



# Taller de GEOLOCATE

---


David Draper  
Nelson Ríos

Taller de GEOLOCATE, Madrid  
24-26 marzo 2009 GBIF

# Objetivos

---

- Presentar la herramienta y sus capacidades
- Definir una sesión básica de georreferenciación con GEOLocate
- Practicar bajo diversas situaciones
- Identificar problemas y necesidades



# Conceptos básicos de cartografía

---

David Draper

Taller de GEOLOCATE, Madrid  
24-26 marzo 2009 GBIF

# 1 Geoide



Se define como, las superficies en las cuales el potencial de gravedad es constante. Para definir un nivel cero de las alturas se toma la superficie promedio de los océanos.

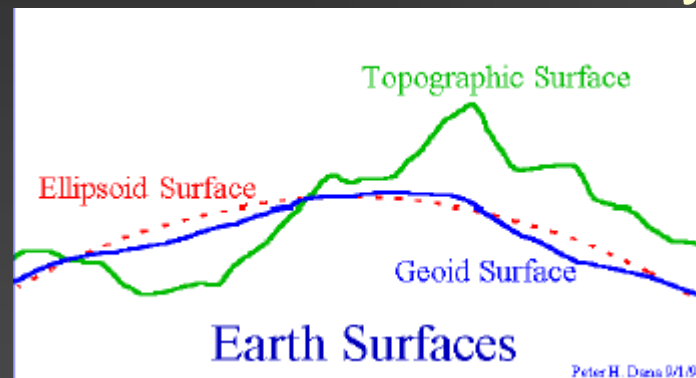
Es la prolongación de la superficie de los océanos por el interior de los continentes

Por la heterogeneidad de los materiales continentales se observan diversas variaciones en la gravedad. EL GEOIDE CONTINUA SIENDO UNA SUPERFICIE IRREGULAR.



## 2 Elipsoide

El elipsoide de revolución (esfera achatada en los polos) es un modelo matemático de la Tierra utilizado para realizar cálculos y que se sitúa lo más cerca



Existen numerosos modelos de elipsoides.  
La mayoría de nuestra cartografía esta representada sobre el Elipsoide de Hayford



# 3 Coordenadas geográficas

Cada punto de la superficie de la tierra se define por una latitud y una longitud

Latitud: es la medida, en grados, del arco del meridiano comprendido entre el ecuador y el paralelo del lugar  
-90° a 90° (o 90°S a 90°N)

Longitud: es la medida, en grados, del arco del paralelo comprendido entre el meridiano de origen y el meridiano del lugar  
-180° a 180° (o 180°E a 180°W)



# 4 Proyecciones

---

- Es la transformación de la red de meridianos y paralelos (red geográfica en el elipsoide o en la esfera) sobre una superficie plana
- Se utiliza una proyección para:
  - representar sobre una superficie plana parte del modelo elipsoidal de la Tierra.
  - obtener valores métricos utilizables más fácilmente que las unidades angulares.
  - hacer más fácil la evaluación de distancias.



# 4 Proyecciones

- Los sistemas de proyecciones cartográficas se analizan por el tipo de superficie adoptada y el grado de deformación.
- Por el tipo de superficie de proyección adoptada, las proyecciones se clasifican en: planas o azimutales, cilíndricas, cónicas, UTM y poliédricas. Según se represente la superficie curva de la Tierra sobre un plano, un cilindro, un cono o un poliedro tangente o secante a la esfera terrestre.
- La selección del tipo de proyección depende de la zona que se quiere cartografiar y de la escala.



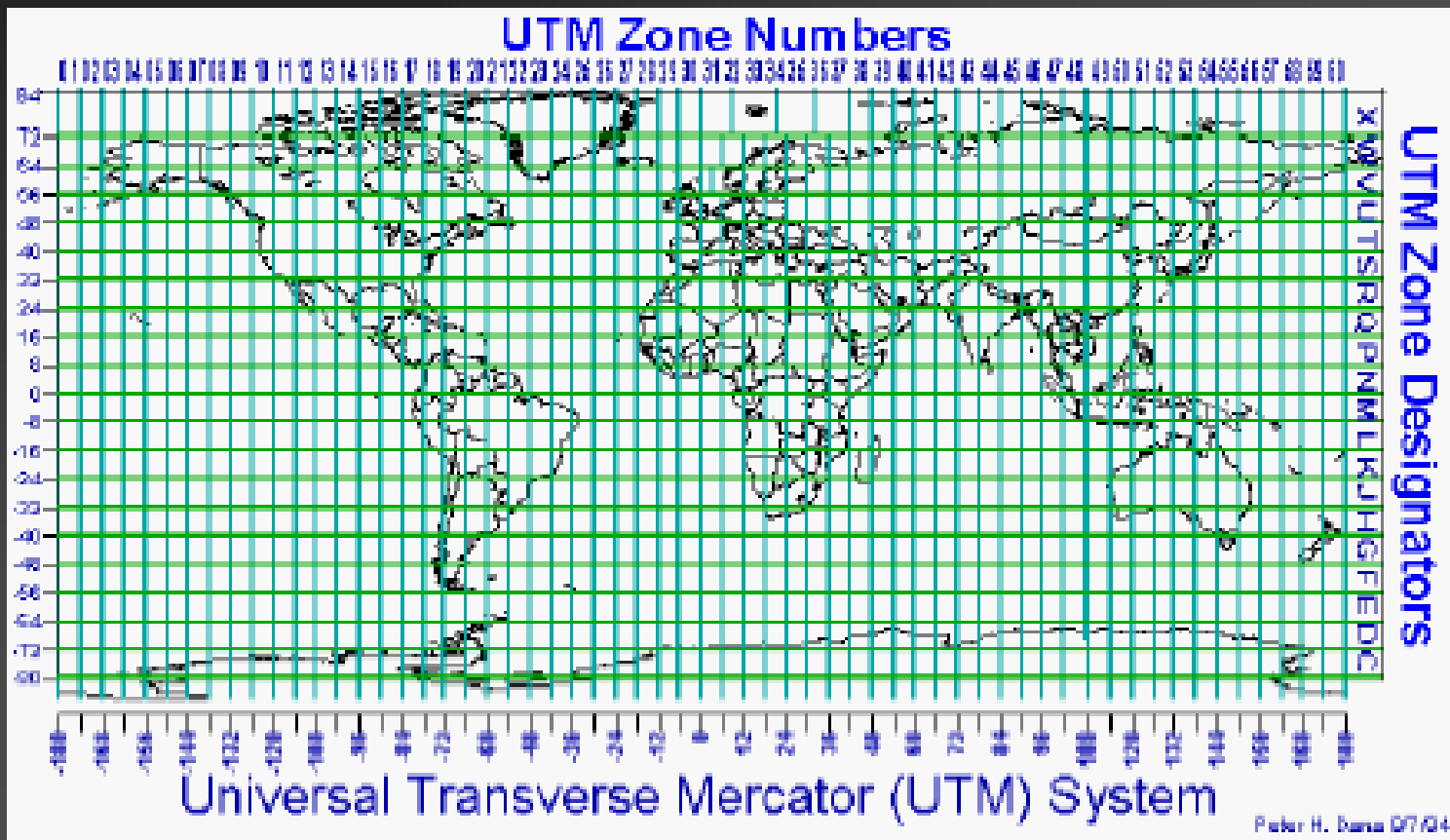


# Sistema UTM

- Surgio tras la 2ª Guerra Mundial por los aliados para normalizar las representaciones cartográficas
- Se trata de una proyección conforme (mantienen las formas)
- Se definen Husos cada 6° de longitud
- Para cada Huso se utiliza un cilindro secante
- Los meridianos y paralelos son ligeramente curvos (a 1:50.000 son prácticamente rectos)

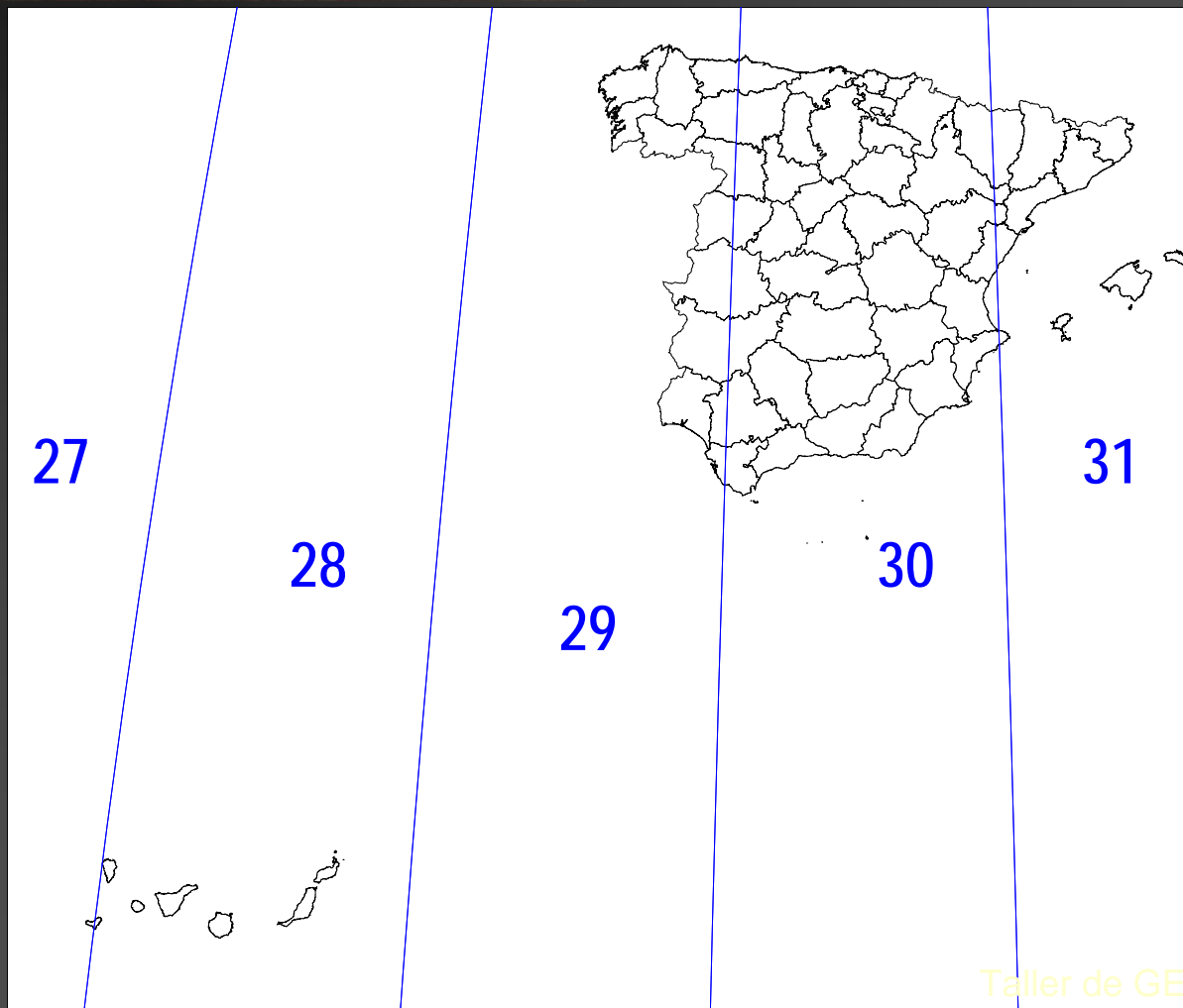


# Sistema UTM



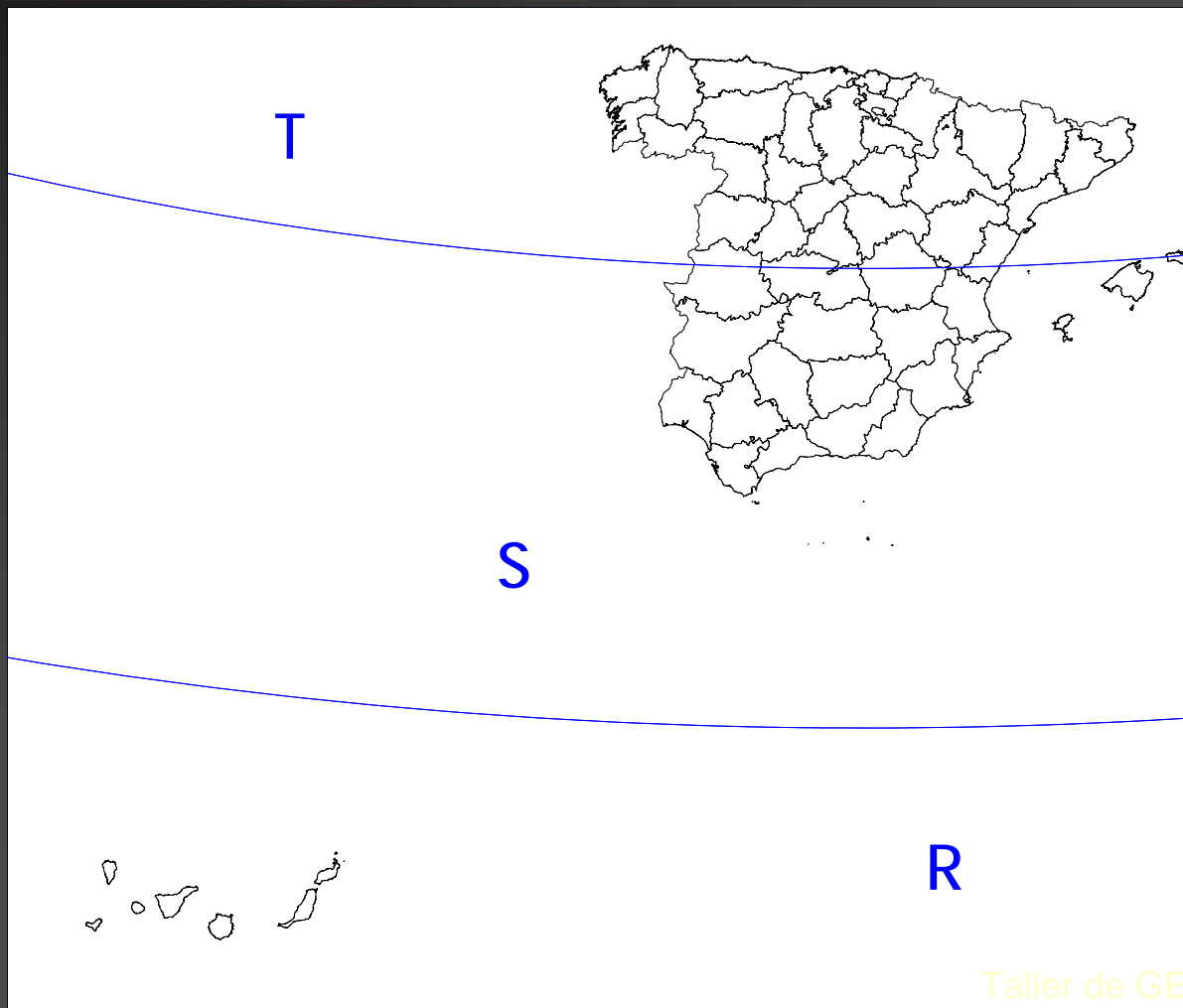


# Sistema UTM: Husos

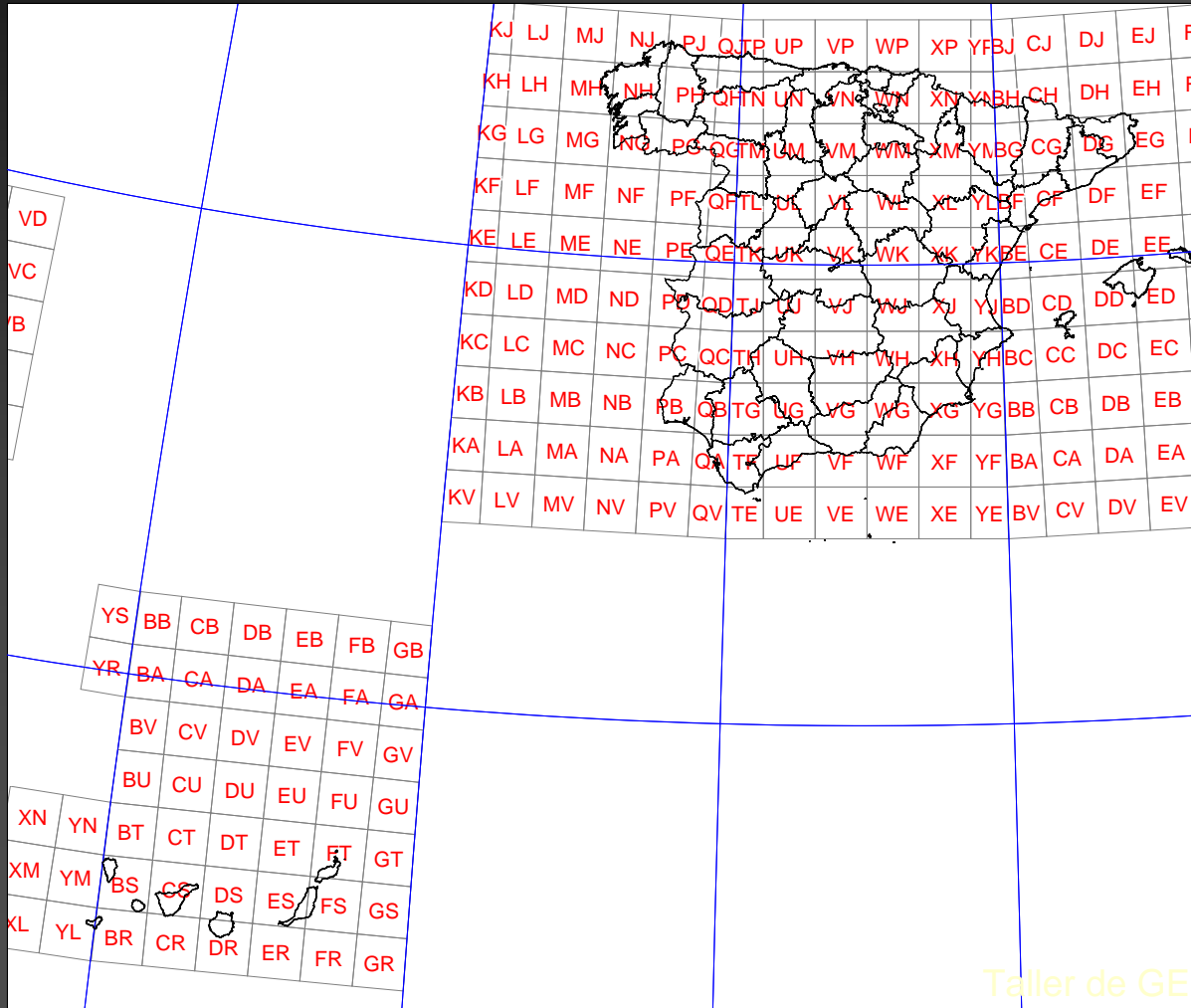




# Sistema UTM: Sectores



# Sistema UTM



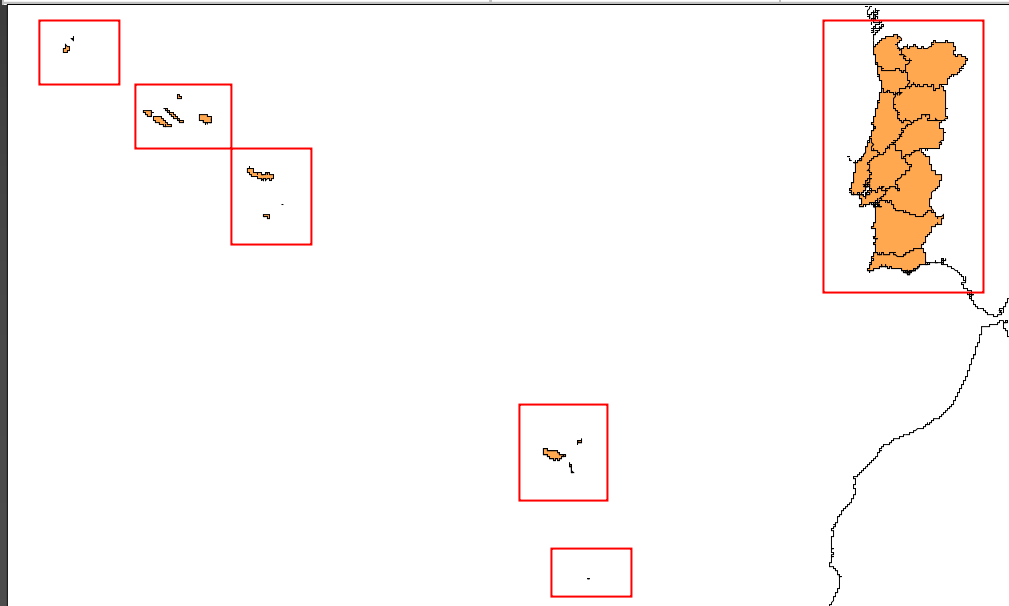
# Datums



- Sistema geodésico de definición local construido a partir de un punto de referencia.
- La selección del datum depende de la zona que se quiere cartografiar.
- En principio corresponde al origen de coordenadas aunque en muchos sistemas se utilizan *falsos orígenes* (UTM).
- Cada Datum se a determinado asociado a un Elipsoide
  - Datum Europeo elipsoide Hayford 1909 (o Internacional 1924)
  - ETRS89, elipsoide GRS80 (~WGS84)
  - WGS84 elipsoide IAG-GRS80 sobre el que se basa la tecnología GPS

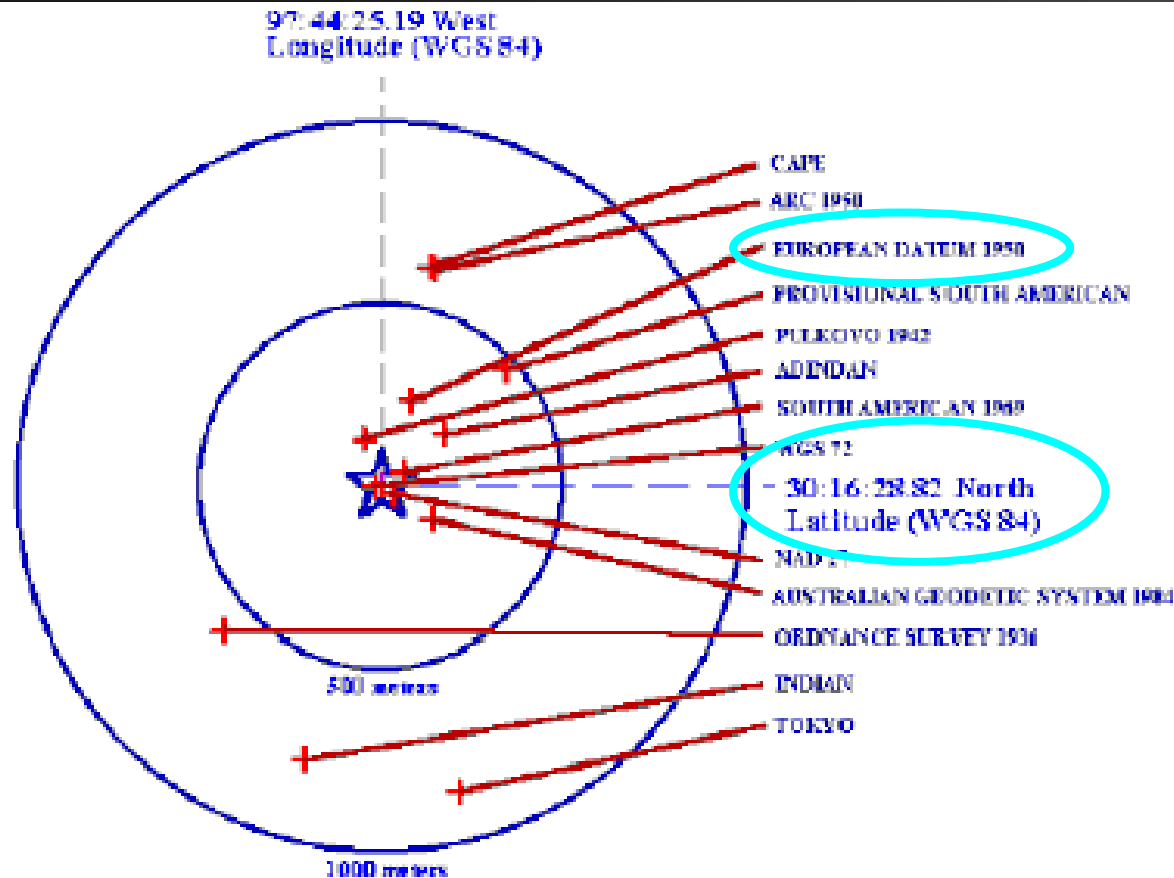
# Datums en Portugal

Name	Map Projection	Area of Use:
Azores Occidental 1939 / UTM zone 25N	UTM zone 25N	Portugal - western Azores - Flores, Corvo.
Azores Central 1948 / UTM zone 26N	UTM zone 26N	Portugal - central Azores - Graciosa, Terceira, Sao Jorge, Pico,
Azores Central 1995 / UTM zone 26N	UTM zone 26N	Portugal - central Azores - Graciosa, Terceira, Sao Jorge, Pico,
Azores Oriental 1940 / UTM zone 26N	UTM zone 26N	Portugal - eastern Azores - Sao Miguel, Santa Maria.
Azores Oriental 1995 / UTM zone 26N	UTM zone 26N	Portugal - eastern Azores - Sao Miguel, Santa Maria.
Porto Santo / UTM zone 28N	UTM zone 28N	Portugal - Madeira, Porto Santo and Desertas islands.
Porto Santo 1995 / UTM zone 28N	UTM zone 28N	Portugal - Madeira, Porto Santo and Desertas islands.
Selvagem Grande / UTM zone 28N	UTM zone 28N	Portugal - Selvagens islands (Madeira province).
Datum 73 / Modified Portuguese Grid	Modified Portuguese	Portugal - onshore.
Datum 73 / UTM zone 29N	UTM zone 29N	Portugal - onshore.
ETRS89 / Portugal TM06	Portugal TM06	Portugal - onshore.
Lisbon (Lisbon)/Portuguese Grid	Portuguese Grid	Portugal - onshore.
Lisbon (Lisbon)/Portuguese National	Portuguese National	Portugal - onshore.
Lisbon 1890 (Lisbon) / Portugal Bonne	Portugal Bonne	Portugal - onshore.
WGS 84 / UTM zone 29N	UTM zone 29N	Between 12 and 6 deg West; northern hemisphere. Portugal -





# Datums



## Position Shifts from Datum Differences

Texas Capital Dome Horizontal Benchmark

Peter H. Dana 8/1/9





# Georreferenciación



David Draper

Taller de GEOLOCATE, Madrid  
24-26 marzo 2009 GBIF

# Punto de la situación

---

- 2.5 billones de especímenes en colecciones biológicas (Cotterill, 1995)
- Se estima que sólo el 5% se encuentran informatizados en bases de datos (Beaman & Conn, 2003)
- Una vez digitalizado, es necesario continuar trabajando para poder utilizar estos especímenes en posteriores análisis
- Un aspecto crucial de esta información es la asociación de unas coordenadas espaciales a los especímenes.

# Punto de la situación

---

- El mayor desafío de las colecciones biológicas es la asignación de las coordenadas espaciales a millones de registros antiguos (Baker & al., 1998)
- El futuro puede ser un poco más alagüeño con la confianza depositada en la tecnología GPS. Con algunas reservas...

# ¿Qué es la georreferenciación?

- La georreferenciación de localidades consiste en la asignación de coordenadas geográficas a la descripción textual de un sitio (Proctor, Blum, Chaplin, 2001; Wieczorek, 2001)
- Asignar coordenadas geográficas a las localidades contenidas en las bases de datos de las colecciones biológicas que carezcan de ellas (posprocesado)
- Georreferenciación no es únicamente asignar unas coordenadas, es necesario determinar la incertidumbre y el grado de precisión de las coordenadas

# Objetivos de la georreferenciación

---

- Asignar coordenadas a partir de la localidad
- Mantener o incrementar la exactitud del registro sin incrementar el error y corrigiendo localidades erróneas
- Mantener o incrementar la precisión evaluando y depurando las localizaciones disponibles para asignar coordenadas espaciales

# Fases del proceso

---

1. Normalización de las localidades
2. Asignación de coordenadas
3. Determinación de la incertidumbre
4. Validación de las coordenadas

# 1 Normalización de las localidades

- La información mínima necesaria que se debe localizar en la base de datos para poder realizar la georreferenciación es la siguiente:
  - País
  - Entidad federativa (división política)
  - Municipio (división administrativa)
  - Descripción del sitio de colecta: rasgo principal de referencia, distancia, dirección o rasgo que indique una dirección
- Identificación de duplicados:
  - **registros redundantes** que son aquellos que pertenecen a distintos especímenes pero que fueron colectados en la misma localidad (por lo tanto la descripción de localidad es la misma)
  - **falsos duplicados** que corresponden a registros de distintos colectores que trabajaron sobre una misma localidad (estos registros presentan una sintaxis distinta pero se refieren al mismo sitio)

# 1 Normalización de las localidades

CONABIO 2006

Redundantes

Redundantes con errores de captura

Falsos duplicados

	Localidad
{	0.25 mi E state boundary, 8 mi SW Calvillo
	0.25 mi E state boundary, 8 mi SW Calvillo
{	0.33 mi NW Ciudad de Los Ninos, 7 mi WSW Aguas Calientes
	0.33 mi NW Cuidad de Los Minos, 7 mi WSW Aguascalientes
	0.33 mi S and 1.25 E Rincon de Romos
	0.33 mi S and 1.25 mi E Rincon de Ramos
{	0.33 mi S, 1.25 mi E Rincon de Romos
	Rincon de Romos, 5 mi W
	0.5MI W RICON DE ROMOS
	0.5 mi S La Labor
	0.5 mi N Presa Calle
	0.5 mi S LaLabor



# 1 Normalización de las localidades

Homogenización de los registros

Localidad	Descripción_homog
0.25 mi E state boundary, 8 mi SW Calvillo	0.25 mi E state boundary, 8 mi SW Calvillo
0.33 mi NW Ciudad de Los Ninos, 7 mi WSW	0.33 mi NW Ciudad de Los Ninos, 7 mi WSW
0.33 mi NW Cuidad de Los Minos, 7 mi WSW	0.33 mi NW Ciudad de Los Ninos, 7 mi WSW
0.33 mi S and 1.25 E Rincon de Romos	0.33 mi S and 1.25 mi E Rincon de Ramos
0.33 mi S and 1.25 mi E Rincon de Ramos	0.33 mi S and 1.25 mi E Rincon de Ramos
0.33 mi S, 1.25 mi E Rincon de Romos	0.33 mi S and 1.25 mi E Rincon de Ramos
Rincon de Romos, 5 mi W	Rincon de Romos, 5 mi W
0.5MI W RICON DE ROMOS	Rincon de Romos, 5 mi W
0.5 mi S La Labor	0.5 mi S La Labor
0.5 mi N Presa Calle	0.5 mi N Presa Calle
0.5 mi S LaLabor	0.5 mi S La Labor
0.5MI W RINCON DE ROMAS	0.5MI W RINCON DE ROMOS,6000FT

Localidad
0.25 mi E state boundary, 8 mi SW Calvillo
0.25 mi E state boundary, 8 mi SW Calvillo
0.33 mi NW Ciudad de Los Ninos, 7 mi WSW Aguas Calientes
0.33 mi NW Cuidad de Los Minos, 7 mi WSW Aguascalientes
0.33 mi S and 1.25 E Rincon de Romos
0.33 mi S and 1.25 mi E Rincon de Ramos
0.33 mi S, 1.25 mi E Rincon de Romos
Rincon de Romos, 5 mi W
0.5MI W RICON DE ROMOS
0.5 mi S La Labor
0.5 mi N Presa Calle
0.5 mi S LaLabor

# 1 Normalización de las localidades

---

- Falta eliminar los **registros redundantes** que no pudieron desecharse previamente debido a que presentaban errores de captura, para esto se realiza una segunda consulta de agrupación basada en la columna duplicada de la descripción
- Para corregir los **falsos duplicados** se procede a desagregar cada descripción de localidad en sus distintos elementos (Atomización). Esta etapa permitirá la eliminación de los falsos duplicados además de estructurar de manera ordenada la información

# 2 Asignación de coordenadas

- La descripción de las localidades se pueden tipificar en 9 clases (Wieczorek & al., 2003):

Tipo	Descripción	Ejemplos
1. Dudosa	El propio registro describe que es dudoso el nombre de la localidad o rasgo de referencia	'Tabarca?'
2. No puede ser ubicada	Los datos de la descripción se perdieron o contienen información que no corresponde con una descripción, o los nombres presentan homónimos que no pueden ser diferenciados, o no se encuentra la localidad o el rasgo en ninguna fuente disponible	'localidad no registrada', 'cumbre'
3. Comprobadamente inexacta	La descripción tiene elementos inconsistentes	'municipio de Poza Rica, Puebla'
4. Coordenadas	La descripción ya posee coordenadas	'UTM 30N 553160 4077280',
5. Localidad o rasgo	La descripción sólo posee el nombre de un elemento geográfico (una localidad, un cerro, una cueva, isla, etc.)	'unión de la carretera 70 con la 45'
6. Distancia	La descripción consiste en una distancia asociada a un elemento geográfico	'A 5 km de Sevilla'
7. Distancia a través de una ruta	Se describe una localidad con una distancia recorrida sobre una carretera, camino, río, etc.	'3.5 km N de la nacional I'
8. Distancias en direcciones ortogonales	Se utilizan dos distancias trazadas en direcciones ortogonales a partir de un elemento geográfico	'6 km N y 4 km W de Madrid'
9. Distancia en una dirección	La descripción contiene una distancia definida en una dirección	'50 km NE de Segovia'

# 2 Asignación de coordenadas: Métodos

- Utilizando Bases de datos:

- Bases toponímicas...

- ➔ Manual (sobre mapas)



- ➔ Apoyada por SIG

- Necesitamos de cartografía de base: municipios, carreteras...

# 3 Determinación de la incertidumbre

- Basado en MaNIS “The point-radius method for georeferencing locality descriptions and calculating associated uncertainty”
- Fuentes potenciales de incertidumbre

Source of uncertainty							
Coordinate source	GPS inaccuracy	locality extent	unknown datum	coordinate imprecision	distance imprecision	map scale	direction imprecision
GPS	X	X	X	X			
locality record		X	X	X			
Map		X	X	X	X	X	X
Gazetteer		X	X	X	X		X

- El método de Punto-radio define una descripción de localidad con dos elementos: el punto de coordenadas y su incertidumbre, representada con una medida de longitud
- Esta distancia define una circunferencia que delimita el área en donde, con mayor probabilidad, se ubica el sitio de colecta

# 3 Determinación de la incertidumbre

- La incertidumbre está compuesta básicamente por un factor de distancia y otro de dirección que interactúan entre sí
- En la distancia se considera:
  - a) La extensión de la localidad de referencia
  - b) El desconocimiento del datum
  - c) La imprecisión en la determinación de la distancia
  - d) La imprecisión en la medición de las coordenadas
  - e) Escala del mapa utilizado
- Dentro del factor de dirección, se considera:
  - f) la imprecisión con que se definió la dirección de referencia

# Casos particulares

---

- Todos los registros no georreferenciados conviene que sean marcados como tales y indicar las causas de la no georreferenciación

# Otros métodos

---

- Shapefile method (Murphey & al., 2004)  
(<http://www.calacademy.org/research/informatics/georef/>)

Se define un polígono que representa el error exacto del registro.

- Ventajas:
  - Se digitaliza la representación real del error
- Desventajas:
  - Incremento en el tiempo de procesamiento
  - Dificultad de comparación entre errores de localización por la complejidad de las siluetas
  - Necesidad de personal especializado

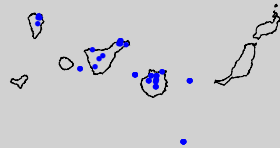
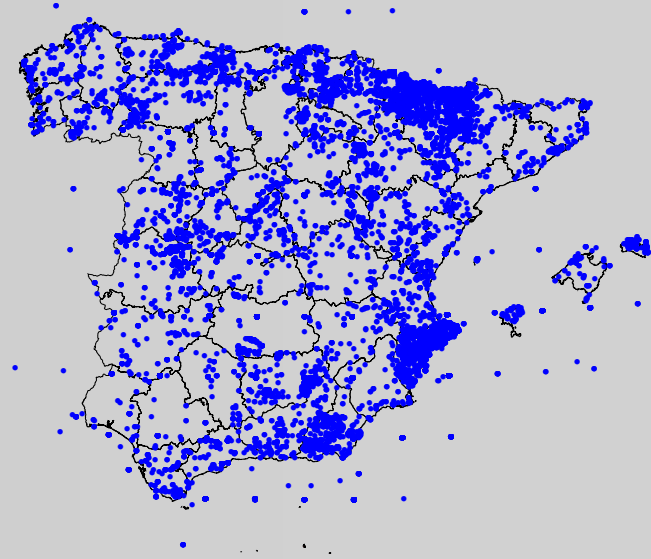


# 4 Validación de las coordenadas

- Validación con el límites administrativos (Hijmans & al., 1999)

- Ejemplo: Familia Leguminosae

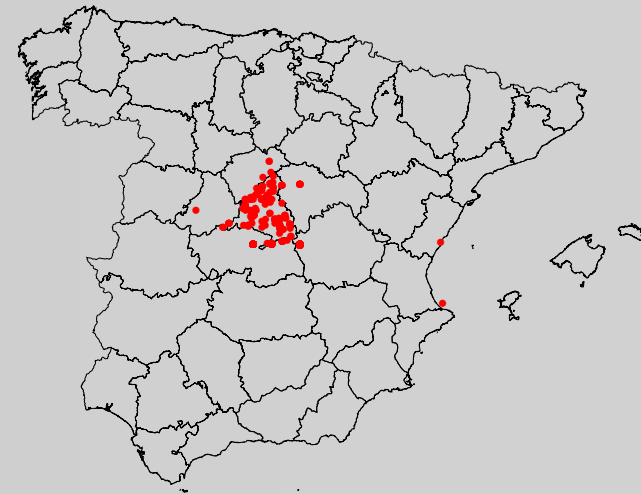
- Fuente: ABH, ARAN, BCN, BIO, COA, FCO, HGI, HUAL, IPE-CSIC, MA, MAF, SALA, SANT y VAL
- Total de registros: 43.613
- Registros georreferenciados: 21.263 (48%)
- Registros fuera del territorio nacional: 1050!!!



# 4 Validación de las coordenadas

⇒ Ejemplo: Provincia de Madrid

- Total de registros: 1353
- Registros georreferenciados: 1062 (48%)
- Registros fuera de la provincia: 147!!!



# Evaluación de métodos de georreferenciación

- Murphey & al., 2004 compararon 4 herramientas:
  - Biogeomancer (Beaman and Conn, 2003), University of Kansas  
(<http://www.biogeomancer.org> )
  - MANIS Georeferencing Calculator, University of California at Berkeley  
(<http://elib.cs.berkeley.edu/manis/gc.html> )
  - GEOLocate, Tulane University  
(<http://www.museum.tulane.edu/geolocate/demo.aspx> )
  - ArcView Georeferencing Extension, California Academy of Sciences  
(<http://www.calacademy.org/research/informatics/georef/> )

# Evaluación de métodos de georreferenciación

Feature/Capability	Biogeomancer	Georeferencing Calculator	GEOLocate	Arc/View Georeferencing Extension
Format	Online	Online	Software download	Software download
Georeferencing Tool	Y	N	Y	Y
Error Estimate	N	Y	Y	Y
Error Expression	N	Point-Radius Method: Numerical error in miles	low, medium, or high precision	polygon (visual), and "span" value
Batch Processing	Y (small batches)	N	Y	N
Single Offsets	Y	Y	Y	Y
Multiple Offsets	N	N	Y	N
Abbreviations	N	N	some	N
PLSS	N	N	Y (no _sec)	N
Place Names Database	Y	N	Y	Y (user supplied)
Street Addresses	N	N	Y	N
Highways	N	N	N	N
Elevation	N	N	N	N
Multiple Bearings (SSE, NNW, etc.)	Y	N	N	N
Text Descriptions	N	N	Y	N
Parse/Ignore Capability	N	Y	N	Y
Uses Map Interface	N	N	Y	Y
Average correct on initial georeferencing:	34%	N/A	48%	25%
Total average time per record:	6.25	5 (including manual georeferencing)	4.1	7
Time: Place Name/Coverages	0	0	0	1
Time: Database Preparation	1 (due to smaller batches)	0.5	0.5	0.5
Time: Georeferencing (coordinates)	0.25	3 (manual georeferencing)	0.1	2.5 (essentially manual georeferencing)
Time: Georeferencing (error)	2 (necessary to regraph coordinates)	0.5	2 (necessary to regraph coordinates on topo map in order to find extents)	1
Time: Quality Checking	3 (many incorrect records returned)	1	1.5 (some incorrect records returned)	2 (very difficult to check because of opaque error shapefiles)

# Localidad

---

- Específica
- Sucinta
- Inequívoca
- Completa
- Exacta

# Localidad

---

- Evitar la incertidumbre debido a la imprecisión de las definiciones mediante la descripción de un lugar en términos de una distancia a lo largo de un camino, o por dos distancias desde el lugar
- Utilice sólo un punto de referencia, debe ser de tamaño pequeño, estable en su posición, en su tamaño, en el tiempo, y fácil de encontrar en los mapas o en nomenclátorees
- Evitar el uso de términos vagos como "cerca"

# Elevación

---

- Proporcionar un valor de elevación
- Cuidado de la elevación del GPS, es menos precisa que altímetro barométrico (cuando calibrado)
- Indicar la fuente de elevación (mapa, altímetro, ...)

# Coordenadas

---

- No son suficientes para describir una localidad
- Las coordenadas originales del GPS son mejores que conversiones posteriores
- Los grados decimales pueden ser más precisos que los GG:MM:SS.



# Coordenadas

Precision	0 degrees Latitude	30 degrees Latitude	60 degrees Latitude	85 degrees Latitude
1.0 degree	156,904 m	146,962 m	124,605 m	112,109 m
0.1 degree	15,691 m	14,697 m	12,461 m	11,211 m
0.01 degree	1,570 m	1,470 m	1,246 m	1,121 m
0.001 degree	157 m	147 m	125 m	112 m
0.0001 degree	16 m	15 m	13 m	12 m
0.00001 degree	2 m	2 m	2 m	2 m
1.0 minute	2,615 m	2,450 m	2,077 m	1,869 m
0.1 minute	262 m	245 m	208 m	187 m
0.01 minute	27 m	25 m	21 m	19 m
0.001 minute	3 m	3 m	3 m	2 m
1.0 second	44 m	41 m	35 m	32 m
0.1 second	5 m	5 m	4 m	4 m
0.01 second	1 m	1 m	1 m	1 m

**Table 4.** Table showing metric uncertainty due to precision of coordinates based on the WGS84 datum at varying latitudes. Uncertainty values have been round up in all cases. From [Wieczorek \(2001\)](#).

# Datum

---

- El datum es la parte esencial de la descripción de coordenadas
- Ajuste el datum del GPS al mismo datum en el que se encuentra el mapa del trabajo de campo
- Sin datum las coordenadas son ambiguas
- Indique siempre el datum con las coordenadas




# Campos del Darwin Core relacionados con la georreferenciación

---

David Draper

Taller de GEOLOCATE, Madrid  
24-26 marzo 2009 GBIF

- 
- La mayoría de proveedores de datos españoles a GBIF utilizan el estándar Darwin Core
  - Darwin Core 2 es un simple conjunto de elementos que permite estructurar datos de registros de especímenes u observaciones para ser compartidos como un documento XML que puede ser transmitido por Internet.

# El esquema Darwin Core 2

- Apropriado para datos de colecciones y observaciones.
- <http://digir.net/schema/conceptual/darwin/2003/1.0/darwin2.xsd>
- 48 Elementos:

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

# Elementos del Darwin Core 2 y la georreferenciación

- **ContinentOcean**: The continent or ocean from which a specimen was collected.
- **Country**: The country or major political unit from which the specimen was collected. ISO 3166-1 values should be used. Full country names are currently in use. A future recommendation is to use ISO3166-1 two letter codes or the full name when searching
- **StateProvince**: The state, province or region (i.e. next political region smaller than Country) from which the specimen was collected.
- **County**: The county (or shire, or next political region smaller than State/Province) from which the specimen was collected
- **Locality**: The locality description (place name plus optionally a displacement from the place name) from which the specimen was collected.

# Elementos del Darwin Core 2 y la georreferenciación

- **Longitude:** The longitude of the location from which the specimen was collected. This value should be expressed in decimal degrees with a datum such as WGS-84
- **Latitude:** The latitude of the location from which the specimen was collected. This value should be expressed in decimal degrees with a datum such as WGS-84
- **CoordinatePrecision:** An estimate of how tightly the collecting locality was specified; expressed as a distance, in meters, that corresponds to a radius around the latitude-longitude coordinates. Use NULL where precision is unknown, cannot be estimated, or is not applicable.
- **BoundingBox:** This access point provides a mechanism for performing searches using a bounding box. A Bounding Box element is not typically present in the database, but rather is derived from the Latitude and Longitude columns by the data provider

# Referencias geográficas

# Campos Darwin Core

Límites tierra-mar:

- Continentes
- Islas, archipiélagos

Límites administrativos:

- Países
- Autonomías
- Provincias
- Municipios
- Distritos, barrios, etc.

Calles, carreteras (dir. postales)

Ríos, etc.

Montañas, etc.

Sistemas de coordenadas X,Y

- Lat./Long. (grados decimales)
- UTM (metros, ...)
- Etc.

¿Que?: Phylum, Species, Sex, ...

¿Como?: BasisOfRecord, PreparationType

¿Quién?: Collector, ...

¿Cuándo?: YearCollected, ...

¿Donde?:

ContinentOcean

Country (País)

StateProvince

County (Municipio)

Locality (¿referencia original?)

Longitude

Latitude

CoordinatePrecision, BoundingBox

MaximumElevation, MaxDepth, ...