

# Calidad de datos: detección y corrección

Taller sobre calidad en bases de  
datos sobre biodiversidad

25-26 noviembre 2008. Madrid.

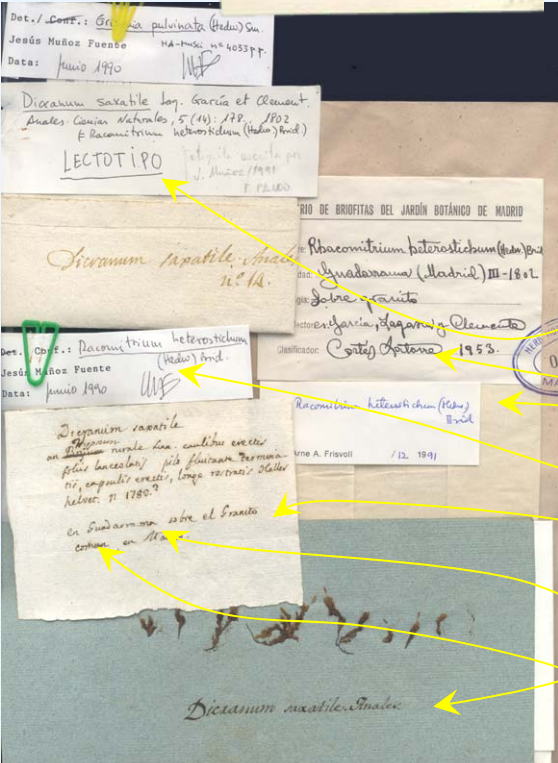
Francisco Pando  
Unidad de Coordinación GBIF-España



# Guión

- **Conceptos**
- **Ciclo de vida de los datos y detección**
- **Introducción de datos: interpretar, distorsionar, errar**
- **Captura de errores: validación y detección**
- **Diseño de base de datos y gestión de errores**
- **Virtualidad y realidad en las BBDD**
- **Casos**
- **La casa por los cimientos. Un ejemplo de Eslovenia**

# Información original y derivada



Herbario - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Refresh Home Search Favorites Media History

Address [/asp/dos.asp?conque=MA-Musci&fnhe=4033&fadic=1](#) Go Google

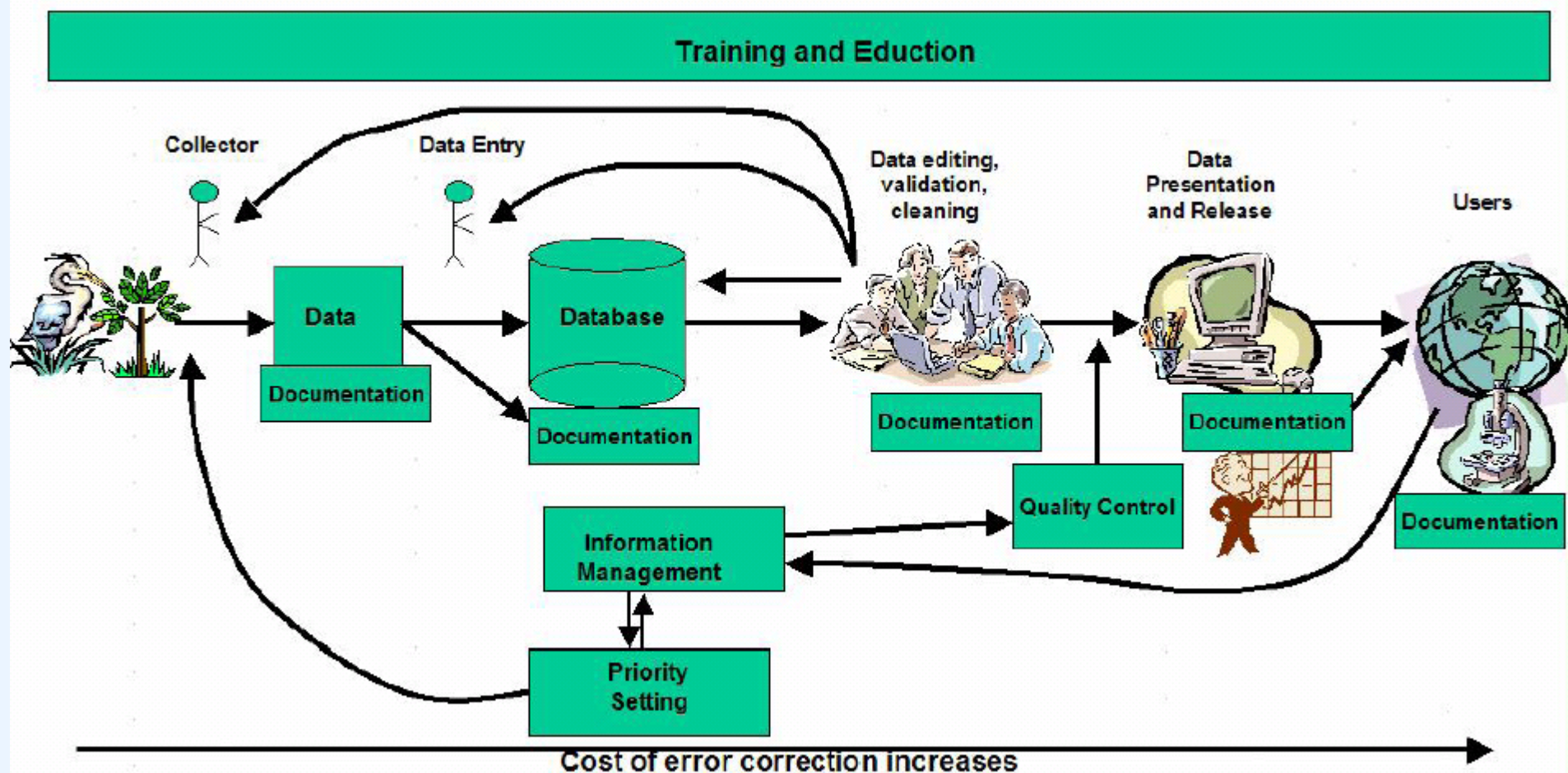
RJB > Colecciones > Herbario > Criptogamia > Bases de datos > Búsqueda

Resultados

Dicranum saxatile	Typus (L)	1
Racomitrium heterostichum (Hedw.) Brid.		2
Det. Cortés Latorre (/1953)		
Racomitrium heterostichum (Hedw.) Brid.		3
Det. J. Muñoz Fuente (05/1990)		
Dicranum saxatile Lagasca, García & Clemente	Typus (L)	4
Det. J. Muñoz (/1991)		
Racomitrium heterostichum (Hedw.) Brid		5
Det. Arne A. Frisvoll (12/1991)		

ESP. M. Guadarrama  
sobre granito  
García, Lagasca & Clemente  
III-1802  
**MA-Musci 4033**  
Nombre aceptado en el Herbario : Racomitrium heterostichum (Hedw.) Brid

# Ciclo de vida de los datos y detección



Chapman, A. D. 2005. *Principles of Data Quality*, version 1.0. Report for GBIF

# Introducción de datos: interpretar, distorsionar, error

**Una base de datos debe conciliar dos requisitos en extremo irreconciliables:**

- 1. Representar la realidad**
- 2. Que la información se encuentre, se pueda comparar y analizar ( $\Rightarrow$  estandarizar y normalizar  $\Rightarrow$  interpretar)**

**Estándares (o referencias) explícitos usados con consistencia**

# Componentes de la validación

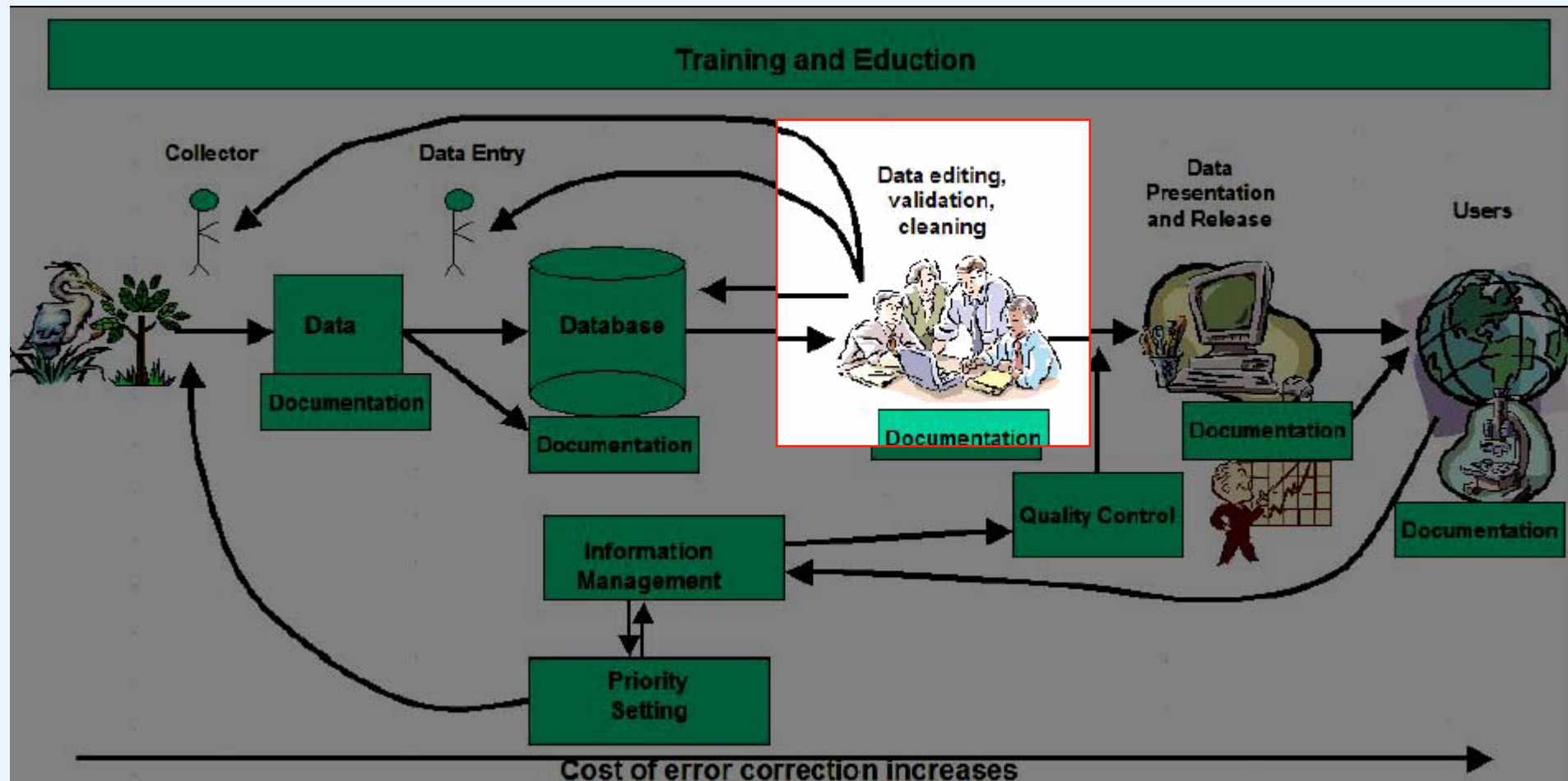
Validación = detectar y corregir:

- Inexactitudes
- Falta de datos
- Potenciales errores (datos no razonables)

La validación trata de garantizar la veracidad del registro



# Captura de errores: validación y detección



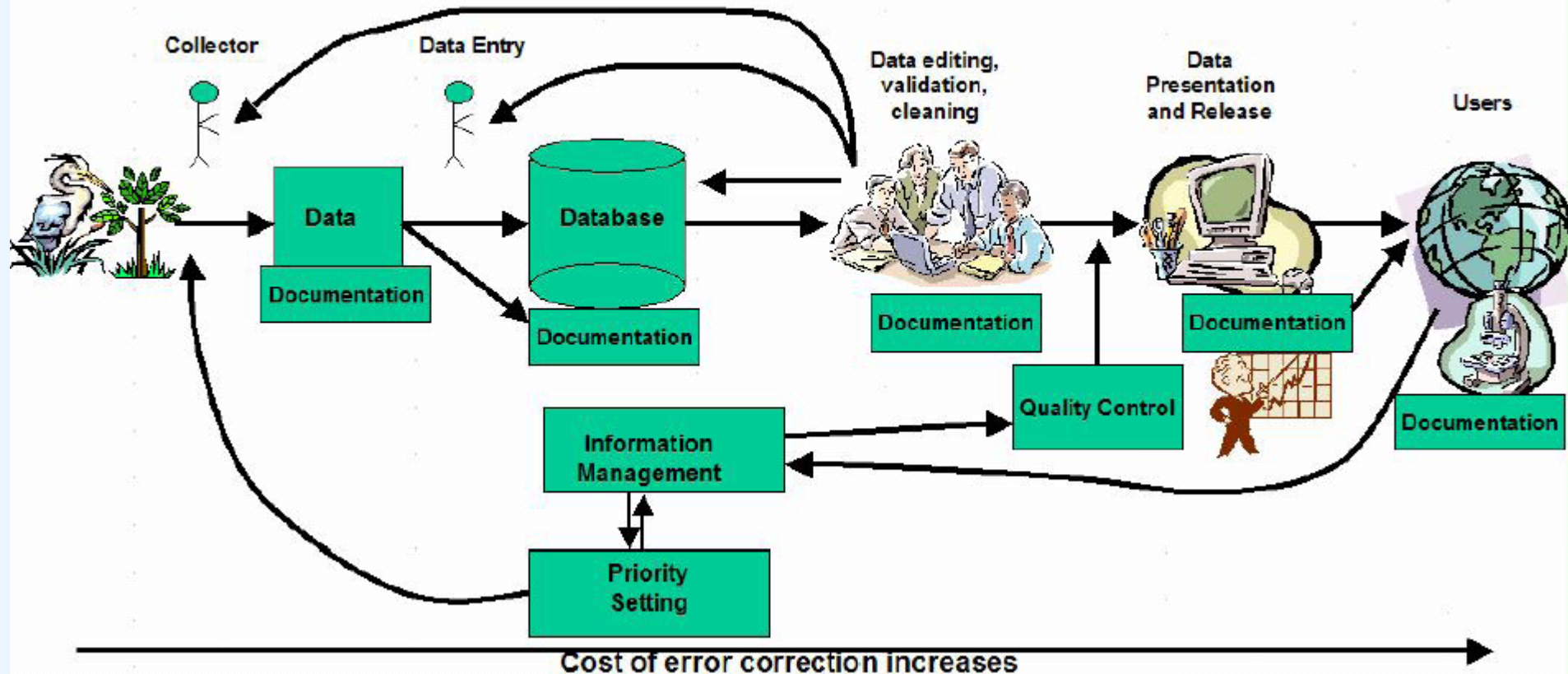
- **Validación:** asegura que los datos se corresponden con el objeto. Incluye también el control de inexactitudes y pérdida de datos

# Procedimiento

- **Establecer procedimientos que no conlleven pérdida de datos**
  - Las reglas de normalización deben cubrir todos los casos
  - Los procedimientos no deben bloquear el progreso del trabajo ante casos no contemplados o dudas

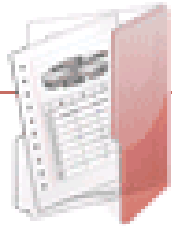


## Training and Education



Procedimientos que permita que los errores detectados se corrijan en la base de datos

# Por ejemplo:



## Occurrence Details: MA MA

*Echinostelium apitectum*

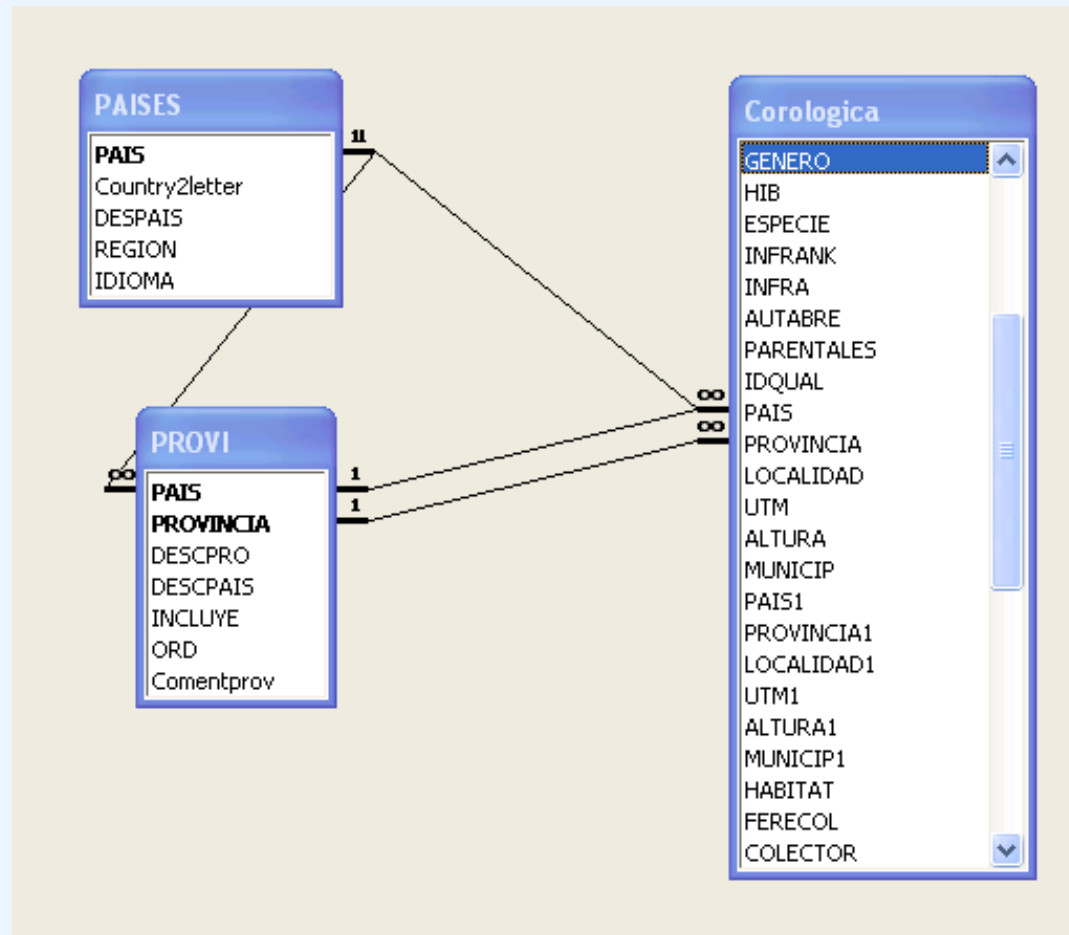
### Actions

- Find:** [Records of Madrid: MA-Fungi fr \(2.0°E, 39.0°N, 3.0°E, 40.0°N\)](#)
- Retrieve:** [Retrieve original record from data](#)
- View:** [View information for \*Echinostel\*](#)
- Send:** [Feedback to GBIF-Spain](#)

Proporcionar al usuario la posibilidad de reportar errores o comentarios

<http://data.gbif.org/occurrences/76377384/>

# Diseño de base de datos y gestión de errores



- Diseños normalizados de bases de datos

# Diseño de base de datos y gestión de errores

- Recomendación: distintos formatos de la misma información debe calcularse de un campo a otro para su uso específico y no duplicar la información en campos paralelos

Campo UTM: 30TUF345871  
Campo UTM10: 30TUF38

... Y si..

Campo UTM: 30TUF345871  
Campo UTM10: 30TUF31

Campo UTM: 30TUF345871

```
-----  
Function UTM10(coord) As String  
Dim Uso, zona, cien, Resto, Equis,  
Dim lenresto As Byte  
If IsNull(coord) Or Len(coord) < 5  
    UTM10 = "no utm"  
    GoTo fin  
End If
```

Campo UTM: 30TUF38  
*temporal  
calculado*

# Virtualidad y realidad en las BBDD

HORTUS REGIUS MATRITENSIS (MA-Fungi)

Martellia sp.

ESP. CIUDAD REAL: Viso del Marqués, Arroyo de la Poveda, bajo Halimium sp., 23-XI-1993, T. Pérez Jarauta, det. F.D. Calonge

MA-FUNGI 32269



*Zelleromyces piennensis* Horno-Arroyo, Viso del Marqués  
F. Calonge

Rev. J. Vidal

Dat. 25-5-2000

GenBank Acc. No. AF215649  
MA-Fungi 32269

Cuestión para debatir

¿Debe el objeto reflejar toda la información acumulada a partir del el?  
por ejemplo: ref. al Genbank, nombre de la camisa (Bot.) o de colección (Zool.)

# Casos (1)

**Daldinia concentrica (Bolton : Fr.) Ces. & de Not.**



ESP. MII: Sporlas  
sobre rama de muerta de *Ulmus minor*

*A. Martínez de Azagra*

Det. ~~Dr. C. Castiella~~

11-sep-1987

**MA-Fungi 22456**

Rev. ~~Dr. C. Castiella, 1987~~

Nombre aceptado en el herb. / herb. current name: *Daldinia concentrica* (Bolton : Fr.) Ces. & de Not.

**“Misma especie, misma localidad, al mismo sobre”**

¡no hacer en casa!



# Casos (2)

Bovista pusilla (Batsch) Pers. 1

Bovista dermoxantha (Vittad.) De Toni 2

Bovista aestivalis (Bonord.) Demoulin 3

Det. ~~F. S. G. G. G.~~ (01/1994)

Bovista pusilla (Batsch) Pers. 4

Det. ~~F. S. G. G. G.~~ (05/1994)

Bovista aestivalis (Bonord.) Demoulin  5

Det. ~~F. S. G. G. G.~~ (05/2000)

ESP. A1: Sierra de las Estancias , 1300 m

*J. A. Oria de Rueda*

28-dic-1987

**MA-Fungi 21820** Rev. ~~F. S. G. G. G.~~

Nombre aceptado en el Herbario : Bovista aestivalis (Bonord.) Demoulin

“Las malas identificaciones son errores que se han de corregir [i.e. eliminar]”

¡no hacer en casa!

# Casos (3)

Alexandria Digital Library Gazetteer Server Client. Using ADL Server [Use ESRI Server](#)

Search Result: 26 matches ([Problem Report](#))

#	names	Class
1	<a href="#">La Cabrera - Cundinamarca, Departamento de - Colombia</a>	populated places
2	<a href="#">La Cabrera - Ciego de Avila, Provincia de - Cuba</a>	populated places
3	<a href="#">Cabrera, Sierra de la - Spain</a>	ridges
4	<a href="#">La Cabrera - Salta, Provincia de - Argentina</a>	housing areas
5	<a href="#">La Cabrera - Spain</a>	populated places
6	<a href="#">Mas de la Cabrera - Spain</a>	populated places
7	<a href="#">La Cabrera, Fila - Venezuela</a>	ridges
8	<a href="#">La Puerta de Cabrera - Durango, Estado de - Mexico</a>	populated places
9	<a href="#">La Cabrera - Cojedes, Estado - Venezuela</a>	populated places
10	<a href="#">La Cabrera, Hacienda - Miranda, Estado - Venezuela</a>	agricultural sites
11	<a href="#">Cabrera - Queretaro de Arteaga, Estado de - Mexico</a>	populated places
12	<a href="#">La Cabrera, Peninsula de - Carabobo, Estado - Venezuela</a>	capes
13	<a href="#">La Cabrera - Santander, Departamento de - Colombia</a>	populated places
14	<a href="#">Cabrera, Sierra de la - Spain</a>	ridges
15	<a href="#">Cabrera Baja, Sierra de la - Spain</a>	mountains
16	<a href="#">La Cabrera, Sierra de - Spain</a>	ridges
17	<a href="#">La Cabrera, Rio de - Spain</a>	streams
18	<a href="#">La Cabrera, Pena - Colombia</a>	mountains
19	<a href="#">La Cabrera - Tachira, Estado - Venezuela</a>	populated places
20	<a href="#">La Cabrera - Miranda, Estado - Venezuela</a>	populated places
21	<a href="#">La Cabrera, Quebrada - Miranda, Estado - Venezuela</a>	streams
22	<a href="#">La Cabrera - Venezuela</a>	populated places
23	<a href="#">Cabrera - Santander, Departamento de - Colombia</a>	populated places
24	<a href="#">La Cabrera - Miranda, Estado - Venezuela</a>	populated places
25	<a href="#">La Cabrera - Choluteca, Departamento de - Honduras</a>	populated places
26	<a href="#">La Cabrera - Spain</a>	populated places



Please set at least one search condition.

Location:  within map  overlaps map  anywhere

Place Name:

Feature Type: ( [Feature Type Thesaurus](#) )

--Any Type--

**General Categories**

administrative areas  
hydrographic features  
land parcels  
manmade features  
physiographic features  
regions

**administrative areas**

military areas

Search

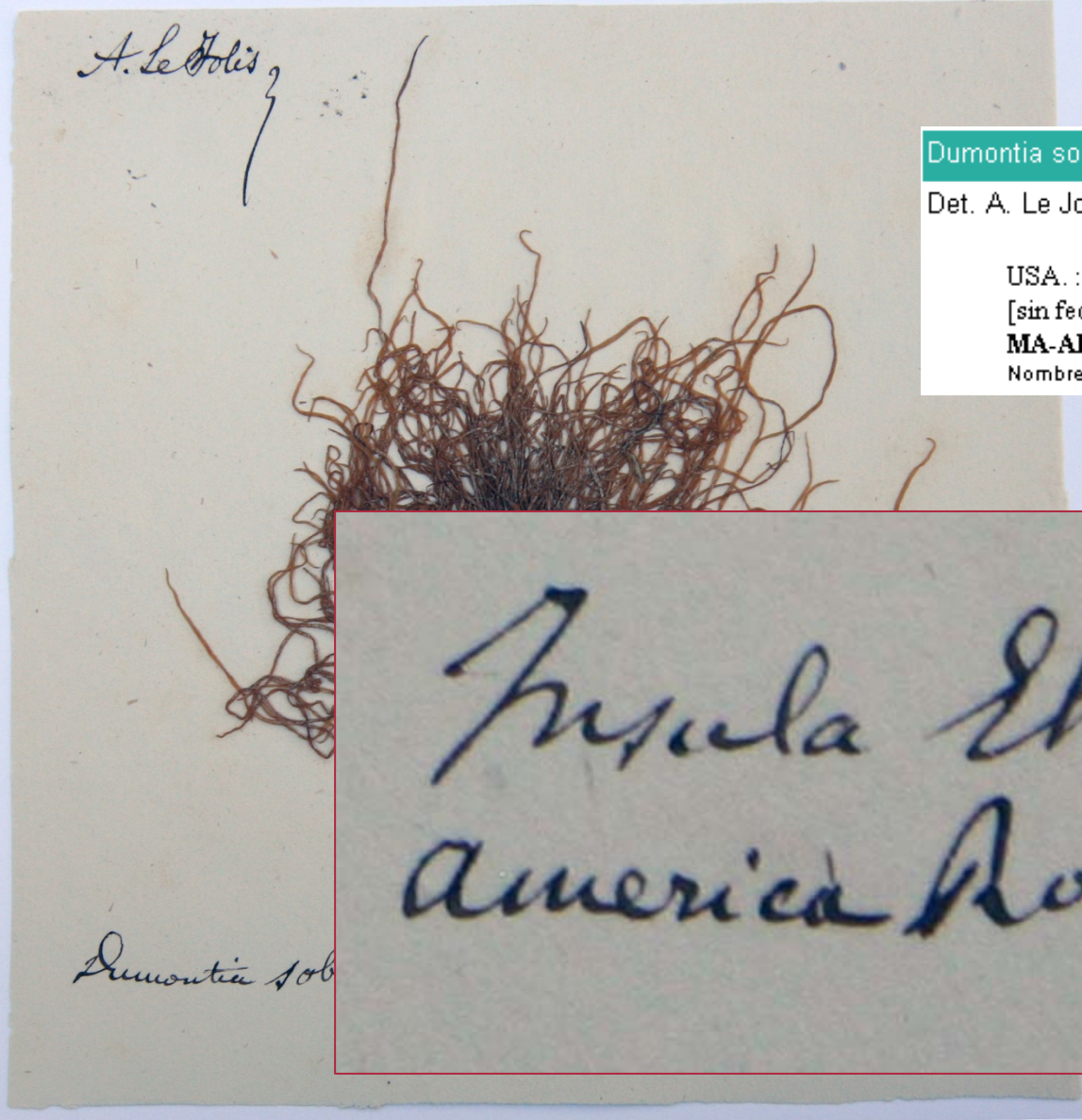
Reset

Place Status:  ANY  Current  Former  Proposed

Identification Code:

[help](#)

# Casos (4)



*Dumontia sobolifera*

Det. A. Le Jolis

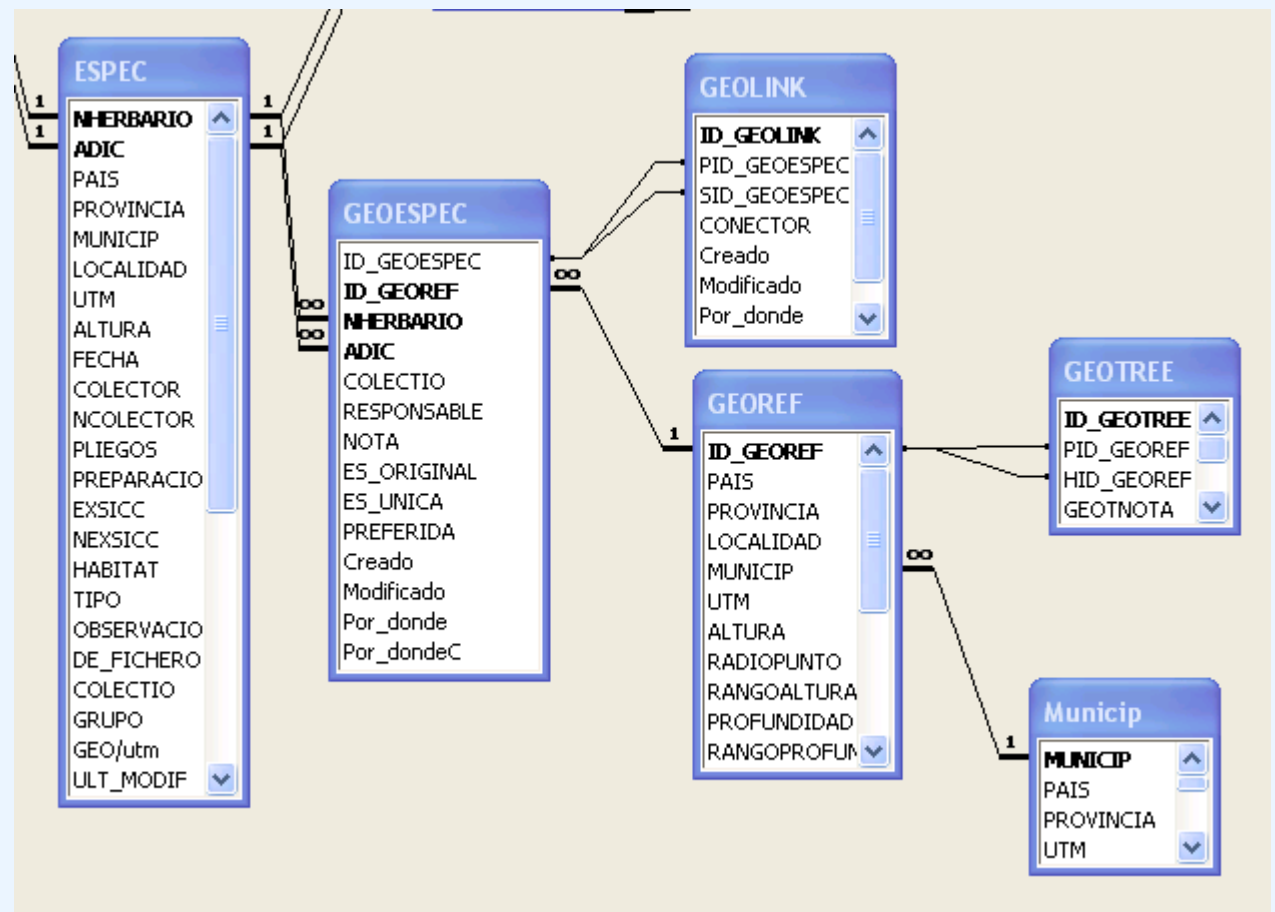
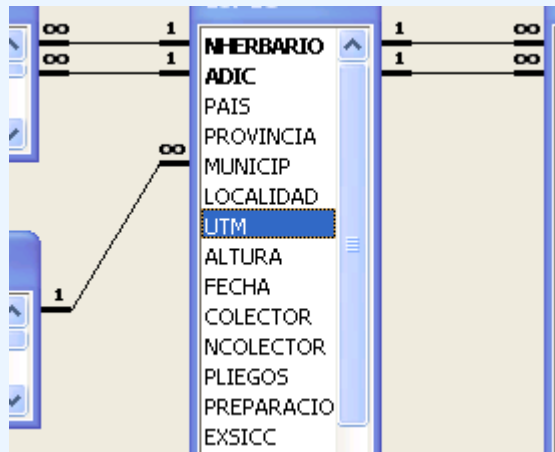
USA. : Alaska: Insula Ehe (America Rossica  
[sin fecha]

**MA-Algae 2415** Herb. A. Le Jolis. - (295/II)

Nombre aceptado en el Herbario : *Dumontia sobolifera*

# Casos(3)

## Herbar 3.4



## Herbar 3.5



# Para terminar: La casa por los cimientos: Un ejemplo de Eslovenia



Kotarec