

Uso de datos de biodiversidad

Secretariado de GBIF

Versión b34f741, 2022-03-30 08:40:29 UTC

Tabla de Contenido

Descripción	1
Audiencia	1
Prerequisitos	2
Objetivos de aprendizaje	2
Certificación	2
Procesamiento de datos	3
Datos aptos para propósitos específicos	3
Avisos e Incidencias	5
Herramientas de Procesamiento de los Datos	6
Manejo de la Incertidumbre Taxonómica	6
Identificación errónea de la especie	7
Sinonimia	7
Nombres nuevos	8
Manejo de la Calidad de los Datos	8
Filtros e Incidencias Geoespaciales	8
Centroides de país	9
Puntos a lo largo del ecuador o meridiano principal	9
Localización incierta	9
Conjuntos de datos en cuadrícula	10
Registros de ausencia	11
Medios de establecimiento	11
Base del Registro	11
Registros Antiguos	12
Duplicados	12
Documentación clave	12
API	13
Computación en la Nube	13
Darwin Core	13
Publicación de datos	13
Publicación de datos: IPT	14
Digitalización	14
GBIF	14
Georreferenciación	14
Especies Invasoras	15
Living Atlases	15
Misceláneo	15
OpenRefine	15
Planificación/Colaboración	15

Red List Assessments	16
Calidad	17
R	17
Especies sensibles	17
Taxonomía	17
Glosario	18
Agradecimientos	21
Diseño del curso e instrucción	21
Traductores	21
Español	21
Francés	21
Recursos	21
Colofón	22
Cita sugerida	22
Colaboradores	22
Licencia	22
URI persistente	22
Control de documentos	22
Imagen de la portada	22

Descripción

Este curso adopta un enfoque modular para la formación en el uso de datos mediados por GBIF, y se basa en el curso de Uso de Datos para la Toma de Decisiones desarrollado como parte del programa de Información sobre la Biodiversidad para el Desarrollo (BID). Se han desarrollado diferentes casos de uso que le permitirán comenzar con algunas de las diferentes maneras en que puede utilizar datos mediados por GBIF.

Este curso está compuesto por contenido en línea asociado con cuestionarios y deberías completarlo primero. Entonces puedes completar el caso de uso pertinente que se basará en lo que has aprendido a través de un conjunto de ejercicios prácticos. Cuando se ofrece como un taller virtual o in situ, se fomenta el trabajo en grupo y la interacción social.

El curso comprende un curso obligatorio en Procesamiento de Datos y luego módulos opcionales que puedes seguir basándose en el camino de aprendizaje más relevante para ti. Los dos primeros módulos opcionales que hemos desarrollado son:

- Modelado de Nicho Ecológico
- Evaluar el Estado de Conservación de una Especie

Los temas incluyen:

- Procesamiento de datos
 - Procesando un conjunto de datos mediados por GBIF y haciéndolo “apto para un propósito”
- Modelado de Nicho Ecológico
 - Introducción al funcionamiento e interpretación de un modelo ecológico básico para determinar la distribución de una especie
 - Explorando el modelado de nicho bajo diferentes condiciones ambientales
- Evaluar el Estado de Conservación de una Especie
 - Utilizando datos mediados por GBIF, para crear mapas de distribución de especies utilizando estándares de mapeo de IUCN
 - Utilizando datos mediados por GBIF, para evaluar el estado de conservación de especies utilizando las Categorías y Criterios de la Lista Roja de la UICN

Audiencia

Este curso está diseñado para personas que trabajan como investigadores o técnicos en investigación sobre biodiversidad o instituciones de políticas. La instrucción proporcionada es particularmente útil para aquellos que tienen la necesidad o el deseo de utilizar datos mediados por GBIF en su propia investigación o análisis.

Prerequisitos

- [Introducción al curso GBIF](#)

Además, para hacer un mejor uso de las actividades en torno a este curso, los participantes deben poseer las siguientes habilidades y conocimientos:

- Competencias básicas en el uso de computadoras e Internet y, en particular, en el uso de hojas de cálculo.
- Conocimiento básico de herramientas de análisis geográfico y estadístico basadas en computadoras, por ejemplo SIG y R, y puede haber ejecutado análisis utilizando estas herramientas.
- Conocimiento básico sobre geografía e informática sobre biodiversidad: conceptos de geografía y mapeo, taxonomía básica y reglas de nomenclatura.
- La voluntad de diseminar el conocimiento aprendido en el taller con socios y colaboradores en su proyecto adaptando los datos sobre biodiversidad a contextos específicos e idiomas, manteniendo al mismo tiempo su valor educativo.
- Un buen dominio del inglés. Aunque se hace un esfuerzo por proporcionar materiales en otros idiomas, la información/los videos serán en inglés.

Objetivos de aprendizaje

- Acceda a los datos mediados por GBIF a través de un rango de puntos de acceso
- Entender problemas comunes de calidad de datos en las descargas de GBIF que pueden afectar al uso de datos
- Aplicar rutinas de procesamiento de datos en las descargas de GBIF para crear conjuntos de datos adecuados para el uso
- Aprende la diferencia entre nichos fundamentales y realizados
- Explica cómo delimitar una región de entrenamiento
- Generar un modelo de nicho simple
- Explicar los resultados del modelo nicho
- Identificar áreas de incertidumbre en la proyección
- Desarrollar una estrategia de comunicación y argumentos convincentes para la integración de la biodiversidad en los procesos de toma de decisiones
- Aplicar el criterio B de las Categorías y Criterios de la UICN para la evaluación del estado de conservación de una especie utilizando un conjunto de datos adecuado
- Aplicar los protocolos de mapeo de IUCN para la producción de mapas de distribución de especies de la Lista Roja

Certificación

Se emitirán certificados, en forma de insignias digitales, a los participantes que completen con éxito

la asignación de casos de uso. Hay varios casos de uso y usted debe completar los más relevantes para su trabajo o que le interesen.

Unresolved directive in index.es.adoc - include::downloads.es.adoc[]

Procesamiento de datos



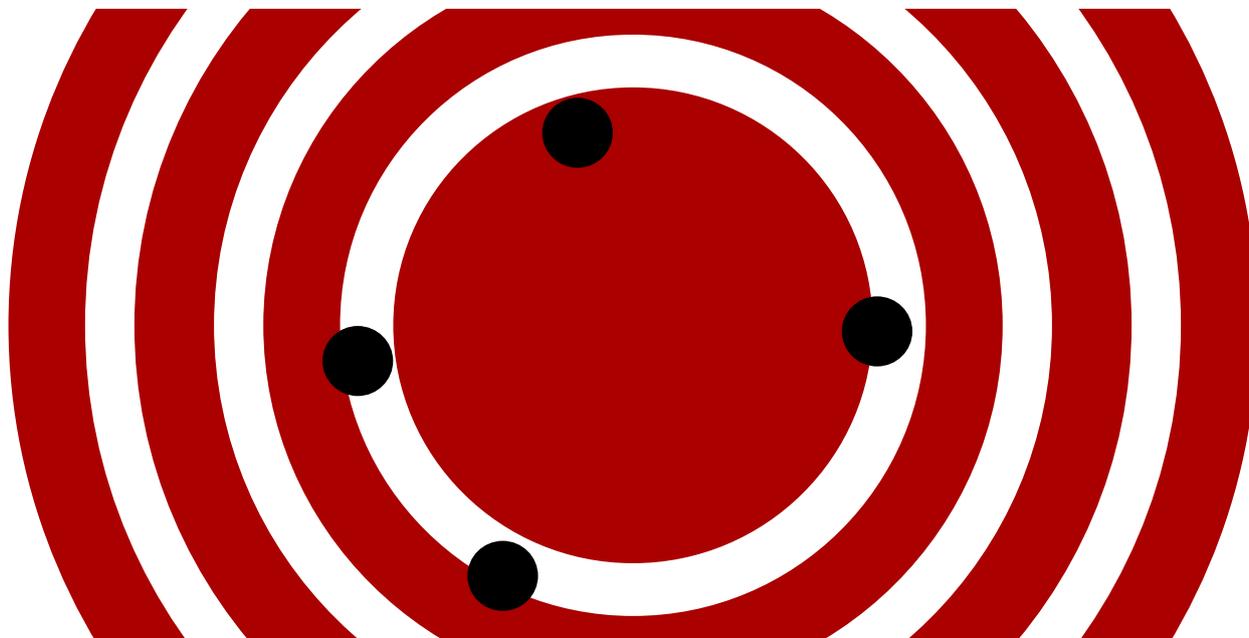
En el módulo de procesamiento de datos, se familiarizará con el concepto de conjuntos de datos aptos para uso y algunos filtros de datos comúnmente utilizados que puede querer considerar para crear su propio conjunto de datos apto para el uso.

Datos aptos para propósitos específicos

Casi siempre querrá pos-procesar su descarga de GBIF de alguna manera que se ajuste a sus propósitos. A veces, tendrá que hacer juicios difíciles para su caso de uso particular. Siempre que esté tratando con miles o millones de registros, nunca sabrá la verdadera calidad de los datos de origen. Es importante tener en cuenta que siempre está mitigando los problemas de calidad de los datos, no eliminándolos.

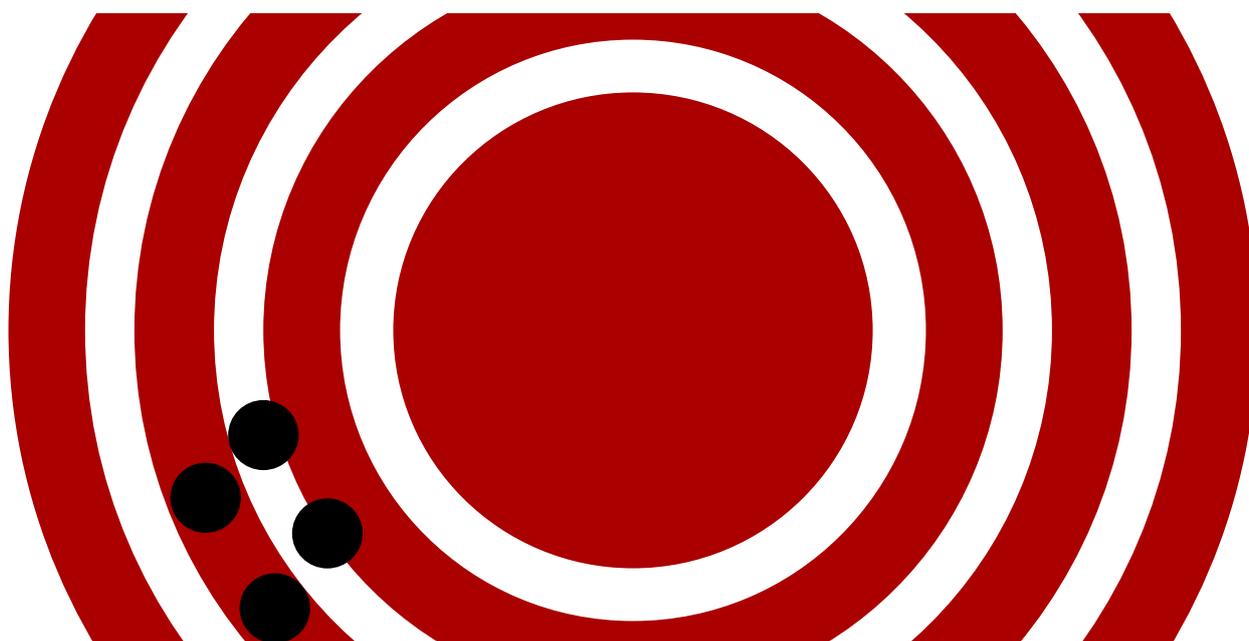
Los datos que obtenemos en la descarga de GBIF contendrán datos de una variedad de fuentes y es probable que los datos varíen en su exactitud y consistencia. La exactitud y la coherencia son dos formas de documentar los errores de datos y son medidas de la calidad de los datos. Estas son medidas de qué tan bien el recolector de datos pudo capturar el valor real que se estaba investigando. La naturaleza del flujo de trabajo de publicación de datos de GBIF significa que la exactitud y consistencia de los datos pueden variar según los publicadores y la fuente de los datos. Conocer estas propiedades de los datos le ayudará a comprender las formas en las que puede y no puede limpiar, validar y procesar los datos.

- Exactitud (precisión) - proximidad de los valores, observaciones o estimaciones medidas al valor real o verdadero, ej.: se ha identificado correctamente la especie o la localidad de la colección.



Por ejemplo, si estamos estudiando biogeografía de plantas en Indonesia, y queremos realizar un análisis específico para sólo una de las islas del archipiélago, entonces una pregunta apropiada podría ser - ¿Las localidades de la isla han sido correctamente georreferenciadas?

- Consistencia (Precisión) - nivel de resolución de los datos, ej.: precisión de las coordenadas, determinación taxonómica.

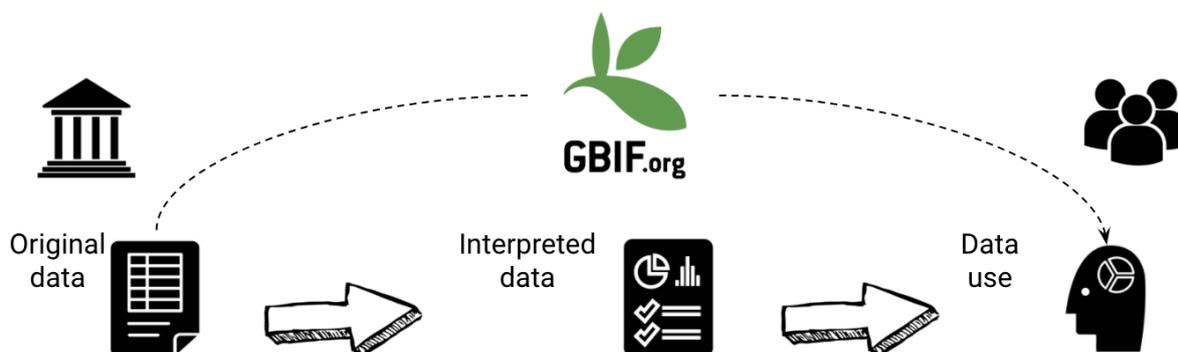


En el ejemplo indonesio, una pregunta apropiada podría ser: ¿La incertidumbre en la estimación de las coordenadas permite al registro de presencia no estar ubicado en la isla?

Por regla general, para la mayoría de análisis se desean datos muy precisos aunque el nivel de precisión pueda variar dependiendo del análisis. GBIF puede ayudar a determinar la precisión de los datos, por ejemplo, mediante filtros y avisos, sin embargo, ¡siempre es necesario hacer doble comprobación!

Avisos e Incidencias

La red de GBIF publica conjuntos de datos, integrándolos en un sistema de acceso común. Aquí los usuarios pueden obtener datos a través de servicios de búsqueda y descarga comunes. Durante el proceso de indexación de los datos en bruto, GBIF incluye incidencias y avisos para los registros con problemas comunes de calidad de datos. Éstos pueden contener información útil para usuarios para crear conjuntos de datos aptos para propósitos específicos.



Las observaciones se muestran en las páginas de los registros individuales para explicar el proceso realizado después de la interpretación:

Excluido - cuando los datos originales no pudieron ser interpretados, por lo cual se excluyen en los campos interpretados.

Alterado - cuando los datos originales son modificados en el proceso de interpretación para ser indexados en GBIF.org.

Inferido - cuando mediante el uso de otra información asociada al registro se puede inferir el dato indexado, en caso de que el original esté vacío.

Excluir todos los registros con un problema en particular no es posible actualmente con la interfaz de búsqueda. Es posible filtrar por tema todos los registros que no interesen seleccionando el tema en particular y haciendo clic en el botón inverso. Sin embargo, la inversión sólo mostrará registros con incidencias y no aquellos que estén libres de problemas. GBIF está trabajando para mejorar este asunto. (en búsqueda de registros de presencia)

Un resumen completo de todas las incidencias y avisos puede encontrarse aquí: <https://data-blog.gbif.org/post/issues-and-flags/>

Herramientas de Procesamiento de los Datos

Mientras que los filtros de GBIF permiten algunos procesamientos de datos, ej.: seleccionar para la descarga sólo aquellos datos que cumplan ciertos criterios, se recomienda realizar procesamientos adicionales de datos para lo que existe una amplia gama de herramientas disponibles. La elección de la herramienta a utilizar estará basada en el acceso personal, la familiaridad y la utilidad de cada una de estas herramientas. Las herramientas que pueden ser utilizadas para este fin son:

- Programas de edición de hojas de cálculo, ej.: Excel, Google Sheets (conjuntos de datos pequeños)
- OpenRefine
- Paquetes y códigos de R, ej.: rgbif, CoordinateCleaner, scrubr y biogeo (procesamiento automático de datos)
- Sistemas de Información Geográfica (GIS), ej.: ArcGIS, QGIS y MapInfo

Siempre es importante incluir un paso de visualización de datos en el procesamiento de datos para poder identificar puntos de datos anómalos que pueden haberse perdido durante la fase de procesamiento. Para aquellos que procesan en R, esto puede hacerse con el mismo R, sin embargo, con OpenRefine o con programas de edición de hojas de cálculo es posible que tenga que utilizar programas de SIG o incluso herramientas como Google Earth para la visualización de los datos.

Manejo de la Incertidumbre Taxonómica

La incertidumbre en torno a la taxonomía de un punto de datos puede surgir por varios motivos:

- Identificación errónea de la especie

- Sinonimia
- Nombres nuevos

Identificación errónea de la especie

La identificación de especies es un proceso complejo, con especies típicamente descritas desde un determinado conjunto de caracteres identificados en una descripción publicada de una especie y enlazadas a un espécimen tipo alojado en una colección científica que se utiliza para validar la identificación de especies. Cuando los taxones son muy similares o un conjunto de rasgos complejos son requeridos para la correcta identificación, pueden requerirse conocimientos taxonómicos específicos que los gestores de datos puede que no posean lo que puede llevar a una identificación errónea de una especie. Como usuarios, debe comprender claramente cómo se hacen las determinaciones taxonómicas para su grupo de interés:

- ¿Cuáles son los caracteres utilizados para definir la especie?
- ¿Son estos caracteres fácilmente confundibles u obtenidos cuando se observa o colecta la especie?
- ¿Existen especies relacionadas que podrían confundirse fácilmente con la especie de interés?

Si cree que existe riesgo de que la especie se haya identificado incorrectamente, puede tomar un enfoque conservador sobre los datos que utiliza y usar solo aquellos datos vinculados a especímenes en colecciones donde la validación taxonómica sería posible y eliminar otras fuentes de datos. Otro enfoque puede ser utilizar datos asociados como por ejemplo información sobre la colecta, medios, secuencias de ADN, etc. para validar la determinación taxonómica.

Sinonimia

La sinónima puede darse cuando la misma especie ha sido descrita varias veces y se da un nuevo nombre a las especies cada vez que se describe, o, cuando se produce un cambio en la taxonomía de una especie, por ejemplo, una especie se traslada de un género a otro. Sólo se puede aceptar un nombre para la especie, y los otros nombres son lo que llamamos sinónimos. Estos sinónimos pueden estar en uso en menor o mayor medida y hay asegurarse al obtener datos de GBIF que éstos se obtienen con el nombre taxonómico que se necesita. El backbone taxonómico de GBIF diferencia entre nombres científicos aceptados y sinónimos, e identificadores únicos en forma de claves taxonómicas. Las búsquedas por especies <https://www.gbif.org/species/search> permiten filtrar nombres aceptados y sinónimos y las claves taxonómicas pueden utilizarse para búsquedas a través de APIs de GBIF.

Claves Taxonómicas Los nombres científicos pueden estar desordenados. Si está accediendo mediante APIs a datos publicados a través de GBIF, y no a través del sitio web, las claves taxonómicas proporcionan una manera efectiva para definir las búsquedas basadas en la taxonomía. Las claves taxonómicas se generan a nivel de especie, género, familia, orden, filo y reino. Los identificadores únicos se generados para los nombres aceptados con sinónimos de aquellos nombres aceptados generados por el mismo identificador. Por lo tanto, puede tener sentido ordenar las especies por sus claves taxonómicas únicas proporcionadas por GBIF durante la indexación del conjunto de datos.

Nombres nuevos

Puede haber casos en los que el nombre científico no coincida con ningún nombre del backbone de GBIF, quizás porque la especie se ha descrito recientemente, o no se encuentra en el listado de especies utilizado por GBIF para construir su backbone. Estos nombres están señalados con el marcador TAXON_MATCH_HIGHERRANK que indica que el nombre científico no ha sido reconocido pero que el registro ha coincidido con un nivel taxonómico superior como por ejemplo el género o la familia. Este marcador puede utilizarse para identificar y filtrar estos datos. Cuando los nombres incluyen errores ortográficos o un formato erróneo, también existe el marcador TAXON_MATCH_FUZZY que puede ser usado para identificar y filtrar nombres que sólo coinciden de manera aproximada con la taxonomía del backbone, siendo una coincidencia no exacta.

Manejo de la Calidad de los Datos

Filtrar los datos permite al usuario obtener los datos más apropiados para un propósito específico. Todas las búsquedas incluyen un conjunto de filtros que pueden ser usados para encontrar los datos que se necesita, y las búsquedas de registros de presencia presentan un conjunto adicional de filtros de búsqueda 'Avanzada' para los usuarios que necesitan hacer un filtrado más avanzado. Mientras que los filtros permiten el filtrado de datos que pueden ser no relevantes, o de menor calidad para sus propósitos, el filtrado adicional puede ser utilizado tanto de forma manual como a través de APIs para abordar problemas adicionales de calidad de los datos que surgen durante el proceso de publicación de datos GBIF. A continuación, se muestran algunos filtros comunes de datos que como usuario podría considerar para hacer los datos más adecuados para su finalidad específica.

Filtros e Incidencias Geoespaciales

Los datos pueden ser filtrados espacialmente en una búsqueda por registros de presencia en una de las siguientes 3 maneras:

- País o área/Continente - los datos son filtrados por país e incluyen datos de la Zona Económica Exclusiva (EEZ, por sus siglas en inglés)
- Área administrativa - este filtro utiliza la base de datos GADM <https://gadm.org/data.html> de áreas administrativas para todos los países del mundo permitiendo a GBIF eliminar incidencias geoespaciales comunes por defecto si se elige obtener datos con una localización.
- Localización - este filtro permite el filtrado de datos con coordenadas y/o dibujar filtros con forma de polígonos o utilizar un archivo GeoJSON para delimitar el filtro de forma. Si se filtran esos datos con coordenadas, se eliminará un buen número de incidencias geoespaciales asociadas con el flujo de publicación de datos. Éstas son:
 - Coordenadas Cero - Las coordenadas son exactamente (0,0) o lo que algunas veces se conoce como "isla nula". Las coordenadas cero-cero es una incidencia geoespacial muy común. GBIF elimina (0,0) cuando hasgeospatialissue es marcado como FALSO.
 - Coordenadas incorrectas de país - Los publicadores de datos a menudo incluyen en GBIF un código de país (US, TW, SE, JP,...). GBIF utiliza el sistema de codificación de dos letras ISO 3166-1 alpha-2 https://en.wikipedia.org/wiki/ISO_3166-1_alpha-2. Cuando un punto no cae en el polígono de un país o de la EEZ pero indica que debería caer en el país, éste será

marcado como "coordenadas incorrectas de país" y será eliminado si los datos son filtrados por localizaciones.

- Coordenadas inválidas - Si GBIF es incapaz de interpretar las coordenadas, ej.: las coordenadas.
- Coordenadas inválidas - Las coordenadas se ubican fuera del rango decimal de valores de latitud/longitud -90,90), (-180,180).

Centroides de país

Los centroides de país hacen referencia al lugar donde la observación se fija en el centro del país en lugar de donde fue observado o colectado el taxón. Los centroides de país normalmente son registros a los que se les otorga un valor de latitud-longitud basado en una descripción textual sobre dónde se localizó el registro original. Los programas de geocodificación utilizan índices de toponimios (nomenclátors), diccionarios geográficos o directorios en conjunto con un mapa o atlas para atribuir coordenadas a los nombres de los lugares. Así pues, si el registro simplemente dice "Brasil", algunos publicadores colocarán el registro en el centro de Brasil. De manera similar, si el registro dice "Texas" o "París", el registro irá en el centro de esas regiones. Esta es una característica casi exclusiva de datos de museo (PRESERVED_SPECIMEN) pero también puede darse con otro tipo de registros.

Actualmente, la identificación de los centroides de país no es posible mediante el uso de los filtros de GBIF, sin embargo, el paquete R CoordinateCleaner se puede utilizar para identificar y filtrar por centroides de país.

Puntos a lo largo del ecuador o meridiano principal

Algunos publicadores consideran que cero y NULL son equivalentes para que los campos vacíos de latitud y longitud de un registro tengan un valor cero. Como resultado, los registros terminan siendo trazados a lo largo del ecuador y las líneas meridianas principales.

Localización incierta

A menudo, queremos estar seguros de que las coordenadas hacen referencia a un lugar determinado y no que éstas están en realidad a miles de kilómetros de donde se observó o colectó el organismo. Hay dos campos - precisión de coordenadas y `coordinateUncertaintyInMeters` - en Darwin Core que obtienes descargando un CSV SIMPLE y que puedes usar para filtrar por "incertidumbre". Sin embargo, estos campos no son muy utilizados por los publicadores ya que consideran que sus registros son bastante precisos (de un GPS) y, por tanto, recomendamos no eliminar del filtrado los valores faltantes.

También hay algunos valores "falsos" de incertidumbre de coordenadas que deben tenerse en cuenta. Estos valores son errores generados por los programas de geocodificación y no representan valores reales de incertidumbre. Estos valores "falsos" son 301, 3036, 999 y 9999. En el caso del valor 301, la incertidumbre es a menudo muchísimo mayor que 301 y en realidad representa un centroide de país.

Conjuntos de datos en cuadrícula

Los conjuntos de datos en cuadrícula son un problema conocido en GBIF. Muchos conjuntos de datos incluyen puntos con una separación idéntica entre ellos que siguen un patrón regular. Estos conjuntos de datos suelen ser encuestas nacionales sistemáticas o datos tomados de algún atlas (los llamados "diseños de colecta rasterizada"). Los registros de presencia referenciados se colocan en un punto central



La mayoría de los publicadores de conjuntos de datos en cuadrícula realmente rellenan una de las siguientes columnas: `coordinateuncertaintyinmeters`, `coordinateprecision`, `footprintwkt`. Así pues, el filtrado en base a estas columnas puede ser una buena manera de eliminar los conjuntos de datos en cuadrícula de nuestra búsqueda. El paquete R `Coordinate cleaner` también incluye una función para eliminar conjuntos de datos en cuadrícula. GBIF tiene una API experimental para identificar conjuntos de datos que muestran cierto grado de "cuadrícula". Puedes leer más sobre el tema aquí: <https://data-blog.gbif.org/post/finding-gridded-datasets/>

Registros de ausencia

De forma predeterminada, los registros de presencia y ausencia se muestran al buscar en www.gbif.org. Los registros de ausencia confirman que una especie no fue encontrada en una localidad específica cuando se muestreó ese área y esta información puede ser útil en, por ejemplo, el desarrollo de modelos de nicho ecológico. Sin embargo, puede darse el caso que solo interese quedarse con los registros de presencia, para lo que se puede filtrar solo en base a ese tipo de registro utilizando el filtro Occurrence Status.

Medios de establecimiento

El término "establishmentMeans" del Darwin Core identifica el proceso mediante el cual los individuos biológicos representados en el registro se establecieron en el lugar. Como tal, puede servir como herramienta de filtrado útil para identificar registros que están fuera del rango nativo de una especie, siendo los términos aceptados para este campo native, nativeReintroduced, introduced, introducedAssistedColonization, vagrant y uncertain. Actualmente, los registros de GBIF se pueden buscar utilizando los términos del vocabulario anterior nativa, introducida, naturalizada, nativa, gestionada e incierto - https://rs.gbif.org/vocabulary/gbif/establishment_means.xml, y estos serán actualizados a finales de 2022. En algunos casos, la eliminación de registros "MANAGED" eliminará registros de zoos.

Utilice este filtro con precaución, ya que como la mayoría de registros no contienen esta información se excluirán de la búsqueda si este filtro está activado. Recomendamos utilizar la información del término Medios de Estabilización para filtrar después de la descarga.

Base del Registro

La base del registro es un término Darwin Core que se refiere a la naturaleza específica del registro y puede referirse a una estas 6 clases:

- Espécimen Vivo - un espécimen que está vivo, por ejemplo, una planta viva en un jardín botánico o un animal vivo en un zoo.
- Espécimen Conservado - un espécimen que ha sido conservado, por ejemplo, una planta en un pliego de herbario o un lote catalogado de peces en un tarro.
- Espécimen Fósil - un espécimen conservado que es un fósil, por ejemplo, un cuerpo fósil, un coprolito, un gastrolito, un ictofósil o un trozo de árbol petrificado.
- Cita de Material - Una referencia a, o cita de, uno, una parte de, o múltiples especímenes en publicaciones científicas, por ejemplo, una cita de un espécimen físico de una colección científica en un tratamiento taxonómico o en una publicación científica o un registro de presencia mencionado en un cuaderno de campo.
- Observación Humana - el resultado de un proceso de observación humana, ej.: evidencia de un registro tomado de notas de campo o literatura, o un registro sin evidencia física ni tomada por una máquina.
- Observación realizada por una Máquina - el resultado de un proceso de observación de una máquina como por ejemplo una fotografía, un vídeo, una grabación de audio, una imagen de control remoto o un registro de presencia basado en telemetría.

La base del registro debería permitir a los usuarios filtrar los individuos de colecciones ex-situ tales como zoos y jardines botánicos o los fósiles, así como filtrar aquellos registros en función de si son un espécimen o una observación, que pueden estar sujetos a validación taxonómica. Hay que tener en cuenta que, aunque este puede ser un filtro útil, los publicadores de datos no siempre rellenan correctamente el campo base del registro, o, puede haber matices en los datos que pueden no ser tan obvios un usuario, ej.: <https://data-blog.gbif.org/post/living-specimen-to-preserved-specimen-understanding-basis-of-record/> y siempre se deberá hacer una doble comprobación de los datos antes de su uso.

Registros Antiguos

GBIF tiene muchos registros de museos que podrían ser más antiguos que lo deseado para algunos estudios.

Duplicados

La duplicación de registros puede ocurrir cuando se hacen varios registros para un mismo individuo. Esto puede ocurrir, por ejemplo, cuando un investigador deposita varios especímenes del individuo de un árbol en herbarios de todo el mundo que a su vez publican esos datos en GBIF, o cuando un individuo ha sido depositado en una colección de historia natural y el individuo también fue muestreado por su ADN. En este caso, habrá un registro del espécimen en las colecciones y otro para su secuencia de ADN.

GBIF ha introducido recientemente una función de agrupación en su búsqueda avanzada que permite a los usuarios identificar grupos de registros, p. ej. registros que parecen proceder de la misma fuente. Esto permite a los usuarios identificar posibles datos duplicados y filtrarlos para excluirlos de la descarga. Tenga en cuenta que si filtra los registros que están en un grupo, perderá todos los registros encontrados en ese grupo y perderá datos potencialmente útiles. El filtro puede usarse mejor para indicar hasta qué punto hay duplicidad en el conjunto de datos, o para descargas independientes de los conjuntos de datos en grupo y sin agrupar a modo de comparación.

Unresolved directive in index.es.adoc - include::R_Data_Processing_Pipeline.es.adoc[]

Unresolved directive in index.es.adoc - include::Ecological-Niche-Modelling.es.adoc[]

Unresolved directive in index.es.adoc - include::Use_Case_ENM.es.adoc[]

Unresolved directive in index.es.adoc - include::Red-List-Assessments.es.adoc[]

Unresolved directive in index.es.adoc - include::Use_Case_Conservation_Status.es.adoc[]

Documentación clave



Las siguientes referencias proporcionan más detalles sobre tratados en este curso. Todos los enlaces se abren en una nueva ventana/pestaña.

API

- [Guía para principiantes sobre APIs de GBIF](#)
- [\(Casi\) todo lo que necesitas saber sobre las APIs de especies de GBIF](#)

Computación en la Nube

- [Tutorial sobre GBIF y Apache-Spark en Microsoft Azure](#)
- [Tutorial sobre GBIF y Apache-Spark en AWS](#)

Darwin Core

- [Términos Darwin Core: Una guía rápida de referencia](#)
- [DarwinCore Simple](#)
- [Preguntas & Respuestas sobre Darwin Core](#)
- [Extensiones Darwin Core registradas con GBIF](#)

Publicación de datos

- [Guía rápida para publicar datos a través de GBIF](#)
- [Cómo publicar datos de biodiversidad a través de GBIF.org](#)
- [Cómo publicar datos de biodiversidad a través de GBIF.org](#)
- [Buenas Prácticas en la Publicación de Datos de Biodiversidad de Evaluaciones de Impacto Ambiental + Secretariado de GBIF & IAIA: International Association for Impact Assessment \(2020\).](#)
- [Guía para que las empresas privadas se conviertan en publicadores de datos a través de GBIF: Documento de plantilla para apoyar el proceso interno de autorización para convertirse en un publicador de GBIF + Rui Figueira, Pedro Beja, Cristina Villaverde, Miguel Vega, Katia Cezón, Tainan Messina, Anne-Sophie Archambeau, Rukaya Johaadien, Dag Endresen & Dairo Escobar \(2020\).](#)
- [Publishing DNA-derived data through biodiversity data platforms + Anders F. Andersson, Andrew Bissett, Anders G. Finstad, Frode Fossøy, Marie Grosjean, Michael Hope, Thomas S. Jeppesen, Urmas Kõljalg, Daniel Lundin, R. Henrik Nilsson, Maria Prager, Cecilie Svenningsen y Dmitry Schigel \(2020\).](#)
- [Tipos de juegos de datos soportados por GBIF](#)
- [Requisitos de calidad de datos de GBIF para su publicación](#)
- [licencia de datos de GBIF](#)
- [Modelos de listas de especies centrales](#)
- [Modelos de Registros de Presencia centrales](#)

- Modelos de eventos de muestreo centrales
- Buenas prácticas con datos de eventos de muestreo
- Compartiendo imágenes, sonidos y videos en GBIF
- Artículos de Datos
- Artículos de datos publicados

Publicación de datos: IPT

- The GBIF Integrated Publishing Toolkit: Facilitating the Efficient Publishing of Biodiversity Data on the Internet + Robertson et al. (2014)
- Instalar el IPT o no instalarlo
- Centros de alojamiento de datos del IPT
- Webinar sobre la Instalación del IPT
- Vídeo sobre la instalación del IPT
- Vídeo tutorial sobre el uso del IPT

Digitalización

- Recursos de Digitalización de iDigBio
- Flujos de Trabajo en la Digitalización de Colecciones de iDigBio
- Protocolos y Flujos de Trabajo en la Digitalización de Colecciones de iDigBio
- Guía de captación de imágenes de especímenes de iDigBio
- Guía en 10 pasos de Canadensys para gestionar imágenes con tus datos de biodiversidad

GBIF

- Qué es GBIF
- Plan Estratégico
- Ser miembro de GBIF
- Revisión Científica
- Establecimiento de un nodo participante de GBIF efectivo: conceptos y consideraciones generales + Secretariado de GBIF (2019).

Georreferenciación

- Georeferencing Best Practices + Arthur D. Chapman & John R. Wieczorek (2020).
- Georeferencing Quick Reference Guide + Paula F. Zermoglio, Arthur D. Chapman, John R. Wieczorek, Maria Celeste Luna & David A. Bloom (2020).

- [Georeferencing Calculator Manual](#) + David A. Bloom, John R. Wieczorek & Paula F. Zermoglio (2020).
- [Recursos de georreferenciación](#)

Especies Invasoras

- [GRISS - Registro Mundial de Especies Invasoras e Introducidas](#)
- [TrIAS - Monitoreo de Especies Exóticas Invasoras](#)

Living Atlases

- [Living Atlases](#)
- [Documentación técnica clave de ALA](#)

Misceláneo

- [Guía de VertNet para abrir archivos de texto en Excel](#)
- [Guía de VertNet para las licencias de datos](#)

OpenRefine

- [Documentación OpenRefine](#)
- [Expresiones comunes de OpenRefine](#)
- [Guía para la limpieza de datos sobre biodiversidad con OpenRefine](#) + Paula F. Zermoglio, Camila A. Plata Corredor, John R. Wieczorek, Ricardo Ortiz Gallego & Leonardo Buitrago (2021).
- [Using Google Refine and taxonomic databases \(EOL, NCBI, uBio, WORMS\) to clean messy data](#) + iPhylo blog post. Rod Page 2012.
- [Reconciling author names using Open Refine and VIAF](#) + iPhylo blog post. Rod Page 2013.
- [Validating scientific names with the GBIF Portal web service API](#) + Guest post fue escrito por Gaurav Vaidya, Victoria Tersigni y Robert Guralnick 2013.
- [iDigBio Cleaning data with OpenRefine](#) + iDigBio 2013.
- [Have We Got the Names "Right"?](#) + Canadensys 2014.
- [Cleaning data with OpenRefine](#) + Desmet y Brosens 2016 TDWG.
- [EasyOpen Redlist](#) + Consultar la Lista Roja de IUCN, usando una lista de especies, OpenRefine, y algún código preescrito. Olly Griffin Julio 2019.

Planificación/Colaboración

- [Agile](#)

- [Qué es SCRUM](#)
- [Marco de trabajo de SCRUM](#)
- [Metodología Kanban](#)
- [Guía Scrum](#)
- [GitHub](#)

Red List Assessments

- **IUCN Red List Categories and Criteria**

- [IUCN Red List Categories and Criteria v3.1 - ENGLISH](#)
- [Categorías y Criterios de la Lista Roja de la UICN v3.1 - ESPAÑOL](#)
- [Catégories et Critères de la Liste Rouge de l'UICN v3.1 - FRANÇAIS](#)

- **Red List Guidelines**

- [Guidelines for Using the Red List Categories and Criteria version 15 - ENGLISH](#)
- [Directrices de uso de las Categorías y Criterios de la Lista Roja de la UICN Versión 14 - ESPAÑOL](#)
- [Lignes directrices pour l'utilisation des Catégories et Critères de la Liste rouge de l'UICN Version 14 - Français](#)

- **Criteria Summary Sheet**

- [Criteria summary sheet - ENGLISH](#)
- [Resumen de los Criterios - ESPAÑOL](#)
- [Résumé des Critères - FRANÇAIS](#)

- **Mapping Standards**

- [Mapping Standards - ENGLISH](#)
- [Standard Attributes for Spatial Data - ENGLISH](#)
- [Attribute codes for Presence](#)

- **Regional and National Levels Guidelines**

- [Guidelines for application of IUCN Red List Criteria at Regional and National Levels Version 4 - ENGLISH](#)
- [Directrices para el uso de los Criterios de la Lista Roja de la UICN a nivel regional y nacional Versión 4 - ESPAÑOL](#)
- [Lignes directrices pour l'application des Critères de la Liste rouge de l'UICN aux niveaux régional et national Version 4. - FRANÇAIS](#)

- [Spatial Tools and Data for Red List Assessments](#) + Includes brief overviews of some tools developed to aid mapping of spatial data and estimation of Red List metrics such as EOO and AOO, plus developer information, links to help files, support networks and associated research publications.

- [Supporting information guidelines](#) + The Documentation Standards and Consistency Checks for IUCN Red List Assessments and Species Accounts (also known as the Supporting Information Guidelines) provides guidance on the required and recommended supporting information for Red List assessments. It also provides guidance on the writing style and format that should be used for all IUCN Red List assessments.

Calidad

- [Principios de Calidad de Datos](#) + Arthur Chapman 2005.
- [Principles and Methods of Data Cleaning: Primary Species and Species-Occurrence Data](#) + Arthur Chapman 2005.
- [Be careful with dates in Excel](#) + DataOne 2014.
- [Codificación de caracteres para principiantes](#)
- [Guide MVZ para Registrar Localidades en Notas de Campo](#)

R

- [Data Carpentry-Data Analysis and Visualization in R for Ecologists](#)
- [Datacamp- Range of courses in R, Python and SQL](#)
- [Introducción al Manual de R](#)
- [Manual sobre rgbif](#)
- [Manual sobre CoordinateCleaner](#)
- [Descarga de registros desde una lista larga de especies en R y Python](#)
- [Cosas comunes a tener en cuenta cuando procesemos datos de GBIF tras la descarga](#)
- [Encontrar Conjuntos de Datos en Cuadrícula & Cargar Conjuntos de Datos en Cuadrícula](#)
- [Centroides de país](#)
- [Utilizar shapefiles en datos de GBIF con R](#)
- [No es una descarga de aves](#)

Especies sensibles

- [Current Best Practices for Generalizing Sensitive Species Occurrence Data](#) + Arthur D. Chapman 2020.

Taxonomía

- [Conjuntos de datos de listados de especies de GBIF y vacío de datos](#)
- [GBIF Labs - Names Parser](#)
- [GBIF Labs - Species Matching](#)

- [Global Names Resolver](#)
- [Estrategia de Asignación de Nombres Marinos para controles de calidad de taxonomía](#)
- [Nomenmatch](#)

Glosario

ALA

Atlas of Living Australia. El nodo australiano de GBIF, que desarrolló un portal de datos de código abierto ampliamente utilizado en la actualidad por la comunidad y socios de GBIF para portales nacionales de biodiversidad.

API

Interfaz de programación de aplicaciones. Un conjunto de métodos de comunicación claramente definidos entre varios componentes de software.

BID

Información sobre Biodiversidad para el Desarrollo. Un proyecto financiado por la UE y coordinado por GBIF cuyo objetivo es aumentar la capacidad de movilización de datos en las regiones de África, el Caribe y el Pacífico.

BIFA

Fondo para la Diversidad Biológica de Asia.

Licencias CC

Creative Commons. Se trata de una serie de licencias establecidas por la organización Creative Commons que permiten compartir y reutilizar la creatividad y el conocimiento mediante la provisión de herramientas legales gratuitas. Tres de ellas se pueden asignar a conjuntos de datos compartidos por GBIF: CC0, CC BY y CC BY-NC.

Vocabulario controlado

Se trata de un conjunto restringido de términos que se utilizan como valores posibles para un campo determinado. Uno puede pensar en ello como una lista de búsqueda o un menú desplegable para un campo en particular. Por ejemplo, el campo DwC basisOfRecord solo debe contener uno de estos valores: "PreservedSpecimen", "FossilSpecimen", "LivingSpecimen", "HumanObservation", "MachineObservation". Diríamos que la lista de valores es un vocabulario controlado para ese campo.

DwC

Darwin Core es un estándar de datos de biodiversidad, mantenido por TDWG y ampliamente utilizado dentro de la comunidad y socios de GBIF. Es un conjunto de términos estandarizados (o nombres de campo) y sus definiciones, que se utilizan para compartir información sobre biodiversidad.

DOI

Digital Object Identifier. Identificador persistente o resolutor que se utiliza para identificar objetos de forma única. Los DOI se utilizan principalmente para identificar información académica,

profesional y gubernamental, como artículos de revistas, informes de investigación y conjuntos de datos, y publicaciones oficiales.

DwC-A

Archivo Darwin Core. Un archivo comprimido (zip) que contiene toda la información necesaria para compartir con GBIF, para un recurso en particular. Cada zip contiene tres tipos de archivos:

1. los datos reales, en uno o más archivos de texto: occurrence.txt/event.txt/measurmentoffact.txt, etc.
2. un archivo de mapeo: rtf.xml
3. un archivo de metadatos (EML): eml.xml Cuando publica utilizando el IPT, crea un Archivo de Darwin Core, que se comparte con GBIF. Además, cuando descarga datos del sitio web de GBIF, también puede elegir un formato DwC-A.

GUID

Identificador Global Único

IPT

Integrated Publishing Toolkit. Es una aplicación web (software) gratuita y de código abierto para publicar datos de biodiversidad. El software en sí se aloja en un servidor (ya sea en su institución o en otro lugar) que debe tener acceso a Internet las 24 horas del día, los 7 días de la semana. Se utiliza para crear y gestionar archivos de Darwin Core que pueden ser compartidos y utilizados por cualquier persona, incluido GBIF.

Préstamo

En el contexto de las colecciones de historia natural, este es el procedimiento de préstamo de especímenes entre instituciones.

LSID

Life Sciences Identifier. Son identificadores persistentes, globales y únicos para objetos biológicos.

Publicación de datos

Con respecto a GBIF, tenemos una definición muy específica de publicación de datos. Se refiere a lograr que los conjuntos de datos de biodiversidad sean accesibles y reconocibles públicamente, en una forma estandarizada, a través de un punto de acceso, generalmente una dirección web (una URL).

Recurso

Un Recurso es el término colectivo utilizado para referirse a un conjunto de datos en particular y sus metadatos una vez que se ha cargado a una infraestructura de IPT.

TDWG

Taxonomic Databases Working Group, ahora renombrado Biodiversity Information Standards.

URN

Número de Recurso Uniforme

UUID

Identificador Universalmente Único

Agradecimientos

Diseño del curso e instrucción

El éxito de este curso depende en gran medida del apoyo brindado a los participantes de la red de mentores de mejora de capacidad de GBIF. Visite la página de GBIF en [tutorización sobre mejora de capacidad](#) para leer más sobre estas personas y sus contribuciones.

Los siguientes individuos son reconocidos por sus contribuciones significativas al diseño, materiales e instrucción del curso:

- Nadine Bowles Newark
- Andrea Baquero
- Kate Ingenloff
- Hannah Owens
- Mélianie Raymond
- Andrew Rodrigues
- Laura Anne Russell
- John Tayleur
- John Waller

Agradecimientos especiales...

Traductores

Español

- Anabela Plos
- Paula Zermoglio

Francés

- Patricia Mergen
- Anne Sophie Archambeau

Recursos

- R
- [Wallace Ecological Niche Modelling Vignette](#)
- [Lista Roja de Especies Amenazadas de la UICN](#)

Colofón

Cita sugerida

Curso de Movilización de Datos de Biodiversidad GBIF (2021) GBIF. 12ª edición. Secretariado de GBIF. <https://doi.org/10.15468/ce-wkk4-2w26>. [Fecha de curso.]

Colaboradores

El *Curso de Uso de Datos sobre Biodiversidad de GBIF* se desarrolló originalmente como parte de **Información de Biodiversidad para el Desarrollo**, un programa financiado por la **Unión Europea**. El plan de estudios original fue creado por Andrew Rodrigues, Hannah Owens y John Tayleur, con contribuciones adicionales de formadores, mentores y estudiantes de GBIF.

Licencia

Curso de Uso de Datos sobre Biodiversidad de GBIF posee la licencia **Creative Commons Atribución 4.0**.

URI persistente

<https://doi.org/10.15468/ce-wkk4-2w26>

Control de documentos

Quinta edición, noviembre 2021

Imagen de la portada

Di Marco M, Ferrier S, Harwood TD, Hoskins AJ and Watson JEM (2019) Wilderness areas halve the extinction risk of terrestrial biodiversity. *Nature*. Springer Science and Business Media LLC 573(7775): 582–585. Available at: <https://doi.org/10.1038/s41586-019-1567-7>