

Arquitectura, herramientas y estándares para el intercambio de información sobre biodiversidad

María Mora
Administradora del Nodo GBIF
Costa Rica

Contenido

- El reto de manejar la información sobre biodiversidad.
- Importancia de utilizar estándares.
- Arquitectura para la integración de datos.
- Algunos de los estándares disponibles para el manejo de información sobre biodiversidad.

El manejar la información sobre la biodiversidad representa un reto enorme

- Número esperado de especies: 10 millones.
 - 1.7 millones han sido descritas y nombradas.
- Número total de especímenes en las colecciones del mundo de 1-3 mil millones.
 - Sin contar un gran número que corresponden a especies aún no descritas.
- 18 000 nuevas especies se describen cada año.
 - El porcentaje no ha mejorado en los últimos 40 años.
- De 1 000 a 10 000 especies se extinguen cada año
 - La tasa es 1000 veces más rápida comparada con la tasa natural.

Fuente www.gbif.org

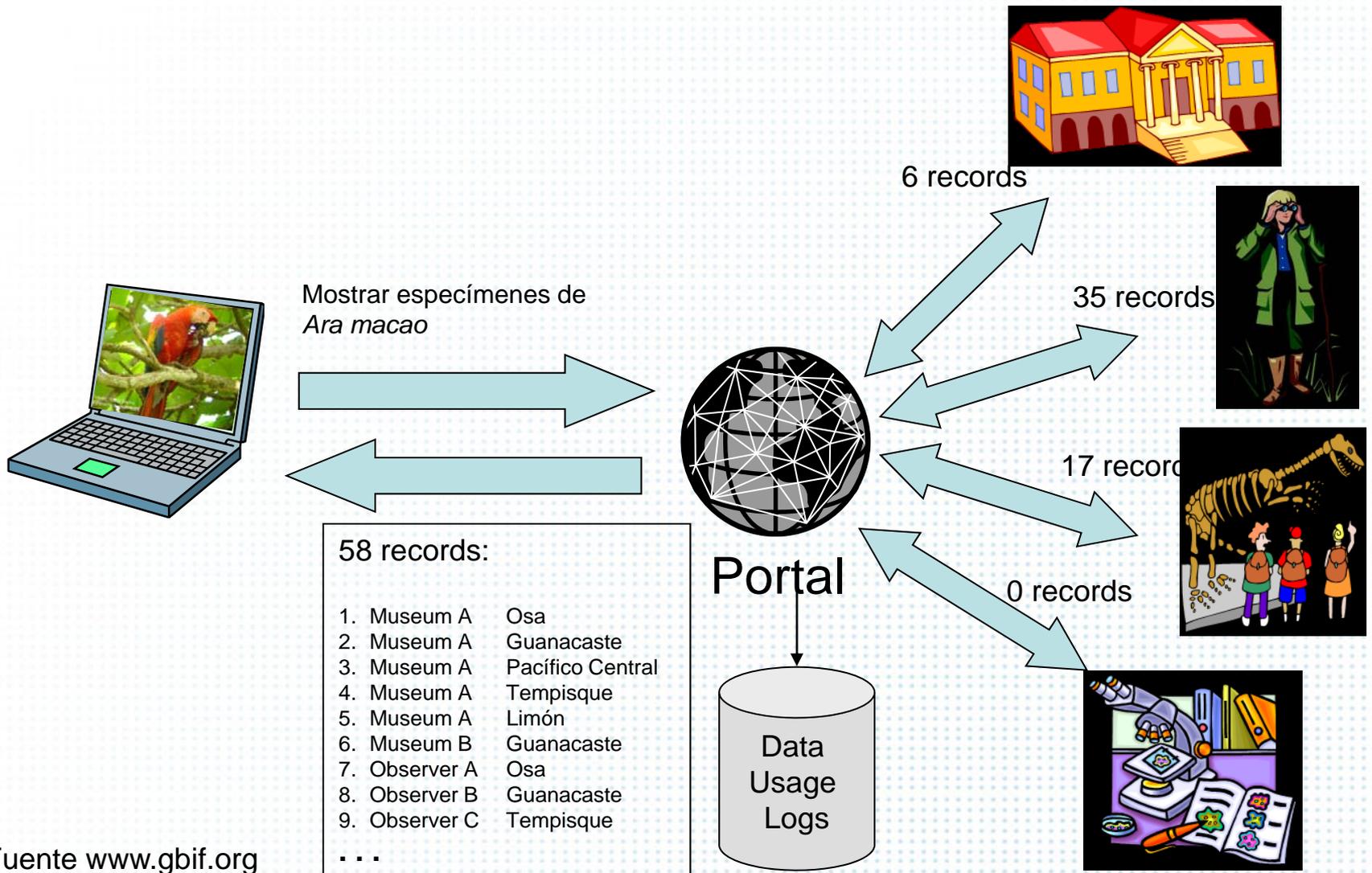
Importancia de utilizar estándares

- Hacer los datos de biodiversidad (genes, especies y ecosistemas) universalmente disponibles para el desarrollo de la ciencia, la sociedad y el futuro sostenible.
- Evitar la duplicación de esfuerzo.
- No importa la plataforma tecnológica (SO y BD) que exista en la institución los estándares son aplicables.
- No dependencia de un proveedor.
- Más personal calificado disponible.

Principales estándares para el intercambio de información sobre biodiversidad

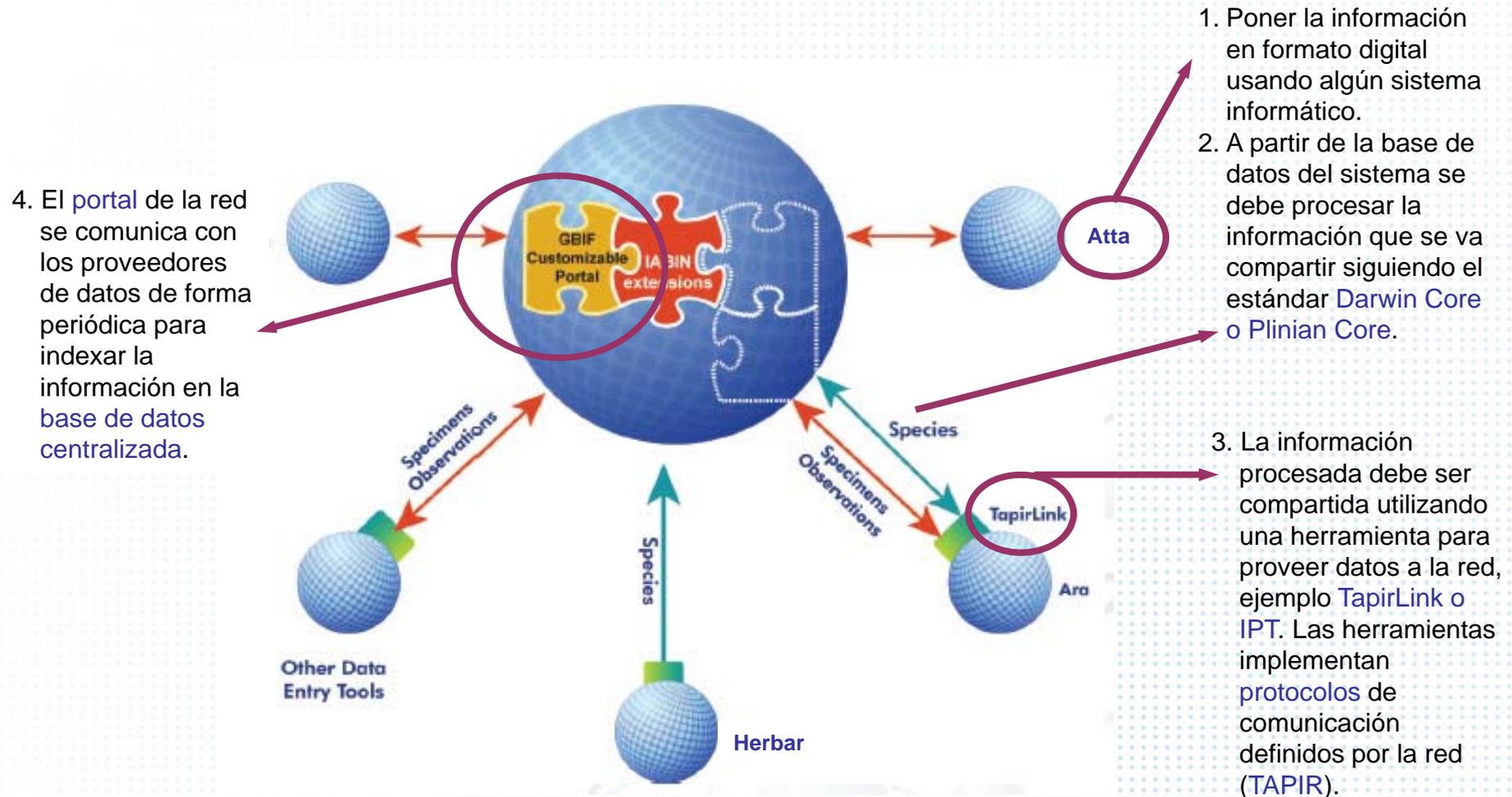
- Estándares de información y comunicación que proveen conectividad e interoperabilidad.
- Usuarios capaces de realizar búsquedas en múltiples recursos distribuidos independientemente de su localización por medio de una interfaz centralizada.
- La información se presenta de forma integrada siguiendo los estándares de intercambio de datos de la red.
- Los protocolos y estándares están sujetos a revisión y evolución continua.

Arquitectura para la integración de datos



Fuente www.gbif.org

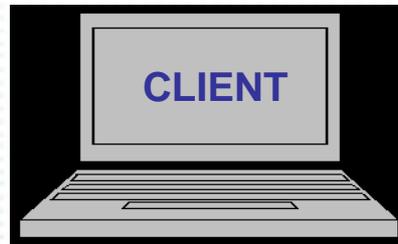
Arquitectura para la integración de datos



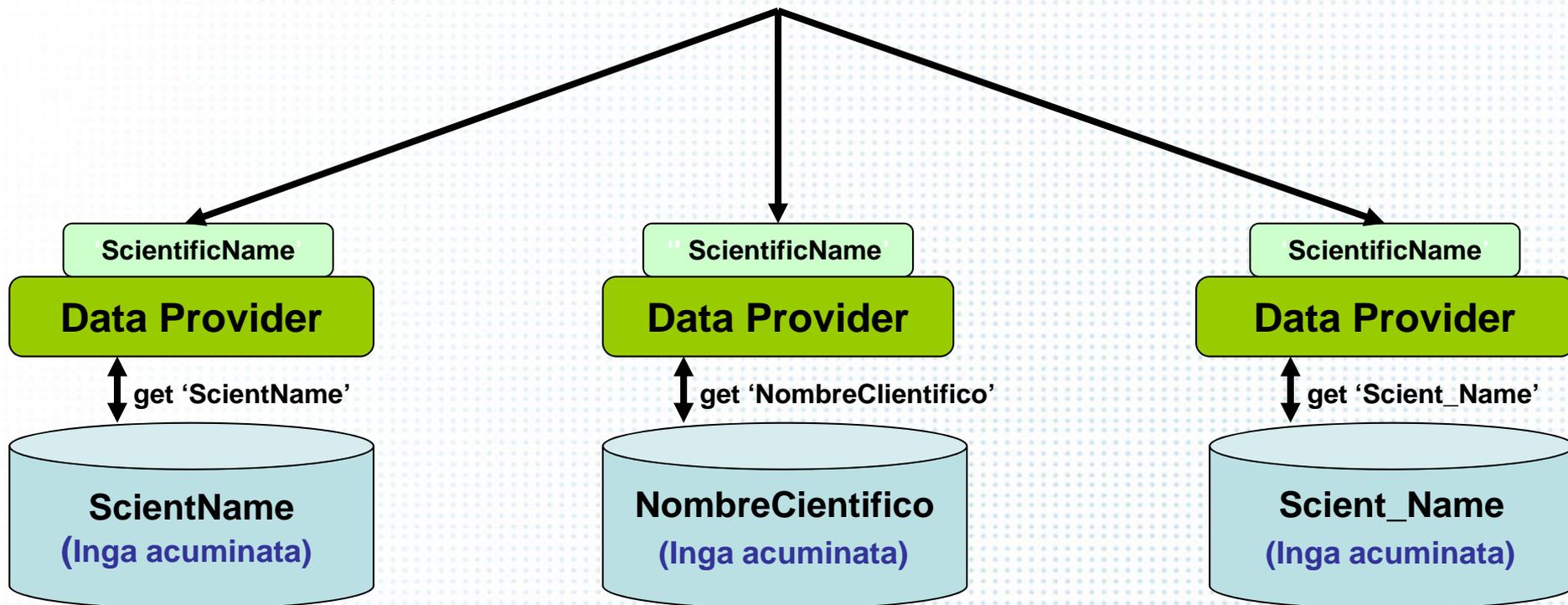
Arquitectura para la integración de datos

Atención a consultas de usuarios

Una consulta típica:
Obtener una lista de
especímenes por nombre
científico

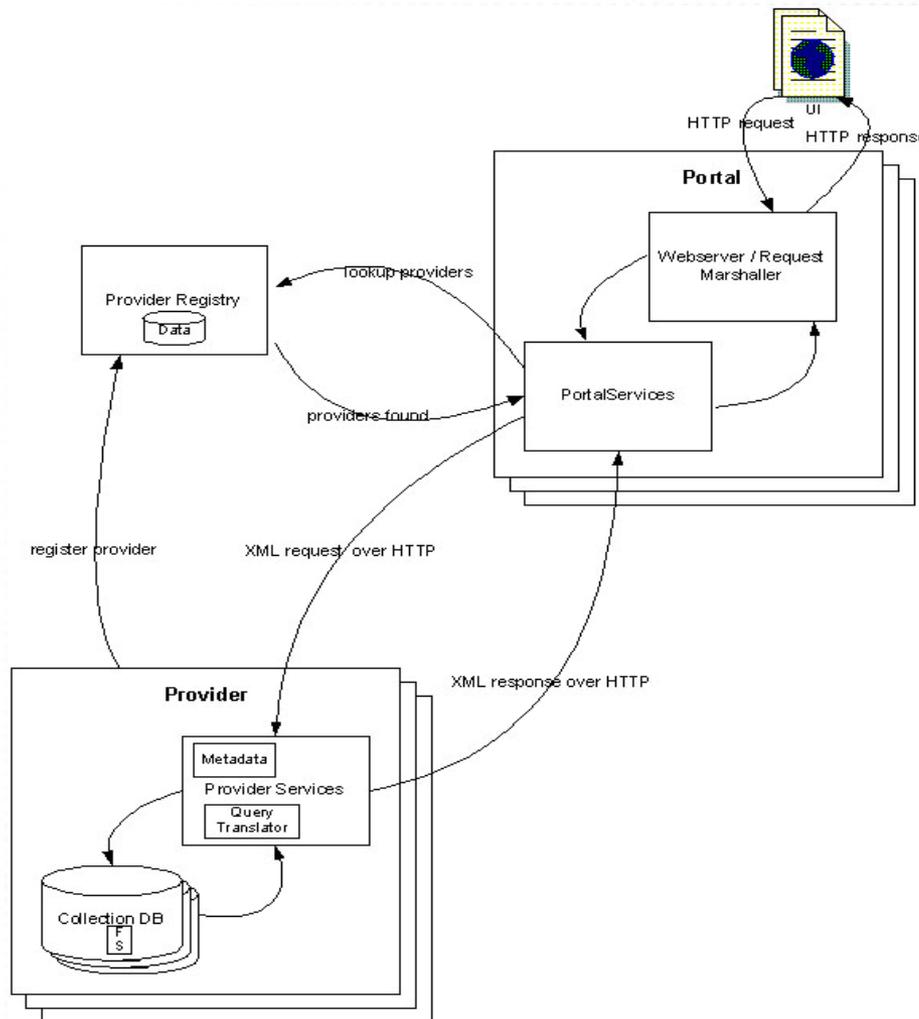


get 'ScientificName'



Arquitectura para la integración de datos

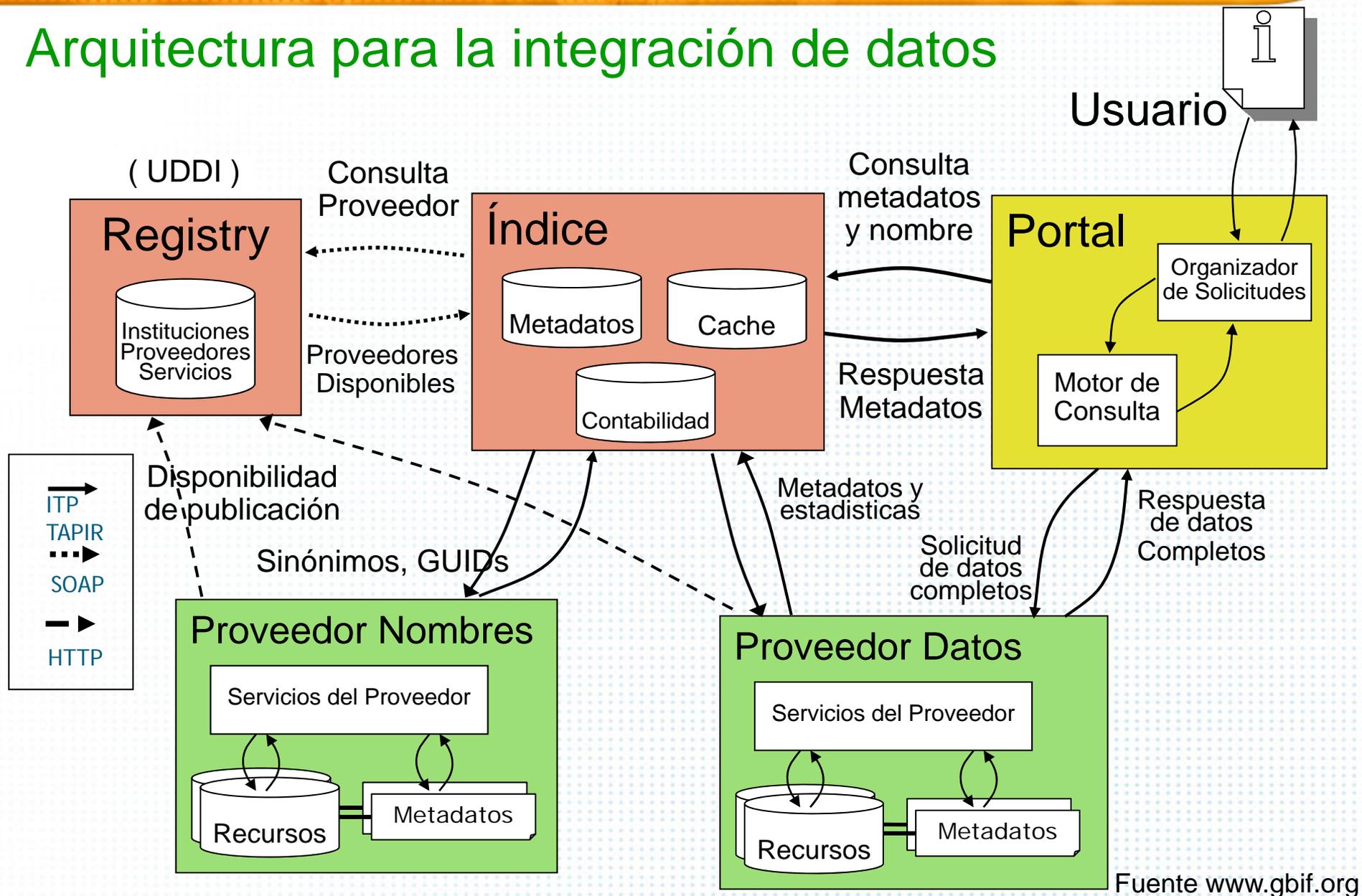
Atención a consultas de usuarios



- **Proveedor:** Recibe la consulta del portal. Las consultas son procesadas por el sistema y los datos son obtenidos de la base de datos para satisfacer las consultas recibidas del portal.
- **Portal:** Es el punto de entrada de los usuarios o las aplicaciones. El portal envía las consultas a los proveedores correspondientes. Presenta la información en un solo formato.
- **Registro:** Provee las páginas amarillas con los proveedores de datos existentes.

Fuente www.gbif.org

Arquitectura para la integración de datos



Protocolos y estándares básicos

Parte de la arquitectura	Estándar o protocolo
Protocolos de comunicación	BioCAsE, DiGIR, TAPIR
Datos:	
Para colecciones, especímenes y observaciones	Darwin Core, ABCD Schema
Para especies	Plinian Core

XML

Extensible Markup Language

- **Metalinguaje**
- **Independiente de la plataforma**
- **Basado en marcas (tags)**
- **Extensible: No especifica el conjunto de marcas ni la gramática.**

[Volver](#)

Documento XML

Ejemplo de un documento XML (instancia):

Elemento

Atributo

```
<product effDate="2001-04-02">  
  <number>557</number>  
  <size>10</size>  
</product>
```

Elemento hijo

[Volver](#)

XML Schema

- **Documento XML que describe y establece restricciones a un tipo de documentos XML.**
- **Permiten validar documentos XML (>bien formados)**
- **Contienen declaraciones:**
 - **Elementos, atributos y notaciones.**
- **y definiciones:**
 - **Tipos de datos, grupos de modelos, grupos de atributos.**
- **Tipos de datos**
 - **Simple y complejos** (con elementos hijos o atributos).
 - **Con nombre** (único, definidos globalmente) **o anónimos.**
 - **Derivación de tipos** (extensiones o restricciones)

XML Schema

Declaración de elemento

```
<xsd:schema xmlns:xsd=http://www.w3.org/2001/XMLSchema>
```

```
<xsd:element name="product" type="ProductType"/>  
</xsd:element>
```

Grupo de Modelo Secuencia

```
<xsd:complexType name="ProductType">
```

```
<xsd:sequence>
```

```
<xsd:element name="number" type="xsd:integer"/>
```

```
<xsd:element name="size" type="SizeType"/>
```

```
</xsd:sequence>
```

```
<xsd:attribute name="effDate" type="xsd:date"/>
```

```
</xsd:complexType>
```

Declaración de atributo

```
<xsd:simpleType name="SizeType">
```

```
<xsd:restriction base="xsd:integer">
```

```
<xsd:minInclusive value="2"/>
```

```
<xsd:maxInclusive value="18"/>
```

```
</xsd:restriction>
```

```
</xsd:simpleType>
```

Definición de tipo de dato complejo con nombre

Restricción de Tipo

```
</xsd:schema>
```

Definición de tipo de dato simple con nombre

El protocolo



- Es parte de la tecnología que inicialmente utilizó GBIF para establecer la red (no recomendado).
- Envío de mensajes codificados en XML sobre HTTP
 - Utilizado como medio de comunicación entre los nodos de datos y los usuarios.
 - Es más liviano y especializado que SOAP
- Permite la creación de un único punto de acceso (portal o motor de búsquedas) a recursos distribuidos.
 - Recursos: una colección de objetos que utilizan un esquema común (BD, documentos en XML).
 - Los recursos distribuidos están acorde con un esquema
- Permite la búsqueda y recuperación de datos estructurados.
 - Búsqueda de valores en los datos en un contexto (semántica)
 - Los resultados son presentados con un conjunto de datos estructurados.
- Hacen que la localización y las características técnicas de un recurso sean transparentes a los usuarios.
- El protocolo **D**istributed **G**eneric **I**nformation **R**etrieval fue creado por el TDWG/CODATA subgrupo de datos de colecciones biológicas.

Fuente www.gbif.org

Interfase para intercambio de datos en el web con DiGIR

- **D**istributed **G**eneric **I**nformation **R**etrieval is un protocolo cliente/servidor para la recuperación de información de recursos distribuidos.
- Usa HTTP como mecanismo de transporte y XML para la codificación de los mensajes entre el cliente y el servidor:
 - Tres tipos de mensajes:
 - **Metadata**: permite obtener los metadatos del proveedor y los recursos que sirve.
 - **Search**: permite realizar búsqueda de registros de especímenes y observaciones basado en un criterio de búsqueda.
 - **Inventory**: permite obtener un conjunto de valores distintos asociados a un concepto, por ejemplo especie.
- Utiliza el esquema Darwin Core 1.2 (apropiado para el intercambio de datos de especímenes y observaciones).

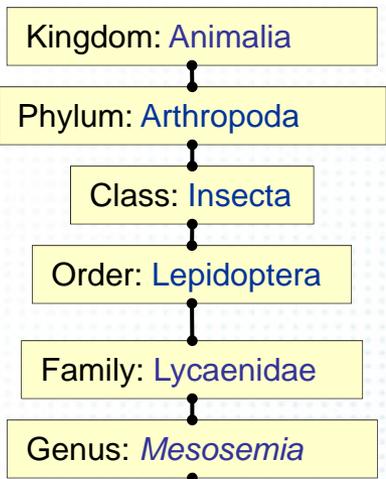
Fuente www.gbif.org

Esquemas para integrar y compartir datos de especímenes y observaciones

- Cada base de datos debe ser mapeada a un esquema común que puede ser:
 - Darwin Core
 - Simple
 - Permite estructurar datos de registros de especímenes y observaciones para ser compartidos como un documento XML que puede ser transmitido por Internet.
 - Apropiado para datos de colecciones y observaciones.
 - <http://wiki.tdwg.org/twiki/bin/view/DarwinCore/WebHome>
 - ABCD
 - Mas complejo (+ 600 conceptos)
 - <http://wiki.tdwg.org/twiki/bin/view/ABCD/WebHome>

Especímenes y observaciones (Darwin Core)

Taxonomic Names



ScientificName: *Mesosemia carissima* Bates, 1866

LifeStage: Adult

Sex: Male

Geospatial data

Country:	Costa Rica
StateProvince:	Guanacaste
County:	La Cruz
Locality:	Estación Pitilla, 9 Km de Santa Cecilia.
Latitud:	10.992609
Longitud:	-85.429477

DateCollected: June 28, 2003

Collector: Isidro Chacón

FieldNumber: 180

IdentifiedBy: Isidro Chacón

DateIdentified: June 28, 2003



DateLastModified	November 5, 2009,
InstitutionCode:	INB
CollectionCode:	Insecta
CatalogNumber:	3217
BasisOfRecord:	PreservedSpecimen

Fuente www.gbif.org

El esquema Darwin Core 1.2

- Darwin Core 1.2 es un simple conjunto de elementos
- GBIF utiliza las versiones 1.2 o 1.4.
- A manera de ejemplo, algunos de los conceptos de la versión 1.2:

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

<http://wiki.tdwg.org/twiki/bin/view/DarwinCore/WebHome>

Fuente www.gbif.org

Elementos del Darwin Core 1.2 (1/8)

- **DateLastModified:** Fecha Última Modificación. Estampa de tiempo compatible con el ISO 8601 indicando la fecha y hora en UTC(GMT) cuando el registro fue modificado o última vez. Ejemplo: el instante "5 de Noviembre, 1994, 8:15:30 am, US Eastern Standard Time" sería representado como "1994-11-05T13:15:30Z"
- **InstitutionCode:** Código de Institución. Un código identificador "estándar" que identifica la institución a la cual pertenece. No existe un registro global para asignación de códigos institucionales. Utilice el código que es estándar en su disciplina.
- **CollectionCode:** Código de Colección. Un valor alfanumérico que identifica la colección dentro de la institución.
- **CatalogNumber:** Número de Catálogo. Un valor alfanumérico que identifica un registro individual dentro de la colección. Se recomienda que este valor provea una llave por la cual el espécimen mismo pueda ser identificado. Si el espécimen tiene diferentes items, tales como varios tipos de preparación, entonces este valor debería indentificar el componente individual del espécimen.
- **ScientificName:** Nombre Científico. El nombre completo del taxón de más bajo nivel al que el Item Catalogado puede ser identificado como miembro, incluye el nombre del género, epíteto específico y epíteto subespecífico (zool.) o la abreviación del rango infraespecífico y el epíteto infraespecífico (bot.) Use el nombre del taxón supragenérico (p.ej. Nombre de familia) si el Item Catalogado no puede ser identificado a un género, especie o infraespecífico.
- **BasisOfRecord:** Base del Registro. Una abreviación que indica si el registro representa una observación (O), organismo vivo (L), espécimen (S), germoplasma/semilla (G), etc.

Fuente www.gbif.org

Elementos del Darwin Core 1.2 (2/8)

- **Kingdom:** Reino. El reino al cual el organismo pertenece
- **Phylum:** Filo. El filo (o division) al cual el organismo pertenece
- **Class:** Clase. El nombre de la clase del organismo
- **Order:** Orden. El nombre del orden del organismo
- **Family:** Familia. El nombre de la familia del organismo
- **Genus:** Género. El nombre del género del organismo
- **Species:** Especie. El epíteto específico del organismo
- **Subspecies:** Subespecie. El epíteto sub-específico del organismo
- **ScientificNameAuthor:** Autor del Nombre Científico. El autor de un nombre científico. La hilera del autor conforme se aplica al nombre aceptado. Puede ser más de un autor (hilera concatenada). Debería estar formada de acuerdo con las convenciones de la disciplina taxonómica aplicable.

Fuente www.gbif.org

Elementos del Darwin Core 1.2 (3/8)

- **IdentifiedBy**: Identificado por. El o los nombre(s) de la(s) persona(s) que aplicaron el Nombre Científico actualmente aceptado al Item Catalogado.
- **YearIdentified**: Año de Identificación. La porción del año de la fecha en que el Item Colectado fue identificada; en cuatro dígitos [-9999..9999], p.ej., 1906, 2002.
- **MonthIdentified**: Mes de Identificación. La porción del mes de la fecha en que el Item de Colección fue identificado; en dos dígitos [01..12].
- **DayIdentified**: Día de Identificación. La porción del día de la fecha en que el Item de Colección fue identificado; en dos dígitos [01..31].
- **TypeStatus**: Estado del Tipo. Indica el tipo nomenclatural que un espécimen representa. En particular, el estado del tipo puede no aplicarse al nombre contenido en el nombre científico, i.e. la identificación actual. En casos raros, un único espécimen puede ser el tipo de más de un nombre
- **CollectorNumber**: Número de Colector. Un “número” (en realidad una hilera) identificadora aplicada a los especímenes (en algunas disciplinas) al momento de la recolección. Establece ligas a diferentes partes/preparaciones de un mismo espécimen y entre notas de campo y el espécimen.

Fuente www.gbif.org

Elementos del Darwin Core 1.2 (4/8)

- **FieldNumber**: Número de Campo. Un "número" (realmente una hilera) creada en el momento de la colecta para identificar todo el material que resulta de un evento de colecta.
- **Collector**: Colector. El o los nombre(s) de los colector(es) responsable(s) de recolectar el espécimen o realizar la observación
- **YearCollected**: Año Colectado. El año (expresado como un entero) en el que el espécimen fue recolectado. El año completo debe ser expresado (p.ej. 1972 debe ser expresado como "1972" no "72").
- **MonthCollected**: Mes Colectado. El mes del año en que el espécimen fue recolectado del campo. Posible rango de valores entre 01 y 12 inclusive.
- **DayCollected**: Día Colectado. El día del mes en que el espécimen fue recolectado del campo. El rango de valores posible varía entre 01 y 31 inclusive.
- **JulianDay**: Día Juliano. El día ordinal del año, i.e. el número de días desde el primero de enero del mismo año. (Enero 1 es día Juliano 1.)

Fuente www.gbif.org

Elementos del Darwin Core 1.2 (5/8)

- **TimeOfDay:** Hora del Día. La hora del día en que un espécimen fue recolectado, expresada como horas decimales desde el medianoche, hora local (p.ej. 12.0 = mediodía, 13.5 = 1:30pm)
- **ContinentOcean:** Continente u Océano. El continente u océano de donde el espécimen fue recolectado.
- **Country:** País. El país o unidad política mayor de donde el espécimen fue colectado. Los valores ISO 3166-1 deberían ser usados. Actualmente se utilizan los nombres completos de los países. Una futura recomendación es utilizar los códigos de dos letras ISO3166-1 o el nombre completo cuando se está buscando.
- **StateProvince:** EstadoProvincia. El estado, provincia o region (i.e. la siguiente región política más pequeña que País) desde donde el espécimen fue recolectado.
- **County:** Condado. El condado (o shire o siguiente región política más pequeña que el Estado/Provincia) desde donde el espécimen fue recolectado.
- **Locality:** Localidad. La descripción de la localidad (nombre de lugar mas un desplazamiento opcional desde el nombre de lugar) desde donde el espécimen fue recolectado. Cuando un desplazamiento desde un lugar es dado, debería ser en unidades de medida no-proyectadas.

Fuente www.gbif.org

Elementos del Darwin Core 1.2 (6/8)

- **Longitude:** Longitud. La longitud de una localidad desde donde el espécimen fue recolectado. Este valor debería ser expresado en grados decimales con un datum , tal como WGS-84.
- **Latitude:** Latitud. La latitud de la localidad desde donde el espécimen fue recolectado. Este valor debería ser expresado en grados decimales con un datum , tal como WGS-84.
- **CoordinatePrecision:** Precisión de la Coordenada. Un estimado de cuán tightly fue especificada la localidad de recolección; expresada como una distancia, en metros, que corresponde a un radio alrededor de las coordenadas de latitud-longitud. Utilize NULL donde la precisión es desconocida, no puede ser estimada o no es aplicable.
- **BoundingBox:** Caja Limítrofe. Este punto de acceso provee un mecanismo para realizar búsquedas utilizando una caja limítrofe. Un elemento Bounding Box no está disponible usualmente en la base de datos, sino que es derivado de las columnas de Latitud y Longitud por el proveedor de datos
- **MinimumElevation:** Elevación Mínimo. La distancia mínima en metros sobre (positiva) o debajo (negativo) del nivel del mar de la localidad de recolección.
- **MaximumElevation:** Elevación Máxima. La distancia máxima en metros sobre (positiva) o debajo (negativa) del nivel del mar de la localidad de recolección.

Fuente www.gbif.org

Elementos del Darwin Core 1.2 (7/8)

- **MinimumDepth**: Profundidad mínima. La distancia mínima en metros bajo la superficie del agua al cual la recolección fue hecha; todo el material recolectado estaba al menos a esta profundidad. Positivo bajo la superficie, negativo sobre la misma (p.ej. recolectando sobre el nivel del mar en el área intermareal).
- **MaximumDepth**: Profundidad máxima. La distancia máxima en metros bajo la superficie del agua al cual la recolección fue hecha; todo el material recolectado se encontraba a esta profundidad como máximo. Positivo bajo la superficie, negativo sobre la misma (p.ej. Recolectando sobre el nivel del mar en el área intermareal).
- **Sex**: Sexo. El sexo de un espécimen. El dominio debería ser un conjunto controlado de términos (códigos) basados en consenso de la comunidad. Valores Propuestos: M=Masculino; F=Femenino; H=Hermafrodita; I=Indeterminado (examinado but no se pudo determinar; U=Desconocido (Unknown, no examinado); T=Transicional (entre sexos; útil para hermafroditas secuenciales)
- **PreparationType**: Tipo de Preparación. El tipo de preparación (skin, slide, etc). Probablemente sea mejor añadir esto como un elemento de registro en vez de un punto de acceso. Debería ser una lista de preparaciones para un único registro de colección.
- **IndividualCount**: Cuenta individual. El número de individuos presentes en el lote o contenedor. No un estimado de abundancia o densidad en la localidad de recolección.

Fuente www.gbif.org

Elementos del Darwin Core 1.2 (8/8)

- **PreviousCatalogNumber:** El número de catálogo previo (completamente calificado) del Item Catalogado si el ítem identificado anteriormente por otro Número de Catálogo, sea en el catálogo actual o en otro catálogo/Institución. Un Número de Catálogo completamente calificado está precedido 'por el Código de la Institución y el Código de la Colección, con un espacio separando cada subelemento. El referenciar un Número de Catálogo previo no implica que un registro para el ítem referenciado está o no presente en el catálogo correspondiente, o siquiera que el catálogo referenciado aun exista. Este punto de acceso pretende brindar una forma de obtener este registro por medio de un identificador previo, que pudo haber sido utilizado en literatura. En versiones futuras de este esquema, este atributo debería tener un conjunto de valores.
- **RelationshipType:** Tipo de Relación. Un nombre o código que identifica el tipo de relación entre este Item de Colección y el Item de Colección referido. Los nombres incluyen: "parásito de", "epífita en", "progenie de", etc. En versiones futuras de este esquema, este atributo debería tener un conjunto de valores.
- **RelatedCatalogItem:** Item de Catálogo Relacionado. El identificador completamente calificado de un Item de Catálogo relacionado (una referencia a otro espécimen); Código de Institución, Código de Colección, Número de Catálogo del Item Catalogado relacionado, donde los tres subelementos están separados por un espacio.
- **Notes:** Notas. Notas de texto libre adjuntas al registro de espécimen.

Fuente www.gbif.org

DiGIR/DwC ejemplo de consulta

```
<request>
  <header>
    <version>1.0</version>
    <sendTime>2003-03-09T19:14:58-05:00</sendTime>
    <source>216.91.87.102</source>
    <destination resource="atta">http://osa.inbio.ac.cr/DiGIR/DiGIR.php</destination>
    <type>search</type>
  </header>
  <search>
    <filter>
      <like>
        <darwin:ScientificName>Inga vera%</darwin:ScientificName>
      </like>
    </filter>
    <records limit="10" 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: Ejemplo

```

<?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>

```

Fuente www.gbif.org

DiGIR y Darwin Core2: Ejemplo

```

<?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>
        ...
      </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>

```

[Volver](#)

Fuente www.gbif.org

DiGIR y Darwin Core2: Ejemplo

```
...
  <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>
</content>
<diagnostics>
  <diagnostic code="MATCH_COUNT" severity="info">42763</diagnostic>
  ...

```

ABCD/BioCASE

El esquema ABCD (Access to Biological Collection Data):

- Es parte del trabajo realizado por TDWG.
- El objetivo del grupo de trabajo es dar acceso a los datos de colecciones biológicas del mundo realizando una estandarización de los datos y metadatos asociados a estas.
- Es altamente comprensivo y estructurado (jerárquico), provee campos para albergar los datos completos de una amplia variedad de bases de datos con información de especímenes (vivos o muertos) y observaciones.
- Es compatible con varios estándares de datos. Usado en Europa.

TAPIR

- TAPIR (TDWG Access Protocol for Information Retrieval)
- Combina y extiende características de los protocolos BioCASE y DiGIR para crear un protocolo de comunicación entre las aplicaciones cliente y los proveedores de datos.
- Implementaciones de referencia: PyWrapper, TapirLink, TapirDotNet.
- Fue desarrollado para trabajar con información de colecciones biológicas, pero es aplicable a otros dominios de aplicación.
- Toda la información referente al protocolo se encuentra en: <http://www.tdwg.org/activities/tapir/>

[Back](#)

TAPIR

- Algunas ventajas:
 - Brinda mayor interoperabilidad entre diferentes redes.
 - Estimula la participación de nuevos miembros.
 - Incrementa el acceso a datos sobre biodiversidad.
 - Facilita el desarrollo de herramientas.

Manejo de esquemas en TAPIR

- En DiGIR, un esquema conceptual es representado como una lista plana de elementos. El esquema por defecto que utiliza DiGIR es el Darwin Core (aunque con facilidad se le pueden adaptar otros mas, aunque siempre planos → Gran desventaja)
- BioCASE introdujo la posibilidad de definir esquemas conceptuales jerárquicos.

Ejemplo: [Species/StandardInformation/AcceptedScientificName](#)

[Species/Distribution/Countries](#)

- BioCASE puede referenciar esquema externos. DiGIR requiere que todo se defina en uno solo.
- TAPIR es capaz de referenciar conceptos de esquemas externos, dando la posibilidad de que los datos puedan utilizar diferentes esquemas para “mapear” sus elementos.

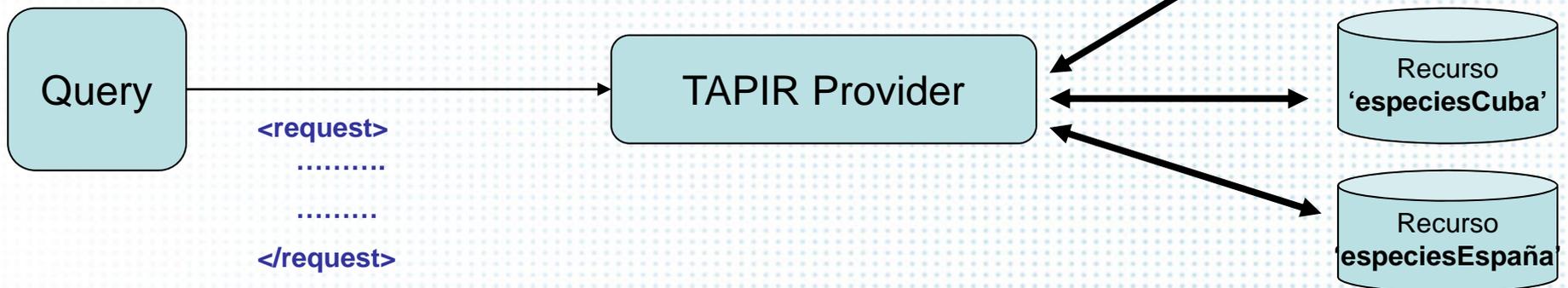
Manejo de recursos en TAPIR

- En DiGIR, los recursos no existen individualmente. Una URL representa al proveedor de datos, y dentro de este proveedor pueden existir N recursos.
- En BioCASE, los recursos pueden consultarse individualmente. Una URL representa al recurso directamente.
- Para el registro en un UDDI, es preferible que cada recurso se acceda individualmente, pues un solo proveedor podría tener diferentes temáticas.
- TAPIR exhibe sus recursos individualmente, tomando la buena práctica implementada por el BioCASE.

Manejo de recursos en TAPIR

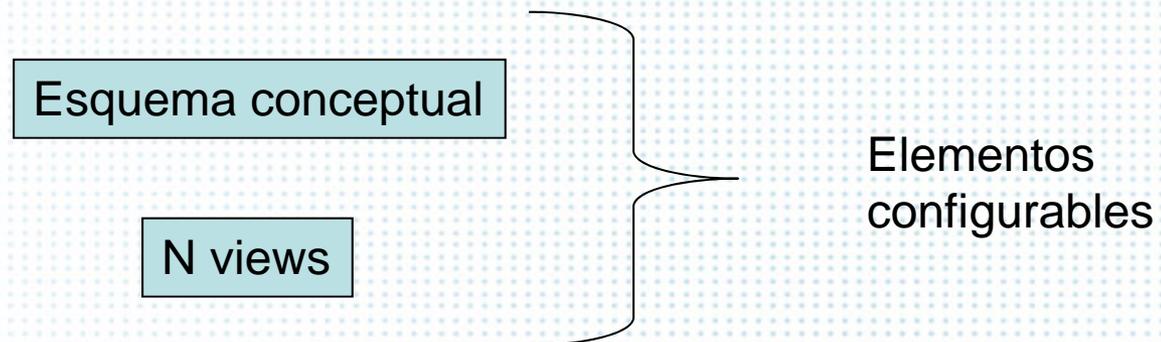


`http://localhost:8082/tapir/dsa_info.cgi?dsa=especiesCR`
`http://localhost:8082/tapir/dsa_info.cgi?dsa=especiePeru`
`http://localhost:8082/tapir/dsa_info.cgi?dsa=especieUSA`



Manejo de esquemas en TAPIR

- Una de las ventajas del TAPIR es que permite devolver resultados de la manera que el usuario los solicite.
- Se utiliza un archivo view con el cual se llega a un acuerdo en el formato en que la información será devuelta al proveedor.
- Un proveedor podría definir N views y devolver los mismos elementos del recurso en formatos diferentes.



Ejemplo de esquema conceptual

```
<schema xmlns="http://www.w3.org/2001/XMLSchema">  
  <element name="Specie">  
    <complexType>  
      <sequence>  
        <element name="ScientificName" type="string"/>  
        <element name="Kingdom" type="string"/>  
        <element name="Habitat" type="string"/>  
      </sequence>  
    </complexType>  
  </element>  
</schema>
```

Specie/ScientificName

Specie/Kingdom

Specie/Habitat

Vistas

```
<view xmlns="http://www.tdwg.org/schemas/tapir/datasource/1.0">
  <structure>
    <xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema">
      <xs:element name="ViewEspecie" minOccurs="0" maxOccurs="-1">
        <xs:complexType>
          <xs:sequence>
            <xs:sequence>
              <xs:element name="ViewNombre" type="xs:string"/>
              <xs:element name="ViewReino" type="xs:string"/>
              <xs:element name="ViewHabitat" type="xs:string"/>
            </xs:sequence>
          </xs:complexType>
        </xs:element>
      </xs:schema>
    </structure>
```

Definición de la
vista
viewSpecies.xml

```
<indexingElement path="/ViewEspecie/ViewNombre"/>
<mapping xmlns:esp="http://www.inbio.ac.cr">
  <node path="/ViewEspecie/ViewNombre">
    <concept path="esp:/Specie/ScientificName"/>
  </node>
  <node path="/ViewEspecie/ViewReino">
    <concept path="esp:/Specie/Kingdom"/>
  </node>
  <node path="/ViewEspecie/ViewHabitat">
    <concept path="esp:/Specie/Habitat"/>
  </node>
</mapping>
</view>
```

Mapeo entre
los elementos
del esquema
conceptual con
elementos del
view definido
arriba

Consulta al proveedor

```
<?xml version='1.0' encoding='UTF-8'?>
<request>
  <header />
  <search count='true' start='0' limit='5'>
    <view location='viewSpecie'/>
    <filter xmlns:esp='http://www.inbio.ac.cr'>
      <like>
        <concept path='esp:/Specie/ScientificName' />
        <literal value='*' />
      </like>
    </filter>
  </search>
</request>
```

Respuesta del proveedor

```
<search>
  <ViewEspecie>
    <ViewNombre>Aspasia principissa</ViewNombre>
    <ViewNombre>Balantiopteryx plicata</ViewNombre>
    <ViewNombre>Caluromys derbianus</ViewNombre>
    <ViewNombre>Cecropia peltata</ViewNombre>
    <ViewNombre>Didelphis virginiana</ViewNombre>
    <ViewReino>Plantae</ViewReino>
    <ViewReino>Animalia</ViewReino>
    <ViewReino>Animalia</ViewReino>
    <ViewReino>Plantae</ViewReino>
    <ViewReino>Animalia</ViewReino>
    <ViewHabitat>Bosques muy humedos.</ViewHabitat>
    <ViewHabitat>Bosques secos.</ViewHabitat>
    <ViewHabitat>Bosques humedos.</ViewHabitat>
    <ViewHabitat>Bosques secos y humedos.</ViewHabitat>
    <ViewHabitat>Habitan sitios dentro de los bosques</ViewHabitat>
  </ViewEspecie>
  <summary start="0" totalMatched="7" next="5" />
</search>
```

Integrated Publishing Toolkit (IPT)

- Herramienta para compartir datos en la red. El desarrollo lo lidera GBIF.
- Provee mecanismos de administración local de los datos, capacidad para generar estadísticas, ampliación de metadatos respecto a TapirLink, entre otras mejoras.
- Provee una interfaz de usuario.
- Mejora la capacidad para actualizar el índice centralizado de la Red GBIF.

Esquemas para integrar y compartir información de especies

- Cada base de datos debe ser mapeada a un esquema común que puede ser:
 - Plinian Core
 - Simple
 - <http://www.pliniancore.org>
 - EOL
 - Tiene concepto homólogos al PLIC
 - <http://www.eol.org>
 - Otras (TDWG, FishBase, SIB-Colombia, entre otros)

Plinian Core

Objetivo

- Especificar los conceptos básicos necesarios para integrar y recuperar información acerca de especies que se encuentra distribuida en bases de datos que administran instituciones alrededor del mundo.
- Se pretende que el esquema sea simple pero con el grado de detalle suficiente para recuperar la información requerida por los públicos meta prioritarios.

[Volver](#)

Plinian Core

Públicos meta

- **Tomadores de decisiones.**
- **Investigadores de biodiversidad:** Científicos, taxónomos, biólogos, técnicos especialistas, bioquímicos, profesionales de salud, educadores y estudiantes universitarios o afines;
- **Educadores, estudiantes y formadores de opinión de distintos niveles:** Profesores, monitores, periodistas, religiosos o afines;
- **Productores y explotadores que hacen un uso sustentable del recurso;**
- **Profesionales de campos fuera de la biodiversidad:** Autores literarios, historiadores, etnógrafos y usuarios gráficos (Ilustradores, pintores, diseñadores) o afines;

[Volver](#)

Plinian Core

Metas de diseño

- No se espera que el esquema se utilice como estructura de datos para desarrollar herramientas de captura de información.
- Debe ser simple.
- Debe reutilizar los estándares internacionales existentes al máximo.
- Debe soportar la combinación de datos provenientes de múltiples fuentes en forma de un conjunto de datos estructurados.

[Volver](#)

Plinian Core

Metas de diseño (continuación)

- Debe permitir compartir información estructurada aunque esta no se encuentre debidamente normalizada.

DocumentedUses

- Text
- DocumentedUse
 - TypeOfUse
 - Description
 - References
 - Region

[Volver](#)

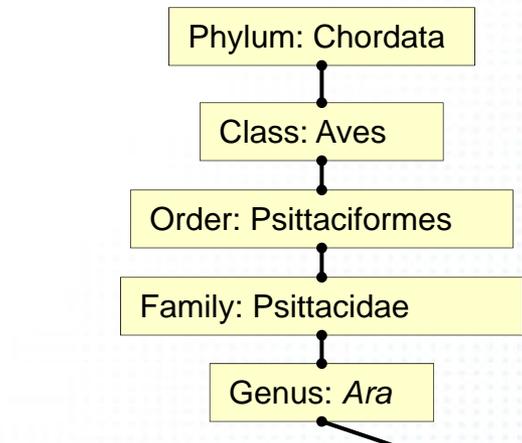
Plinian Core

Metas de diseño (continuación)

- Soporte a información en múltiples idiomas.
- Integración de datos de uso relativamente común:
 - Conceptos utilizados en múltiples disciplinas.
 - Que se encuentran comúnmente en bases de datos relevantes.
 - Que sean de utilidad para los públicos meta seleccionados.
- Permitir el manejo de versiones de registros de información asociada a una especie.
- Modularización (extensiones).

[Volver](#)

Información de especies (Plinian Core)



Taxonomic Names

Species: *Ara ambiguus*

Abstract
The green macaw is a macaw species native to the forests of Central America, Colombia and Ecuador. It is currently an endangered species....



Collection: INBio - Atta
Record id: INB -200
Country: Costa Rica

Record metadata

Natural history

Life cycle

According to data of the Great Green Macaw Research and Conservation Project in the North Zone of Costa Rica, macaws lay their eggs in January and February, incubating them for a period of 30 days....

Habitat

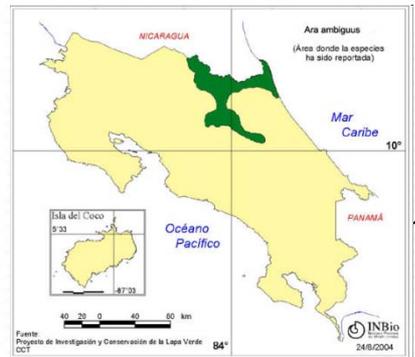
Macaws frequent the canopy of humid lowland forests.

Interactions

Their diet is very specific, as well as the place they select for their nest. In both cases they use the mountain almond tree (*Dipteryx panamensis*)....

Taxonomic Description

It is the second largest Psittacidae of the New World, measuring 79 cm and weighing approximately 1,442 grams. Its tale is relatively shorter and more robust than Scarlet macaw's (*Ara macao*), with much heavier bill...



In Costa Rica, according to Stiles and Skutch (1989), historically the green macaw nested along Caribbean lowlands of Costa Rica....

Distribution

Plinian Core

Conceptos incluidos

Fuente de información

- Código de la institución anfitriona
- Sección de la institución anfitriona
- Código de la institución proveedora de la información
- Sección de la institución proveedora.

Información taxonómica (TCS)

- Nombre científico
- Sinónimos
- Nombre comunes
- Clasificación (jerarquía taxonómica)
- Información del tipo
- Referencias

Información de la página o registro

- Identificador del registro
- Idioma
- Autores
- Colaboradores
- Fecha de publicación
- Fecha de la última modificación
- Público meta
- Versión y código de la versión anterior

[Volver](#)

Plinian Core

Conceptos incluidos

Descripción

- Resumen
- Hábito
- Ciclo de vida
- Reproducción
- Estacionalidad
- Descripción científica
- Descripción breve
- Fenología
- Alimentación
- Comportamiento
- Relaciones
- Datos moleculares
- Número cromosómico N

Demografía y conservación

- Población
- Legislación nacional
- Legislación regional
- Hábitat
- Territorio

Distribución

- Países
- Descripción general
- Endemismo

[Volver](#)

Plinian Core

Conceptos incluidos (continuación)

Usos y folclor

- Usos documentados
- Usos tradicionales
- Folclor

Otros

- Referencias generales
- Información no estructurada
- Información multimedia

[Volver](#)

Preguntas

Muchas Gracias