

Herramientas útiles para la calidad y limpieza de datos

Katia Cezón

katia.gbif@gmail.com

Gbif.Es

Taller online GBIF.ES: Estandarización y publicación de datos de biodiversidad de regiones polares. Mayo 2021.



CSIC

Gbif.Es



Introducción

La calidad de los datos debe ser abordada durante las diferentes etapas de manipulación de los datos, que van desde la captura y registro de datos en el momento de recolectarlos en el campo, hasta su publicación y diseminación a través de distintos medios para su uso.

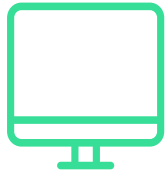


Introducción

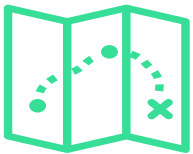
En esta sesión veremos algunas consideraciones previas y herramientas útiles para mejorar la calidad de nuestros datos, **focalizándonos en aquellas que pueden ser especialmente útiles para la normalización y la estandarización de los datos de cara a su publicación en los portales de GBIF.** Estas se pueden agrupar en los siguientes grupos.



Introducción



Herramientas genéricas para el almacenamiento y gestión de datos



Herramientas para el tratamiento de nombres científicos, fechas y coordenadas



Herramientas para la limpieza y validación de datos

- Excel, Access, Open Office, etc.
- Herramientas para la gestión de nombres científicos (atomización, herramientas para comprobar status, búsqueda de autores, etc.)
- Herramientas geográficas (visualización, comprobación de coordenadas, conversión, etc.)
- Herramientas para el tratamiento de las fechas
- Open Refine
- Darwin Test
- Darwin Core Archive Validator

Aspectos a tener en cuenta a la hora de elegir una herramienta

Ninguna herramienta es perfecta y hay que tener en cuenta varios aspectos a la hora de elegir una u otra para realizar determinadas tareas.

Es difícil realizar todas las tareas utilizando una única herramienta, y al mismo tiempo, **varias herramientas pueden solaparse en sus funciones**.



No existe una receta

Aspectos a tener en cuenta a la hora de elegir una herramienta

Es fundamental **identificar a alguien en el equipo** que se sienta cómodo con la manipulación de datos, ya que tendrás que hacer correcciones, migrar desde / hasta diferentes formatos, etc. Cada situación es diferente y necesitará hacerse con un conjunto de herramientas que se aplique a su contexto.

Elección de responsable



Aspectos a tener en cuenta a la hora de elegir una herramienta

- Precio, disponibilidad, licencia
- Requerimientos técnicos
- Facilidad de uso
- Documentación y soporte
- Flexibilidad
- Automatización
- Etc.



The image features three crossed wrenches on a plain white background. The wrenches are arranged in a 'V' shape, with one at the top and two at the bottom. They are made of a dark, possibly black or dark grey, metal. The text is centered over the wrenches in a white, bold, sans-serif font.

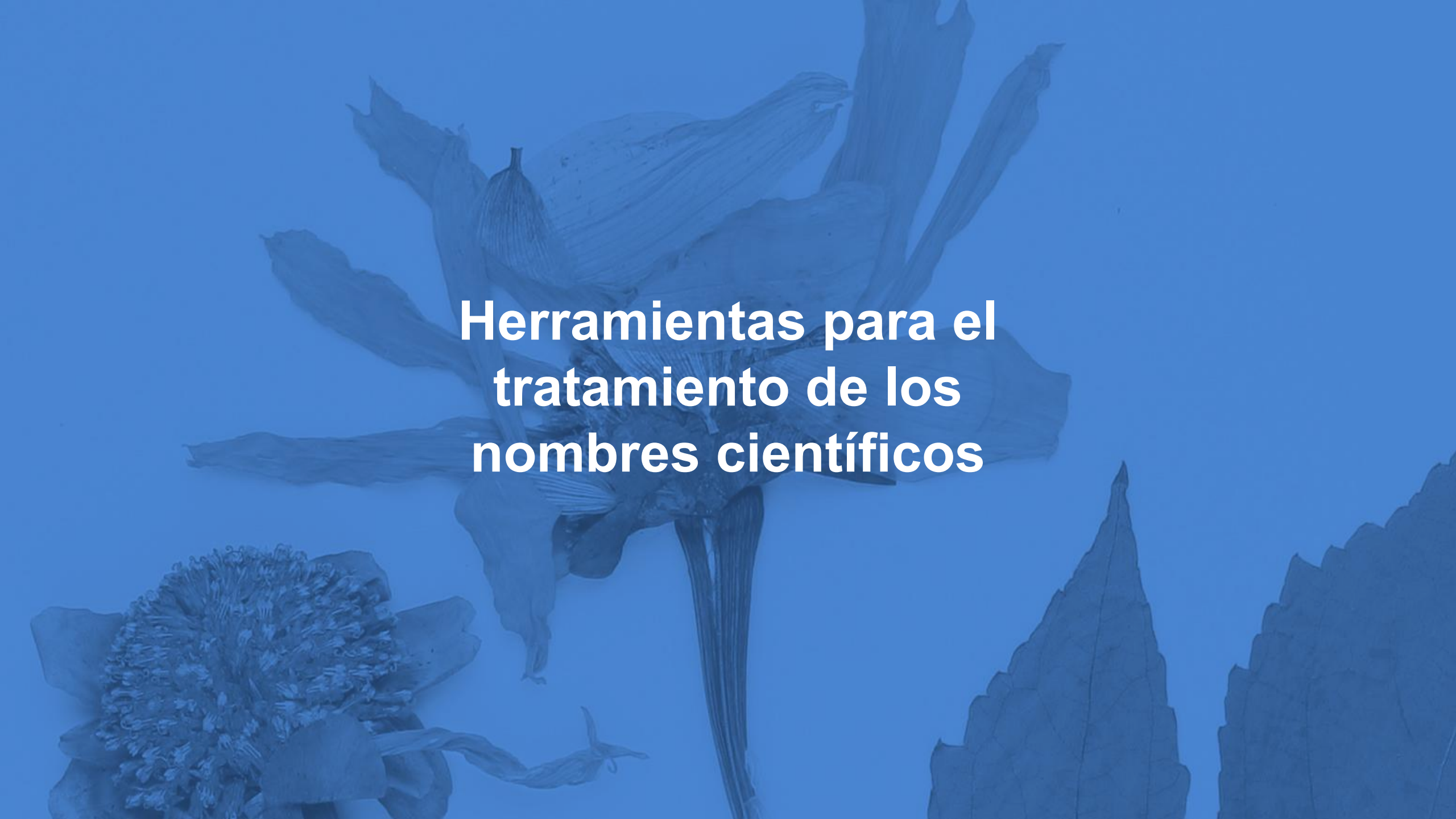
Herramientas genéricas para el almacenamiento y gestión de datos

Herramientas de almacenamiento y gestión de datos

Muchas de las **herramientas de software sencillas** que utilizamos para el **almacenamiento y la gestión** de datos en el día a día pueden utilizarse para depurar, normalizar los datos y modificar su formato.

Permiten manejar los datos mediante tablas (formadas por filas o registros y columnas o variables), crear relaciones entre tablas, elaborar consultas y formularios para introducir datos o informes para extraer la información.



A botanical illustration of a flower and its parts, including a large flower with a prominent central cluster of stamens, a smaller flower, and several leaves, all rendered in a light blue color against a darker blue background. The text is overlaid on the central part of the illustration.

Herramientas para el tratamiento de los nombres científicos

Herramientas para el tratamiento de los nombres científicos

Permiten realizar tareas como la **atomización** de sus componentes (género, epíteto específico, etc.), permiten la **resolución de nombres científicos** de cualquier grupo taxonómico, generan la jerarquía taxonómica, etc.

Atomización

[GBIF - Name parser](#)

[Name Parser GBIF](#)

[España](#)

Resolución

[Species matching GBIF](#)

[List Matching Service](#)

[Global Names Resolver](#)

[T-REX](#)

[TNRS \(plantas\)](#)



Herramientas que se verán en profundidad en apartados siguientes

A person in a white coat stands on a map of Europe, pointing upwards. The map is overlaid with a green-to-blue gradient. The text 'Herramientas para el tratamiento de los datos geográficos' is centered over the map.

Herramientas para el tratamiento de los datos geográficos

Herramientas para el tratamiento de los datos geográficos

La información geográfica es fundamental para los datos de biodiversidad. Es importante contar con herramientas que permitan **analizar, visualizar y transformar datos** en base a criterios geográficos. Los Sistemas de Información Geográfica (ej., software GIS, QGIS, DivaGIS) pueden ser muy útiles, pero no son fáciles de usar para todos los públicos. Alternativas más simples como las que se enumeran a continuación pueden ayudar a realizar las operaciones más frecuentes:

Comprobación

[Info XY \(species link tools\)](#)

[Excel to kml \(Earth point\)](#)

[Google Earth, Google Maps,](#)

[Carto](#)

Sistemas de información geográfica

Conversión

[Canadensys coordinates conversion](#)

Conversor de coordenadas GBIF.ES

[Geotrans](#)

 Herramientas que se verán en profundidad en apartados siguientes

Herramientas para el tratamiento de las fechas



Herramientas para el tratamiento de las fechas

Herramientas que permiten la **unificación** y **transformación** de distintos formatos de fechas.


Atomización y unificación de fechas

Canadensys date parsing

Función “fechasa” disponible en diversas aplicaciones de GBIF.es

13-VI-1980	→	1980-06-13
13 Junio 1980	→	1980-06-13
13-06-1980	→	1980-06-13

Herramientas que se verán en profundidad en apartados siguientes



**Herramientas de
depuración y
validación para la
publicación**

Herramientas de validación y depuración para la publicación

Herramientas específicas que comprueban la **validez de los datos para su publicación** una vez que estos ya están adaptados al estándar.



Darwin Test



DATA VALIDATOR



Herramientas que se verán en profundidad en apartados siguientes

Taller online GBIF.ES: Estandarización y publicación de datos de biodiversidad de regiones polares

¡Gracias!

katia@gbif.es

