

## 1.- Introducción

### Introducción al aprendizaje en línea



### Introducción al aprendizaje en línea

La educación en línea es un entorno de aprendizaje que se caracteriza por novedosas formas de comunicación y uso de recursos tecnológicos en donde el estudiante aprende a través de vídeos, podcast, wikis, materiales multimedia, foros de discusión, etc. Estas herramientas representan un ambiente educativo interactivo en el que participan docentes, especialistas y por supuesto, los alumnos.

El aprendizaje en línea tiene muchos beneficios (por ejemplo, la posibilidad de volver a visitar los materiales en cualquier momento, más flexibilidad, facilidad de acceso, etc.), pero también plantea desafíos. Un reto importante para el estudiante online son las estrategias para hacer eficiente el estudio, para lo cual tendrás que hacer uso de tus mejores habilidades de planificación, tener claridad en los objetivos y motivaciones, y elaborar estrategias para hacer frente cuando surjan problemas. Afortunadamente, hay mucha ayuda para que puedas completar las actividades en línea de este taller ¡Además cuentas con instructores y compañeros para ayudarte!

### Recomendaciones para el aprendizaje en línea

Estas son algunas recomendaciones a considerar para completar con éxito las actividades en un entorno de aprendizaje en línea.

- **Motivación:** Es muy importante que tengas clara la motivación que te llevó a matricularte en este taller. Normalmente ésta suele ser una combinación de motivación intrínseca (ej., deseo de desarrollar habilidades, de crecer intelectualmente, interés personal en el tema) y motivación extrínseca (ej. obligaciones laborales, desarrollo de carrera profesional). Si bien ambos tipos de motivación son positivos, la motivación intrínseca generalmente le mantendrá interesado por más tiempo y dará como resultado una experiencia de aprendizaje más profunda y duradera.
- **Planificación:** Aprender en línea requiere tanto tiempo (o más) que el aprendizaje presencial. Asegúrate de que estudias la agenda del taller y revisa tu calendario. Reserva el tiempo necesario (¡hazlo ahora!) para completar las actividades. Reserva un poco más de tiempo en caso de que tengas que enfrentarte a un trabajo inesperado o a compromisos personales.
- **Crear un buen ambiente de aprendizaje:** asegúrate de que tienes el espacio físico y temporal para trabajar con los materiales de aprendizaje sin interrupción. Si bien algunos de los materiales ofrecidos en el curso pueden ser consultados en cualquier lugar (ej., a través de un dispositivo móvil), reserva tiempo suficiente para reflexionar, volver a revisar los materiales, escribir notas personales, etc., o lo que funcione mejor para ti dependiendo de tu

estilo de aprendizaje personal.

- **Trabajar por adelantado:** Trata de completar las tareas mucho antes de la fecha límite y evita dejar el trabajo para el último minuto. Pueden surgir dificultades inesperadas (técnicas o conceptuales) que requieran dedicar un tiempo extra e investigación para ser resueltas.
- **Aprender socialmente:** Aprender en línea no significa aprender solo. Asegúrate de conectarte con tus compañeros y formadores. Utiliza los mecanismos de comunicación en línea que dejamos a tu disposición ; ). Esto te proporcionará una experiencia de aprendizaje mucho más rica y satisfactoria, te ayudará a mantenerte en el buen camino y servirá de apoyo cuando tengas problemas.
- **Conecta tu vida y tu trabajo:** Al repasar los materiales en línea, detente para reflexionar la manera en que el tema estudiado se relaciona con tu entorno de trabajo. ¿Cómo podrías hacer uso de los nuevos conocimientos presentados? ¿Se puede aplicar directamente? Si no lo hace, ¿hay buenas ideas sobre las que podría basarse para que resulten útiles?

Si tienes dificultades y / o problemas para mantenerte al día con el calendario general del taller, no dudes en ponerte en contacto con el instructor y pedir consejo. ¡Estamos aquí para ayudarte!

## Para saber más

Otros enfoques sobre el mismo tema (en inglés):

- <http://onlinelearningtips.com/2013/06/04/self-motivation-in-an-online-environment/>
- <http://www.emergingedtech.com/2012/01/5-tips-for-staying-motivated-while-studying-online/>

## Introducción a GBIF

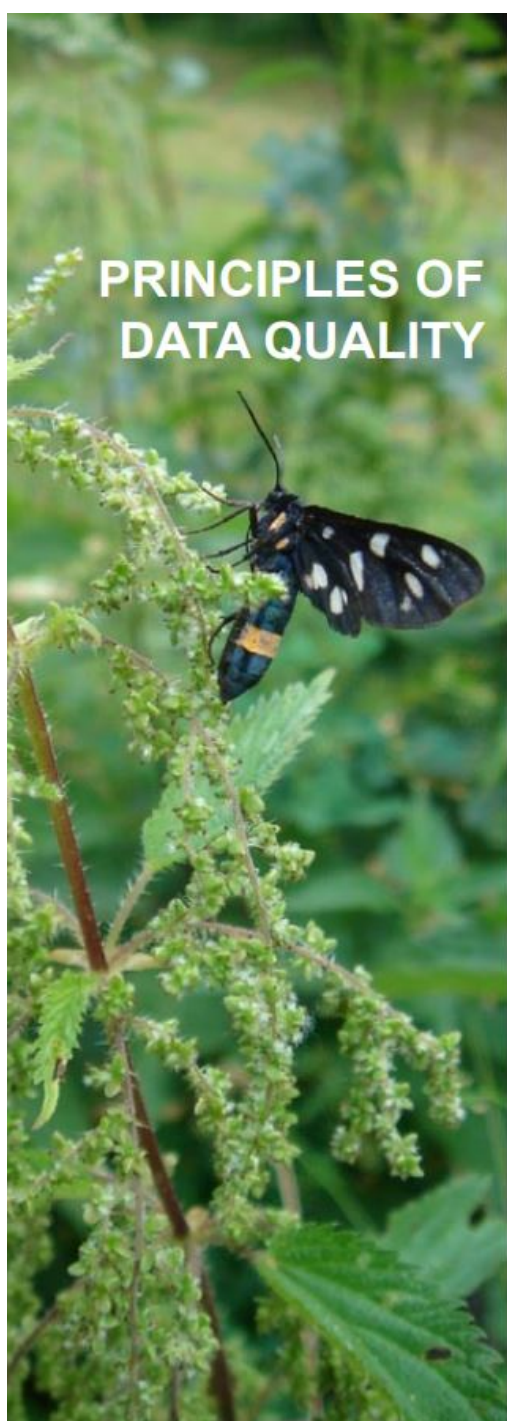
### Introducción a los contenidos del curso

## Introducción

Los datos de biodiversidad almacenados en herbarios, colecciones de historia natural, y procedentes de inventarios, son un recurso muy valioso. No sólo proporcionan información sobre la presencia de una especie en un momento y lugar determinado, si no que además son la base de numerosos proyectos de investigación y juegan un papel muy importante en la toma de decisiones relativas a la conservación y el uso sostenible de la biodiversidad. Por esta razón, la implementación de procesos que garanticen su calidad debe formar parte de cualquier sistema y/o base de datos de biodiversidad.

Este taller tiene como objetivo sentar las bases teóricas y presentar diferentes herramienta para el control de calidad de los datos que afectan a las distintas etapas de la gestión de un dato: desde su captura hasta su publicación y uso en el contexto de GBIF, pasando por su sistematización, identificación taxonómica, documentación, almacenamiento y preservación de la información.

Se muestran los conceptos más importantes relacionados con la calidad de bases de datos sobre biodiversidad, siguiendo especialmente las obras *Principles of Data Quality* (Principios de la calidad de datos) y *Principles and Methods of Data Cleaning* (Principios y métodos de limpieza de datos) de Arthur D. Chapman (2005) disponibles a continuación y en la [carpeta de documentos](#) de la plataforma de aprendizaje.



**Arthur D. Chapman<sup>1</sup>**

*Although most data gathering disciples treat error as an embarrassing issue to be expunged, the error inherent in [spatial] data deserves closer attention and public understanding ...because error provides a critical component in judging fitness for use. (Chrisman 1991).*



<sup>1</sup> Australian Biodiversity Information Services  
PO Box 7491, Toowoomba South, Qld, Australia  
email: papers.digit@gbif.org

**Documento:** Chapman, A. D. 2005. Principles of Data Quality, version 1.0. (Inglés). También disponible [aquí](#).





## PRIMARY SPECIES AND SPECIES-OCCURRENCE DATA

Arthur D. Chapman<sup>1</sup>

*Error qui non resistitur, approbatur*

An error not resisted is approved.

(Ref. Doct. & Stud. c. 770).

*Keywords:*

*Data Cleaning, Data Cleansing,*



<sup>1</sup> Australian Biodiversity Information Services  
PO Box 7491, Toowoomba South, Qld, Australia  
email: papers.digit@gbif.org

**Documento:** Chapman, A. D. 2005. Principles and Methods of Data Cleaning – Primary Species and Species- Occurrence Data, version 1.0. (Inglés). También disponible [aquí](#).

## Referencias

- Chapman, A. D. 2005. Principles and Methods of Data Cleaning – Primary Species and Species- Occurrence Data, version 1.0. Report for the Global Biodiversity Information Facility, Copenhagen.
- Chapman, A. D. 2005. Principles of Data Quality, version 1.0. Report for the Global Biodiversity Information Facility, Copenhagen.

## 2.- Principios de calidad de datos, calidad en todo el proceso

## Cadena de información de la calidad de los datos

### Introducción

Este primer apartado trata de responder algunas preguntas básicas relativas a los datos de biodiversidad y su calidad ¿A qué llamamos datos primarios de biodiversidad? ¿Para qué los usamos? ¿A qué nos referimos cuando hablamos de calidad de datos? Se muestran además las distintas etapas por las que va pasando un dato desde que es tomado en el campo hasta su difusión en internet o en publicaciones científicas.

### Vídeo

Su navegador no soporta VIDEO.  
Por favor, descargue el fichero: [video/mp4](#)

**Vídeo:** Cadena de información de la calidad de los datos - Katia Cezón (10:33 minutos).

También puedes ver el vídeo [aquí](#)

### Presentación

[Presentación: Cadena de información de la calidad de datos](#)

### Lecturas recomendadas\*

- Chapman, A. D. 2005. Principles of Data Quality, version 1.0. Report for the Global Biodiversity Information Facility, Copenhagen. Disponible [aquí](#).
- Chapman, A.D. 2005. Uses of Primary Species-Occurrence Data, version 1.0. Report for the Global Biodiversity Information Facility, Copenhagen. Disponible [aquí](#).

\*Todas disponibles para descargar en la [carpeta de documentos](#).

## Principios generales de la calidad de datos

### Introducción

En esta sección se dan una serie de principios generales para conseguir una adecuada calidad de datos. Estos principios están basados en la obra de Arthur Chapman [Principles of Data Quality](#).

### Vídeo

Su navegador no soporta VIDEO.  
Por favor, descargue el fichero: [video/mp4](#)

**Vídeo:** Principios generales de la calidad de datos - Katia Cezón (6:36 minutos).

También puedes ver el vídeo [aquí](#)

## Presentación

[Presentación: Principios generales de la calidad de datos](#)

### Toma de datos en el campo

## Introducción

La toma de datos en el campo constituye uno de los primeros pasos en la cadena de información de la biodiversidad y es por tanto un hito crítico que puede repercutir en la calidad de los mismos. En este apartado se exponen una serie de buenas prácticas que tienen como objetivo evitar la pérdida de la calidad y maximizar el potencial de uso de los datos en este punto clave de la cadena de la información.

## Vídeo

Su navegador no soporta VIDEO.  
Por favor, descargue el fichero: [video/mp4](#)

**Vídeo:** Toma de datos en el campo - Katia Cezón (10:01 minutos).

También puedes ver el vídeo [aquí](#).

## Presentación

[Presentación: Toma de datos en el campo](#)

### Calidad en los datos espaciales y temporales

## Introducción

En este apartado se explican los principios de calidad aplicables a los datos de localización y datación a lo largo de todo su proceso de captación y tratamiento.

## Vídeo

Su navegador no soporta VIDEO.  
Por favor, descargue el fichero: [video/mp4](#)

**Vídeo:** Calidad en los datos espaciales y temporales - Katia Cezón (25:30 minutos).

También puedes ver el vídeo [aquí](#).

## Presentación

[Presentación: Calidad en los datos espaciales y temporales](#)

## Lecturas recomendadas\*

- Chapman, A.D. and J. Wiecek (eds). 2006. Guide to Best Practices for Georeferencing. Copenhagen: Global Biodiversity Information Facility. Disponible [aquí](#).
- Muñoz López, E., Hernández Barrios, J.C., Colin López, J. 2004. Georreferenciación de localidades de colectas biológicas de la Conabio. Biodiversitas 54:8-15. Disponible [aquí](#).
- Wiecek, 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. Disponible [aquí](#).
- Otegui, J., Ariño, A., & al, 2013. On the dates on the GBIF mobilised primary biodiversity data records. Biodiversity Informatics 8: 173-184. Disponible [aquí](#).

\*Todas disponibles para descargar en la [carpeta de documentos](#).

## Calidad en los datos taxonómicos

### Introducción

La mayoría de datos biológicos incluyen información taxonómica cuya calidad puede hacer la diferencia en el nivel de uso y reúso de los mismos. En este vídeo veremos en qué consisten los datos taxonómicos, las fuentes de errores más comunes y cómo mejorar su calidad.

### Vídeo

Su navegador no soporta VIDEO.  
Por favor, descargue el fichero: [video/mp4](#)

**Vídeo:** Calidad en los datos taxonómicos - Francisco Pando (16:29 minutos).

También puedes ver el vídeo [aquí](#).

## Presentación

[Presentación: Calidad en los datos taxonómicos](#)

## Lectura recomendada\*

- Ytow, N, Morse DR, Roberts DM. (2001) Nomenclator: a nomenclatural history model to handle multiple taxonomic views. *Biological Journal of Linnean Society* 73 (1):81-98. Disponible [aquí](#).

\*Disponible para descargar en la [carpeta de documentos](#).

## Calidad en los datos descriptivos

## Introducción

En esta charla se trata la calidad de los datos descriptivos, que son aquellos que se utilizan para caracterizar los elementos de biodiversidad. Estos incluyen caracteres taxonómicos, como color de las hojas, tamaño de los tallos, etc., descriptores: estatus de conservación, usos de la especie, etc.

## Vídeo

Su navegador no soporta VIDEO.  
Por favor, descargue el fichero: [video/mp4](#)

**Vídeo:** Calidad en los datos descriptivos - Francisco Pando (12:29 minutos).

También puedes ver el vídeo [aquí](#).

## Presentación

[Presentación: Calidad en los datos descriptivos](#)

### Calidad de datos durante la digitalización y el almacenamiento

## Introducción

En esta charla veremos como la manera en la que se digitalizan los datos de biodiversidad y el diseño de la base de datos pueden condicionar su calidad.

Además (sólo en la presentación), se incluyen algunas pautas o recomendaciones sobre el almacenamiento y copias de seguridad de los datos.

## Vídeo

Su navegador no soporta VIDEO.  
Por favor, descargue el fichero: [video/webm](#)

**Vídeo:** Calidad de datos durante la digitalización y el almacenamiento - Francisco Pando (20:41 minutos).

También puedes ver el vídeo [aquí](#).

## Presentación

[Presentación: Calidad de datos durante la digitalización y el almacenamiento](#)

## Lectura recomendada\*

- CONABIO. Digitisation and data quality control of Mexican and Central American Botanical Specimens held at the Missouri Botanical Garden Herbarium. v 1.2. Disponible [aquí](#).



\*Todas disponibles para descargar en la [carpeta de documentos](#).

### 3.- Estándares para la información sobre biodiversidad

#### Estándares para la información sobre biodiversidad

## Introducción

En esta sección se presenta una breve introducción a qué son los estándares y cómo los utilizamos cuando tratamos con datos de biodiversidad. En particular, nos enfocaremos en el estándar Darwin Core, ampliamente utilizado por la comunidad para compartir los datos.

En el primer video se recuerdan algunos conceptos básicos tales como qué son los estándares y para qué son útiles. En el segundo video se describe la estructura del estándar Darwin Core y cómo se utiliza para compartir datos.

## Vídeo 1

**Vídeo:** Estándares para la información sobre biodiversidad: un poco de contexto – Paula Zermoglio (08:15 minutos).

También puedes ver el vídeo [aquí](#).

### Presentación 1

[Presentación: Estándares para la información sobre biodiversidad: un poco de contexto](#)

## Vídeo 2

**Vídeo:** Estándares para la información sobre biodiversidad: Darwin Core – Paula Zermoglio (15:26 minutos).

También puedes ver el vídeo [aquí](#).

## Presentación 2

[Presentación: Estándares para la información sobre biodiversidad: Darwin Core](#)

**NOTA:** Para el video número 2 y su correspondiente presentación se utiliza la página web de Darwin Core en una versión posterior a la actualmente disponible (Jan 2018). Si compara el video con la página actual verá que el formato de la página es diferente, pero los contenidos son los mismos. La nueva versión de la página web estará disponible en los próximos meses.

## Luego de ver los videos...

Luego de haber visto ambos videos, visita la [página web](#) del estándar Darwin Core e identifica qué términos del estándar te serían útiles para compartir datos de biodiversidad. Nota que algunos términos pueden tener nombres que no son intuitivos (e.g., para colector, el término correspondiente en Darwin Core es recordedBy).

Te proponemos además que, si tienes preguntas sobre el estándar o sobre su utilización, pruebes enviar tus preguntas a través del [formulario](#) de la iniciativa [Darwin Core Questions & Answers](#).

## Lectura recomendada

Wieczorek et al. (2012). Darwin Core: An Evolving Community-Developed Biodiversity Data Standard. PLoS ONE 7(1): e29715. <https://doi.org/10.1371/journal.pone.0029715>

## Enlaces útiles

Biodiversity Information Standards (TDWG): <http://www.tdwg.org/>

Darwin Core Standard: <http://rs.tdwg.org/dwc/terms/index.htm>

Darwin Core Questions & Answers Site: <https://github.com/tdwg/dwc-qa>

### 4.- Limpieza, formato y transformación de datos. Herramientas útiles para mejorar la calidad de los datos

#### Herramientas útiles para la calidad y limpieza de los datos

#### Herramientas útiles para la calidad y limpieza de los datos

## Introducción

En esta sección se proponen algunas herramientas que pueden resultar útiles para garantizar la calidad de los datos de biodiversidad y limpiar los mismos. Se incluyen tanto herramientas generales de gestión y almacenamiento, como herramientas específicas para cubrir los diferentes componentes de los datos de biodiversidad (herramientas para el tratamiento de nombres científicos, datos geográficos, fechas, etc.).

Además de la charla, se incluyen breves tutoriales sobre el funcionamiento de algunas de ellas.

## Vídeo

Su navegador no soporta VIDEO.  
Por favor, descargue el fichero: [video/mp4](#)

**Vídeo:** Herramientas útiles para la calidad y limpieza de los datos - Katia Cezón (10:16 minutos).

También puedes ver el vídeo [aquí](#).

## Presentación

[Presentación: Herramientas útiles para la calidad y limpieza de los datos](#)

## Tutoriales

En esta sección se muestran tutoriales de algunas de las herramientas proporcionadas por GBIF.org o GBIF.es. Si hay alguna herramienta que te interesa y no sabes cómo funciona, ponte en contacto con nosotros : )

### Species Matching de GBIF.org

Este servicio que ofrece GBIF.ORG <https://www.gbif.org/tools/species-lookup> permite normalizar los nombres científicos ubicados en un archivo .csv frente al "backbone" taxonómico de GBIF.

El "Backbone" de GBIF es la base taxonómica que utiliza GBIF para integrar la información de nombres científicos de diferentes recursos, lo cual proporciona consistencia al portal y facilita las búsquedas taxonómicas en el mismo. Para construirlo el portal se nutre de Catalogue of Life y de otras 54 fuentes que pueden consultarse aquí: <https://www.gbif.org/dataset/d7dddbf4-2cf0-4f39-9b2a-bb099caae36c>.

Su navegador no soporta VIDEO.

Por favor, descargue el fichero: [video/mp4](#)

**Vídeo tutorial:** Funcionamiento del servicio "Species Matching" de GBIF.org.

### Name Parser de GBIF.es

Se trata de una aplicación en base de datos MS Access para desglosar nombres científicos en sus componentes básicos: género, especie, autor de la especie, rango infraespecífico, epíteto infraespecífico, autor infraespecífico y año.

Descargable [aquí](#).

Su navegador no soporta VIDEO.

Por favor, descargue el fichero: [video/mp4](#)

**Vídeo tutorial:** Funcionamiento del servicio "Name Parser" de GBIF.es.

### Conversor de coordenadas de GBIF.es

Se trata de una mini aplicación en base de datos MS Access que sirve para convertir coordenadas expresadas en cualquier formato (MGRS, UTM, geográficas sexagesimales) a coordenadas geográficas decimales, que es el formato de publicación de GBIF.

Descargable [aquí](#).

Su navegador no soporta VIDEO.

Por favor, descargue el fichero: [video/mp4](#)

**Vídeo tutorial:** Funcionamiento del servicio de conversión de coordenadas de GBIF.es.

### Conversor de fechas de GBIF.es

Se trata de una mini aplicación en base de datos MS Access que sirve para procesar y homogeneizar fechas expresadas en distintos formatos.

Descargable [aquí](#).

Su navegador no soporta VIDEO.  
Por favor, descargue el fichero: [video/mp4](#)

**Vídeo tutorial:** Funcionamiento del servicio de conversión de fechas de GBIF.es.

## Ejercicio 1: Validación de errores en el proceso de calidad I

### Objetivo

## Ejercicio 2: Validación de errores en el proceso de calidad II

### Objetivo

El objetivo de este ejercicio es la identificación y corrección de errores de tipo nomenclatural, geográfico y de formato de fecha, presentes en la base de datos de prueba. Para ello puedes utilizar libremente las herramientas que desees, aunque recomendamos la práctica con algunas de las mostradas durante el taller.

### Metodología

El ejercicio consiste en la identificación y corrección de errores de consistencia en un conjunto de datos. Para el ejercicio se utilizará un dataset sobre el cual se realizará la limpieza, y una hoja de ruta con una planilla en la cual los alumnos capturarán los errores detectados y las acciones tomadas para remediarlos.

El conjunto de datos a utilizar ha sido modificado específicamente para el ejercicio. El conjunto de datos original corresponde a la colección [Fishes in MZNA-VERT: freshwater fishes of Hidalgo state \(Mexico\)](#) de la Universidad de Navarra.

<sup>1</sup>Museum of Zoology M, Galicia Paredes D (2016). Fishes in MZNA-VERT: freshwater fishes of Hidalgo state (Mexico). University of Navarra – Department of Environmental Biology. Occurrence Dataset <https://doi.org/10.15468/4rg5de> accessed via GBIF.org on 2018-01-19.

### Evaluación

La evaluación de este ejercicio se realizará analizando el conjunto de datos final y la hoja de resultados provistos por cada alumno. Se contemplarán como elementos de evaluación la aplicación de los conceptos y técnicas aprendidos durante el curso y cómo son aplicados para la corrección de errores.

Una vez finalizado el ejercicio envía los dos archivos de resultados un archivo comprimido a través del sistema de [tareas de la plataforma](#).

Recuerda que puede revisar el material del curso cuantas veces lo necesites y que puedes consultar a los docentes si surgen dudas.

## Documento de descripción del ejercicio

## Taller GBIF.ES: Calidad en bases de datos de biodiversidad

### Ejercicio 2: Validación de errores en el proceso de calidad II



## CONTENIDOS

[INTRODUCCIÓN](#)

[NECESITARÁS](#)

[INSTRUCCIONES](#)

[EVALUACIÓN](#)

[Acceso al documento con las descripciones del ejercicio](#)

[Acceso a los materiales para el ejercicio](#)



## Validación de datos con Darwin Test

## Validación de datos con Darwin Test

# Resumen de contenidos

En esta sección se presenta **Darwin Test**, una aplicación diseñada para comprobar la calidad de datos de las colecciones de historia natural. Se detalla cómo descargar e instalar este programa y se muestran sus principales funcionalidades a través de vídeos explicativos.

Al final de los contenidos teóricos se incluye un documento con los ejercicios prácticos a realizar.

## Introducción

Darwin Test es una aplicación diseñada para la validación y chequeo de los registros almacenados en tablas con formato [estándar Darwin Core](#) (1.2, Darwin Core 1.4 y Darwin Core Archive), utilizado en GBIF para hacer accesibles sus datos vía web.

Estas son algunas de sus características más importantes son:

- Permite detectar errores en los campos de los archivos con formato estándar Darwin Core (1.2, Darwin Core 1.4 y Darwin Core Archive), mediante consultas predeterminadas configurables por el usuario
- Ofrece la posibilidad de corregir los errores detectados en los datos desde los propios formularios de validación. Dichos cambios pueden integrarse de forma permanente en los archivos con formato Darwin Core de origen o reflejarse en una tabla intermedia para su posterior corrección
- Permite la activación/desactivación de las consultas existentes
- Permite la creación de nuevas consultas de validación adaptadas a las necesidades de cada usuario
- Permite el chequeo de los nombres científicos con la Tabla de Nombres SCIENTIFIC\_NAMES del consorcio [The Catalogue Of Life: JTIS](#) y [Species 2000](#), o con otra base de datos con nombres científicos que cumpla unos requisitos mínimos.
- Permite convertir coordenadas UTM, MGRS y coordenadas sexagesimales en coordenadas geográficas en grados decimales, utilizadas en el formato Darwin Core
- Chequea la existencia de caracteres ASCII anómalos y los reemplaza por caracteres válidos que puede definir el usuario
- Permite calcular un [Índice de Calidad Aparente \(ICA\)](#) para el seguimiento de la mejora de la calidad de los datos
- Permite la creación y gestión de filtros de usuario para eliminar registros o generalizar coordenadas selectivamente de las tablas con formato Darwin Core
- Permite el intercambio de datos entre los formatos Darwin Core 1.2, Darwin Core 1.4 y Darwin Core Archive (exportación entre versiones).

## Antes de empezar

Darwin Test está basado en MS-Access, funciona con sistema operativo Windows y MS-Office XP instalado.

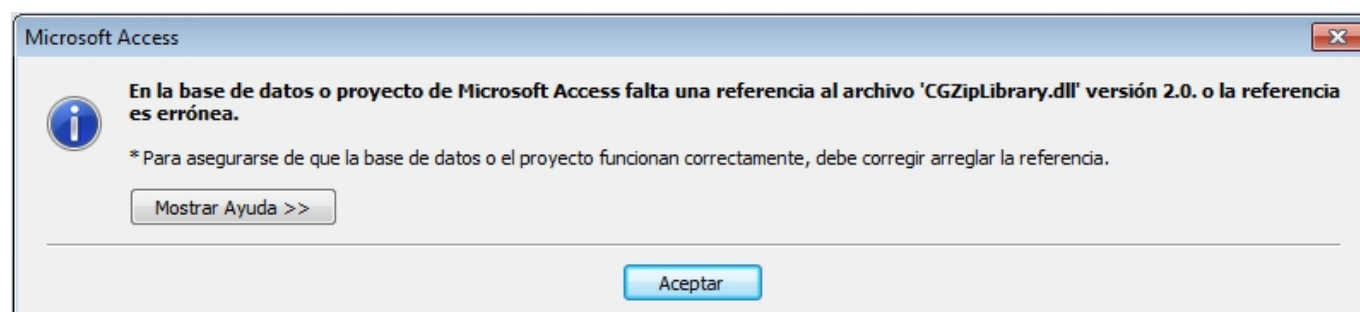
Consta de una única base de datos llamada *Darwin\_test(versión).mdb* donde se encuentran los menús y la programación.

Desde la página web: [http://www.gbif.es/darwin\\_test/Darwin\\_Test.php](http://www.gbif.es/darwin_test/Darwin_Test.php) se accede al punto de descarga de la aplicación, donde se pueden encontrar todas las versiones de la aplicación disponibles, tanto para Access 2003 como para Access 2007 o superiores, si bien, **para el desarrollo del taller en el siguiente apartado proporcionaremos una misma versión para todos los alumnos.**

## Instalación y manejo

1. Descargamos el fichero **Darwintest33-2k2.ZIP** desde [aquí](#).
2. Creamos un directorio donde instalar la aplicación (recomendamos el directorio C:\Darwintest) y descomprimos aquí el fichero ZIP que acabamos de descargar.
3. Para entrar en la aplicación, hacemos doble *click* sobre **Darwin\_test33.mdb**, en ese momento aparecerá el menú principal desde el que se gestionan las principales operaciones. En los siguientes tutoriales se explican las principales funcionalidades de la aplicación. Para más información consulte el [manual del usuario](#) en formato *pdf*.

**Nota importante:** en algunas ocasiones al abrir la aplicación por primera vez puede salir un mensaje de Microsoft Access advirtiéndole de la falta de una o varias *referencias*, tal como se muestra en la imagen que sigue.



Éstos, son unos archivos imprescindibles para que funcionen algunas opciones de la aplicación y es necesario que se instalen en el equipo. Si el alumno se encuentra en ese caso debe seguir las instrucciones que siguen:

1. Descarga las siguientes referencias pulsando [aquí](#).
2. Descomprime el archivo y pega las librerías en C:\Windows\System32 de tu equipo.
3. Ejecuta la aplicación para comprobar que no aparece ningún mensaje de aviso.

Si a pesar de esto el problema no se hubiese solucionado, contacta con nosotros a través de alguno de los medios especificados en la metodología para resolver el problema.

## 1. Configuración de validaciones

En el vídeo tutorial de esta sección veremos cómo **configurar las validaciones** (comprobaciones) que realiza la aplicación para detectar posibles errores en las bases de datos con formato Darwin Core y cómo **vincular la tabla de nombres científicos de Catalogue of Life (Species2000)** que la aplicación puede utilizar como referencia para comprobar los nombres científicos.

Su navegador no soporta VIDEO.  
Por favor, descargue el fichero: [video/mp4](#)

**Vídeo tutorial:** Configuración de las Validaciones en Darwin Test (2:11 min). También disponible [aquí](#).

## 2. Detectar y corregir errores

En este tutorial veremos cómo comprobar la calidad de los datos de una tabla en formato Darwin Core y cómo corregir los errores detectados desde los formularios de validación. Además se explica el procedimiento para detectar y eliminar los caracteres ASCII anómalos, que son caracteres como tabuladores, saltos de párrafo, etc., que pueden interferir en la correcta indexación de los datos en internet.

Su navegador no soporta VIDEO.  
Por favor, descargue el fichero: [video/mp4](#)

**Vídeo tutorial:** Detectar y corregir errores con Darwin Test (3:40 min). También disponible [aquí](#).

### 3. Índice de Calidad Aparente (ICA)

En el siguiente vídeo se muestra el formulario que resume la calidad de nuestros datos.

Su navegador no soporta VIDEO.  
Por favor, descargue el fichero: [video/mp4](#)

**Vídeo tutorial:** Cálculo del Índice de Calidad Aparente (ICA) en Darwin Test (2:03 min). También disponible [aquí](#).

### 4. Generalización de coordenadas

En este vídeo veremos en detalle la opción de Darwin Test para generalizar y eliminar coordenadas, que se puede realizar si no se desea informar de la verdadera localización del espécimen.

Su navegador no soporta VIDEO.  
Por favor, descargue el fichero: [video/mp4](#)

**Vídeo tutorial:** Generalización de coordenadas en Darwin Test (3:40 min). También disponible [aquí](#).

### 5. Conversión de coordenadas

En el siguiente vídeo se explica cómo transformar a coordenadas geográficas decimales (ej. 42.06 -1.35) un conjunto de coordenadas expresadas en cualquier otro formato utilizando la aplicación Darwin Test.

Su navegador no soporta VIDEO.  
Por favor, descargue el fichero: [video/mp4](#)

**Vídeo tutorial:** Conversión de coordenadas en Darwin Test (3:23 min). También disponible [aquí](#).

## Ejercicio 3: Validación de datos con Darwin Test

### Objetivo

En esta sección se incluyen 4 casos prácticos que tienen como objetivo comprobar la calidad y corregir errores de bases de datos de biodiversidad presentadas en formato estándar Darwin Core (Darwin Core 1.2, Darwin Core 1.4 y Darwin Core Archive), utilizando la aplicación Darwin Test.

### Metodología

A continuación se proporciona un documento donde encontrarás los materiales e instrucciones necesarios para completar las tareas. Para realizarlas necesitarás instalar Darwin Test siguiendo las instrucciones del apartado teórico anterior.

### Evaluación

La evaluación de este ejercicio se realizará analizando el conjunto de datos final. Para que los instructores puedan revisar los resultados, una vez finalizados debes enviar tu propio Darwin\_Test31.mdb comprimido (archivo .zip o .rar) a través del sistema de [tareas de la plataforma](#).

Recuerda que puedes revisar el material del curso cuantas veces lo necesites y que puedes consultar a los docentes si te

surgen dudas.

## **Documento de descripción del ejercicio**

## Taller GBIF.ES: Calidad en bases de datos de biodiversidad

### Ejercicio 3: Validación de datos con Darwin Test



## CONTENIDOS

### [CONTENIDOS](#)

### [INTRODUCCIÓN](#)

### [ANTES DE EMPEZAR](#)

### [Práctica 1. Validación de datos](#)

### [Práctica 2. Corrección de campos validados](#)

### [Práctica 3. Transformación de coordenadas](#)

### [Práctica 4. Generalización de coordenadas](#)

[Acceso al documento con las descripciones del ejercicio](#)

[Acceso a los materiales para el ejercicio](#)



## Limpieza de datos utilizando OpenRefine

### Introducción a OpenRefine. Detectando y corrigiendo errores con OpenRefine

# Introducción

[OpenRefine](#) (antes GoogleRefine) es una herramienta potente para la limpieza y transformación de datos. Es mantenida y desarrollada por la comunidad, permite trabajar con diferentes formatos de archivos, y presenta la ventaja de que el archivo de datos original no es afectado por los procesos de limpieza. Si bien la herramienta se ejecuta en un navegador, no requiere conexión a Internet (excepto cuando interactúa con otros servicios), y ello permite utilizarla desde cualquier ordenador. OpenRefine presenta un manejo general que es bastante intuitivo, aunque algunas funciones específicas requieren un poco de práctica.

## Utilizando OpenRefine

Para comprender cómo utilizar OpenRefine, seguiremos un ejercicio modelo, utilizando un conjunto de datos que ha sido modificado para demostrar las funciones básicas de la herramienta.

En primera instancia, deberemos instalar el programa en nuestro ordenador. Para ello, sigue estas [instrucciones](#).

Una vez instalado el programa, sigue los pasos detallados en la **Guía Básica de Uso de OpenRefine** que se proporciona a continuación para aprender a manejar la aplicación.

Guía de Uso



# Guía de Uso Básico de

# **Refine**<sup>OPEN</sup>

Zermoglio P, Wieczorek J. 2018

1

[Acceso al documento](#)

Ten en cuenta que la guía es larga y que practicar todas las funciones puede llevar algún tiempo. Dependiendo de tu experiencia previa con este tipo de herramientas, planifica tiempo suficiente para seguir la guía completa (estima no menos de dos horas).

# Enlaces útiles

- Sitio web: <http://openrefine.org/>, descarga: <http://openrefine.org/download.html>
- Repositorio en GitHub: <https://github.com/OpenRefine/OpenRefine/wiki> (en este enlace encontrarás detalles sobre la utilización del programa y sus distintas funciones).

## Ejercicio 4: Limpieza de datos utilizando Open Refine

### Objetivo

Aplicar los conceptos aprendidos durante el curso para la limpieza de un conjunto de datos utilizando OpenRefine.

### Metodología

El ejercicio consiste en la identificación y corrección de errores en un conjunto de datos utilizando OpenRefine. Para el ejercicio se utilizará un conjunto de datos sobre el cual se realizará la limpieza, y una hoja de ruta en la cual los alumnos capturarán los errores detectados y las acciones tomadas para remediarlos.

El conjunto de datos a utilizar ha sido modificado específicamente para el ejercicio. El conjunto de datos original corresponde a la colección de ornitología del Instituto Alexander vonHumboldt, Colombia <sup>1</sup>.

<sup>1</sup>Borja-Acosta K (2017). Colección de Aves de Colombia del Instituto Alexander von Humboldt. Version 30.0. Instituto de Investigación de Recursos Biológicos Alexander von Humboldt. Occurrence Dataset <https://doi.org/10.15472/maeoiz> accessed via GBIF.org on 2018-01-10.

### Evaluación

La evaluación de este ejercicio se realizará analizando el conjunto de datos final y la hoja de ruta provistos por cada alumno. Se contemplarán como elementos de evaluación la aplicación de los conceptos y técnicas aprendidos durante el curso y cómo son aplicados para la corrección de errores.

Una vez finalizado, envía tus resultados en un archivo comprimido a través del sistema de [tareas de la plataforma](#).

Recuerda que puede revisar el material del curso cuantas veces lo necesites y que puedes consultar a los docentes si surgen dudas.

## Documento de descripción del ejercicio

## Taller GBIF.ES: Calidad en bases de datos de biodiversidad

### Ejercicio 4: Limpieza de datos utilizando OpenRefine



## CONTENIDOS

### CONTENIDOS

#### INTRODUCCIÓN

- [1. Organización del trabajo.](#)
- [2. Verificación de registros: números de catálogo.](#)
- [3. Errores de formato y rango](#)
- [4. Unificación de valores similares.](#)
- [5. Evaluación y limpieza taxonómica.](#)

*Ejercicio preparado por P. Zermoglio*

---

[Acceso al documento con las descripciones del ejercicio](#)

[Acceso a los materiales para el ejercicio \(csv\)](#)

¡Ya estás listo para empezar!

## Encuesta de evaluación

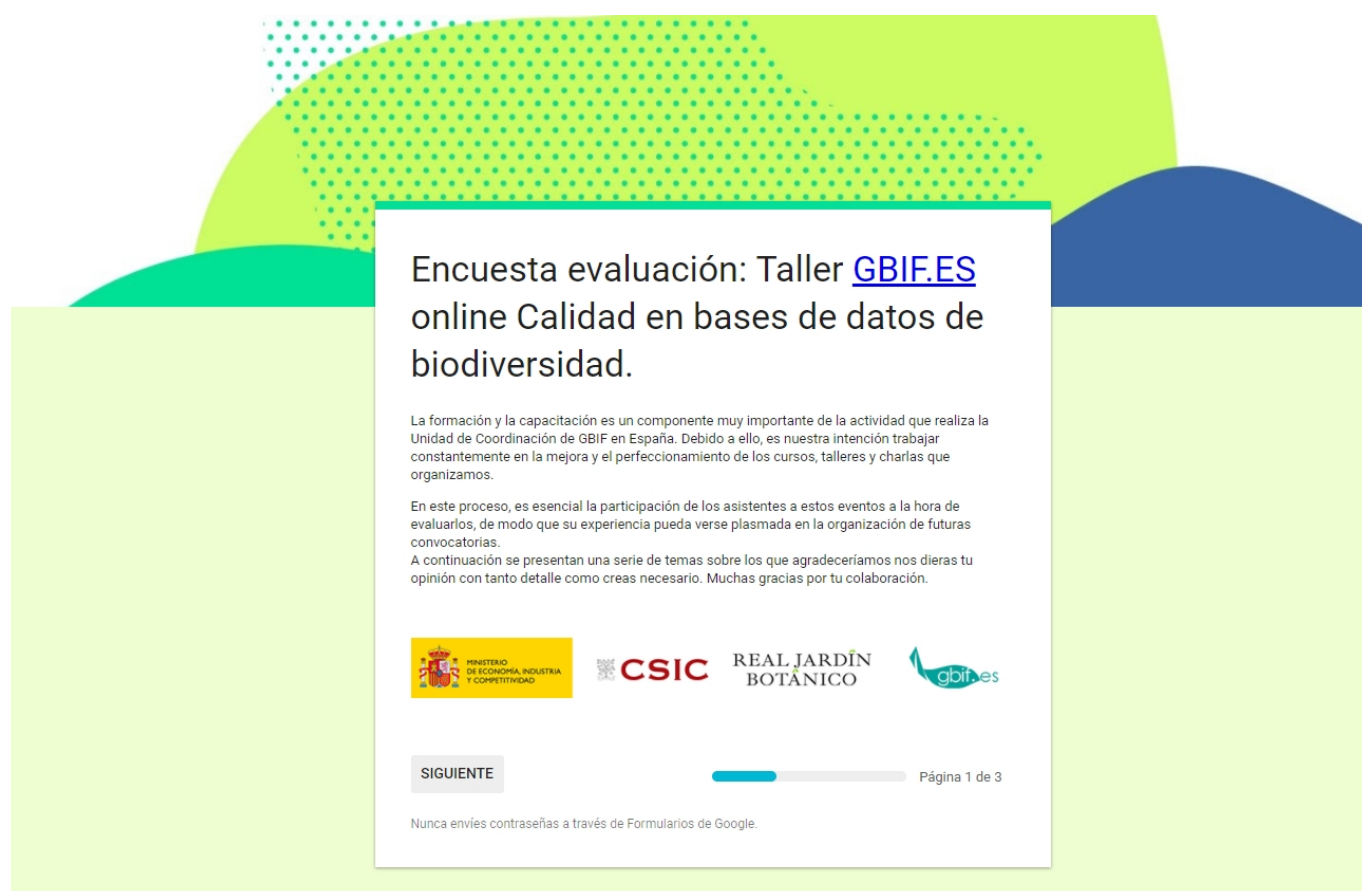
### Encuesta de evaluación

La formación y la capacitación es un componente muy importante de la actividad que realiza la Unidad de Coordinación de GBIF en España. Debido a ello, es nuestra intención trabajar constantemente en la mejora y el perfeccionamiento de los cursos, talleres y charlas que organizamos.

En este proceso, es esencial la participación de los asistentes a estos eventos a la hora de evaluarlos, de modo que tu experiencia pueda verse plasmada en la organización de futuras convocatorias : )

A continuación se presentan una serie de temas sobre los que agradeceríamos nos dieras tu opinión con tanto detalle como creas necesario.

¡Muchas gracias por tu colaboración!







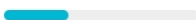
**Encuesta evaluación: Taller [GBIF.ES](#) online Calidad en bases de datos de biodiversidad.**

La formación y la capacitación es un componente muy importante de la actividad que realiza la Unidad de Coordinación de GBIF en España. Debido a ello, es nuestra intención trabajar constantemente en la mejora y el perfeccionamiento de los cursos, talleres y charlas que organizamos.

En este proceso, es esencial la participación de los asistentes a estos eventos a la hora de evaluarlos, de modo que su experiencia pueda verse plasmada en la organización de futuras convocatorias.

A continuación se presentan una serie de temas sobre los que agradeceríamos nos dieras tu opinión con tanto detalle como creas necesario. Muchas gracias por tu colaboración.

**SIGUIENTE**  Página 1 de 3

Nunca envíes contraseñas a través de Formularios de Google.

[Haz clic aquí para acceder a la encuesta de evaluación](#)