

ESTÁNDARES PARA EL MANEJO E INTERCAMBIO DE INFORMACIÓN SOBRE BIODIVERSIDAD

Francisco Pando - GBIF España

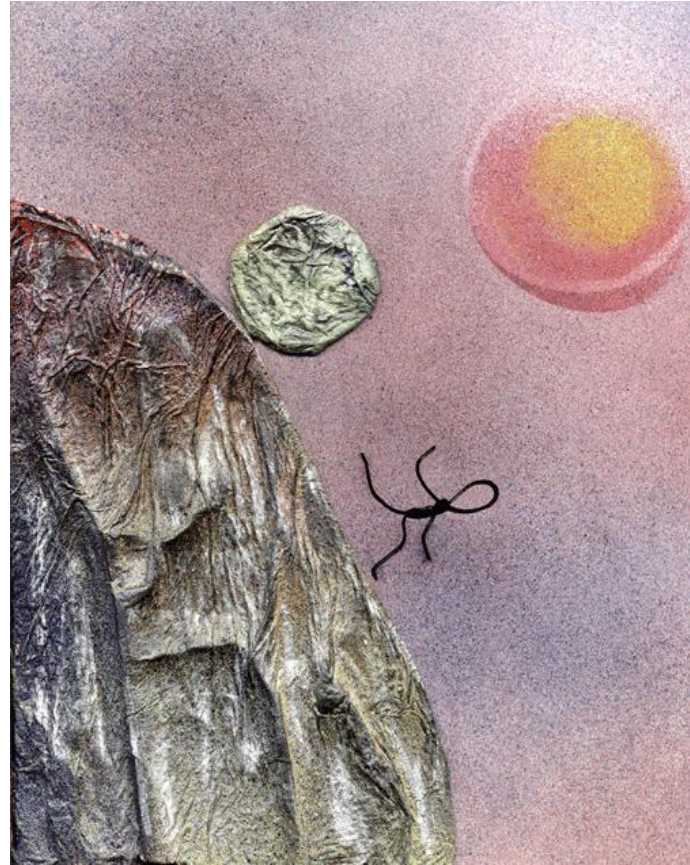
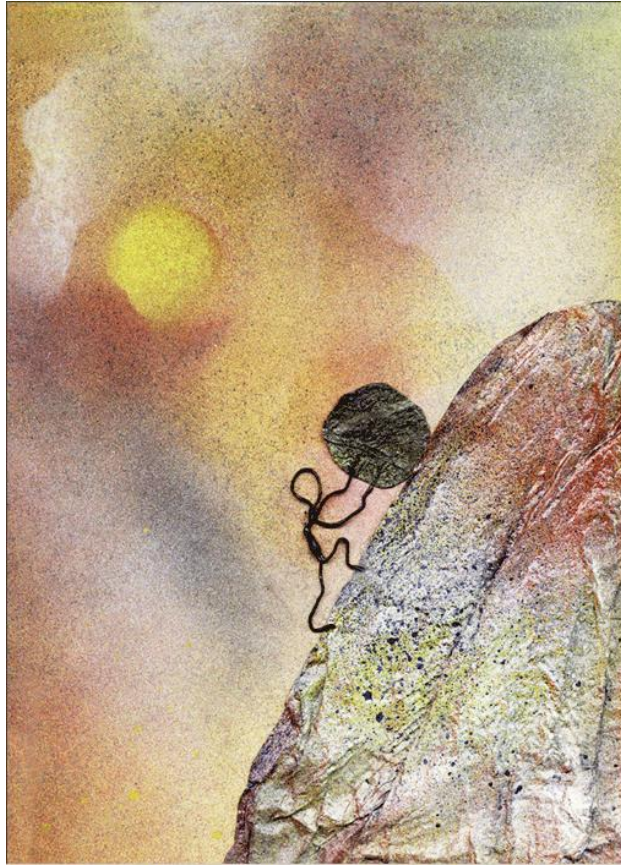
Métodos y Calidad en la Digitalización de Colecciones
CENPAT Puerto Madryn, Argentina 2012



Sumario

- **Compartir, para qué**
- **Estandarizar, qué**
- **Estandarización, para qué**
- **Tipos de estándares**
- **Qué hay**
- **Qué va a haber**
- **Consideraciones finales**

Compartir para qué



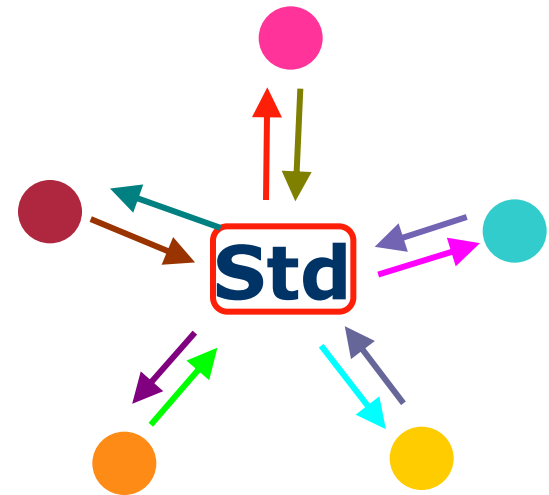
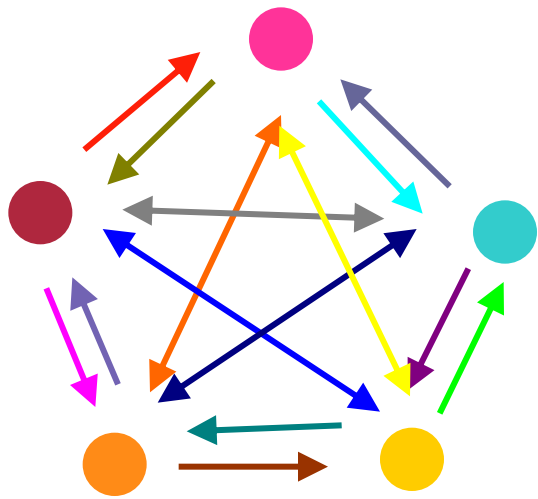
© The Art of
Steven Nelson

HEYWOOD, V.H. (1974). Systematics-the stone of Sisyphus. Biol. J. Linn. Soc. 6(2): 169-178.

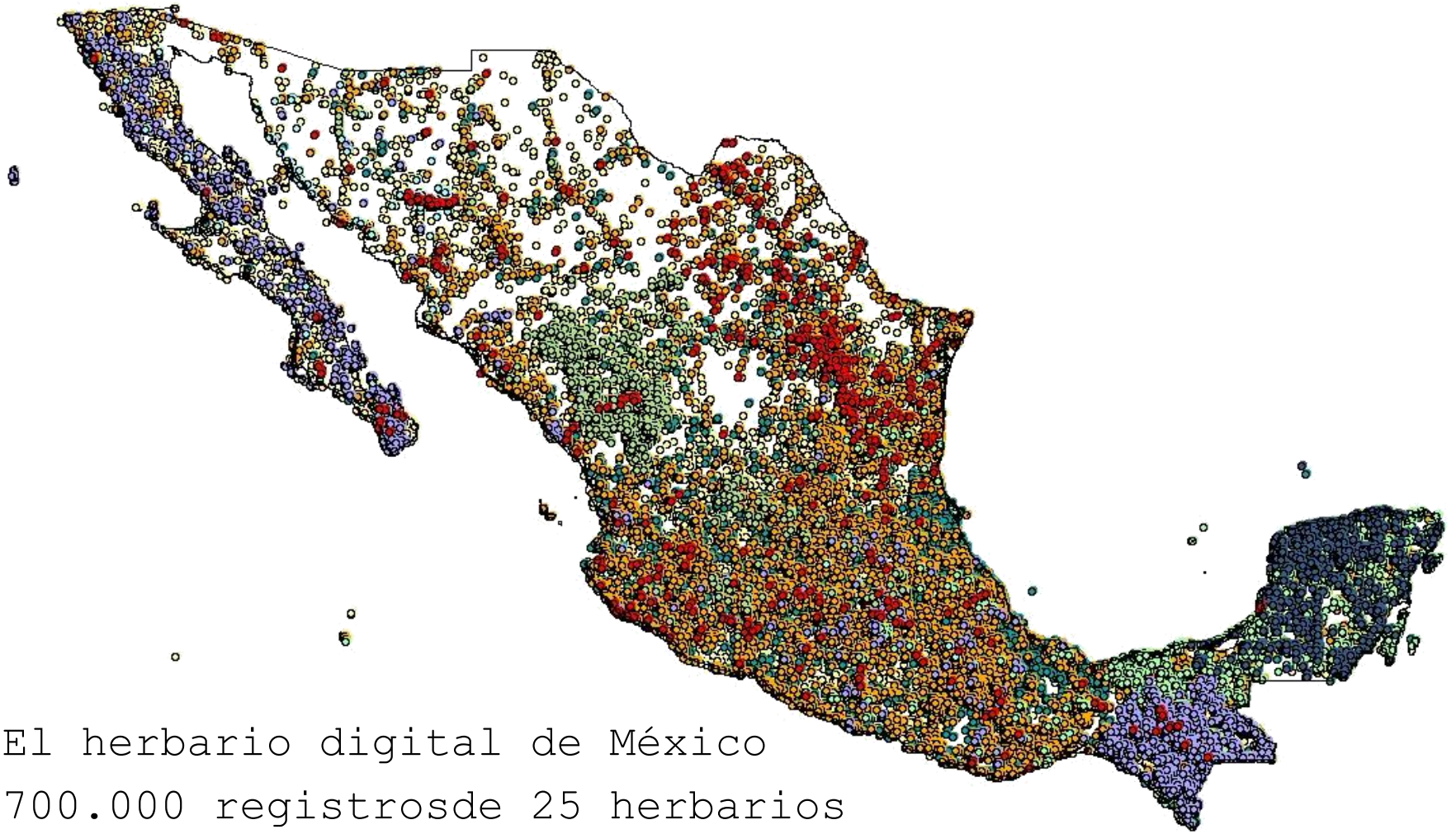
Estándarizar para qué

- Estándarizar para intercambiar
- Estándarizar para utilizar
- Estándarizar no para producir

Estandarizar para intercambiar



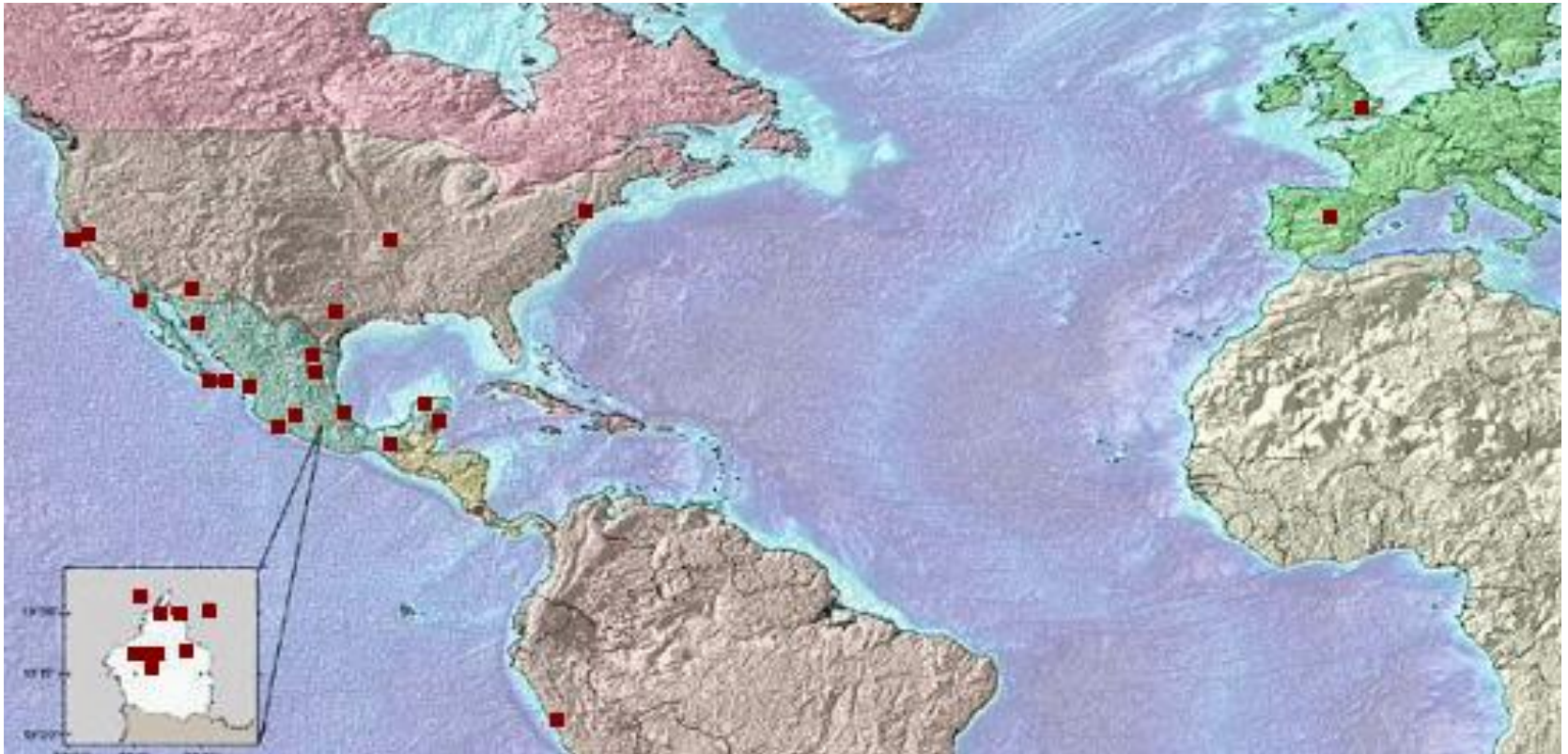
Estandarizar para utilizar



El herbario digital de México
700.000 registros de 25 herbarios
En México y E.E.U.U.

J. Soberón, 2004

Red mundial de información sobre biodiversidad (CONABIO)



Estandarizar no para producir

Un estándar de intercambio no es una buen guía para diseñar una base de datos

- ¿Dos sitios para el nombre científico?
- ¿Quién ha hecho la identificación?
- ¿Imágenes?
- ¿Enlaces a datos moleculares?
- ¿Datos de hábitat?

Una base de datos tiene que servir a nuestros fines primero, después viene compartir y diseminar

Attr#	Name	Description
1	ScientificName	Genus [+ " " + species [+ " " + subspecies]
2	Kingdom	The kingdom to which the organism belongs.
3	Phylum	The phylum (or division) to which the organism belongs.
4	Class	The class name of the organism.
5	Order	The order name of the organism.
6	Family	The family name of the organism.
7	Genus	The genus name of the organism.
8	Species	The species name of the organism.
9	Subspecies	The subspecies name of the organism.
10	InstitutionCode	A unique identifier for your institution.
11	CollectionCode	Unique identifier for the collection within the institution.
12	CatalogNumber	Unique identifier for the specimen record.
13	Collector	The name of the collector or collectors (last name, initials, and observation) from the field.
14	Year	The year (four digit) in which the specimen was collected.
15	Month	The month of the year (1..12) in which the specimen was collected.
16	Day	The day of month that the specimen was collected.
17	Country	The country or major political unit (oceanic island).

Tipos de estándares

- **Vocabularios controlados (léxicos)**
 - Que valores son válidos
 - (listas de géneros, ...)
- **Estructurales (sintácticos)**
 - Como se codifica la información
 - (v.gr.: DwC , ABCD)
- **Definiciones (semánticos)**
 - para humanos, para máquinas
 - (v.gr.: HISPID)

Vocabularios controlados



Ficha nomenclatural

Realizar nueva búsqueda

Eremophilus mutisii

Autor: Humboldt, 1805

Reino: Animalia

Nivel taxonómico: Especie

Estado taxonómico: Válido

Taxones relacionados:

Estado de revisión/verificación: Revisado

Fecha de revisión/verificación:

Jerarquía taxonómica:

En rojo aparecen los nombres no válidos / no aceptados

Reino: [Animalia](#)

Phylum: [Chordata](#)

Subphylum: [Vertebrata](#)

Superclase: [Osteichthyes](#)

Clase: [Actinopterygii](#)

Subclase: [Neopterygii](#)

Infraclase: [Teleostei](#)

Superorden: [Ostariophysi](#)

Orden: [Siluriformes](#)

Familia: [Trichomycteridae](#)

Género: [Eremophilus](#)

Especie: *Eremophilus mutisii*

Referencias

Fuente de la descripción :

Original: Humboldt, E.H.A. von (1805) Mémoire sur l' Eremophilus et Astroblepus.

Estructurales (sintácticos)

<record>

<dwc:DateLastModified>Mar 7, 2004</dwc:DateLastModifie

<dwc:InstitutionCode>Univeristy of Alaska Museum of th
North</dwc:InstitutionCode>

<dwc:CollectionCode>Mammal

<dwc:CatalogNumber>15985</dw

<dwc:ScientificName>Puma con

<dwc:Kingdom>Animalia</dwc:I

<dwc:Phylum>Chordata</dwc:Pl

<dwc:Class>Mammalia</dwc:CL

<dwc:Order>Carnivora</dwc:O

<dwc:Family>Felidae</dwc:Fa

<dwc:Genus>Puma</dwc:Genus>

<dwc:Species>concolor</dwc:Sp

<dwc:Subspecies xsi:nil='tr

<dwc:ScientificNameAuthor>(

<dwc:YearCollected>1985</dw

<dwc:MonthCollected>1</dwc:M

<dwc:DayCollected>27</dwc:D

<dwc:ContinentOcean xsi:nil

<dwc:Country>UNITED STATES<

CODATA/TDWG working on Biological Collection Data (supported by BioCASE)

ABCD 2.06

1 number of levels shown **REFRESH**

show complex types **SHOW SELECTED**

show nesting level **ENTRANCE POINTS**

show attributes **PREVIOUS**

Structured Element Annotation

- Copyrights -

Documentation Container element for one to several statements, normally representing different language representations of the same content.

AppInfo Nothing found.

XPath =
/xs:schema/xs:complexType[@name="IPRStatements"]/xs:sequence/xs:element[@name="Copyrights"]

Definiciones (semánticos)

HISPID

Herbarium Information Standards and Protocols for Interchange of Data

Genus Name

Transfer code: **gen**

TDWG Short name: GENUS

Description: The name of the genus of a plant or ,if part of a hybrid formula or intergrade, then the name 'parent' of that formula or intergrade, entered in full.

Relevant standards: ABIS, ITF, MFN, Bisby (1994).

Domain/Range/Values: Alpha; any valid genus name, capitalisation of the first letter only.

Comments: This field, combined with the following **Species Epithet** and **Infraspecies Epithet** fields, contains minimum taxonomic information for the identity of the specimen.

This field may contain an embedded hyphen, in which case the second word is in lowercase

1. This field *must* contain one of the following:

1.1 A validly published generic name under the ICBN or a manuscript name.

If the **Hybrid Flag** field is *not* filled (not transferred), then this field must contain:

1.1.2 A nonhybrid name, validly published under the ICBN or a manuscript name.

If the name of the genus of the plant is unknown, then:

Qué hay (la situación mundial)

- TDWG
 - “Taxonomic Databases Working Group”
 - Biodiversity Information Standards--
- Ámbitos más reducidos, estándares más ricos (estándares nacionales, temáticos, ..)

[Home](#)
[About TDWG](#)
[Getting Started](#)
[Standards](#)
[Activities](#)
[Membership](#)

[Conference 2012](#)

[Biodiversity Information
Projects of the World](#)

[Biodiversity Information
Networks Database](#)

[Biodiversity Informatics
Events Database](#)

[Wiki](#)

[OJS](#)

[Mailing Lists](#)

Basic Standards Recommendations

The most widely deployed formats for biodiversity occurrence data are [Darwin Core \(wiki\)](#) and [ABCD \(wiki\)](#). New deployments of these and other XML based formats should use the [TAPIR](#) exchange protocol.

The TDWG community's priority is the deployment of Life Science Identifiers (LSID), the preferred [Globally Unique Identifier](#) technology and transitioning to RDF encoded metadata as defined by a set of simple [vocabularies](#). All new projects should address the need for tagging their data with LSIDs and consider the use or development of appropriate vocabularies.

TDWG's activities within the biodiversity informatics domain can be found in the [Activities](#) section of this website.

2012 TDWG annual conference Announcement

The Biodiversity Information Standards (TDWG) Executive Committee is pleased to announce that the 2012 TDWG annual conference will be held on October 22-26, 2012 at the China People's Palace in Beijing, China.

Latest News

[30-May-2012 Deadline to apply to host TDWG meeting in 2013 expression of interest extended to the 15th of June 2012](#)

[\[more\]](#)

[29-May-2012 Travel Awards for African Participation in 2012 Biodiversity Information Standards \(TDWG\) Conference](#)

TDWG Standards

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

TDWG's standards are listed below. For information about current work within TDWG, we would suggest that you look at the [TDWG Wiki](#). For further information, please [contact us](#).

TDWG Current Standards (*)

Title	Activity (Task or Interest Group)	Category	Status	Date Published	
GUID and Life Sciences Identifiers Applicability Statements		Applicability Statements	Current	30-Jan-2011	Download
Darwin Core	DarwinCore Task Group (DwC)	Technical Specification	Current	09-Oct-2009	Download
TAPIR - TDWG Access Protocol for Information Retrieval	TAPIR Task Group	Technical Specification	Current	09-Sep-2009	Download

TDWG Current (2005) Standards (*)

Title	Activity (Task or Interest Group)	Category	Status	Date Published	
Access to Biological Collection Data - version 2.06	Access to Biological Collections Data	Technical Specification	Current (2005)	16-Sep-2005	Download
Structured Descriptive Data	Biological	Technical	Current	16-Sep-2005	Download

TDWG Draft Standards (*)

Title	Activity (Task or Interest Group)	Category	Status	Date Published	
TDWG Standards Documentation Specification	TDWG Infrastructure Project	Technical Specification	Draft		Download (**)
Natural Collections Descriptions (NCD): A data standard for exchanging data describing natural history collections	Natural Collections Descriptions Interest Group	Technical Specification	Draft		Download (**)
MRTG Submission		Technical Specification	Draft		Download (**)
ABCDDNA ♦ DNA extension for Access to Biological Collection Data	Access to Biological Collections Data	Technical Specification	Draft		Download (**)

TDWG Prior Standards ^(*)

Title	Activity (Task or Interest Group)	Category	Status	Date Published	
HISPID3 - Herbarium Information Standards and Protocols for Interchange of Data	Observation and Specimen Records	Technical Specification	Prior	01-Oct-1996	Download
Economic Botany Data Collection Standard	Economic Botany Interest Group	Best Current Practice	Prior	01-Oct-1995	Not available for download
Plant Occurrence and Status Scheme		Status and Categories	Prior	01-Oct-1995	Download
Plant Names in Botanical Databases		Best Current Practice	Prior	01-Oct-1995	Download
Authors of Plant Names		Status and Categories	Prior	01-Oct-1992	Not available for download
World Geographical Scheme for Recording Plant Distributions		Status and Categories	Prior	01-Oct-1992	Download
XDF - A Language for the Definition and Exchange of Biological Data Sets		Technical Specification	Prior	01-Oct-1991	Not available for download
Botanico-periodicum-huntianum/supplementum		Status and Categories	Prior	01-Oct-1991	Not available for download
Index Herbariorum. Part I: The Herbaria of the World		Status and Categories	Prior	01-Oct-1990	Not available for download
International Transfer Format for Botanic Garden Plant Records		Technical Specification	Prior	01-Oct-1987	Download
Floristic Regions of the World		Status and Categories	Prior	01-Oct-1986	Not available for download
Taxonomic Literature, ed. 2 and its Supplements		Status and Categories	Prior	01-Oct-1976	Not available for download
Botanico-periodicum-huntianum		Status and Categories	Prior	01-Oct-1975	Not available for download

Sobre Darwin Core

Biodiversity
Information
Standards
TDWG

Introduction

References

Quick Reference Guide

Term Index

Record-level Terms

Occurrence

Event

Location

GeologicalContext

Identification

Taxon

ResourceRelationship

MeasurementOrFact

Term Definitions

Simple Darwin Core

Type Vocabulary

Namespace Policy

XML Guide

Text Guide

Darwin Core Terms: A quick reference guide

Title: Darwin Core Terms: A quick reference guide

Date Issued: 2009-02-12

Date 2011-10-26

Modified:

Abstract: This document is a quick reference for all recommended Darwin Core terms. For complete historical term information, including version changes and pre-standard terms, see [[HISTORY](#)]. For a comparative table of elements from pre-standard versions of Darwin Core to the current terms in the standard, see [[VERSIONS](#)].

Contributors: John Wieczorek (MVZ), Markus Döring (GBIF), Renato De Giovanni (CRIA), Tim Robertson (GBIF), Dave Vieglais (KUNHM)

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

Part of TDWG Standard: <http://www.tdwg.org/standards/450/>

Creator: Darwin Core Task Group

Identifier: <http://rs.tdwg.org/dwc/2011-10-26/terms/>

Latest Version: <http://rs.tdwg.org/dwc/terms/>

Replaces: <http://rs.tdwg.org/dwc/2009-12-07/terms/>

Document Status: Current Standard

Darwin Core: Términos

<http://rs.tdwg.org/dwc/terms/>

Record-level Terms

[dcterms:type](#) | [dcterms:modified](#) | [dcterms:language](#) | [dcterms:rights](#) | [dcterms:rightsHolder](#) | [dcterms:accessRights](#) | [dcterms:basisOfRecord](#) | [dcterms:institutionID](#) | [dcterms:collectionID](#) | [dcterms:datasetID](#) | [dcterms:institutionCode](#) | [dcterms:collectionCode](#) | [dcterms:datasetName](#) | [dcterms:ownerInstitutionCode](#) | [dcterms:basisOfRecord](#) |

Occurrence

[occurrenceID](#) | [catalogNumber](#) | [occurrenceRemarks](#) | [recordNumber](#) | [recordedBy](#) | [individualID](#) | [individualCount](#) | [sex](#) | [lifeStage](#) | [associatedReferences](#) | [associatedOccurrences](#) | [associatedSequences](#) | [associatedTaxa](#)

Event

[eventID](#) | [samplingProtocol](#) | [samplingEffort](#) | [eventDate](#) | [eventTime](#) | [startDayOfYear](#) | [endDayOfYear](#) | [year](#) | [month](#) | [day](#) | [vert](#)

dcterms:Location

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

GeologicalContext

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

Identification

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

Taxon

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

Auxiliary Terms

ResourceRelationship

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

MeasurementOrFact

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

DwC: múltiples versiones

Darwin Core Versions

The following versions of the Darwin Core are of historical significance:

- **DwC 1.2 schema** -- first deployed version (used in GBIF, also known as [DarwinCoreV2?](#))
- **DwC 1.21 schema** -- revised version (used in GBIF, MaNIS, HerpNet, OrNIS, and FishNet2)
- **DwC 1.3 schema** -- draft standard of the Darwin Core accepted at TDWG Meeting 2004, Christchurch, New Zealand. (unused)
- **DwC 1.4 schema** -- draft standard under discussion (not for use)(but used in GBIF, and recommended to use in GBIF see: <http://www.gbif.org/DataProviders/HowTo>)
- **OBIS** -- based on DwC 1.2 used in GBIF, Ocean Biogeographic Information System)
- **PaleoPortal** -- based on DwC 1.2 (used in The Paleontology Portal)

Biodiversity
Information
Standards
T D W G

Introduction

References

Quick Reference Guide

Simple Darwin Core

Type Vocabulary

Namespace Policy

XML Guide

Text Guide

Complete History

Decision History

Mapping to ABCD

Mapping to Old Versions

Darwin Core

Title: Darwin Core

Date Issued: 2009-02-12

Date 2009-10-08

Modified:

Abstract: This document is a cover page, an entry-level document to the Darwin Core standard. It describes the purpose of the standard and orients the reader to the documents that cover specific topics within the standard, such as the quick guide to the list of terms.

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

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

Part of TDWG <http://www.tdwg.org/standards/450/>

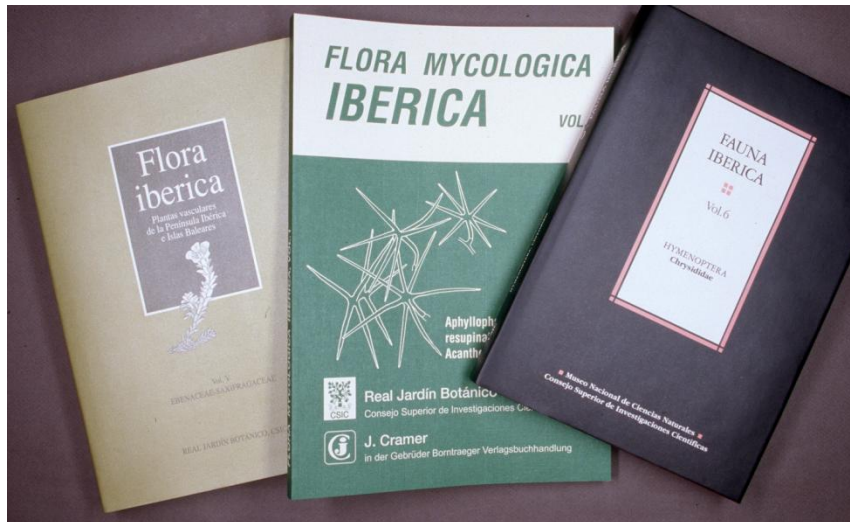
Estándares temáticos y/o nacionales

Manual de las bases de datos nomenclaturales de *Flora Mycologica Iberica* y *Flora iberica*

Francisco Pando, Félix Muñoz Garmendia & Carlos Aedo
Real Jardín Botánico

El objetivo principal de este manual –disponible aquí– está concebido para usarse con los programas NOME el de servir de guía para fichar, en una base de datos organismos, e información relacionada, como p publicación, de su tipo, de su estado nomenclatural, e

http://www.rjb.csic.es/bibmaste/manu_n.html



SiB Sistema de Información
sobre Biodiversidad de Colombia

Enlaces

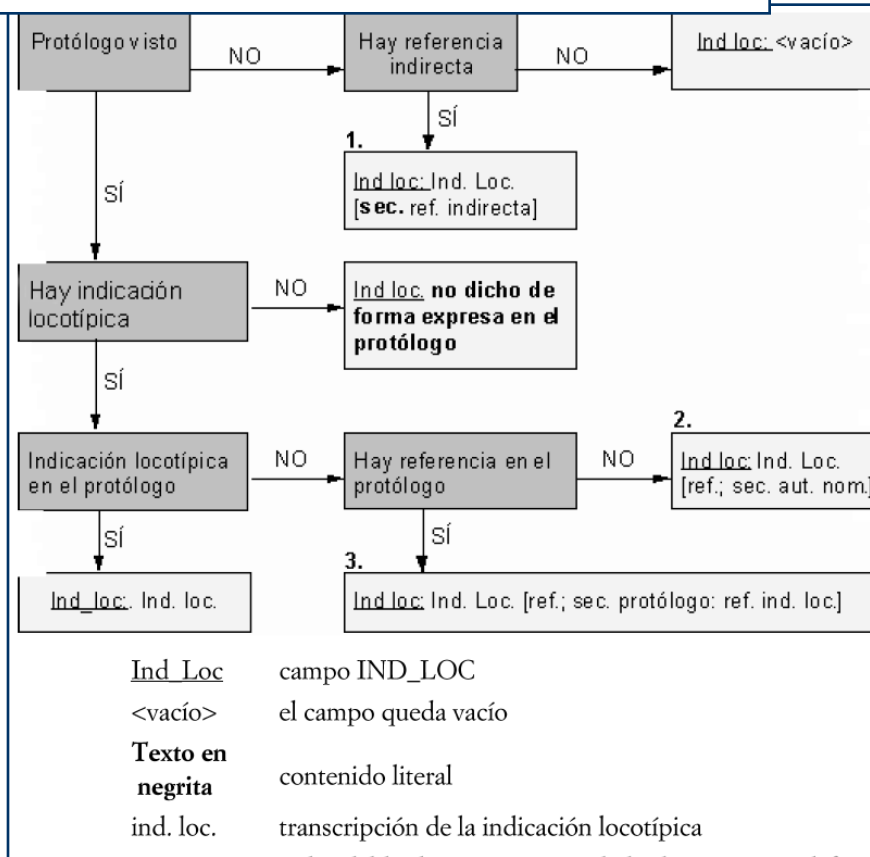
- [Archivos de autoridad taxonómica](#)

AAT disponibles

1. [Vitolopez A. L. \(2004\) Archivo de autoridad taxonómica de esc \(Coleoptera: Cicindelidae\) de Colombia](#)
2. [Díaz-Ruiz M., Díaz-Pulido G. \(2004\) Archivo de autoridad taxonómica de macroalgas verdes \(Chlorophyta\) del Caribe colombiano](#)
3. [Quiroga-Cárdenas S. Y., Bolaños-Rodríguez D. M. \(2003\) Polyclad Tropical Occidental](#)
4. [Martínez C. & G. Ball \(2003\) Los Platynini \(Coleoptera: Carabidae\)](#)
5. [Calderón Saenz E., Farfán Camargo J. \(2004\) AAT de 13 géneros Colombianas](#)
6. [Murillo, J. \(2003\) Euphorbiaceas de Colombia](#)
7. [Sendoya S., Fernández F. \(2004\) AAT de hormigas \(Hymenoptera\) neotrópico](#)
8. [Rodríguez-P M. E. & Y. Muñoz-Saba \(2004\) Murciélagos de la familia Colombia](#)
9. [Calderón-Sáenz E., Tobón I. C. \(2004\) Catálogo taxonómico de los Cycnoches, Embreea, Lycaste, Otoglossum, Phragmipedium, Psychotria, Schlimia y Selenipedium \(Orchidaceae\) para Colombia](#)
10. [Brummitt R. K. \(1992\) Vascular plant families and genera: a list](#)

Estándares más ricos, también más restringidos

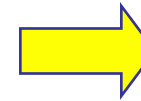
Estándar proyectos *Flora iberica & Flora Mycologica Iberica*



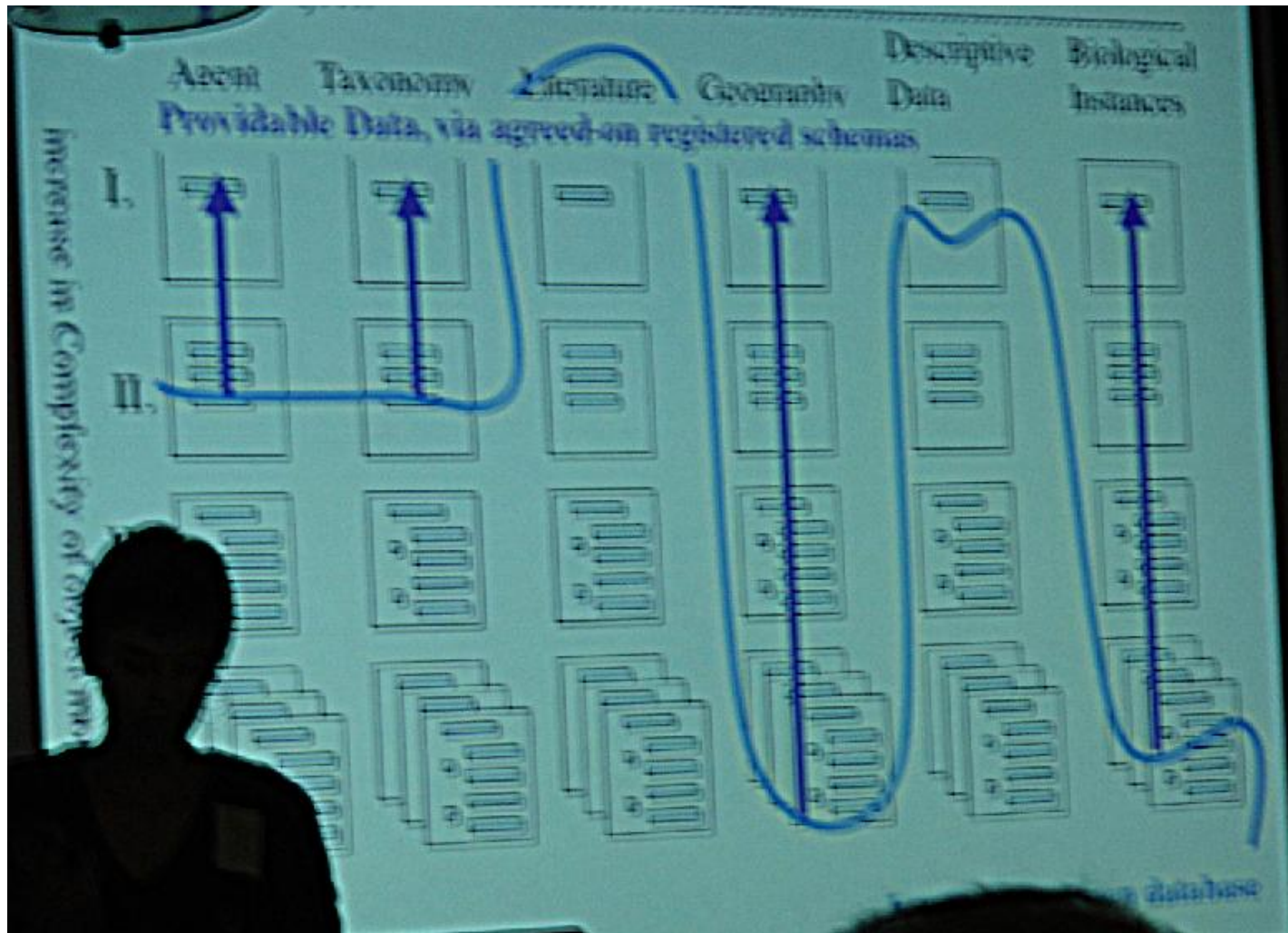
Estándar TDWG & Red de datos del GBIF

Darwin Core 1.2

Type: is the specimen a nomenclatural type or not



La clave: agregar o atomizar de manera compatible



TDWG & GBIF

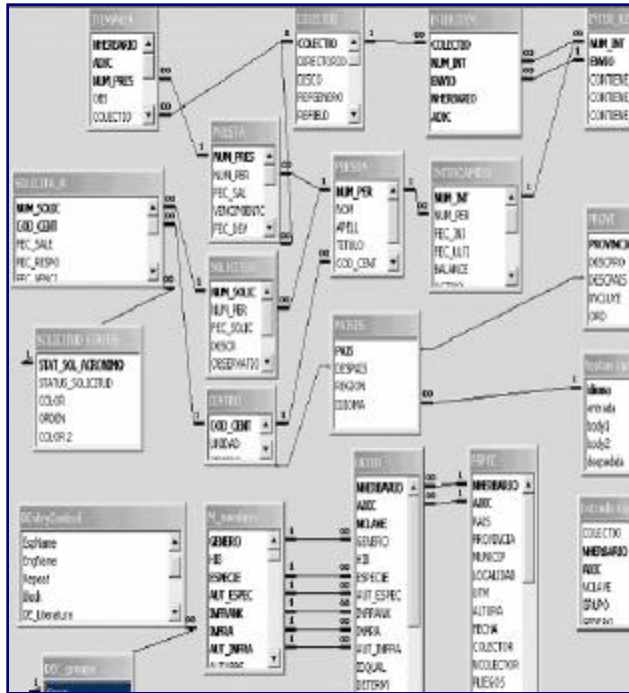
**TDWG proporciona estándares y
comunidad de expertos para las
redes de datos del GBIF**

**GBIF marca la dirección y las
necesidades de desarrolladores y
usuarios**

Que va a haber

- Web semántica
- Ontologías
- esquemas xml,
- triple storage
- ...

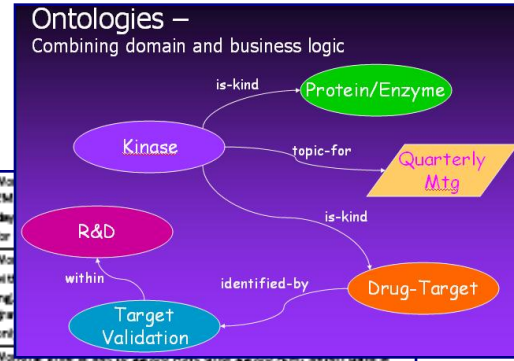
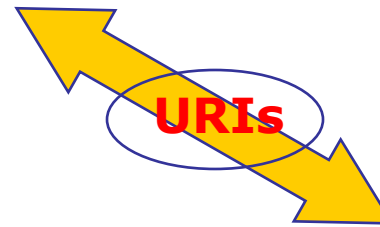
Web semántica: comunicar contenidos



Las definiciones de las bases de datos se almacenan (o hacen uso) en un sistema:

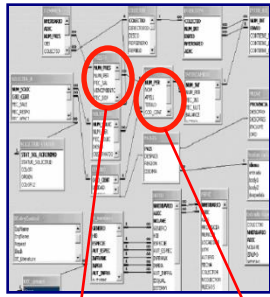
- *Accesible vía Web*
- *Ininteligible por “maquinas”*
- *Relaciona conceptos, relaciones e instancias*

→ **ontologías**



Standard	Work 8 AM to 5 PM, 5 days a week
Night Shift	Work 11 PM to 7 AM, 5 days a week
24 Hours	Working time is split into three shifts: day, evening and night. Each shift is 8 hours long, so that work never stops. This base calendar is typically used for projects in a manufacturing situation, for example, which might run two or three back-to-back shifts every day of the week.

Web semántica: Triple storage



<i>Sujeto</i>	<i>Predicado</i>	<i>objeto</i>
25322	genero	Inga
25322	especie	alba
25322	pais	COL
COL	Pais largo	Colombia

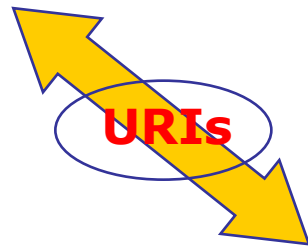
XML

Especimenes

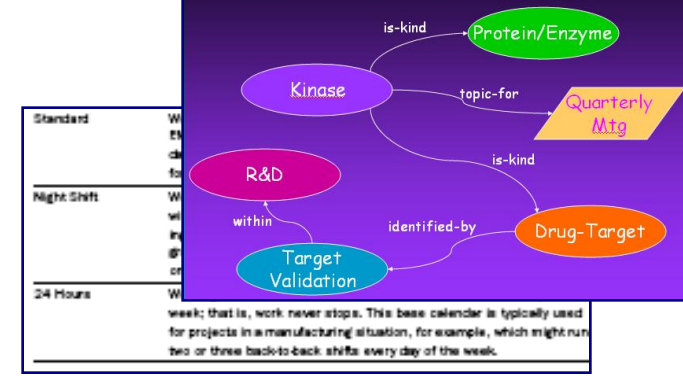
<i>Campo</i>	<i>valor</i>
Nmr_ejemplar	25322
<i>genero</i>	<i>Inga</i>
<i>especie</i>	<i>alba</i>
<i>pais</i>	<i>COL</i>

Pais

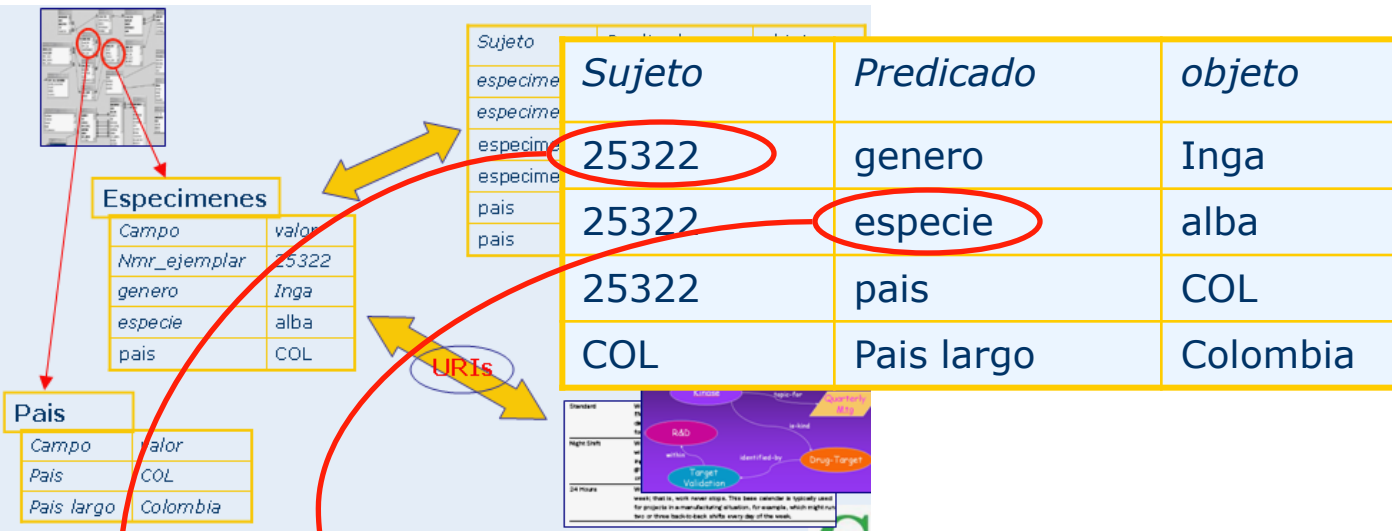
<i>Campo</i>	<i>valor</i>
Pais	COL
<i>Pais largo</i>	<i>Colombia</i>



Ontologies – Combining domain and business logic



Web semántica: identificadores únicos

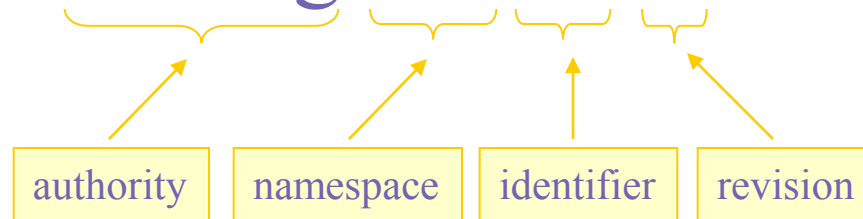


• **Identificadores únicos de conceptos (ontología)**

• **Identificadores únicos de objetos**

- Life Science Identifiers (LSID) son un tipo de identificadores globales
 - Facilitar comunicación e intercambio
 - Más robustos que URL
 - Se pueden resolver (esto es, los entienden las máquinas)

urn:lsid:esa.org:esa:26:3



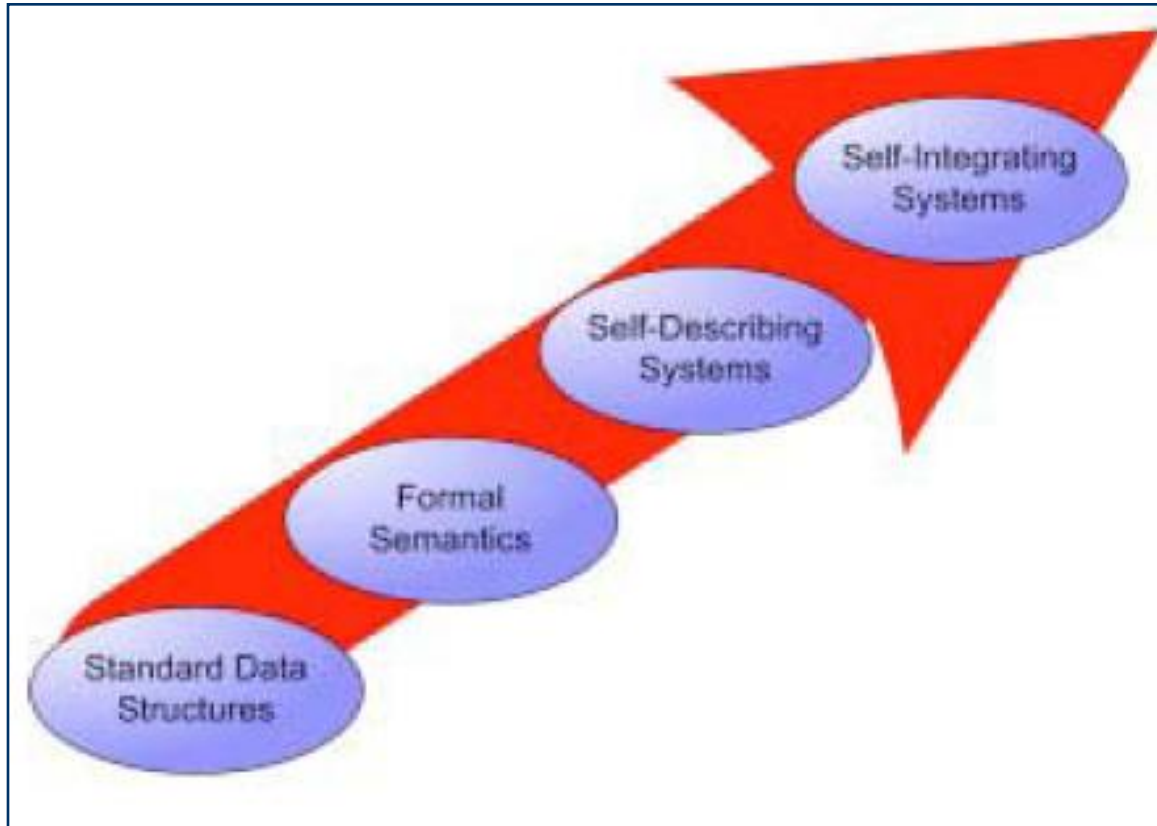
Ver: Adoption of Persistent Identifiers for Biodiversity Informatics

Web semántica: navegar los datos como el metro de una ciudad



Bases de datos conectadas por ontologías y la Web
Conseguir que podamos inferir hechos a partir de datos
distribuidos

Hacia dónde vamos



La idea es que la web funciona como un único sistema de información

Web semántica en acción: "Linked (open) data"

¿Qué es Linked Data?

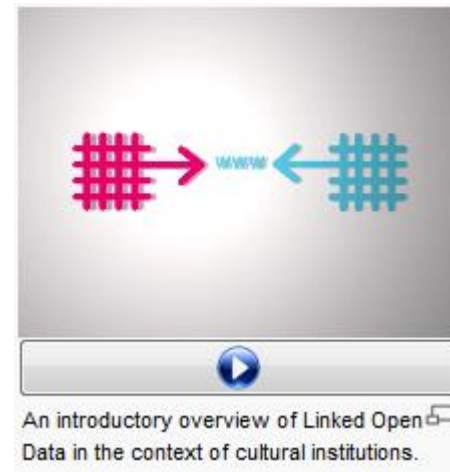
Los Datos Enlazados es la forma que tiene la [Web Semántica](#) de vincular los distintos datos que están distribuidos en la Web, de forma que se referencian de la misma forma que lo hacen los enlaces de las páginas web.

La Web Semántica no se trata únicamente de la publicación de datos en la Web, sino que éstos se pueden vincular a otros, de forma que las personas y las máquinas puedan explorar la web de los datos, pudiendo llegar a información relacionada que se hace referencia desde otros datos iniciales.

¿Cómo funciona?

Los Datos Enlazados, como parte de la Web Semántica, se basa en la aplicación de ciertos principios básicos y necesarios, que fomentarán el crecimiento de la Web, tanto a nivel de los documentos [HTML](#) (vista clásica de la Web), como a nivel de los datos expresados en [RDF](#) (vista de la Web Semántica).

1. Usar [URIs](#) para identificar las cosas
2. Usar [URIs HTTP](#)
3. Ofrecer información sobre los recursos usando [RDF](#)
4. Incluir enlaces a otros URIs



- <http://www.youtube.com/watch?v=uju4wT9uBIA>

Estándares y web semántica

- Los estándares del TDWG son compatibles con la Web semántica
- Resolver ontologías
 - No nos libramos de los mapeos
- Resolver LSIDs
 - Puede haber más de uno...
- Lo que hagamos debe tener en cuenta el escenario de Web semántica

Consideraciones finales

- No inventar la rueda
- Es mejor un estándar ampliamente usado que uno más rico de menor implantación
- Un estándar de intercambio no debe ser tomado como una norma de implementación en un sistema de gestión/producción

A vuestra disposición

Francisco Pando

Responsable

GBIF España, U. de Coordinación

Real Jardín Botánico-CSIC

Pza. de Murillo, 2

28014 Madrid, España

pando@gbif.es

Tel.: + 34 91 420 3017

Fax: + 34 91 429 2405

www.gbif.es

www.gbif.org