

TALLER GBIF.ES CALIDAD EN BASES DE DATOS DE BIODIVERSIDAD

Madrid, 2-4 Diciembre 2015



Katia Cezón
Unidad de Coordinación – GBIF.ES
Nodo Nacional de Información en
Biodiversidad en España

CALIDAD DE DATOS ESPACIALES Y TEMPORALES

Madrid, 2-4 Diciembre 2015



Katia Cezón
Unidad de Coordinación – GBIF.ES
Nodo Nacional de Información en
Biodiversidad en España



Chapman, A.D. and J. Wieczorek (eds). 2006. Guide to Best Practices for Georeferencing. Copenhagen: Global Biodiversity Information Facility. Available online at http://www.gbif.org/orc/?doc_id=1288.

Muñoz López, E., Hernández Barrios, J.C., Colin López, J. 2004. Georreferenciación de localidades de colectas biológicas de la Conabio. *Biodiversitas* 54:8-15.

Wieczorek, J., Guo, Q. & Hijmans, R.J. 2004. The point-radius method for georeferencing locality descriptions and calculating associated uncertainty. *International journal of geographical information science*.

Otegui, J., Ariño, A., & al, 2013. On the dates on the GIBIF mobilised primary biodiversity data records. *Biodiversity Informatics* 8: 173-184.

DATOS ESPACIALES

GEORREFERENCIACIÓN RETROSPECTIVA



La **georreferenciación retrospectiva** de localidades consiste en la asignación de coordenadas geográficas a partir de la descripción textual del lugar de recolección.

La descripción de una localidad debe contener el *mayor número de elementos geográficos* que nos permita localizar con la mayor precisión posible el sitio de recolección.



Muchas colecciones de museos y herbarios tienen una *información muy básica* de la localidad, y supone un gran esfuerzo convertir estas descripciones a coordenadas geográficas.

En el caso de las colecciones históricas:

- los colectores en general carecían de mapas detallados de la zona que muchos nombres de localidades ya no están en uso.
- no aparecen en los mapas actuales ni en los gaceteros publicados.

ELEMENTOS DE UNA LOCALIDAD



La descripción de una localidad debe contener el *mayor número de elementos geográficos* que nos permita localizar con la mayor precisión posible el sitio de recolección.

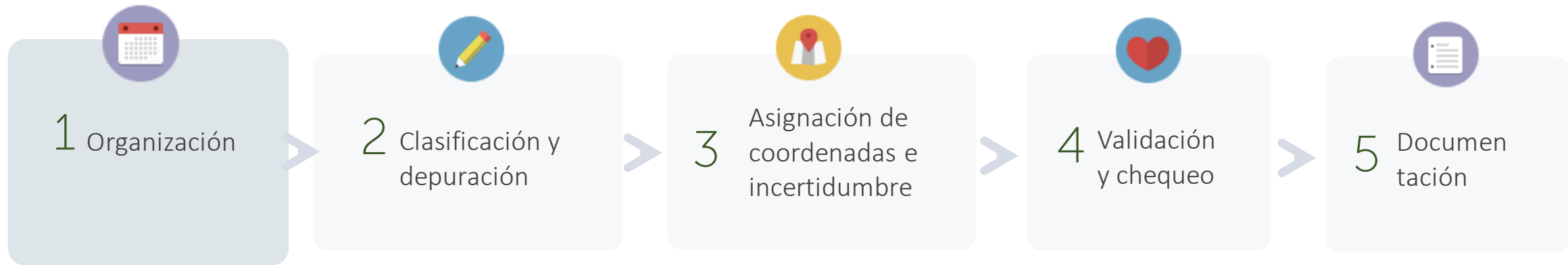
ESPAÑA. Segovia: Basardilla, 2,5 Km. al SO por la Ctra. Sg-160, muy próximo al poblado de Don Toribio, en el cruce del puente sobre el río Pirón, 1470 m, 41°2'1.78"N 4° 1'0.56"W, WGS84.

Un posible estándar podría ser:

- | | | |
|---|---|---|
| 1 | País | |
| 2 | Provincia | |
| 3 | Municipio | |
| 4 | Localidad principal | |
| | <ul style="list-style-type: none">• Distancia• Dirección• Vía de acceso | |
| 5 | Altitud-prof (punto o rango) | <ul style="list-style-type: none">• Localidad de referencia• Referencia complementaria |
| 6 | Coordenadas | |
| 7 | Datum | |
| 8 | Incertidumbre (RadioPunto) | |

GEORREFERENCIACIÓN RETROSPECTIVA

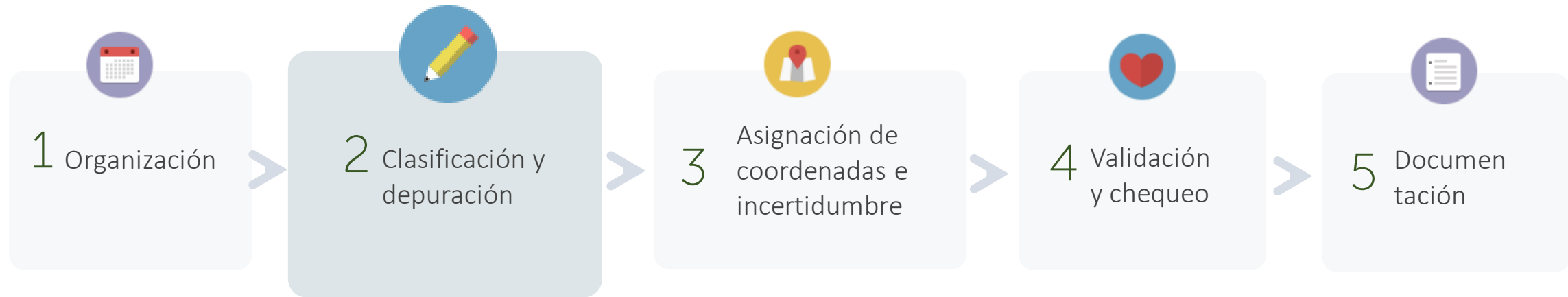
PROCEDIMIENTOS



- Definir los **estándares** y métodos a utilizar.
- Separar la informatización de la colección de la tarea de georreferenciación.

GEORREFERENCIACIÓN RETROSPECTIVA

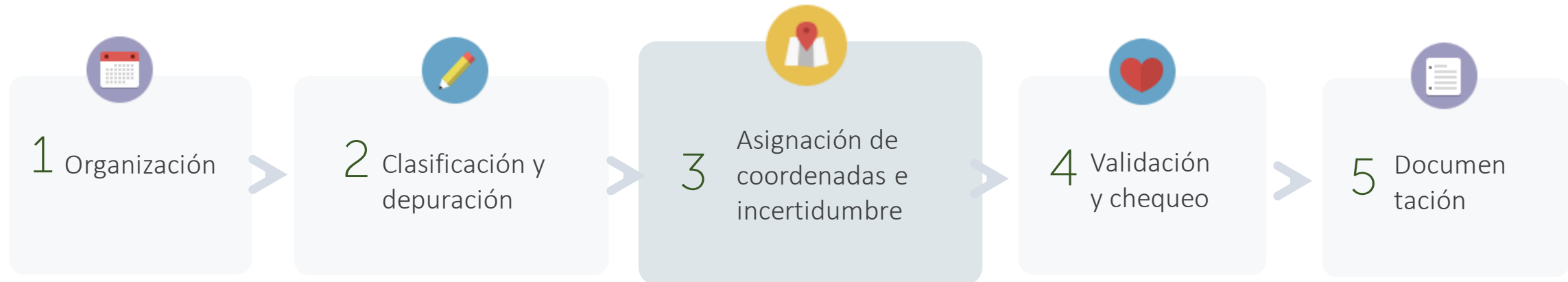
PROCEDIMIENTOS



- Ordenar y estandarizar las localidades.
- Ver si hay registros de la **misma localidad** ya georreferenciados y validados.
- **Excluir** las localidades sin elementos suficientes para ser georreferenciados.

GEORREFERENCIACIÓN RETROSPECTIVA

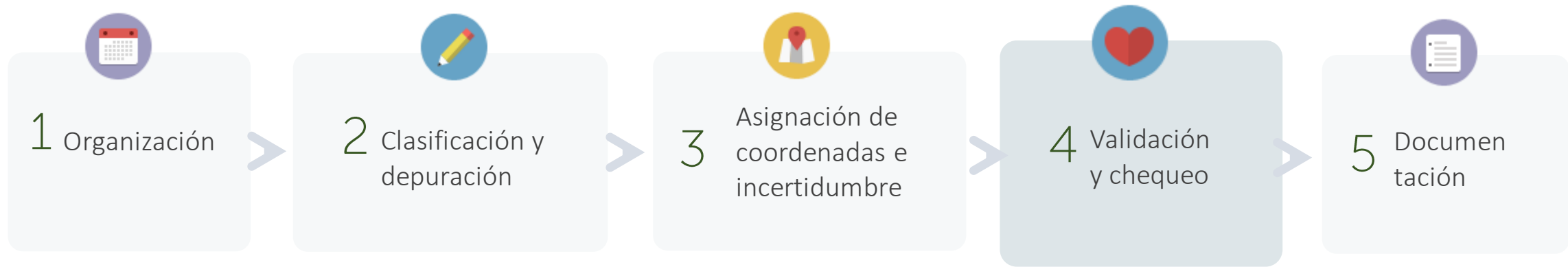
PROCEDIMIENTOS



- Con **cartografía** impresa o cartografía digital y uso de SIG.
- Calcular las **incertidumbres**.

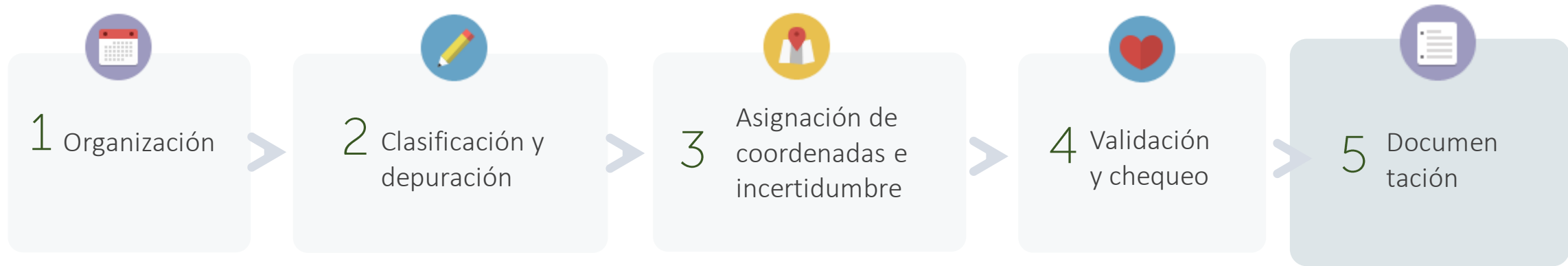
GEORREFERENCIACIÓN RETROSPECTIVA

PROCEDIMIENTOS



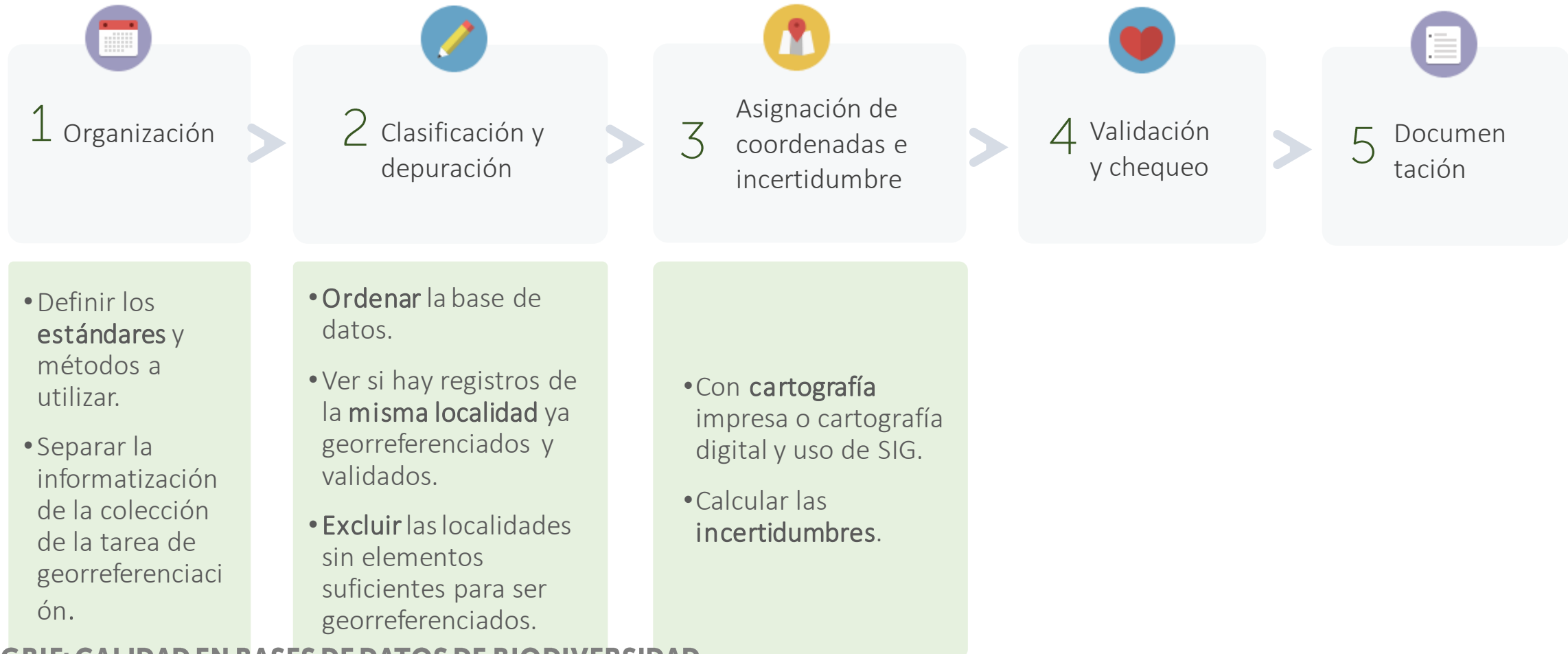
GEORREFERENCIACIÓN RETROSPECTIVA

PROCEDIMIENTOS



GEORREFERENCIACIÓN RETROSPECTIVA

PROCEDIMIENTOS



GEORREFERENCIACIÓN RETROSPECTIVA

CLASIFICACIÓN DE LOCALIDADES

Tipos de **localidades que no son aptas** para ser georreferenciadas:

- 1 Dudosa > ¿Formentera?
- 2 No se puede localizar > En el valle, junto río.
- 3 Demostrablemente inexacta > Segovia. Río Tajo.
- 4 Sólo existen coordenadas > 42.4532 84.8429

GEORREFERENCIACIÓN RETROSPECTIVA

ASIGNACIÓN DE COORDENADAS E INCERTIDUMBRES

MÉTODOS DE
GEORREFERENCIACIÓN

- 1 **Punto** x, y . Es un elemento adimensional.
- 2 El **polígono**. Preciso, pero de muy difícil manejo.
- 3 El **Bounding Box** o rectángulo de coordenadas. Más impreciso que el polígono y requiere calcular la incertidumbre de las coordenadas en dos dimensiones.
- 4 El **Radiopunto**. Es muy eficaz, porque las incertidumbres pueden combinarse fácilmente en un sólo atributo.

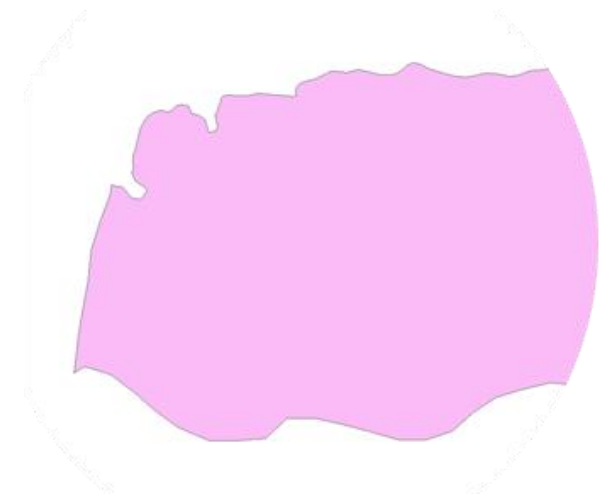


GEORREFERENCIACIÓN RETROSPECTIVA

ASIGNACIÓN DE COORDENADAS E INCERTIDUMBRES

MÉTODOS DE
GEORREFERENCIACIÓN

- 1 **Punto** x, y . Es un elemento adimensional.
- 2 El **polígono**. Preciso, pero de muy difícil manejo.
- 3 El **Bounding Box** o rectángulo de coordenadas. Más impreciso que el polígono y requiere calcular la incertidumbre de las coordenadas en dos dimensiones.
- 4 El **Radiopunto**. Es muy eficaz, porque las incertidumbres pueden combinarse fácilmente en un sólo atributo.

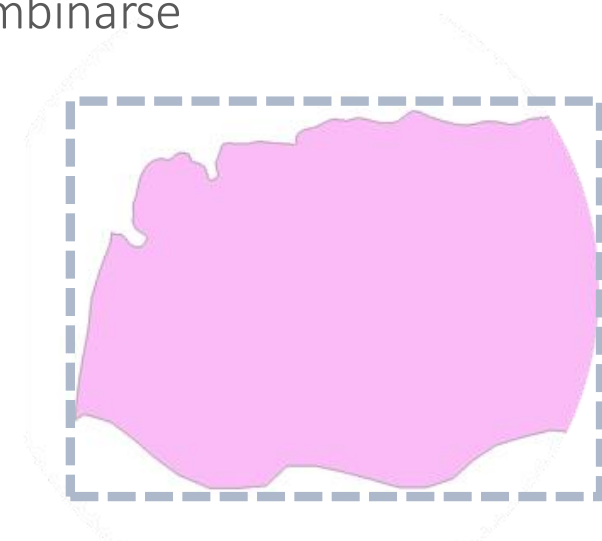


GEORREFERENCIACIÓN RETROSPECTIVA

ASIGNACIÓN DE COORDENADAS E INCERTIDUMBRES

MÉTODOS DE
GEORREFERENCIACIÓN

- 1 **Punto** x, y . Es un elemento adimensional.
- 2 El **polígono**. Preciso, pero de muy difícil manejo.
- 3 El **Bounding Box** o rectángulo de coordenadas. Más impreciso que el polígono y requiere calcular la incertidumbre de las coordenadas en dos dimensiones.
- 4 El **Radiopunto**. Es muy eficaz, porque las incertidumbres pueden combinarse fácilmente en un sólo atributo.

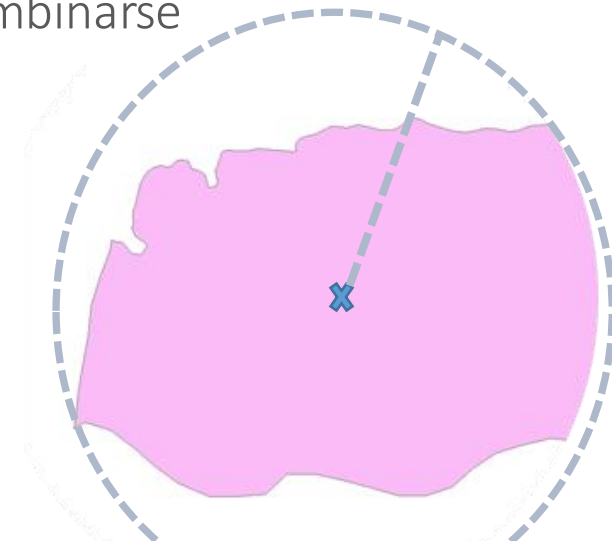


GEORREFERENCIACIÓN RETROSPECTIVA

ASIGNACIÓN DE COORDENADAS E INCERTIDUMBRES

MÉTODOS DE
GEORREFERENCIACIÓN

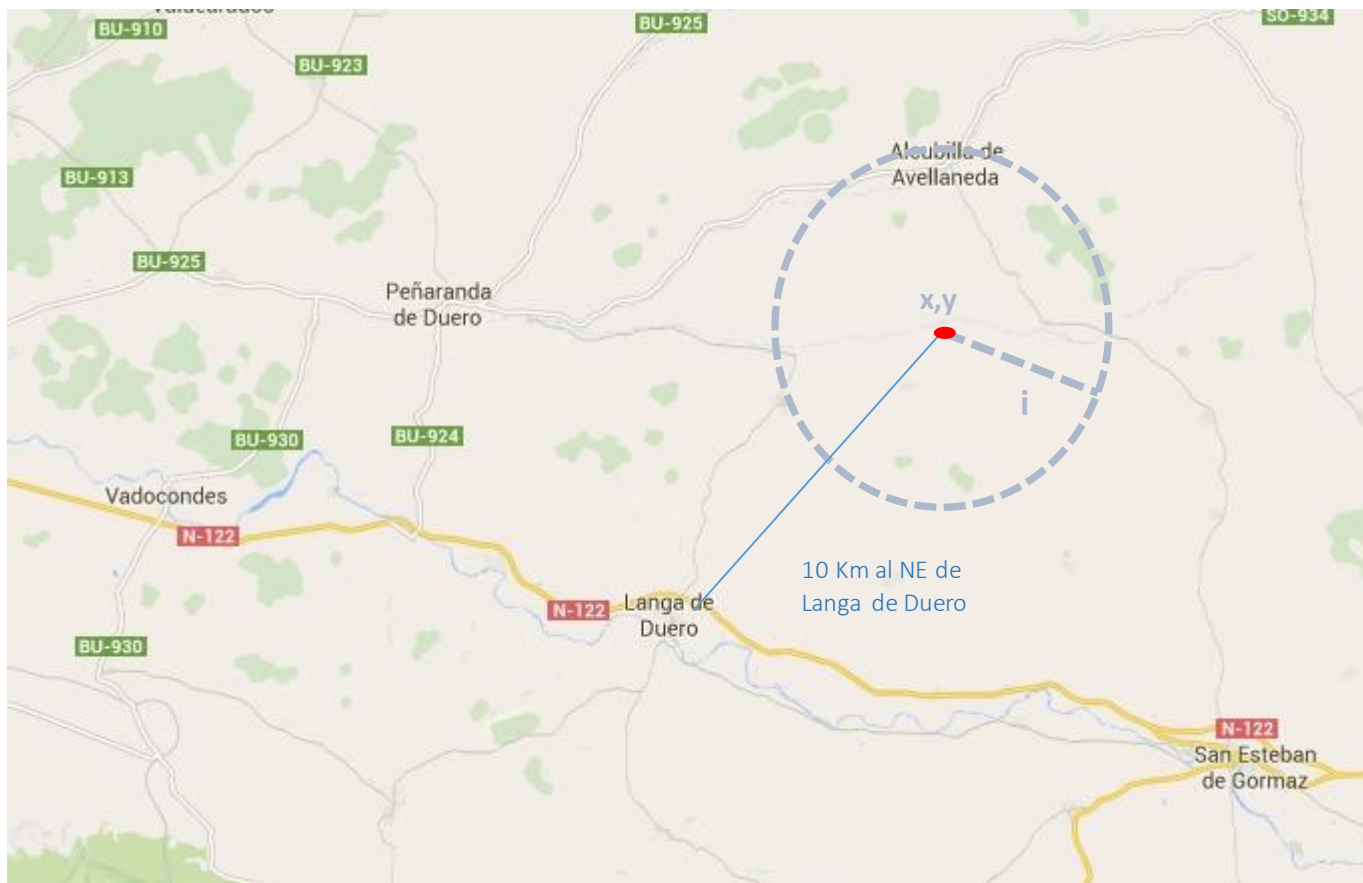
- 1 **Punto** x, y . Es un elemento adimensional.
- 2 El **polígono**. Preciso, pero de muy difícil manejo.
- 3 El **Bounding Box** o rectángulo de coordenadas. Más impreciso que el polígono y requiere calcular la incertidumbre de las coordenadas en dos dimensiones.
- 4 El **Radiopunto**. Es muy eficaz, porque las incertidumbres pueden combinarse fácilmente en un sólo atributo.



GEORREFERENCIACIÓN RETROSPECTIVA

ASIGNACIÓN DE COORDENADAS E INCERTIFUMBRES

MÉTODOS DE
GEORREFERENCIACIÓN



A partir de la descripción de la localidad obtenemos un par de coordenadas (x,y) asociadas a una medida de longitud que será su incertifumbre (i) o Radiopunto:

“Distancia que define el radio del área más probable en donde se encontraría el sitio de recolección.”

GEORREFERENCIACIÓN RETROSPECTIVA

ASIGNACIÓN DE COORDENADAS E INCERTIDUMBRES

MÉTODO DEL
RADIOPUNTO

Como calcular la incertidumbre

La magnitud de la **incertidumbre** depende de la **precisión** con la que se ha descrito el sitio. Depende de estas variables:

- 1 Extensión de la localidad
- 2 Datum desconocido
- 3 Imprecisión en la medición de la distancia
- 4 Imprecisión en la medición de la dirección
- 5 Imprecisión en la toma de coordenadas
- 6 Escala del mapa

Todos estos elementos se cuantifican de **manera individual** para después **sumar** sus valores y obtener un **único valor de incertidumbre máxima**.

GEORREFERENCIACIÓN RETROSPECTIVA

ASIGNACIÓN DE COORDENADAS E INCERTIDUMBRES

MÉTODO DEL
RADIOPUNTO

Incertidumbre debida a la extensión de la localidad:

- Cualquier localidad que se tome como referencia (población, río, valle) tiene una extensión determinada.
- La incertidumbre debida a la extensión es la distancia máxima entre dos puntos dentro de la población.
- En colecciones históricas se debe tener en cuenta que la extensión de la población ha cambiado con el tiempo.



WILLER GBIF: CALIDAD EN BASES DE DATOS DE BIODIVERSIDAD

GEORREFERENCIACIÓN RETROSPECTIVA

ASIGNACIÓN DE COORDENADAS E INCERTIDUMBRES

MÉTODO DEL
RADIOPUNTO

Incertidumbre debida a Datum desconocido:

- Se denomina Datum a un conjunto de parámetros cuyos valores, una vez definidos, permiten la referenciación precisa de localizaciones sobre la superficie terrestre.
- Aplicable a la cartografía o a las coordenadas capturadas con GPS.
- La incertidumbre por no conocer el datum puede variar de 0 a 500 m. En zonas concretas, 3.500 m.
- Si se conoce el datum, la incertidumbre es 0.

GEORREFERENCIACIÓN RETROSPECTIVA

ASIGNACIÓN DE COORDENADAS E INCERTIDUMBRES

MÉTODO DEL
RADIOPUNTO

Incertidumbre debida a la distancia:

Está definida por el grado de precisión con que se registra una distancia, con o sin decimales (ej. 5.33 Km. NE de Conil, Cádiz)

La incertidumbre se calcula como 1 dividido por el denominador de la precisión fraccional.



- 9 Km = Precisión fraccional 1 / 1 → 1 Km
- 9.5 Km = Precisión fraccional 1 / 2 → 0.5 Km
- 9.75 Km = Precisión fraccional 3 / 4 → 0.25 Km
- 9.6Km = Precisión fraccional 6 / 10 → 0.1 Km

Para medidas de números múltiplos de 10, se multiplica 0.5 por la máxima potencia de 10 que tiene dicho número:

- 140 Km = 0.5 x 10 = 5 Km
- 200 Km = 0.5 x 100 = 50 Km
- 2000 Km = 0.5 x 1000 = 500 Km

GEORREFERENCIACIÓN RETROSPECTIVA

ASIGNACIÓN DE COORDENADAS E INCERTIDUMBRES

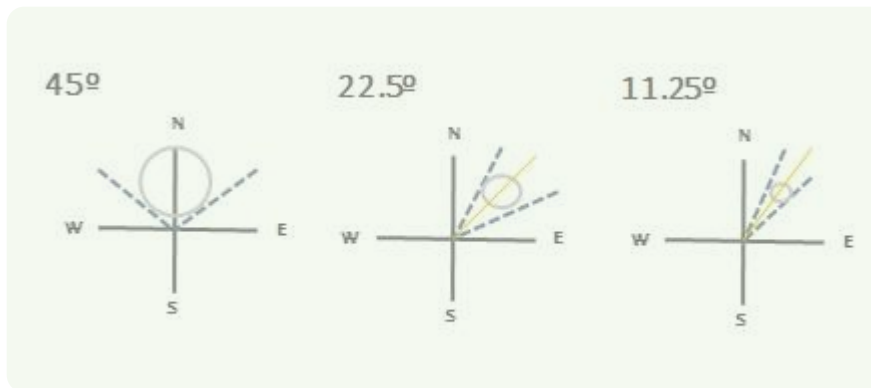
MÉTODO DEL RADIOPUNTO

Incertidumbre debida a la dirección:

- Distancia de desvío
15 Km de Camargo (por aire), Recife
- Dirección de desvío
N de Matalascañas, Huelva
- Desvío desde un origen
25 Km W of Albuquerque



Es importante registrar si la distancia se toma “por carretera” o “por aire” (en línea recta y sin tener en cuenta el relieve).



GEORREFERENCIACIÓN RETROSPECTIVA

ASIGNACIÓN DE COORDENADAS E INCERTIDUMBRES

MÉTODO DEL
RADIOPUNTO

Incertidumbre debida a la imprecisión de las coordenadas:

- Definida por el nivel de precisión con el que fueron registradas un par de coordenadas: Incluir siempre el mayor número de decimales posible

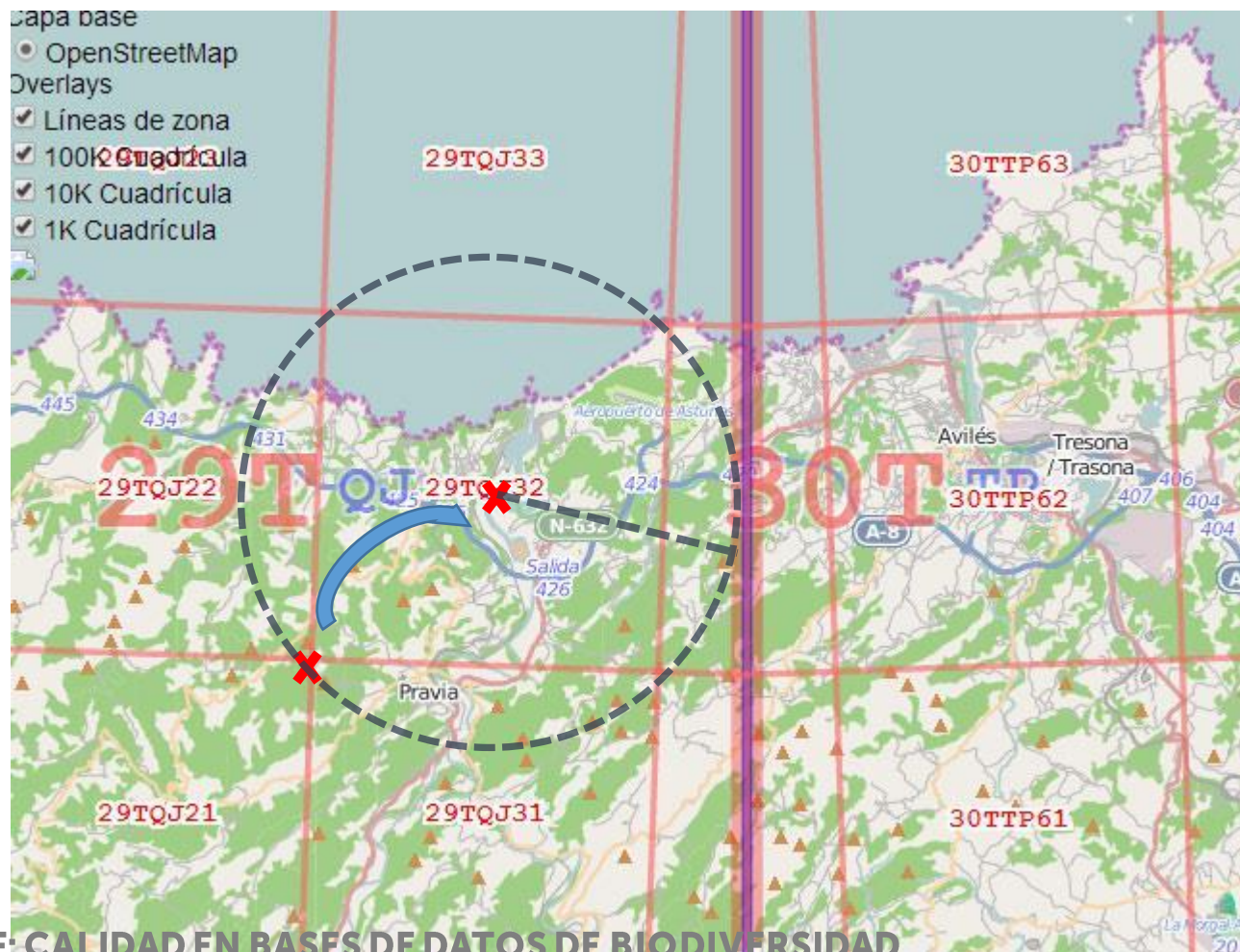
42.51830 N, 10.45673 E Incertidumbre=2m

- Chapman y Wieczorek recomiendan el sistema de coordenadas decimales y registrar hasta 5 decimales.
- Si usamos UTM, siempre debemos registrar la Zona.

GEORREFERENCIACIÓN RETROSPECTIVA

ASIGNACIÓN DE COORDENADAS E INCERTIDUMBRES

MÉTODO DEL
RADIOPUNTO



Para calcular el radio de incertidumbre, recentramos 29TQ32 a coordenada UTM:

Teorema de Pitágoras
Radiopunto= (Lado del cuadrado de la UTM (en metros) x $\sqrt{2}$) / 2

GEORREFERENCIACIÓN RETROSPECTIVA

ASIGNACIÓN DE COORDENADAS E INCERTIDUMBRES

MÉTODO DEL
RADIOPUNTO

Incertidumbre debida a la escala del mapa:

- El nivel de precisión utilización de un mapa está definido por su escala.

En el ejemplo, precisión asociada a una línea de 0.5 mm

Scale of Map	Map Horizontal Uncertainty (Geosciences Australia ¹¹)
1:1000	0.5 m
1:10,000	5 m
1:25,000	12.5 m
1:50,000	25 m
1:75,000	
1:100,000	50 m
1:250,000	160-300 m
1:500,000	
1:1 million	500 m

GEORREFERENCIACIÓN RETROSPECTIVA

VALIDACIÓN DE LAS COORDENADAS

Testeo de las coordenadas ya asignadas:

- Chequear la localidad contra localidades ya existentes y validadas en la base de datos.
- Chequear contra una referencia externa: ¿es la localidad coherente con las localidades de recolección del colector?
- Chequear contra una referencia externa usando un SIG ¿se sitúa la localidad en tierra o en el mar?
- Chequear los puntos que quedan fuera de los límites del espacio geográfico.
- Chequear los puntos que quedan fuera de los límites en el espacio medioambiental.

GEORREFERENCIACIÓN RETROSPECTIVA

DOCUMENTACIÓN

Otros datos que conviene añadir son:

- Estado de verificación de las coordenadas.
- Fuente de referencia utilizada en la georreferenciación.
- Responsable de la georreferenciación.
- Fecha en la que ha sido georreferenciada.
- Notas.



Es importante **no perder la información original**, si no añadir un nuevo registro con las coordenadas y los datos calculados. Esto nos **permite enriquecer la base de datos** y tener elementos de chequeo con la información original.

DATOS TEMPORALES

DATOS TEMPORALES



- *Fecha de captura, recolección u observación* del espécimen
- *Hora del día* puede ser útil para futuros estudios (fauna).
- *Fecha de identificación* o revisión del ejemplar.

- Ordenar los registros por colector y fecha permitirá detectar las localidades improbables para ese colector en ese día.
- Con localidades históricas, la fecha permitirá investigar cual es el nombre actual y podremos acceder a mapas datados en aquella poca para georreferenciarlas.

DATOS TEMPORALES

- El año de recolección en las colecciones históricas permite conocer los cambios que han podido sufrir países (fronteras), ciudades (extensión), carreteras (cambios de trazado), cauces de río, etc. a lo largo del tiempo, importante a la hora de calcular incertidumbre de la localidad.
- Los historiadores han llevado a cabo recientes desarrollos de *itinerarios* de colectores según la información de la localidad y la fecha. Son muy útiles a la hora de detectar errores en otros registros de ese colector.

DATOS TEMPORALES

PROBLEMAS FRECUENTES

- Fechas ausentes o incompletas
- No se ajustan a un estándar
- No son uniformes, están representadas de múltiples formas en los campos



15 - Enero- 02

15/01/2002

15-I-2002

2002/01/15

15-Jan-2002

Etc.

DATOS TEMPORALES

PROBLEMAS FRECUENTES

Biodiversity Informatics, 8, 2013, pp. 173-184

ON THE DATES OF THE GBIF MOBILISED PRIMARY BIODIVERSITY DATA RECORDS

JAVIER OTEGUI (1), ARTURO H. ARIÑO (1)*, VISHWAS CHAVAN (2) AND SAMY GAJI (2)

(1) *Department of Zoology and Ecology, University of Navarra, Pamplona, Spain.*

(2) *Global Biodiversity Information Facility Secretariat, Universitetsparken 15, 2100, Copenhagen, Denmark.*

**corresponding author*

Abstract — There are more than 390 million primary biodiversity data records published by hundreds of data publishers through the GBIF network. Thus, the GBIF network is the single most comprehensive index for this kind of data. Ensuring or, at least assessing data quality is of capital importance for the reliability and usability of this data. While conducting a time data gap analysis on this mass of data, we have detected some issues with the way date information is processed and shared. Dates can be obscured or altered under certain circumstances, when a specific combination of publisher's error or date handling features, and faulty or inadequate date parsing and processing routines gets chained together. The extent of the date unreliability (either at the source or through GBIF portal) is relatively low, and problems are concentrated in a few data publishers. The types of errors and misprocessing in dates through the sources and the published records are analysed, impact on the overall data quality of the published index was assessed, and corrective measures are suggested.

Keywords — Primary biodiversity data, GBIF, dates, data quality, fitness for use.

Gracias