    </dc:title>
    <dc:creator>
      <rdf:Seq>
        <rdf:li>Autor del ejemplo</rdf:li>
      </rdf:Seq>
    </dc:creator>
    <dc:description>
      <rdf:Alt>
        <rdf:li xml:lang="x-default">Descripción del archivo</rdf:li>
      </rdf:Alt>
    </dc:description>
    <xmp:CreateDate>2023-04-06T00:00:00Z</xmp:CreateDate>
    <xmp:ModifyDate>2023-04-06T00:00:00Z</xmp:ModifyDate>
    <xmpRights:UsageTerms>
      <rdf:Alt>
        <rdf:li xml:lang="x-default">Derechos de uso</rdf:li>
      </rdf:Alt>
    </xmpRights:UsageTerms>
    <photoshop:Headline>Encabezado del Photoshop</photoshop:Headline>
  </rdf:RDF>
</x:xmpmeta>
```

# El contexto del contexto



Biodiversity Information Standards **TDWG**

## Biodiversity Information Standards (TDWG)

We are a non-profit organization and a community dedicated to developing biodiversity information standards.

Image by Jennifer Latuperisa-Andresen

### Actively maintained

#### ABCD

Access to Biological Collection Data (ABCD) Schema

#### Audiovisual Core

Audiovisual Core Multimedia Resources Metadata Schema

#### Darwin Core

Darwin Core

#### GUID Applicability Statements

GUID and Life Sciences Identifiers Applicability Statements

#### Latimer Core

Latimer Core, a standard for describing collections

#### SDS

TDWG Standards Documentation Standard (SDS)

#### TAPIR

TDWG Access Protocol for Information Retrieval (TAPIR)

#### TCS

Taxon Concept Schema (TCS)

#### VMS

Vocabulary Maintenance Standard (VMS)

#### WGSRPD

World Geographical Scheme for Recording Plant Distribution

### Not under active maintenance

#### Economic Botany

Economic Botany Data Collection Standard

#### Floristic Regions

Floristic Regions of the World

#### HISPID3

Herbarium Information Standards and Protocols for Interchange of Data (HISPID3)

#### ITF2

International Transfer Format for Botanic Garden Plant Records (ITF2)

#### Plant Names

Plant Names in Botanical Databases

#### POSS

Plant Occurrence and Status Scheme (POSS)

#### SDD

Structured Descriptive Data (SDD)

#### XDF

XDF - A Language for the Definition and Exchange of Biological Data Sets

### Externally maintained

See this page

# Audiovisual Core Multimedia Resources Metadata Schema

The Audiovisual Core (AC) is a set of vocabularies designed to represent metadata for biodiversity multimedia resources and collections. These vocabularies aim to represent information that will help to determine whether a particular resource or collection will be fit for some particular biodiversity science application before acquiring the media. Among others, the vocabularies address such concerns as the management of the media and collections, descriptions of their content, their taxonomic, geographic, and temporal coverage, and the appropriate ways to retrieve, attribute and reproduce them.

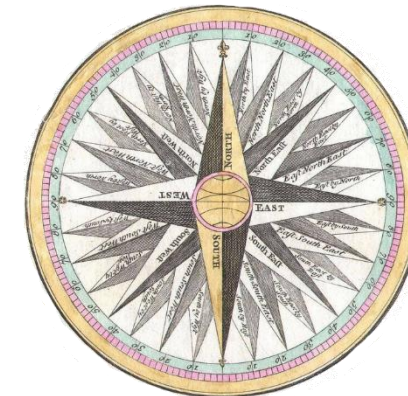
[Website](#)[GitHub](#) 

GBIF/TDWG Multimedia Resources Task Group. 2013. Audiovisual Core Multimedia Resources Metadata Schema. Biodiversity Information Standards (TDWG)

<http://www.tdwg.org/standards/638>



# AC, conceptos claves



AC es un conjunto de vocabularios de metadatos para describir recursos y colecciones multimedia relacionados con la biodiversidad

Los recursos multimedia son artefactos digitales o físicos que normalmente comprenden más que texto

imágenes, obras de arte, dibujos, fotografías, sonido, video, animaciones, materiales de presentación y medios interactivos en línea

Una colección multimedia es un conjunto de tales objetos

Reutiliza:

- Dublin Core (DC),
- Darwin Core (DwC),
- International Press and Telecommunications Council (IPTC),
- Adobe Extensible Metadata Platform (XMP),
- Metadata Working Group (MWG) schema,
- Natural Collections Schema (NCD)

Un registro AC es una descripción de un recurso multimedia

Cada término se identifica mediante un identificador uniforme de recursos (URI) único

Representation-neutral

Representations? > CSV, XML, RDF, JSON

# AC: elementos obligatorios

1. **Identifier** (`dcterms:identifier`): An arbitrary code that is unique for the resource, with the resource being either a provider, collection, or media item. Whereas the identifier must be globally unique for providers and collections (e. g. a URI), identifiers for media items may be unique only within the context of a collection or provider.
2. **Type** (`dcterms:type`): Any dcmi type term from <http://dublincore.org/documents/dcmi-type-vocabulary/#section-7-dcmi-type-vocabulary> may be used. Recommended terms are Collection, StillImage, Sound, MovingImage, InteractiveResource, and Text.
3. **Metadata Language** (`ac:MetadataLanguage`): Language of description and other metadata .
4. **Copyright Statement** (`dcterms:rights`): Information about rights held in and over the resource. A full-text, readable copyright statement, as required by the national legislation of the copyright holder. On collections, this applies to all contained objects, unless the object itself has a different statement. When available, it is also recommended to provide the Copyright Owner using `xmpRights:Owner`

elementos = términos = propiedades

# AC: vocabularios

- 
1. **Management Vocabulary**
2. **Attribution Vocabulary**
3. **Agents Vocabulary**
4. **Content Coverage Vocabulary**
5. **Geography Vocabulary**
6. **Temporal Coverage Vocabulary**
7. **Taxonomic Coverage Vocabulary**
8. **Resource Creation Vocabulary**
9. **Related Resources Vocabulary**
10. **Service Access Point Vocabulary**
11. **Region of Interest Vocabulary**
- Metadatos de los metadatos
- IPR & credits, quien participó y como
- Qué, donde, cuando
- Exif, de qué
- Exif, acceso, licencia, ROIs

•137 términos

<https://ac.tdwg.org/termlist/>

# Ejemplo

<b>ac:associatedSpecimenReference</b>	04d9efc4-8b47-4545-b9d9-781508392b44
<b>dcterms:identifier</b>	00a1388e-bc5c-46f2-ab3f-d5b7ce4d6655
<b>dc:type</b>	StillImage
<b>ac:subtypeLiteral</b>	
<b>ac:title</b>	PP29675_image_1
<b>xmp:MetadataDate</b>	09/04/2019
<b>ac:metadataLanguageLiteral</b>	eng
<b>ac:providerManagedID</b>	744562
<b>ac:hasServiceAccessPoint</b>	<a href="https://mm.fieldmuseum.org/00a1388e-bc5c-46f2-ab3f-d5b7ce4d6655">https://mm.fieldmuseum.org/00a1388e-bc5c-46f2-ab3f-d5b7ce4d6655</a>
<b>dc:rights</b>	[Copyright] Field Museum of Natural History - CC BY-NC
<b>xmpRights:Owner</b>	Field Museum of Natural History
<b>xmpRights:WebStatement</b>	<a href="https://www.fieldmuseum.org/field-museum-natural-history-conditions-and-suggested-norms-use-collections">https://www.fieldmuseum.org/field-museum-natural-history-conditions-and-suggested-norms-use-collections</a>
<b>photoshop:Credit</b>	Please cite this as: (c) The Field Museum (DATE) CC-BY-NC
<b>dc:creator</b>	Field Museum of Natural History - Botany Department   Robert North : Field Museum of Natural History

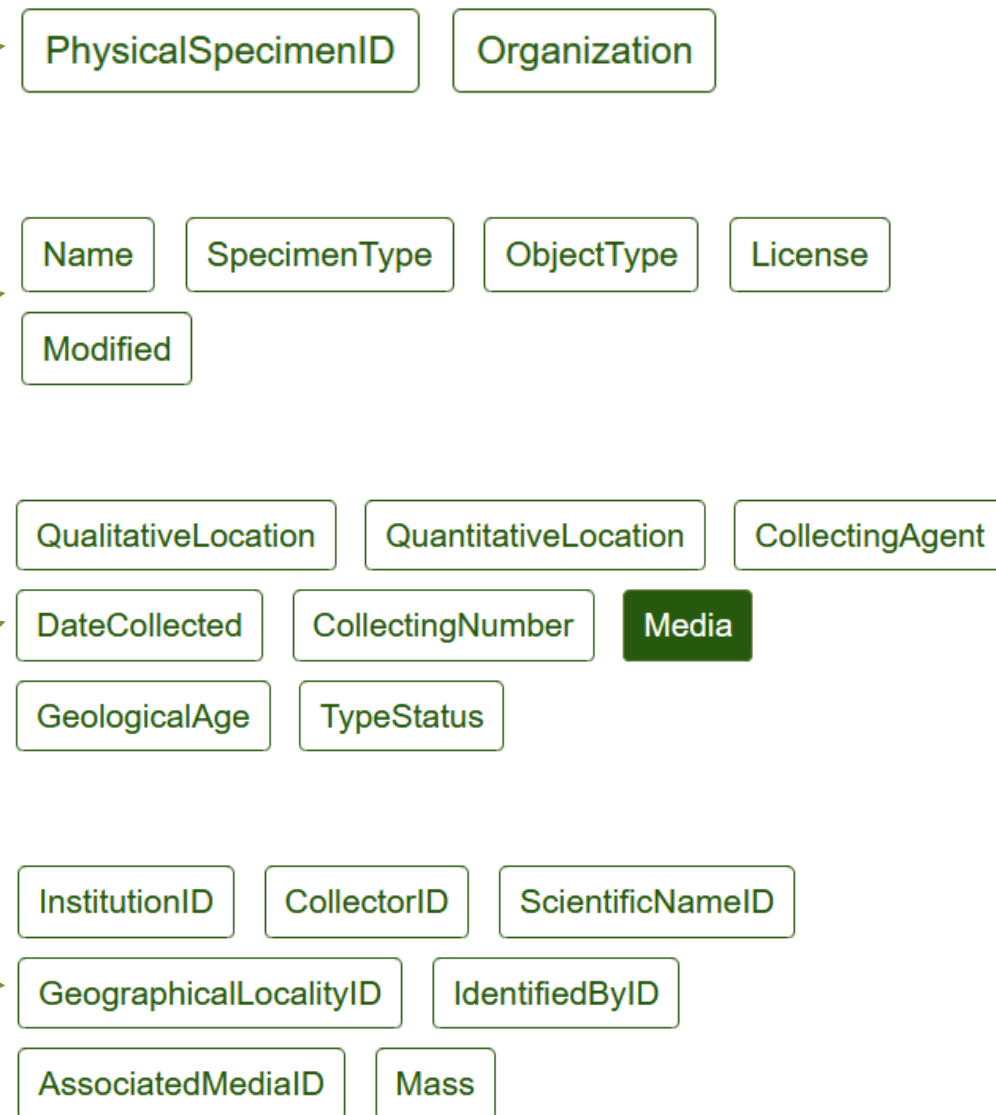
<b>ac:providerLiteral</b>	Field Museum of Natural History
<b>dcterms:description</b>	PP 29675 A+B [HS, M] Pinnularia, Moscovian / Desmoinesian, Francis Creek Shale Member, United States of America, Illinois, Will, Mazon Creek Region
<b>ac:tag</b>	Fossil Ferns
<b>xmp:CreateDate</b>	22-jul-14
<b>ac:IDofContainingCollection</b>	<a href="http://biocol.org/urn:lsid:biocol.org:col:34795">http://biocol.org/urn:lsid:biocol.org:col:34795</a>
<b>ac:accessURI</b>	<a href="https://fm-digital-assets.fieldmuseum.org/744/562/PP29675_image_1.jpg">https://fm-digital-assets.fieldmuseum.org/744/562/PP29675_image_1.jpg</a>
<b>dc:format</b>	jpeg
<b>ac:variantLiteral</b>	mediumQualityFurtherInformationURL
<b>ac:hashFunction</b>	MD5
<b>ac:hashValue</b>	9b7ad87adb8f5ada19e91a202f4339b1
<b>exif:PixelXDimension</b>	1500
<b>exif:PixelYDimension</b>	1908

**ac | dc @ dcTerms | xmp | xmpRights | photoshop | exif**

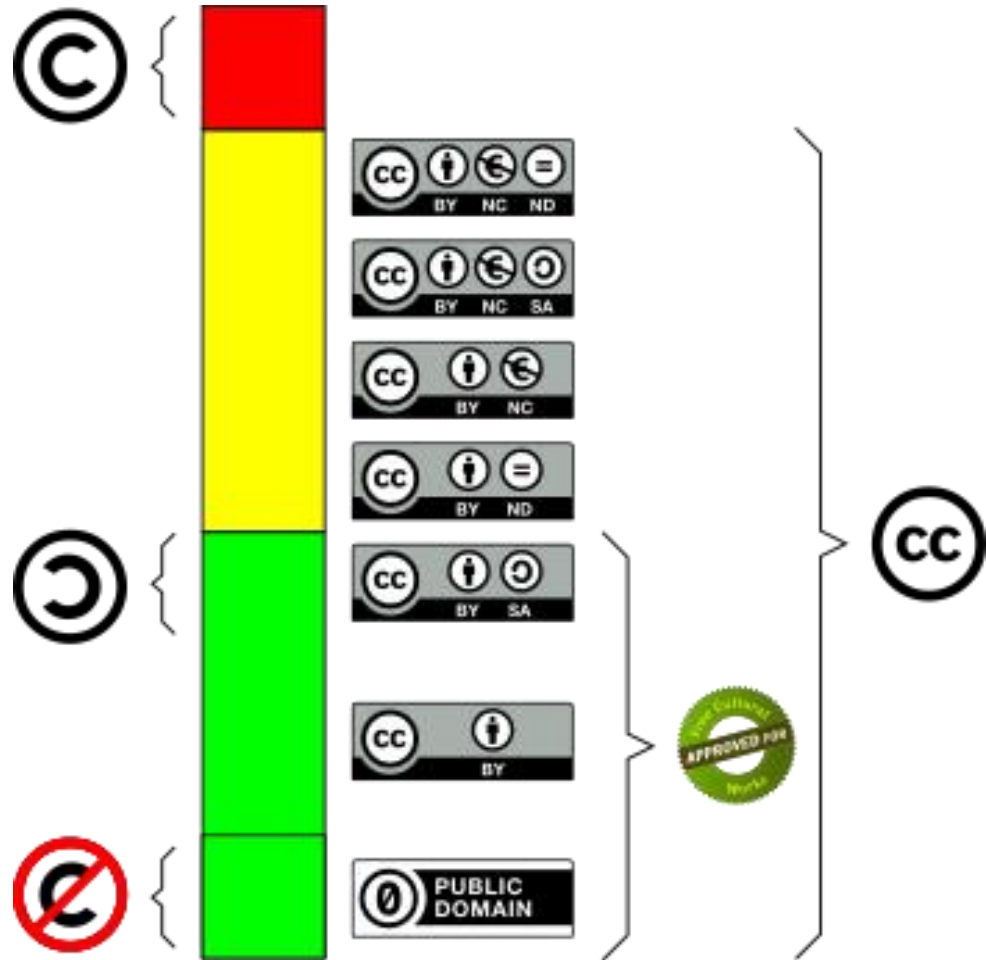
# MIDS

- [The Minimum Information about a Digital Specimen \(MIDS\) specification](#)
  - <https://github.com/tdwg/mids>
- <https://collections.naturalsciences.be/cpb/cetaf-passport-and-collections-registry-manual/collections/collections-registry/mids>

1. MIDS-0: **bare catalogue** level information, unique identifiers, etc.; images if available;
2. MIDS-1: **basic** information derived from the collection, such as a (higher) taxon and geography;
3. MIDS-2: **regular** information of what (valid name), where, when, by whom, and how; as derived from specimen labels;
4. MIDS-3: **extended** information, enriched with external sources, not directly available from labels.



# IPR y Copyright



## Derechos de propiedad intelectual (Intellectual Property Rights, IPR)

- Derechos morales: Su finalidad proteger a la persona del autor a través de su obra. Son irrenunciables e inalienables. (p.ej. Reconocimiento de autoría; respeto a la integridad de la obra)
- Derechos de carácter patrimonial: se pueden ceder o transferir: Explotación (reproducción, distribución, comunicación, acceso, transformación...) y compensatorios (p.ej. por ceder d. de reproducción)
- Si hay una licencia –aunque sea CC, o incluso CC0– tiene que poderse identificar quien la concede. Eso se hace bien con indicación de ©, bien indicando el autor
- Una base de datos (=juego de datos) puede tener un tipo de licencia y su contenido, o parte de su contenido, otro (p.ej., una BBDD con imágenes)

Francisco Pando

Unidad de Coordinación de GBIF.ES

CSIC

Joaquín Costa 22, 28002 Madrid

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



<https://creativecommons.org/licenses/by/4.0/deed.es>

GBIF-ES es el Nodo Nacional de Información en Biodiversidad patrocinado por el Ministerio de Ciencia, Innovación y Universidades, gestionado por el Consejo Superior de Investigaciones Científicas (CSIC). <https://www.gbif.es/> <https://datos.gbif.es/>, <https://elearning.gbif.es>



MINISTERIO  
DE CIENCIA, INNOVACIÓN  
Y UNIVERSIDADES



**CSIC**  
CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS

