



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

**GLOBAL BIODIVERSITY  
INFORMATION  
FACILITY**

**Como ser un  
Proveedor de Datos  
de GBIF**

***Silvia Lusa***

***GBIF España***

*lusa@gbif.es*

***Lisboa, 7 de Noviembre de 2006***

# INFRAESTRUCTURA MUNDIAL DE INFORMACIÓN EN BIODIVERSIDAD

## GBIF

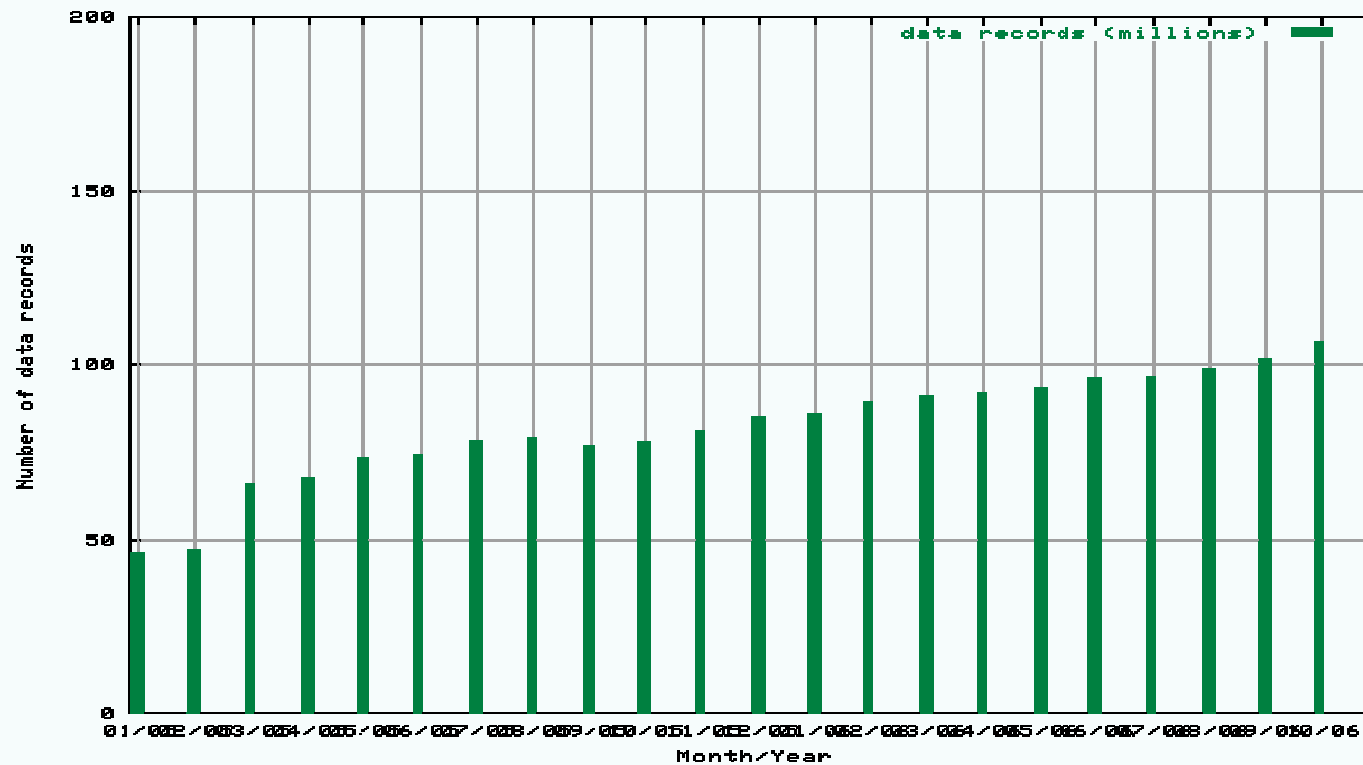
- **GBIF** es una iniciativa internacional que se propone hacer accesible a través de Internet de forma gratuita toda la información disponible sobre los organismos vivos conocidos a nivel mundial
- **Concebida** a instancias de la OCDE (Organización para la Cooperación y el Desarrollo Económico)
  - 1996, MegaScience Forum Working Group- encargado de lanzar iniciativas científicas de interés fundamental pero que por su escala no eran abordables por país alguno.
  - En este foro surge el concepto de GBIF con la idea de aplicar la informática como mecanismo para facilitar y administrar información proveniente de la naturaleza
- **Los miembros** son estados, economías y organizaciones internacionales
  - Los estados miembros se comprometen a establecer uno ó más nodos a través de los que compartir su información sobre biodiversidad contribuyendo además económicamente a la iniciativa

# TRABAJO Y PRINCIPIOS DE GBIF

- **Trabaja** para construir la principal red de bases de datos sobre biodiversidad, herramienta fundamental en el desarrollo científico de los países y que contribuirá a una mejor protección y uso de la biodiversidad en el planeta
- **Los proveedores de datos** tienen el control sobre los mismos, sus bases de datos tienen entidad por sí mismas y se reconocen los derechos de propiedad intelectual
- **Promueve** una arquitectura **no centralizada**, y el uso de software de "código abierto"
- **Esta red se construye a partir** de fuentes heterogéneas relativas a datos primarios:
  - Especímenes, Observaciones y Nombres

# Proveedores de Datos

- Actualmente:
  - **186** Proveedores de datos
  - **856** Colecciones
  - Más de **106 millones** de registros de especímenes/observaciones.



# ¿Como se proporcionan los datos a GBIF?

A través de un **modelo** en el que:

- Los proveedores puedan:
  - Controlar y gestionar el acceso a los datos
  - Proporcionar los metadatos tanto de las instituciones como de las colecciones
- Que permita compartir los datos a través de:
  - un formato de intercambio común  
(para poder comparar y combinar la información proveniente de distintas fuentes)
  - de estructura independiente del formato en el que cada institución tenga sus bases de datos

# Paquetes de Software

- GBIF recomienda el uso de diversos **paquetes de software** (provider package) compuestos por:
  - Un formato de intercambio de datos común (estándar)
  - Un protocolo que sea capaz de reconocer y manejar dicho formato a través de internet

→ DiGIR protocol (<http://www.digir.net/>)

+

Darwin Core (<http://darwincore.calacademy.org/>)

→ BioCASE protocol (<http://www.biocase.org/>)

+

ABCD (<http://www.bgbm.org/TDWG/CODATA/default.htm>)

→ ...TAPIR

# Paquetes de Software

- **DiGIR Provider (php)**

- Desarrollado a partir del “data provider” de [digir.sourceforge.net](http://digir.sourceforge.net) y modificado para que sea capaz de trabajar con el registro UDDI
- Soportado por el Secretariado Internacional de GBIF (workshops, helpdesk)

- **GBIF Data Repository Tool**

- Herramienta para Nodos que permite crear y gestionar un servicio de alojamiento (almacena datos subidos en formato documento, como hojas excel, documentos de Word, etc)
- DiGIR, Python, Zope y MySQL
- Soportado por el Secretariado Internacional de GBIF

**Disponibles desde:**

GBIF tools downloads: (<http://www.gbif.org/serv/gbif-tools>)

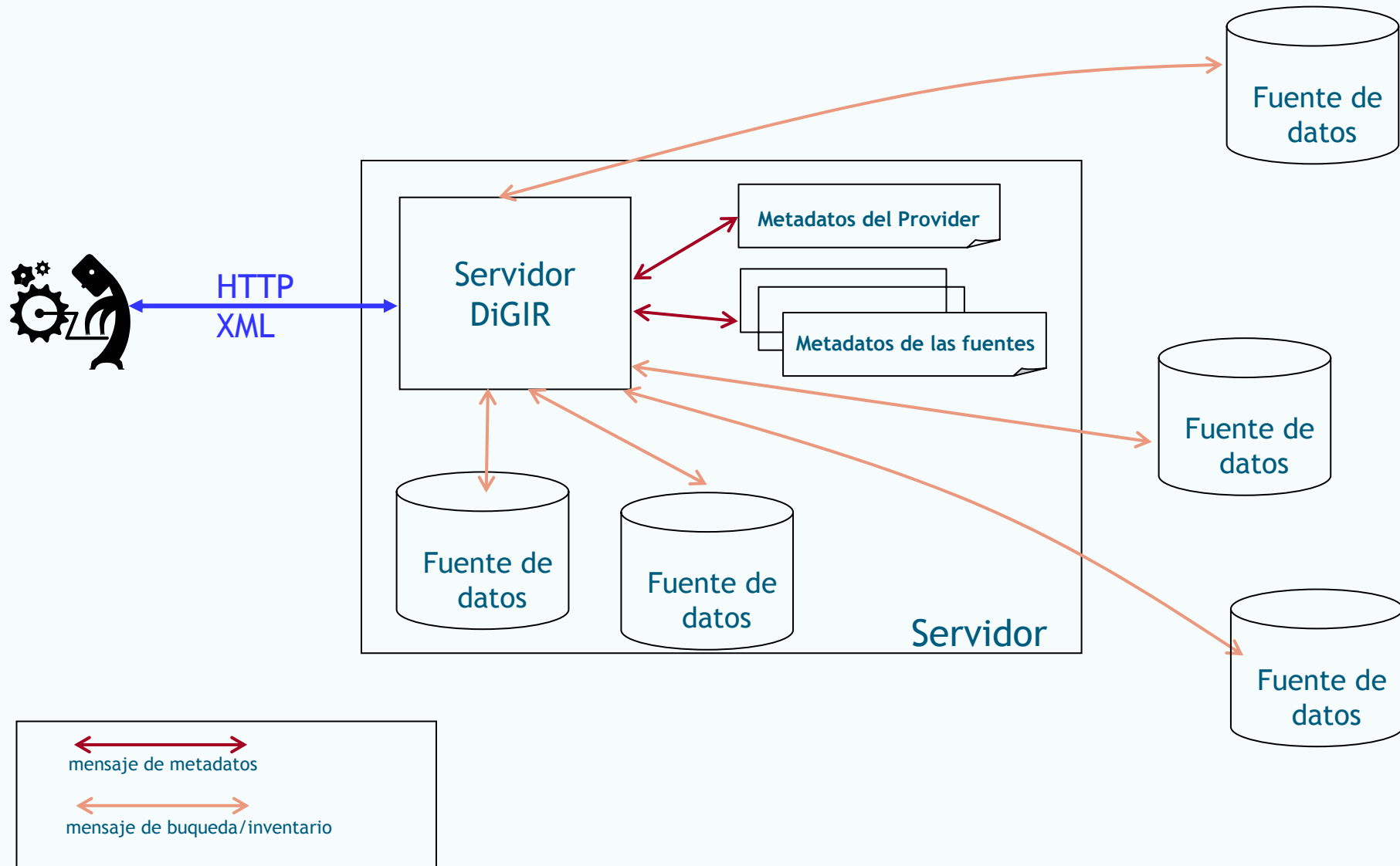
- **BioCASE - BioCASE wrapper** (<http://www.biocase.org/provider/>)

# **DiGIR:** Distributed Generic Information Retrieval

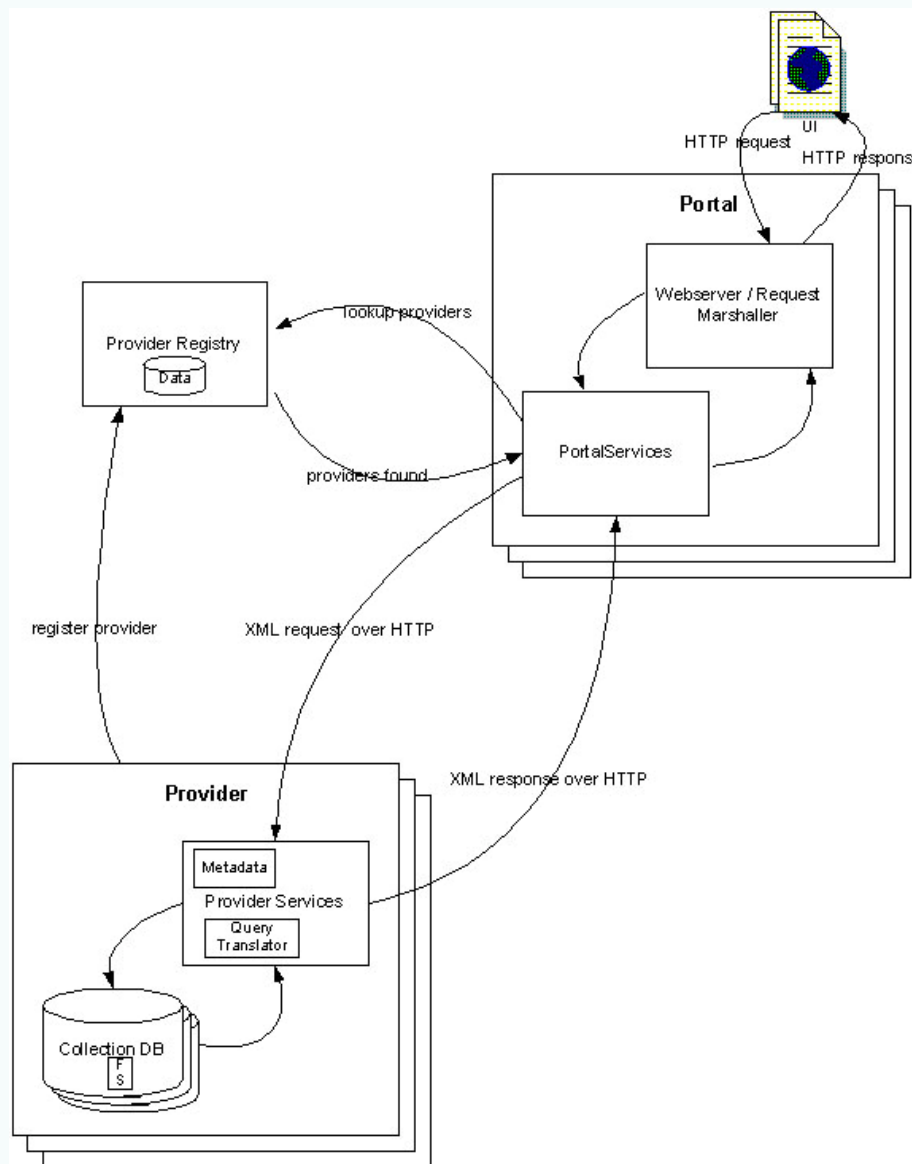
- Es un protocolo cliente/Servidor para recopilar datos desde fuentes distribuidas.
- Usa HTTP como protocolo de transporte y XML para la codificación de los mensajes entre cliente y servidor.
  - Tres tipos de mensajes:
    - **Metadata:** Obtiene los metadatos de los proveedores y de las fuentes de datos servidas.
    - **Search:** Encuentra registros de especímenes y observaciones que cumplen un criterio de búsqueda, por ejemplo: el nombre de una especie y/o un rectángulo que define un área de la superficie de la tierra y/o ...
    - **Inventory:** Obtiene el conjunto de valores asociados a un concepto. Por ejemplo: Especies.
- Proporciona un único punto de acceso a varias fuentes distribuidas de información
- Trabaja con el estándar de intercambio Darwin Core2 (para especímenes y observaciones), estableciendo la correspondencia entre las bases de datos de las colecciones y este formato.



# DiGIR Provider: ¿Como Trabaja?



# DiGIR Provider: ¿Dónde se integra?



- **Proveedor (DiGIR):** recibe peticiones del portal y las procesa devolviendo los registros almacenados en las bases de datos que cumplan la petición de búsqueda realizada por el portal.
- **Portal:** Este es el punto de entrada para los usuarios y/o aplicaciones. El portal remitirá las peticiones a los correspondientes proveedores de forma totalmente transparente al usuario presentando la información en un formato común.
- **Registro:** Provee las páginas amarillas de los proveedores existentes

# Formato de Intercambio de Datos: Darwin Core 2

- Darwin Core 2 es un estandar que permite que los datos estén estructurados y sean compartidos en documentos XML, a través de Internet.
  - Adecuado para datos de colecciones y observaciones.
  - Versiones y evolución del estandar (1.4, imágenes, extensiones) ([http://darwincore.calacademy.org/Documentation/DarwinCore2Draft\\_v1-4\\_HTML](http://darwincore.calacademy.org/Documentation/DarwinCore2Draft_v1-4_HTML))
  - XML Schema version 1.2 (48 Campos): <http://digir.net/schema/conceptual/darwin/2003/1.0/darwin2.xsd>

<b>DateLastModified *</b>	<b>InstitutionCode *</b>	<b>CollectionCode *</b>	<b>CatalogNumber *</b>
<b>ScientificName *</b>	<b>BasisOfRecord</b>	<b>Kingdom</b>	<b>Phylum</b>
<b>Class</b>	<b>Order</b>	<b>Family</b>	<b>Genus</b>
<b>Species</b>	<b>Subspecies</b>	<b>ScientificNameAuthor</b>	<b>IdentifiedBy</b>
<b>YearIdentified</b>	<b>MonthIdentified</b>	<b>DayIdentified</b>	<b>TypeStatus</b>
<b>CollectorNumber</b>	<b>FieldNumber</b>	<b>Collector</b>	<b>YearCollected</b>
<b>MonthCollected</b>	<b>DayCollected</b>	<b>JulianDay</b>	<b>TimeOfDay</b>
<b>ContinentOcean</b>	<b>Country</b>	<b>StateProvince</b>	<b>County</b>
<b>Locality</b>	<b>Longitude</b>	<b>Latitude</b>	<b>CoordinatePrecision</b>
<b>BoundingBox</b>	<b>MinimumElevation</b>	<b>MaximumElevation</b>	<b>MinimumDepth</b>
<b>MaximumDepth</b>	<b>Sex</b>	<b>PreparationType</b>	<b>IndividualCount</b>
<b>PreviousCatalogNumber</b>	<b>RelationshipType</b>	<b>RelatedCatalogItem</b>	<b>Notes</b>



# Estructuración de los campos de Darwin Core2

- **Identificación:**

ScientificName, Kingdom, Phylum, Class, Order, Family, Genus, Species, Subspecies, ScientificNameAuthor, IdentifiedBy, YearIdentified, MonthIdentified, DayIdentified, TypeStatus.

- **Preparación y Conservación:**

InstitutionCode, CollectionCode, CatalogNumber, BasisOfRecord, Sex, PreparationType, IndividualCount.

- **Recolección y Notas:**

Collector, CollectorNumber, FieldNumber, Notes, PreviousCatalogNumber

- **El cuándo:**

YearCollected, MonthCollected, DayCollected, JulianDay, TimeOfDay.

- **El dónde:**

ContinentOcean, Country, StateProvince, County, Locality, Longitude, Latitude, CoordinatePrecision, BoundingBox, MinimumElevation, MaximumElevation, MinimumDepth, MaximumDepth

- Documento traducido al español:

[http://www.gbif.es/ficheros/DarwinCore2\\_esp.pdf](http://www.gbif.es/ficheros/DarwinCore2_esp.pdf)

# Campos de Darwin Core2

- **DateLastModified:** La fecha y hora de la última modificación del registro según norma ISO 8601 en UTC(GMT). Ejemplo: "5 de Noviembre 5, 1994, 8:15:30, Madrid (GMT+1:00)" sería "1994-11-05T7:15:30Z"
- **InstitutionCode:** Un código que identifica la institución o centro a la que la colección pertenece. No existe ningún registro global para la asignación de los códigos; utiliza el código que es estándar dentro de tu disciplina.
- **CollectionCode:** Un valor alfanumérico único que identifica la colección dentro de la institución.
- **CatalogNumber:** Un texto alfanumérico único que identifica un registro individual dentro de una colección. Se recomienda que este valor proporcione una clave por la que el espécimen en cuestión puede ser identificado. Si el espécimen tiene varios elementos como por ejemplo varios tipos de preparación, este valor debería identificar la componente individual del espécimen.
- **ScientificName:** nombre completo de menor rango especificado en la identificación del organismo.
- **BasisOfRecord:** Una descripción que indica si el registro representa una observación (O), un organismo vivo (L), un espécimen testigo (S), germoplasma/semilla (G) etc.
- **Kingdom:** El nombre del reino al que pertenece el organismo.
- **Phylum:** El nombre de la división al que pertenece el organismo.
- **Class:** El nombre de la clase al que pertenece el organismo.
- **Order:** El nombre del orden al que pertenece el organismo.
- **Family:** El nombre de la familia al que pertenece el organismo.
- **Genus:** El nombre del género al que pertenece el organismo.
- **Species:** El nombre de la especie al que pertenece el organismo.
- **Subspecies:** El nombre de la subespecie al que pertenece el organismo.
- **ScientificNameAuthor:** El autor o autores del nombre científico. Se debe ajustar a las convenciones de la disciplina taxonómica correspondiente.

# Campos de Darwin Core2

- **IdentifiedBy:** El/los nombre/s de la/s persona/s que asignaron el nombre científico, actualmente aceptado, al objeto catalogado.
- **YearIdentified:** El año de la identificación, con cuatro dígitos, p. ejemplo 1906, 2002 etc.
- **MonthIdentified:** El mes de la identificación, con dos dígitos [01..12].
- **DayIdentified:** El día del mes de la identificación, con dos dígitos [01..31].
- **TypeStatus:** Indica la designación más reciente del ejemplar como tipo nomenclatural.
- **CollectorNumber:** Un número identificador (en formato texto) aplicado al ejemplar en el momento de la recolección. Establece enlaces entre diferentes partes/preparaciones de un único espécimen y entre notas de campo y el espécimen.
- **FieldNumber:** Un número identificador (en formato texto) aplicado en el momento de la recolección a un conjunto de material que se ha generado durante una única recolección.
- **Collector:** El/los nombre/s del/los recolector/es responsable/s de la recolección o que hizo la observación.
- **YearCollected:** El año durante el que el ejemplar ha sido recolectado en el campo, siempre con cuatro dígitos, p.ej. 1972.
- **MonthCollected:** El mes del calendario durante el que el ejemplar ha sido recolectado en el campo, siempre con dos dígitos, [01...12].
- **DayCollected:** El día del mes durante el que el ejemplar ha sido recolectado en el campo, siempre con dos dígitos, [01...31].
- **JulianDay:** El día ordinario del año (o sea, el número de días pasados desde el día 31 de Diciembre del año anterior, el 1 de Enero es JulianDay 1) en que el ejemplar ha sido recolectado. Puede ser calculado mediante los campos YearCollected, MonthCollected, y DayCollected.

# Campos de Darwin Core2

- **TimeOfDay:** La hora en que el ejemplar ha sido recolectado, expresado como hora decimal desde medianoche, hora local (p.ej.: 12.0 y 13.5).
- **ContinentOcean:** El nombre completo del continente u océano donde el ejemplar ha sido recolectado.
- **Country:** El país o la unidad política mayor donde el espécimen ha sido recolectado, expresado según la norma ISO 3166-1 con código de dos letras, p.ej. España = ES.
- **StateProvince:** El nombre completo de la provincia del que el ejemplar ha sido recolectado.
  
- **County:** El nombre completo del municipio del que el ejemplar ha sido recolectado.
- **Locality:** Descripción de la localidad de recolección.
- **Longitude:** La longitud de la localización donde el ejemplar ha sido coleccionado, expresado en grados decimales.
- **Latitude:** La latitud de la localización donde el ejemplar ha sido coleccionado, expresado en grados decimales.
- **CoordinatePrecision:** Una estima de cómo de cerca ha sido especificada la localidad de recolección; expresada en la distancia en metros que corresponde al radio alrededor de la localidad. Utiliza NULL si el error es desconocido, si no se puede estimar o si no aplica.
- **BoundingBox:** Proporciona un mecanismo para ejecutar búsquedas utilizando un cerco. Un cerco normalmente no está presente en la base de datos, pero se puede generar a partir de los campos Latitude y Longitude. (este dato NO es devuelta en las búsquedas)
- **MinimumElevation:** La distancia mínima en metros por encima (positivo) o por debajo (negativo) del nivel del mar en que se encuentra la localidad de recolección.
- **MaximumElevation:** La distancia máxima en metros por encima (positivo) o por debajo (negativo) del nivel del mar en que se encuentra la localidad de recolección.

# Campos de Darwin Core2

- **MinimumDepth:** La distancia mínima en metros por debajo de la superficie del agua donde la recolección ha sido realizada; todo el material recolectado tiene al menos esta profundidad. Positivo por debajo de la superficie, negativo por encima (p. ejemplo: recolecciones por encima del nivel del mar en zonas de marea).
- **MaximumDepth:** La distancia máxima en metros por debajo de la superficie del agua donde la recolección ha sido realizada; todo el material recolectado tiene al menos esta profundidad. Positivo por debajo de la superficie, negativo por encima (p.ejemplo: recolecciones por encima del nivel del mar en zonas de marea).
- **Sex:** Código que representa el sexo del organismo (propuesto: M=macho, F=hembra, H=hermafrodita, I=sin determinar (examinado pero no pudo ser determinado), U=desconocido (no examinado), T=transicional (entre sexos, útil para hermafroditas secuenciales)
- **PreparationType:** Una lista de métodos de preparación y conservación del ejemplar (piel, cráneo, esqueleto, animal entero (ETHO), preparación microscópica, etc.). Se refiere a un único registro de colección.
- **IndividualCount:** El número de individuos presentes en un lote o contenedor a que el número de catalogo se refiere. No es una estima de la abundancia o densidad en la localidad de recolección.
- **PreviousCatalogNumber:** Lista de códigos de colección asignados previamente al ejemplar. Tiene que ser compuesto por el código del centro y del código de la colección, aunque quizás ya no existan.
- **RelationshipType:** Un nombre o código que identifica el tipo de relación entre el registro de la colección y el registro de la colección referenciada. Valores que puede tener: "parasite of", "epiphyte on", "progeny of", etc. En futuras versiones del esquema este atributo podría ser interesante.
- **RelatedCatalogItem:** El identificador completo del registro con el cual está relacionado (una referencia a otro espécimen); Compuesto por: Institution Code, Collection Code y Catalogue Number del registro con el que se relaciona (los tres elementos separados por un espacio).
- **Notes:** Cualquier anotación que se desee realizar sobre el registro (especimen u observación).



# DiGIR y Darwin Core2

- Request:

```
<?xml version="1.0" encoding="UTF-8"?>
  <request xmlns="http://digir.net/schema/protocol/2003/1.0"
    xmlns:darwin="http://digir.net/schema/conceptual/darwin/2003/1.0"
    xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xsi:schemaLocation="http://digir.net/schema/protocol/2003/1.0
http://digir.sourceforge.net/schema/protocol/2003/1.0/digir.xsd
http://digir.net/schema/conceptual/darwin/2003/1.0
http://digir.sourceforge.net/schema/conceptual/darwin/2003/1.0/darwin2.xsd">

    <header>
      <version>0.95</version>
      <sendTime>2005-06-14T18:30:19+02:00</sendTime>
      <source>127.0.0.1</source>
      <destination resource="biotella">http://giorgos.gbif.org:80/digir/DiGIR.php</destination>
      <type>search</type>
    </header>

    <search>
      <filter>
        <like>
          <darwin:InstitutionCode>bioshare.com</darwin:InstitutionCode>
        </like>
      </filter>
      <records limit="3" start="0">
        <structure
schemaLocation="http://digir.sourceforge.net/schema/conceptual/darwin/brief/2003/1.0/darwin2brief.xsd"/>
      </records>
      <count>true</count>
    </search>

  </request>
```



# DiGIR y Darwin Core2

## ● Response

```
<?xml version="1.0" encoding="utf-8" ?>
<responseWrapper>
  <response xmlns='http://digir.net/schema/protocol/2003/1.0'>
    <header>
      <version>$Revision: 1.10 $</version>
      <sendTime>11-09-2003 16:33:53+0200</sendTime>
      <source resource="biotella">http://giorgos.gbif.org:80/digir/DiGIR.php</source>
      <destination>192.38.103.181</destination>
    </header>
    <content xmlns:darwin='http://digir.net/schema/conceptual/darwin/2003/1.0'
      xmlns:xsd='http://www.w3.org/2001/XMLSchema'
      xmlns:xsi='http://www.w3.org/2001/XMLSchema-instance'>
      <record>
        <darwin:DateLastModified>19930717T225000Z</darwin:DateLastModified>
        <darwin:InstitutionCode>bioshare.com</darwin:InstitutionCode>
        <darwin:CollectionCode>pyy</darwin:CollectionCode>
        <darwin:CatalogNumber>4</darwin:CatalogNumber>
        <darwin:ScientificName>Diarsia mendica</darwin:ScientificName>
      </record>
      <record>
        <darwin:DateLastModified>19950526T220000Z</darwin:DateLastModified>
        <darwin:InstitutionCode>bioshare.com</darwin:InstitutionCode>
        <darwin:CollectionCode>pyy</darwin:CollectionCode>
        <darwin:CatalogNumber>6</darwin:CatalogNumber>
        <darwin:ScientificName>Lycia lapponaria</darwin:ScientificName>
      </record>
      <record>
        <darwin:DateLastModified>19950526T220000Z</darwin:DateLastModified>
        <darwin:InstitutionCode>bioshare.com</darwin:InstitutionCode>
        <darwin:CollectionCode>pyy</darwin:CollectionCode>
        <darwin:CatalogNumber>7</darwin:CatalogNumber>
        <darwin:ScientificName>Plutella maculipennis</darwin:ScientificName>
      </record>
    </content>
    <diagnostics>
      <diagnostic code="MATCH_COUNT" severity="info">42763</diagnostic>
      <diagnostic code="RECORD_COUNT" severity="info">3</diagnostic>
      <diagnostic code="END_OF_RECORDS" severity="info">>false</diagnostic>
    </diagnostics>
  </response>
</responseWrapper>
```

(obtenida según:)

<http://digir.sourceforge.net/schema/conceptual/darwin/brief/2003/1.0/darwin2brief.xsd>

```
<xsd:complexType>
  <xsd:sequence>
    <xsd:element ref="darwin:DateLastModified" />
    <xsd:element ref="darwin:InstitutionCode" />
    <xsd:element ref="darwin:CollectionCode" />
    <xsd:element ref="darwin:CatalogNumber" />
    <xsd:element ref="darwin:ScientificName" />
  </xsd:sequence>
</xsd:complexType>
```

# Paquete GBIF DiGIR Provider

- Incluye:
  - El software del DiGIR Provider.
  - Servidor Web Apache2
  - Y librerías de PHP.
- Solo se necesita tener conocimientos básicos del sistema operativo.
- Disponibles para:  
([http://circa.gbif.net/Public/irc/gbif/ict/library?l=/digir\\_provider\\_package](http://circa.gbif.net/Public/irc/gbif/ict/library?l=/digir_provider_package))
  - Linux (RedHat 7.3, 8, 9)
  - MS Windows (2000, XP)
- Bases de datos soportadas:
  - MySQL, PostgreSQL
  - MS SQL Server, MS Access (solo en el paquete de DiGIR para MS Windows)
  - Oracle 8i/9i
- Permite registrar nuestro proveedor de datos de forma automática en el Servidor UDDI de GBIF (GBIF UDDI Registry:  
<http://registry.gbif.net>)

# Paquete GBIF DiGIR Provider

Instalación en 4 pasos:

1. Instalación del paquete de GBIF DiGIR Provider.
  2. Definición de los metadatos del proveedor.
  3. Definición de las fuente(s) de datos.
  4. Registrarse en el Servidor UDDI de GBIF.
- Ir a [http://circa.gbif.net/Public/irc/gbif/ict/library?l=/digir\\_provider\\_package](http://circa.gbif.net/Public/irc/gbif/ict/library?l=/digir_provider_package), acceder al directorio del S. Operativo (Windows, Linux, Solaris) que tengan y descargar la "guía del usuario".

.....

# Descarga del Paquete de DiGIR Provider

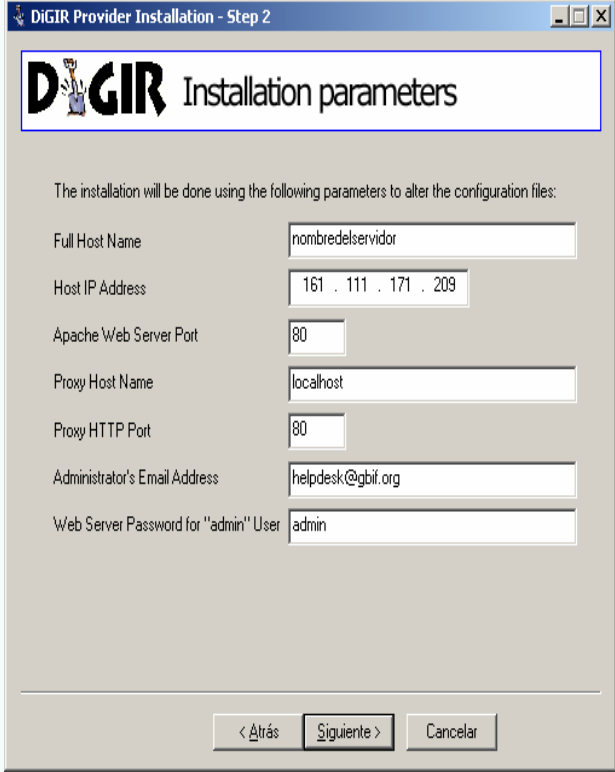
- Crear el directorio **TallerGBIF** en la unidad C:
- Seguir el enlace (GBIF tools download):  
<http://www.gbif.org/serv/gbif-tools> -> [DiGIR Provider](#) -> [Windows](#)
- Descargar y guardar en C:\TallerGBIF los ficheros:
  - [BiotellaTrainDB](#)
  - [DiGIR Provider package \(Windows 2000, XP\) - FOR TRAINING - with Perl libraries](#)

.....

# Instalación del Paquete de DiGIR Provider

- Descomprimir el fichero **WinPackageDiGIR.zip** a en el disco C:
- Ejecutar el archivo **DiGIRProviderInstall.exe**
- Rellenar los diferentes campos.
  - Usar el puerto 80 para el servidor Web.
  - Poner como contraseña del administrador: "admin"
  - Parámetros de configuración (en sg).  
(Ver Guía de usuario)

.....



The screenshot shows a window titled "DiGIR Provider Installation - Step 2" with the "DiGIR Installation parameters" header. Below the header, a message states: "The installation will be done using the following parameters to alter the configuration files:". The parameters are listed in a table-like format with input fields:

Full Host Name	nombredelservidor
Host IP Address	161 . 111 . 171 . 209
Apache Web Server Port	80
Proxy Host Name	localhost
Proxy HTTP Port	80
Administrator's Email Address	helpdesk@gbif.org
Web Server Password for "admin" User	admin

At the bottom of the window, there are three buttons: "< Atrás", "Siguiete >", and "Cancelar".

# Instalación del Paquete de DiGIR Provider

- Parar/Iniciar el Proveedor usando los respectivos archivos .bat que hay en la carpeta bin de nuestra instalación:
  - DiGIRProvider start service.bat/DiGIRProvider stop service.bat Si lo instalamos como servicio
  - DiGIRProvider start.bat/DiGIRProvider stop service.bat si no lo instalamos como servicio
- Verificar que el DiGIR Provider está funcionando correctamente accediendo a la URL: <http://10.115.64.XXX/digir>
- Acceder a la url de nuestro punto de acceso:  
<http://10.115.64.XXX/digir/digir.php>

.....

```
- <header>
  <version>$Revision: 1.14 $</version>
  <sendTime>2006-08-10T15:34:08+0200</sendTime>
  <source>http://localhost:5000/digir/DIGIR.php</source>
  <destination>127.0.0.1</destination>
  <type>metadata</type>
</header>
- <content>
- <metadata>
- <provider>
  <name>localhost Provider</name>
  <accessPoint>http://localhost:5000/digir/digir.php</accessPoint>
  <implementation>$Revision: 1.14 $</implementation>
- <host>
  <name>localhost Provider</name>
  <code>Unique code for institution</code>
  <relatedInformation>http://localhost:5000</relatedInformation>
- <contact type="administrative">
  <name>Host administrative contact.</name>
  <title>Title of contact</title>
  <emailAddress>NOSPAM</emailAddress>
  <phone>full phone number</phone>
  </contact>
+ <contact type="technical">
  <abstract>New provider installation.</abstract>
  </host>
</provider>
</metadata>
</content>
- <diagnostics>
  <diagnostic code="STATUS_INTERVAL" severity="info">600</diagnostic>
  <diagnostic code="STATUS_DATA" severity="info">1,0,0</diagnostic>
</diagnostics>
</response>
```

# Definir los Metadatos del Proveedor

- Acceder a la URL: <http://10.115.64.XXX/digir/admin/setup.php> e identificarse con el usuario "admin" y la contraseña establecida en el paso 2 de la instalación del DiGIR Provider
- Hacer click en el enlace de la parte superior de la página debajo de la palabra **provider**.
- Rellenar los campos del formulario y pulsar en "save changes".
  - Campo Name dentro de Host poner TEST seguido del nombre de la institución.
  - <Host><**Code**>: Usa el enlace: <http://www.gbif.org/links/codon> para encontrar el código de tu institución.
- Definición de los metadatos de los alumnos en sus instalaciones del DiGIR Provider.
  - Recordar reiniciar el proveedor para que los cambios surtan efecto.
  - Accede a la url de nuestro punto de acceso

.....



# Gestionar las fuentes: Nuestra BD de Ejemplo

- Familiarizándose con la base de datos de entrenamiento:
  - Biotella: Una de las muchas herramientas para gestionar datos de especímenes y observaciones.
    - <http://www.bioshare.net/biotella>
    - "Código Abierto": Aplicación limitada a la licencia de Microsoft Access
    - Puede actuar como fuente de un DiGIR Provider (para instalaciones bajo MS Windows)
- La base de datos de entrenamiento contiene datos de Lepidoptera.

# Correspondencia entre la BD y Darwin Core2

- Alternativas:
  - Mapeando en la Base de Datos (**rápidas consultas con índices**, disponible en Biotella). Mediante una vista, consulta, ... → Aconsejada
  - Mapeando directamente desde DiGIR Provider (no es necesario del trabajo de la base de datos)

Microsoft Access - [DarwinCore : Table]

File Edit View Insert Format Records Tools Window Help

Type a question for help

DateLastModified	InstitutionCode	CollectionCode	CatalogNumber	ScientificName	BasisOfRecord	Kingdom	Phylum	Class	Order
19930717T225000Z	bioshare.com	pyy	4	Diarsia mendica	O	Animalia	Invertebrata	Insecta	Lepidoptera
19950526T220000Z	bioshare.com	pyy	6	Lycia lapponari	O	Animalia	Invertebrata	Insecta	Lepidoptera
19950526T220000Z	bioshare.com	pyy	7	Plutella maculip	O	Animalia	Invertebrata	Insecta	Lepidoptera
19950527T140000Z	bioshare.com	pyy	8	Plutella maculip	O	Animalia	Invertebrata	Insecta	Lepidoptera
19950623T220000Z	bioshare.com	pyy	14	Scopula ternata	O	Animalia	Invertebrata	Insecta	Lepidoptera
19950623T220000Z	bioshare.com	pyy	15	Eupithecia pygr	S	Animalia	Invertebrata	Insecta	Lepidoptera
19950623T220000Z	bioshare.com	nvv	16	Discoloxia blom	S	Animalia	Invertebrata	Insecta	Lepidoptera

Record: 1 of 42763

ISO 8601 compliant stamp indicating the date and time in UTC(GMT) when the record was last modified. Example: the instant "November 5, 1994, 8:15

NUM

# Definición de las fuente(s) del Proveedor.

- Acceder a la URL: <http://10.115.64.XXX/digir/admin/setup.php> e identificarse con el usuario "admin".
- Click en el botón "add resource".
- Rellenar los formularios:
  - **1. Metadata**
    - **RecordIdentifier:** Usa el formato [ParticipantCode:InstitutionCode:CollectionCode](#)
  - **2. Data Source**
  - **3. Tables**
  - **4. Filter**
  - **5. Mapping**
- Definir los metadatos (paso 1.) en las instalaciones de los alumnos.

.....

# Definición de una fuente local

- Creación de una nueva fuente de datos en el DiGIR Provider para la base de datos de **biotellaTrainDB.mdb** (MS Access).
  - Localización: **C:\TallerGBIF**
  - Datasource definición:
    - Type: SQL
    - Driver: -- Microsoft Access/Jet
    - Database encoding: ISO-8859-1
    - Datasource string: Provider=Microsoft.JET.OLEDB.4.0;Data Source="c:\MyDirectory\MyDatabase.mdb"
  - Tables
    - Root Table: DarwinCore
    - Key Field: CatalogNumber
- Definir la(s) fuente(s), mapear (Catalognumber es numerico) en las instalaciones de los alumnos
  - Recordar reiniciar el proveedor para que los cambios surtan efecto.
  - Accede a la url de nuestro punto de acceso

.....

# Definición de una fuente remota

- Ejemplo de cómo crear una nueva fuente en el DiGIR Provider para la base de datos **pyyf** (PostgreSQL).
  - Location: 84.204.46.40
  - datasource string: 84.204.46.40 :10550
  - Port: 10550
  - Driver: PostgreSQL 7 or later
  - Username: giorgos
  
  - Root Table: cache
  - Key Field: id
  
  - DateLastModified ⇔ timereported
  - ScientificName ⇔ taxonname
  - YearCollected ⇔ yeacoll
  - MonthCollected ⇔ moncoll
  - DayCollected ⇔ daycoll

# Testeo de nuestro DiGIR provider

- Trabajar con los tres tipos de mensajes del protocolo DiGIR: metadata, inventory, search

Acceder a las siguientes urls y hacer las comprobaciones necesarias para la fuente local definida:

- [http://10.115.64.XXX/digir/test/eg\\_metadata.php](http://10.115.64.XXX/digir/test/eg_metadata.php)
- [http://10.115.64.XXX/digir/test/eg\\_inventory.php](http://10.115.64.XXX/digir/test/eg_inventory.php)
- [http://10.115.64.XXX/digir/test/eg\\_search.php](http://10.115.64.XXX/digir/test/eg_search.php)

# Registro en el UDDI Registry de GBIF

- Universal Description Discovery & Integration es un directorio especial que provee métodos para publicar y encontrar información de negocios y especificaciones de servicios.
  - UDDI está basado en estándares tales como XML y SOAP.
- Los 4 principales tipos de datos:
  - **businessEntity**: representa la información básica del negocio. Por ejemplo, información de contacto, clasificación, descripciones, etc.
    - **businessService**: describe un servicio proporcionado por el negocio.
      - **bindingTemplate**: contiene la URL de su punto de acceso y una referencia a uno o más tModel
  - **tModel**: descripción abstracta de una especificación o comportamiento particular el cual forma el Servicio Web.

# Registro en el UDDI Registry de GBIF

- **Registro automático en el GBIF UDDI registry.**
  - Utilización de los valores de los elementos definidos como metadatos del proveedor (más alguna información extra).
    - **Business Entity**
      - business name: {el <name> del <host> de la institución}
      - description: {la localización (URL) que apunta al <host> de la institución <related information>}
    - **Business Service**
      - service name: {El <name> común del proveedor}  
(tu.nombre.servidor)}
      - description: {Un <abstract> de la información del <host>}
    - **Binding Template**
      - access point: <http://your.server.name:port/digir/DiGIR.php>
      - description: Punto de acceso del {<host> <abstract>}



# Registro en el UDDI Registry de GBIF

Utilización de la herramienta de registro on-line  
[**recomendado**]

- Acceder a la URL:  
<http://www.gbif.org/DataProviders/registerme>
- Rellenar el primer campo de texto con el punto de acceso de nuestro DiGIR Provider. Por ejemplo  
<http://digir.server.name/digir/DiGIR.php>
- Seleccionar el participante de GBIF de tu área, país y presiona el botón **Accept**.
  - Tu proveedor será registrado y el GBIF Helpdesk ([helpdesk@gbif.org](mailto:helpdesk@gbif.org)) informará al gestor del Nodo del dominio seleccionado.


# Navegación por el GBIF UDDI Registry

GBIF UDDI Registry - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Address <http://registry.gbif.net/uddi/web> Go Links

GBIF UDDI Registry Welcome, Guest: [Login](#)



## Global Biodiversity Information Facility

**Main**

- Home
- Find business
- Find service
- Find binding
- Find tModel
- Direct get
- Publish

**User**

- Login

**Taxonomies**

- Browse
- Validate

**Information**

- Help
- Documentation
- Contact us

**Resources**

- [oasis-open.org](http://oasis-open.org)
- [uddi.org](http://uddi.org)

**Administration** ▶

---

### Global Biodiversity Information Facility UDDI Registry

#### UDDI Registry

The UDDI (Universal Description, Discovery and Integration) registry enables businesses to quickly, easily and dynamically find and transact business with one another using their preferred applications.

#### GBIF UDDI Registry

If you have a data provider up and running, it is not very useful unless users know about it. Therefore, GBIF offers a registry where you can advertise your data and services. Think of it as the one global marketplace of shared biodiversity information that integrates data within and across domains.

GBIF UDDI registry service is available for GBIF Nodes. All DiGIR providers installed using GBIF software distributions can be registered automatically here so that they can be located by portals. Providers can also be registered manually following the link: <http://www.gbif.org/DataProviders/registerme>.

After the registration your provider and the resource(s) served by it should be listed in the [GBIF advanced search portal](#). Contact [GBIF Helpdesk](#) for questions about registration.

GBIF UDDI is powered by Systinet (c) WASP UDDI engine.


---

#### Helpful hints:

Click on the Find business or Find tModel link in the navigation menu to the left and input the '%' wildcard into the Name edit field, then click on the Search button. You should now see entities stored in the UDDI registry. You can now browse any of the entities that you wish.

The Super inquiry methods Find service and Find binding allow you to search for services and bindings directly - you are not constrained by the standard drill-down pattern and you may get the information you want more efficiently.

If you want to try the API over SOAP, please use the <http://registry.gbif.net/uddi/inquiry> URL for the UDDI inquiry and SuperInquiry API, <https://registry.gbif.net/uddi/inquiry> for Secure inquiry.

Powered by  **systinet**

GBIF.org [Contact us](#)

Internet

# Navegación por el GBIF UDDI Registry

- Encontrar la información que ha sido registrada en el UDDI registry concerniente a tu proveedor:
  - Acceder a la URL <http://registry.gbif.net>.
  - Click sobre el enlace Find business situado debajo en la parte derecha de la página.
  - Introduce % en el campo **Business name\***: y pulsa sobre el botón de búsqueda (Search).
  - Entra en el enlace de tu institución (<your institution>).
  - Navega por los diferentes enlaces.

## Navegación por el GBIF UDDI Registry (III)

- Encuentra todas las entradas de negocio correspondientes a Proveedores de datos bajo un nodo Participante:
  - Acceder a la URL <http://registry.gbif.net>.
  - Entrar en el enlace navegación (Browse) bajo el árbol de Taxonomía (**Taxonomies**).
  - Entra en el enlace gbif:nodes.
  - Hacer clic en el enlace de Spain en la parte de Categorías (Categories).
  - Presiona sobre el botón encontrar negocios (Find business).

# Monitorización DiGIR providers- The Big Dig

- The "Big Dig" es un proyecto de ecoforge.net desarrollado en 2006 por el Natural History Museum and Biodiversity Research Center of the University of Kansas
- Servicio que genera de manera automática informes del estado y las características de los DiGIR providers
  - Status: Una vez al día
  - Schema: Una vez a la semana
- Evalúa el estado de las instalaciones de los DiGIR provider
  - Examina registros y reporta las características del software instalado
    - Status, versión, número de colecciones, número de registros, schema, tiempo de respuesta, encoding
  - Ayuda a los administradores ante posibles problemas
- <http://bigdig.ecoforge.net/>

# Uso del Portal de Búsqueda (I)

GBIF - Query Form - Microsoft Internet Explorer

Address <http://afroditi.gbif.org:8080/pres/PresentationServlet?action=menu&loc=http%3A%2F%2Fdigir.sourceforge.net%2Fschema%2Fconceptual%2Fdarwin%2F2003%2F1.0%2Fdarwin2Info> Go Links >>

**GBIF Home** : [Portal Home](#) : **Query Form**

**Select data providers** ⓘ

ABIF DiGIR Provider - Australian Faunal Directory  
Australian National Insect Collection - Australian National Insect Collection  
BeBIF Provider - BCCM/IHEM - BIOMEDICAL FUNGI AND YEASTS COLLECTION  
BeBIF Provider - BCCM/LMG - Laboratory of Microbiology Gent Bacteria Collection  
BeBIF Provider - BCCMTM/MUCL - (AGRO)INDUSTRIAL FUNGI AND YEASTS COLLECTION  
BeBIF Provider - Royal Museum of Central Africa - Metafro-Infosys - Prelude  
BeBIF Provider - Royal Museum of Central Africa - Metafro-Infosys - Xylarium  
BeBIF Provider - University of Ghent - Zoology Museum - Invertebratacollectie  
BeBIF Provider - University of Ghent - Zoology Museum - Nematodendcollectie  
BeBIF Provider - University of Ghent - Zoology Museum - Vertebratacollectie

Select all providers  
Clear all selections

*\* Indicates data provider may be currently unavailable*

**Select query conditions** ⓘ

Select a concept ▼ Select a comparator ▼ Submit query

Add another condition Remove last condition Clear all conditions

**Specify results** ⓘ

Please specify what results you would like:

**Count** Check if you would like a total count of matching records.  
 **Records** Check if you would like records; then specify the record format.

Brief Result Set (view [structure definition](#))  
 Full Result Set (view [structure definition](#))  
 Custom result set

Please select your result data fields:

*\* Some fields may be required result data fields. Required data fields cannot be removed.*

BasisOfRecord	>>	DateLastModified *
Kingdom	>	InstitutionCode *
Phylum	>	CollectionCode *
Class	<	CatalogNumber *
Order	<	ScientificName *
Family	<<	

# Uso del Portal de Búsqueda

- Encuentra todos los registros de una fuente de datos donde el concepto "Genus" de Darwin Core2 contenga la palabra Colias:
  - Acceder a la URL <http://10.115.64.105:10080/pres/PresentationServlet?action=home> y presionar sobre el botón de consulta (Build query).
  - Hacer clic sobre una de las fuentes disponibles en la sección de "Select data providers".
  - Seleccionar la opción "Genus" de la lista de selección "Select a concept" en la sección "Select query conditions".
  - Seleccionar "like" como comparador "Select a comparator" y teclea `colias` en el cuadro de texto más a la izquierda.
  - Presiona el botón enviar consulta (Submit query).

# Portal de Datos -- <http://www.gbif.net>

GBIF Biodiversity Data Portal - Microsoft Internet Explorer

File Edit View Favorites Tools Help

HOME | GBIF | BROWSE TAXONOMY | SEARCH | DATA PROVIDERS | COUNTRIES | DATA USE



Prototype data portal  
Global Biodiversity Information Facility

**BROWSE TAXONOMY**

- Kingdom: [Animalia](#)
- Kingdom: [Bacteria](#)
- Kingdom: [Fungi](#)
- Kingdom: [Monera](#)
- Kingdom: [Plantae](#)
- Kingdom: [Virus](#)

**DATA PROVIDERS**

- [Taxonomic name](#)
- [Specimen/observation](#)

**COUNTRIES**

- [Data by country](#)

**GBIF REGISTRY**

Data providers	104
Records	46703233

- [Become a data provider](#)
- [Advanced search](#)

**ABOUT GBIF**

- [What is GBIF?](#)
- [GBIF Demo Project 2003](#)
- [Current Participants](#)
- [Invitation to Participate](#)
- [GBIF People](#)
- [GBIF History](#)
- [Important Documents](#)
- [Publications about GBIF](#)
- [Ebbe Nielsen Prize](#)
- [GBIF Symposia](#)
- [Work Programme](#)

**ABOUT GBIF**

The Global Biodiversity Information Facility (GBIF) is an international non-profit organisation to provide free and universal access to data regarding the world's biodiversity. To learn more about GBIF itself, visit the GBIF Communications Portal: [www.gbif.org](http://www.gbif.org)

A wide range of countries and organisations participate in GBIF and have made their data available here. These bodies maintain ownership for all of the data they share. Any feedback provided through this web site will be passed back to the data provider concerned.

**ABOUT THIS PORTAL**

This portal is a prototype service providing access to the two types of data which are already being shared through the GBIF Network:

- Taxonomic names.** GBIF developing an 'Electronic Catalogue of Taxonomic Names'. This will provide access to authoritative information about both scientific and common names for all organisms, and will integrate data from a wide range of different organisations. The portal already includes data for over 486,000 scientific names and 217,000 common names, and more data will be connected soon. Some names are listed with the words 'Tentative position in taxonomy'. This indicates that the name is only known to the portal from specimen/observation records and should not be treated as authoritative simply on the basis of being listed here.
- Specimens and Observations.** The GBIF Network already provides access to over 40 million records of occurrences of different organisms. Many of these relate to specimens in natural history museums and herbaria around the world, or to living cultures of micro-organisms, but at least a third come from observations of wild organisms. Wherever possible these records include information about the locality where the organisms were found and are used to generate maps of the distribution of these occurrences. Counts of occurrence records are listed against the organism names to which they apply. For genera and taxa of higher ranks, these counts include only those occurrences which have been identified to the taxon concerned. For species these counts include all occurrences for the species and also for any included infraspecific taxa, as well as for any known synonyms.

This is a work in progress. Please explore what is already present and visit us again in the coming months as we integrate more data and provide more flexible interfaces to search and browse the data. The following icons are used on many pages in the portal:

- Send feedback on a data item to the original data provider
- Get further details
- Download data

**USE OF DATA**

GBIF Participants have made their data available for use according to the terms of the GBIF Data Use Agreement. Please understand these terms before using GBIF data: [GBIF Data Use Agreement](#)

**SEARCH**

Search for name:   Country/Territory: [All countries](#)

Selecting a country limits results to those taxa which specimen/observation records show have been identified from the country concerned.

- Search by scientific name (any rank)
- Find names starting with search string (minimum 3 characters)
- Search by common name in any language
- Find names containing search string (minimum 3 characters)



**Preguntas**

**y Respuestas**

# ...y si hay tiempo

Lluvia de ideas y más para el futuro:

- Portal customizable GBIF
- TAPIR
- WASABI (DiGIR2)
  - Aplicación web para la Arquitectura semántica
  - Basada en SPARQL (Query Language for RDF)
- LSID
  - <http://lsid.sourceforge.net/>