
PROTOCOLO Y PROCEDIMIENTOS PARA LA GESTIÓN DE BASES DE DATOS DE BIODIVERSIDAD, ENFOCADO EN COLECCIONES BOTÁNICAS

**Santiago, Chile
2020**



PROTOCOLO Y PROCEDIMIENTOS PARA LA GESTIÓN DE BASES DE DATOS DE BIODIVERSIDAD, ENFOCADO EN COLECCIONES BOTÁNICAS



AUTORES

Victoria Arévalo^{1,2}, Antonia Serey^{1,2}

Paulette I. Naulin², Leisy Amaya³

1 Consorcio Patrimonio Botánico de la Universidad de Chile (PBUCH)

2 Laboratorio de Biología de Plantas, Facultad de Ciencias Forestales y de la Conservación de la Naturaleza, Universidad de Chile

3 Nodo Nacional de Información en Biodiversidad (GBIF Chile), Ministerio del Medio Ambiente.

Citar como:

Arévalo, V., Serey, A., Naulin, P.I., y Amaya, L. 2020. Protocolo y Procedimientos para la Gestión de Bases de Datos de Biodiversidad, enfocado en Colecciones Botánicas. Consorcio Patrimonio Botánico de la Universidad de Chile. Santiago. Primera versión. pp. 70.

ÍNDICE DE CONTENIDOS

1	INTRODUCCIÓN.....	1
2	ANTECEDENTES	3
2.1	Gestión de herbarios	3
2.2	Gestión de información de colecciones botánicas a través de bases de datos	5
2.3	Global Biodiversity Information Facility (GBIF)	6
2.4	Estándar de biodiversidad para compartir datos de biodiversidad: El Estándar DwC	7
2.5	Tipos y fuentes de datos de biodiversidad.....	8
2.5.1	Tipos de datos.....	8
2.5.2	Fuentes de datos	10
2.6	Calidad de datos.....	10
2.6.1	Principios fundamentales para la calidad de datos.....	11
3	ESTRUCTURACIÓN DE CAMPOS AL ESTÁNDAR DARWIN CORE.....	14
3.1	Estructuración de datos al Estándar darwinCore	14
3.2	Mapeo de Campos	14
3.2.1	Mapeo directo.....	14
3.2.2	Mapeo Indirecto	14
4	HERRAMIENTAS DE GESTIÓN PARA EL MANEJO Y CALIDAD DE BASES ...	16
4.1	Herramientas para el tratamiento de nombres científicos	16
4.1.1	Herramientas de normalización de nombres científicos	16
4.1.2	Catálogos taxonómicos globales en línea para el chequeo taxonómico.....	23
4.2	Herramientas para el tratamiento de coordenadas	24
4.2.1	Conversión de coordenadas.....	24
4.3	Herramientas para el tratamiento de fechas.....	28
4.3.1	Canadensys Date Parsing	28
4.4	Herramientas para corrección de formato	32
4.4.1	Open Refine.....	32
4.5	Configuración de Microsoft Excel para mejorar la gestión de datos	41
4.5.1	Funciones de Excel para la gestión de datos	41
4.5.2	Importar archivos de texto en Excel.....	42
5	HERRAMIENTAS DE PUBLICACION VÍA GBIF: IPT	48
5.1	Proceso de publicación	48

5.2	Carga de datos	48
5.3	Elaboración de metadatos	54
5.3.1	Procedimiento para la elaboración y carga de metadatos	54
5.3.2	Tipos de metadatos	55
5.4	Creación del Archivo Darwin Core	61
5.4.1	Versiones publicadas	61
5.4.2	Estado de la publicación.....	62
6	REFERENCIAS BIBLIOGRAFÍA	64

1 INTRODUCCIÓN

El concepto de Biodiversidad o diversidad biológica hace referencia a la variedad de formas de vida en la tierra (UNESCO, 2017), incluyendo desde variaciones y características genéticas a ecosistemas y paisajes, considerando las interacciones entre sí, y con el medio a distintos niveles de organización (SCDB, 2006; CONAMA, 2008). Cada nivel posee tres componentes: composición, estructura y función (Noss, 1990). La biodiversidad provee numerosos servicios ecosistémicos a la humanidad.

Las personas más afectadas por la pérdida de biodiversidad son las que dependen en mayor medida de los Servicios Ecosistémicos, de la red de seguridad que les brinda la biodiversidad en ecosistemas naturales y de la relación con estos, sin ser capaces de sustituir los bienes y servicios que pierde el ecosistema (Díaz *et al.*, 2006). Estas personas suelen tener poca influencia económico-política. Es por ello, que cuando el desarrollo económico no considera los efectos en estos servicios puede disminuir su calidad de vida, aún si otros segmentos de la sociedad se benefician (Cerdeira y Bidegain, 2018).

Si bien a nivel global se ha mejorado el análisis e interpretación de datos reunidos, hace falta mejorar la coordinación a fin de garantizar las existencias de modelos y tecnologías que puedan integrar estos conocimientos en sistemas aplicados funcionales (SCDB, 2014). Es por ello que, los datos de biodiversidad existentes deben ser visibilizados y compartidos, estandarizando la forma de ingreso hacia futuras fuentes, facilitando el proceso de hacerlos accesibles de forma libre a la red, para el desarrollo de conocimiento y tomas de decisiones mejor fundamentadas.

La protección de las especies terrestres y marinas es fundamental, por lo que es necesario tener conocimiento de ellas. En marco de esto los repositorios biológicos de colecciones científicas como por ejemplo: los herbarios son fundamentales porque se encargan de almacenar datos botánicos para estudios tradicionales de taxonomía y sistemática, mas también para: ecología, bioingeniería, conservación, seguridad alimentaria y elementos humanos sociales y culturales (James *et al.*, 2018). Los datos contenidos en herbarios son referencias históricas y confiables, necesarios para construir modelos predictivos sólidos en respuesta al cambio global a nivel de taxón o grupo funcional (Willis *et al.*, 2017). Además, contienen información valiosa para el estudio y entendimiento de la flora y sus cambios en el tiempo.

Los Datos de Biodiversidad (DB) son registros en forma de texto digital o registro multimedia que detalla los hechos acerca de la presencia de un organismo biológico en un lugar y tiempo determinado, a través de fuentes de datos primarias (Amaya, 2018). Las colecciones de herbario se pueden transformar en DB a través de una serie de procesos de homologación de conceptos botánicos específicos a un lenguaje más general y estructurado para ser compartidos.

Los desafíos principales que han presentado en los DB son, la falta de un sistema coordinado de publicación e integración, la falta de estándares para protocolos de datos y de comunicación, llevando a que los repositorios que los contienen tiendan a aislarse entre sí (Wieczorek *et al.*, 2012). Como una alternativa a mejorar estas dificultades, en 1999 surge “*Global Biodiversity Information Facility*” o GBIF, esta organización intergubernamental es una red internacional de infraestructura de investigación destinada a proporcionar acceso abierto a datos sobre biodiversidad a cualquier persona, en cualquier lugar (GBIF, 2018a).

En Chile se establece un nodo GBIF (Nodo GBIF Chile) en el año 2009. Este busca fomentar la movilización y el libre acceso a datos de biodiversidad, aportando herramientas para la administración de datos biológicos y mejorar su calidad, acceso y usos (GBIF, 2018b) facilitando la transformación de datos de colecciones biológicas a DB.

Parte importante de la transformación de datos de colecciones biológicas a DB, es transformarlas a un lenguaje tipo, con el fin de generar entendimiento entre las partes que publican información y sus posibles usuarios. Tanto GBIF, como otras plataformas usan el estándar darwinCore, cuyo objetivo es crear un lenguaje común para publicar, compartir y documentar datos sobre registros biológicos, facilitando la interoperabilidad de las bases de datos de biodiversidad a nivel mundial (MMA, 2017).

La creación de este lenguaje es una oportunidad para la transformación de datos de colecciones de biodiversidad a DB teniendo la particularidad de poder organizar la información obtenida de distintas prácticas de curación de datos de historia natural, que se han desarrollado local y orgánicamente durante cientos de años en un núcleo formado por 185 campos (James *et al.*, 2018) permitiendo que esta información se pueda comparar y utilizar.

El protocolo tiene como finalidad entregar las herramientas necesarias para que instituciones que albergan datos de biodiversidad, puedan compartirlos de forma abierta con la comunidad, de esta manera se ayuda a visibilizar colecciones e instituciones, además da opción a ser un aporte al conocimiento ciudadano y científico global, permitiendo ayudar a la toma de decisiones con los distintos usos que se le puedan dar a estos datos, a través de la plataforma GBIF. Además incluye desde cómo organizar la información para crear una base de datos, llevarla a un lenguaje estandarizado para que de esta forma pueda ser comprendida por quien revise o requiera la información, dando alternativas para mejorar la calidad de sus datos a través de herramientas de libre acceso.

2 ANTECEDENTES

2.1 Gestión de herbarios

La importancia de los herbarios como una fuente de datos de primera mano es muy significativo, ya que no solo habla de la flora de algún lugar sino también del tipo de ambiente, la comunidad en la que pertenece, la relación con la fauna y los usos antrópicos (Yost *et al.*, 2018). Además, permiten mantener la variabilidad genética de las especies, para su posterior estudio como por ejemplo, la caracterización morfológica y molecular, siendo un gran potencial como una fuente de datos para la documentación de la fenología de las plantas (Willis *et al.*, 2017).

En la actualidad se están efectuando esfuerzos en la digitalización de los datos y así poder hacerlos accesibles para su análisis. La digitalización se debe realizar a través de bases de datos de biodiversidad, encontrándose éstas en varios formatos de distintas plataformas y programas. Un paso importante hacia la comprensión de bases de datos es poder proporcionar una visión estandarizada de las fuentes de datos heterogéneas (Wieczorek *et al.*, 2012). Estas plataformas admiten que la información científica, que en ocasiones es escasa sea, de suma importancia para la política pública de un país (SIB-Colombia, 2018).

Existen varios países y continentes que comparten sus colecciones biológicas tanto de herbarios como de fauna, de forma libre como el Museo de Historia Natural de París-Francia, la plataforma SIB-Colombia, iDigBio de EE. UU, *Biodiversity Information System for Europe* (BISE) de Europa y Banco de Datos de la Naturaleza de España, entre otras.

La documentación y creación de los datos de herbarios son posibles gracias al trabajo del taxónomo vegetal, cuyo principal trabajo es identificar y clasificar a los seres vivos y colocarlos en los “contenedores” correctos. En la actualidad, los taxónomos no solamente describen especies, sino que también completan las descripciones realizadas en el pasado, aplicando técnicas como microscopía electrónica, rayos X y microfotografía (Ortigosa, 2017). Actualmente la mayoría de los taxónomos de tiempo completo trabajan en centros de patrimonio natural en donde distribuyen su tiempo entre labores de curación, que consisten en mantener en buenas condiciones las muestras tanto en su calidad de almacenamiento (cuidados técnicos) como de información taxonómica (correcta identificación) y actualización de especímenes, por último, etiquetar e identificar, describir y escribir acerca de nuevas especies (Villaseñor, 2015). Lamentablemente, en muchos casos, la información que se incorpora a las bases de datos es de dudosa confianza para ser usada en los análisis de la biodiversidad. Es evidente que mientras mejor curada está una colección, más importante será su participación en estas nuevas líneas emergentes de investigación (Villaseñor, 2015).

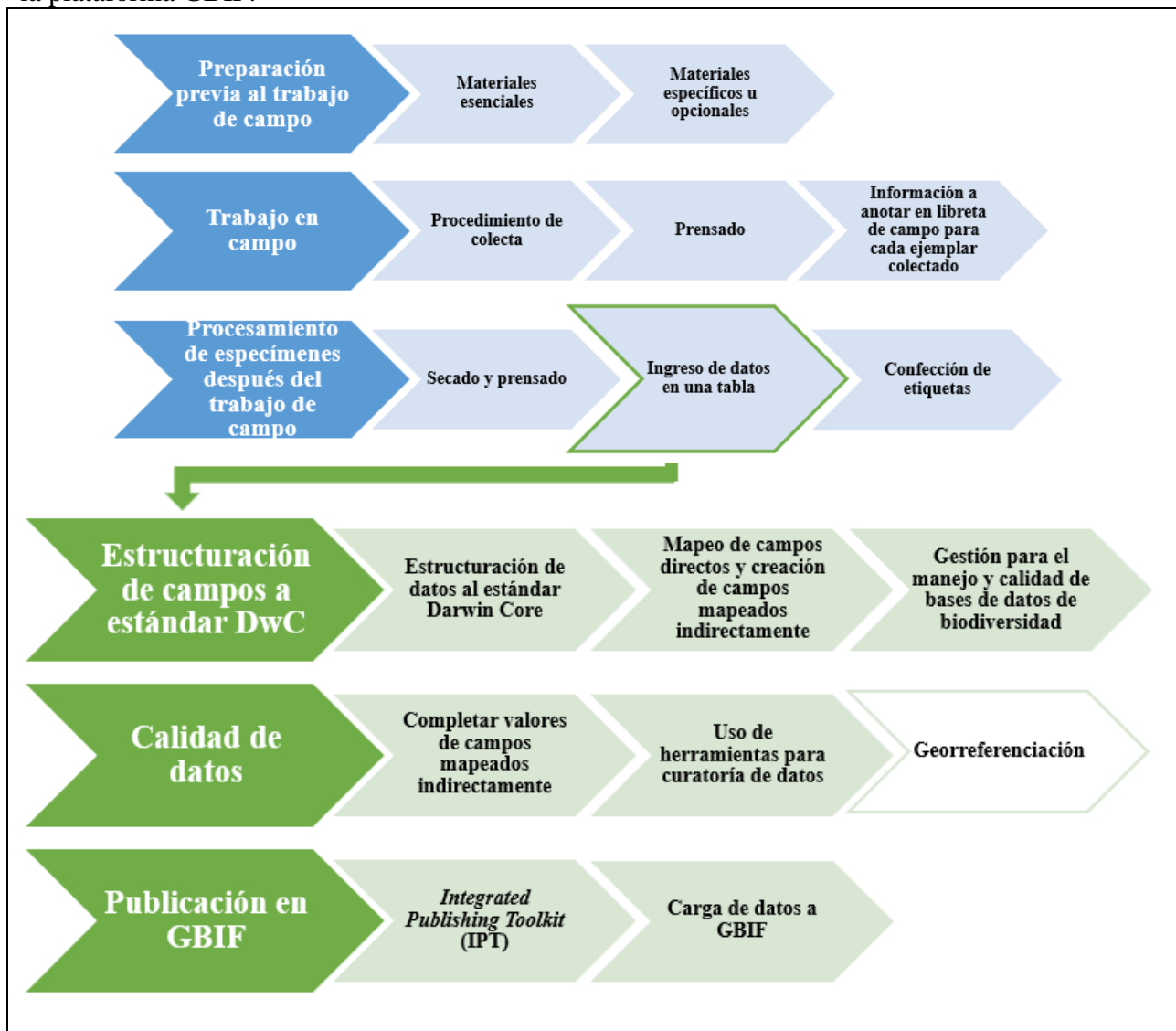
La información descrita sobre un ejemplar en su pliego de herbario empieza con el trabajo de colecta, teniendo claridad sobre su objetivo, como serán tomados los datos en terreno y la forma en que se traspasarán a una base de datos. Para así llevarlos a un estándar y compartirlos de manera que otro usuario pueda entender la información de la misma forma que el colector la plasmó en la etiqueta.

Si sabemos cómo describir los datos desde la colecta, llevarlos al estándar y compartirlos, esto implicará un costo mucho menor en cuanto al trabajo y tiempo posterior que se deberá

proporcionar en la curatoría de bases de datos. Pero gran parte de las colecciones biológicas no solo tienen registros actuales, donde se pueda influir en como tomarán los datos, porque estos ya fueron tomados y depositados en la colección. Estos datos registrados por colectores en el pasado se deben manejar para actualizarlos y llevarlos a un estándar con el fin de compartirlos con la comunidad en la actualidad.

Cuando se habla de estandarización, se refiere a tener un formato único de ingreso para un campo en una base de datos. Por ejemplo, para compartir una fecha, el estándar darwinCore pide que su formato sea: AAAA-MM-DD, si en la base de datos está escrito “15 de Jul 2019” habrá que llevarlo al estándar “2019-07-15”. Mientras que si desde la colecta se escribe el dato como se describe en el estándar, al llevarlo a la base de datos quedará estandarizado directamente. En la Figura 1, se muestra el flujo del proceso de colecta con la curatoría y publicación de datos.

Figura 1. Flujo de información entre la pre-colecta de especímenes y su posterior carga de datos a la plataforma GBIF.



Fuente: Elaboración propia.

2.2 Gestión de información de colecciones botánicas a través de bases de datos de biodiversidad

Desde hace muchos años se han realizado estudios que describen la diversidad de formas de vida de un territorio, aunque en los últimos tiempos se ha incrementado de forma exponencial su número y precisión. La información que recogen estos estudios ha sido archivada en diversos formatos bibliográficos. La posibilidad de utilizar dicha información en estudios actuales tiene un claro interés, no sólo para analizar la evolución temporal de las formas de vida, sino para comprender cómo y porqué se han producido cambios, y para conocer la naturaleza lo más exactamente posible. Su compilación en forma de base de datos proporciona la posibilidad de un acceso rápido y completo a la información que poseen (Martín, 2016).

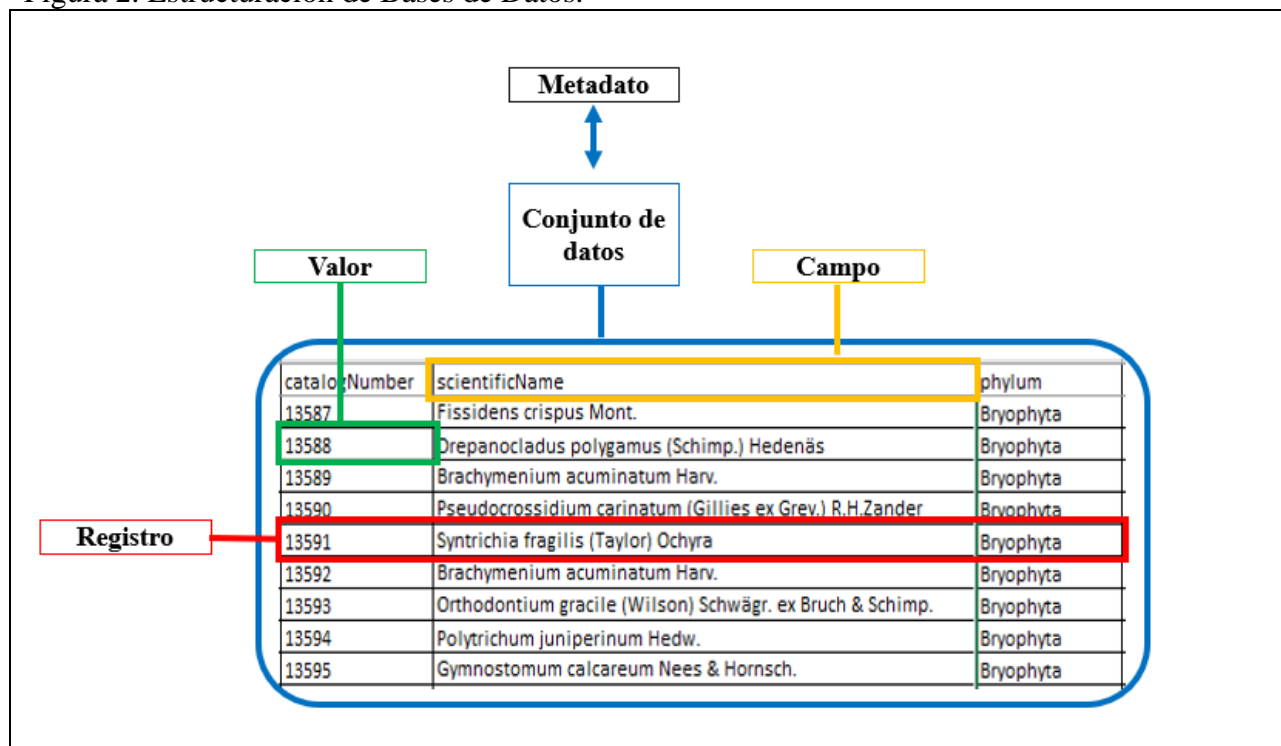
En el mundo existen plataformas que albergan bases de datos de biodiversidad, que tienen por función la recopilación de información sobre numerosos grupos biológicos. Son capaces de proporcionar información del medio biofísico, pues constituye el soporte sobre la que se desarrolla la biodiversidad, siendo imprescindible para comprender su composición y dinámica (Martin, 2016).

Una gran cantidad de iniciativas internacionales, han permitido que repositorios que almacenan colecciones de diversidad compartan sus registros como datos de biodiversidad, siendo posible consultarlos a través de internet. Varios países y continentes comparten sus colecciones biológicas de herbarios o zoológicas, de forma libre, por ejemplo: el Museo de Historia Natural de París-Francia, la plataforma SIB-Colombia, iDigBio y MyCoPortal de EE. UU, las cuales son bases de datos de biodiversidad. Por otro lado, existen las redes de biodiversidad que trabajan en conjunto con instituciones de otros países, como: GBIF, e-Birds, iNaturalist, *Biodiversity Information System for Europe* (BISE) de Europa, Banco de Datos de la Naturaleza de España. La unión de distintas bases de datos dibuja una imagen mucho más completa del conjunto de la biodiversidad de una determinada área o un determinado grupo taxonómico, complementando la información de las otras (González, 2006).

Las bases de datos de biodiversidad (BDB), son la integración de datos de la biodiversidad del territorio marino o terrestre de un país, región o lugar geográfico, en herramientas tecnológicas al servicio de la toma de decisiones, que tienen por objetivo la recopilación de la mayor cantidad de información de las especies presentes. Estas BDB pueden provenir de colecciones, bibliografías, trabajos de campo, etc. La elaboración de BDB requiere de todo un esfuerzo de planificación, ejecución y mantenimiento, cuyo resultado en ocasiones es difícil de ver.

En síntesis las BDB están compuestas por dos partes, los metadatos para saber de qué tratan los datos y un conjunto de datos que corresponde a una sumatoria de registros, a su vez estos registros están formados por valores relativos a los campos descritos (Figura 2).

Figura 2. Estructuración de Bases de Datos.



Fuente: Elaboración propia.

2.3 Global Biodiversity Information Facility (GBIF)

En la actualidad hay plataformas que admiten que la información de biodiversidad de los países se pueda destinar hacia distintas fuentes, como lo hace GBIF. Esta es una red internacional e infraestructura de investigación, nacida en el año 2001 y financiada por los gobiernos del mundo. Destinada a proporcionar a cualquier persona, acceso abierto a datos de todos los tipos de vida en el mundo (GBIF, 2018b).

Los miembros de GBIF son 54 países y 36 organizaciones internacionales, que firman un Memorando de Entendimiento en el que se comprometen, entre otras cosas, a instaurar un nodo de datos desde el cual harán disponible la información de biodiversidad. GBIF trabaja a través de nodos participantes, donde Chile forma parte de GBIF desde el 2009 y anexo desde el 2012, donde el Ministerio del Medio Ambiente tiene a cargo su implementación y funcionamiento (GBIF, 2018c).

Los datos que son incorporados en la red GBIF, en todo momento se mantienen bajo el control de la colección o institución de origen, indicando qué datos y con qué precisión se ponen a disposición de la comunidad. Las distintas colecciones pueden aplicar restricciones al uso de sus datos, de modo que se respeten las políticas de propiedad intelectual de la colección. Al tener una red como GBIF, conlleva a obtener beneficios como (Martin, 2016), por ejemplo: formación y capacitación; los procesos de toma de decisión se pueden fundamentar en bases más sólidas; hace posibles estudios no abordables hasta el momento; repatriación de información a los países de origen; aprovecha al máximo las propuestas existentes; la dimensión de la iniciativa permite aprovechar y difundir las “lecciones aprendidas” mediante colaboraciones entre los nodos más

desarrollados y los menos desarrollados. GBIF realiza esfuerzos importantes para que la información contenida en las bases de datos sea relevante a nivel mundial, con el fin de tener acceso a la mayor cantidad de datos.

GBIF reúne todos los datos a través del estándar de biodiversidad darwinCore que fue diseñado con el propósito de crear un lenguaje común para la publicación y documentación de datos sobre registros biológicos (GBIF, 2018b).

2.4 Estándar de biodiversidad para compartir datos de biodiversidad: El Estándar darwinCore

El estándar darwinCore (DwC) fue creado en el año 1999, como un conjunto de términos formales, que tienen semántica claramente definida y procesos para gestionar los cambios que son necesarios para asegurar la utilidad en la integración de los datos. Es un estándar diseñado con el objetivo de crear un lenguaje común para publicar y documentar datos sobre registros biológicos, listas de especies y catálogos taxonómicos. La creación de este lenguaje ha sido difícil de incorporar en las distintas disciplinas e instituciones, ya que se ha visto con dificultad en el desarrollo local, donde ha habido una escasa cultura de intercambio de datos (Wieczorek *et al.*, 2012).

El estándar cuenta actualmente con 185 términos para la documentación de información de la biodiversidad. Los términos dentro del estándar DwC están organizados en once categorías también denominadas como “clases”. De éstas, nueve cubren aspectos generales como: el nivel de registro, ocurrencia, organismo, muestra de material, evento, ubicación, contexto geológico, identificación, taxón del dominio de la biodiversidad y las dos restantes se refiere a elementos complementarios como la medida o hecho y la relación del recurso sobre información genérica de los registros. Dentro de la planilla de DwC existen campos que son obligatorios, campos opcionales pero requeridos y campos recomendados que permiten documentar información relacionada con el individuo o la especie (Cuadro 1).

Los datos ingresados en DwC son utilizados en una variedad de esquemas de codificación (valores separados por comas, XML, *JavaScript Object Notation*, *Resource Description Framework*, etc.) (Wieczorek *et al.*, 2012).

Los campos son enunciados creados que clasifican la información que se tenga disponible, sin embargo no todos los campos son obligatorios, sino que se debe llenar la mayor cantidad de ellos, para tener información más completa. Cabe resaltar que para compartir una base de datos en GBIF, solo es necesario completar los campos obligatorios que contienen información básica.

Cuadro 1. Ejemplos de algunos campos del Estándar darwinCore.

Campos	Estándar darwinCore	Definición
Obligatorio	occurrenceID	Identificador único del registro biológico.
	catalogNumber	Un identificador (preferiblemente único) para el registro dentro del conjunto de datos o colección.
Opcional pero recomendado	eventDate	Fecha en la que ocurrió el evento
	kingdom	El nombre científico completo del reino al que pertenece el taxón
Recomendado	acceptedNameUsage	El nombre completo, con autoría e información de fecha si se conoce, del taxón actualmente válido (zoológico) o aceptado (botánico)

Fuente: [Darwin Core Quick Reference Guide](#)

2.5 Tipos y fuentes de datos de biodiversidad

2.5.1 Tipos de datos

2.5.1.1 Observaciones y colecciones (*Occurrences*)

Registros de especímenes conservados en herbarios y colecciones biológicas, u observaciones registradas en terreno (Cezón, 2018). Como por ejemplo:

- ✓ Pliegos de herbarios, bancos de germoplasma, semillas, etc.
- ✓ Especímenes conservados en alcohol, formol, montados, etc.
- ✓ Fósiles, muestras de tejidos, ADN, etc.

Campos obligatorios	Campos valorados
occurrenceID	Kingdom
basisOfRecord	decimalLatitude
eventDate	decimalLongitude
scientificName	country

2.5.1.2 Listas de especies (*Checklist*)

Consiste en listas de nombres científicos o taxones que pueden estar restringidas a un grupo taxonómico, región geográfica, temática concreta o a una combinación de las tres (Cezón, 2018).

- ✓ Checklist
- ✓ Listas de referencia taxonómica
- ✓ Listas rojas
- ✓ Geográficas

Campos obligatorios	Campos Valorados
taxonID	parentNameUsageID
scientificName	Kingdom
taxonRank	eventDate

2.5.1.3 Eventos de muestreo (*Sampling event*)

Datos generados a partir de estudios de monitoreo como parcelas de flora y vegetación, cámaras trampa, censos de aves, muestreos en aguas, etc. Indican claramente los métodos, eventos y abundancia relativa de las especies registradas.

Este tipo de conjuntos de datos se permite estructurar y publicar datos como resultado de programas de monitoreo o eventos particulares.

Constituyendo un conjunto de datos complejos que proporcionan mayor detalle sobre el evento de colecta, no solamente ofrecen evidencia de la presencia de especies, sino que también permite evaluar la composición de la comunidad, patrones de abundancia de especies, tendencias de las poblaciones, mediciones asociadas a ejemplares y los lugares de muestreo y colecta entre otras (Cezón, 2018).

Ejemplo:

- ✓ Transecto y parcelas permanentes de vegetación
- ✓ Censos de aves
- ✓ Muestreo de agua dulce o marina

Campos obligatorios	Campos Valorados
eventID	samplingEffort
sampleProtocol	sampleSizeValue
eventDate	sampleSizeUnit
	countryCode
	decimalLatitude
	decimalLongitude

2.5.1.4 Metadatos (Qué/dónde/cuándo/cómo/quién)

Describe y contextualiza conjuntos de datos aún no digitalizados. Esta clase de publicaciones ofrece a los investigadores una herramienta valiosa para dar a conocer un juego de datos que aún no está disponible en línea (Cezón, 2018), permitiendo a otros usuarios descubrir la existencia de estos datos, logrando diferenciar si estos fueran útiles o no.

2.5.2 Fuentes de datos

Las fuentes de datos corresponden a cualquier forma tangible o intangible que contenga información capturada de biodiversidad. Los datos que pueden ser publicados, pueden provenir desde:

- ✓ Libretas de campo
- ✓ Catálogos de especies
- ✓ Evaluación de proyectos
- ✓ Imágenes satelitales
- ✓ Cámaras trampa
- ✓ Repositorios
- ✓ Observaciones casuales

2.6 Calidad de datos

Durante cientos de años la mayoría de los datos que se encuentran almacenados en colecciones biológicas, habían sido utilizados para estudios ecológicos, taxonómicos, biogeográficos y filogenéticos. Actualmente por la expansión de la tecnología los usos han aumentado, siendo accesibles para todos los usuarios.

Hay mucha información proveniente de diversos repositorios disponibles en el mundo, pero existen vacíos de información. Los datos que son publicados en los repositorios digitales, más adelante se convierten en conocimiento y luego ayudan a la toma de decisiones.

En el mundo existen regiones que poseen una baja representatividad de especies aunque tengan alta biodiversidad. Esto se debe a que han sido poco estudiadas, en general por el reducido apoyo económico lo que no permite representar su real diversidad biológica.

En Chile existen varias colecciones que tienen toda la iniciativa o intención de compartir los datos, sin embargo, no poseen personal para que se encargue del traspaso de la información de la etiqueta de una colección biológica hacia una base de datos.

Parte significativa de los registros de los repositorios existentes no son fácilmente accesibles. Muchas veces se cree que por publicar los datos en un *paper* se tiene acceso a ellos en una tabla dentro de él. Sin embargo no es así, pues no todos los datos son abiertos a la comunidad y poseen una baja operabilidad.

Muchas veces los administradores de colecciones biológicas no conocen herramientas que facilitan el trabajo y la manipulación masiva de los conjuntos de datos. A veces realizan el trabajo manual para estandarizar las fechas o limpiar espacios vacíos, realizando un sobre esfuerzo, dado

que al manejar grandes volúmenes de datos editando unitariamente, se está más propenso a ignorar fallas dentro del conjunto de datos (Buitrago, 2018)

2.6.1 Principios fundamentales para la calidad de datos

2.6.1.1 Principio 1: Basura entra → Basura sale

Este principio se refiere a que si ingreso basura a un sistema, pues también va a salir basura. Puede que los datos que se están trabajando, se consideren valiosos y obviamente no son basura, pero la manera en la que se presentan estos datos puede generar “ruido”, pudiendo influir negativamente estudios que utilizaron estos datos y que terminarán en algún artículo científico o en un análisis de modelamiento erróneo.

Claro ejemplo de calidad de datos es cuando, se ingresa una coordenada errónea y con esa coordenada se genera un modelo de distribución de especie, esto conlleva a obtener resultados de mala calidad de análisis y por ende toma de decisiones.

Esto puede surgir en las diversas etapas, siendo la obtención de datos en terreno aquella que necesitaría más esfuerzo en corregir, ya que provocaría una cadena de errores en las etapas posteriores.

Todo esto tiene un impacto directo, pudiendo afectar a los indicadores, análisis, toma de decisiones y políticas directas. Para ellos existen tres definiciones que hay que tomar en cuenta para este principio (Buitrago, 2018):

Definiciones		
1	2	3
Conformidad con los hechos	Conformidad con los requisitos	Los datos deben ser adecuados al uso
Los datos tienen calidad, si la información derivada de los datos representa correctamente el mundo real (hechos).	Satisfacción del consumidor; si un consumidor está satisfecho con un servicio/producto, este servicio o producto tiene calidad para este consumidor.	Un dato tiene calidad si es adecuado para ser usado, siendo útiles para algunos y para otros no. Éste es el concepto de usabilidad.

2.6.1.2 Principio 2: Calidad de datos, concepto idiosincrático

El concepto idiosincrático, corresponde a algo distintivo y propio del individuo y dependerá de la usabilidad de los datos, para ello se definen los siguientes conceptos:

Conceptos		
Uso	Dato	Adecuación
<p>La calidad esta dado por el propósito, como por ejemplo:</p> <p>Modelo de distribución o generar una lista nacional de especies.</p>	<p>Para cada propósito existen determinados datos, por ejemplo:</p> <p>Para hacer modelamiento, necesito datos con coordenadas y sino ese dato no es adecuado.</p>	<p>Se pueden definir los datos: si es usable, si es completo, consistente, es exacto, preciso y es creíble.</p>

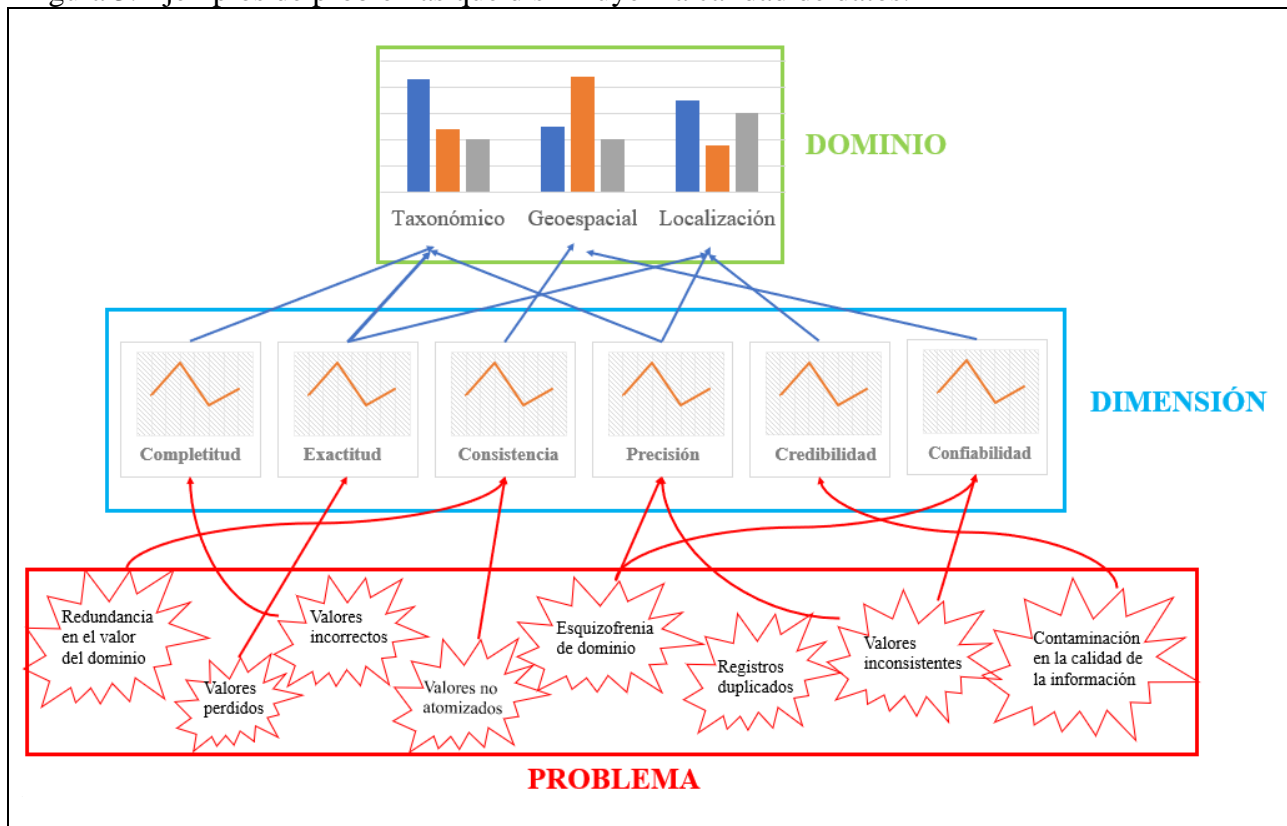
La credibilidad no es un tema menor. Es importante preguntarse ¿de dónde provienen estos datos? ¿quién los transformó? ¿cómo los transformó? y declarar los procesos que se realizaron sobre los datos, para que puedan ser utilizados (Buitrago, 2018).

2.6.1.3 Principio 3: Los datos tienen calidad si no existe ningún problema que la degrade

Existen tres pisos o niveles en los cuales podría generarse problemas y que disminuyen la calidad de datos (Figura 3). Esto se refiere principalmente al trabajo de digitalización de los datos en una colección biológica (Buitrago, 2018).

Niveles que disminuyen la calidad de datos		
Dominio	Dimensión	Problema
<p>Se refiere específicamente al tipo de dato que se necesita dentro de la evaluación: campo taxonómico, campo geoespacial, estandarización.</p>	<p>Aspectos que mejorar (completitud de coordenadas y validación taxonómica, entre otros).</p>	<p>En la base se encuentran los problemas de valores perdidos o faltantes, si los registros son duplicados, valores inconsistentes, etc.</p>

Figura 3. Ejemplos de problemas que disminuyen la calidad de datos.



3 ESTRUCTURACIÓN DE CAMPOS AL ESTÁNDAR DARWIN CORE

3.1 Estructuración de datos al Estándar darwinCore

En el estándar darwinCore es necesario realizar la estructuración de los datos, teniendo un archivo donde estén contenidos los datos de la colección biológica. Por lo tanto se tienen varias opciones de uso, como por ejemplo:

- a) Uso de plantillas estándar, en el que contiene todos los elementos del estándar DwC.
- b) Generar plantillas DwC (*DwC Excel Template Generator*), donde el publicador selecciona solo los campos de interés.
- c) Exploración y conocimiento profundo del estándar DwC, el publicador es capaz de entender cada concepto del estándar y aplicarlo correctamente.

3.2 Mapeo de Campos

El mapeo de campo consiste en realizar una correspondencia y reconocimiento entre los campos de la base de datos original y los del estándar darwinCore. Por lo tanto, es necesario tener en cuenta dos cosas: conocer los datos originales y conocer el estándar DwC.

Luego de identificados los registros de la base de datos original, estos se trasladan a las columnas que sean pertinentes a herbarios en el estándar DwC, debido a que este estándar trabaja con datos de biodiversidad.

3.2.1 Mapeo directo

Consiste en realizar la observación de los campos, que luego son migrados directamente al campo que se reconoció del estándar.

3.2.2 Mapeo Indirecto

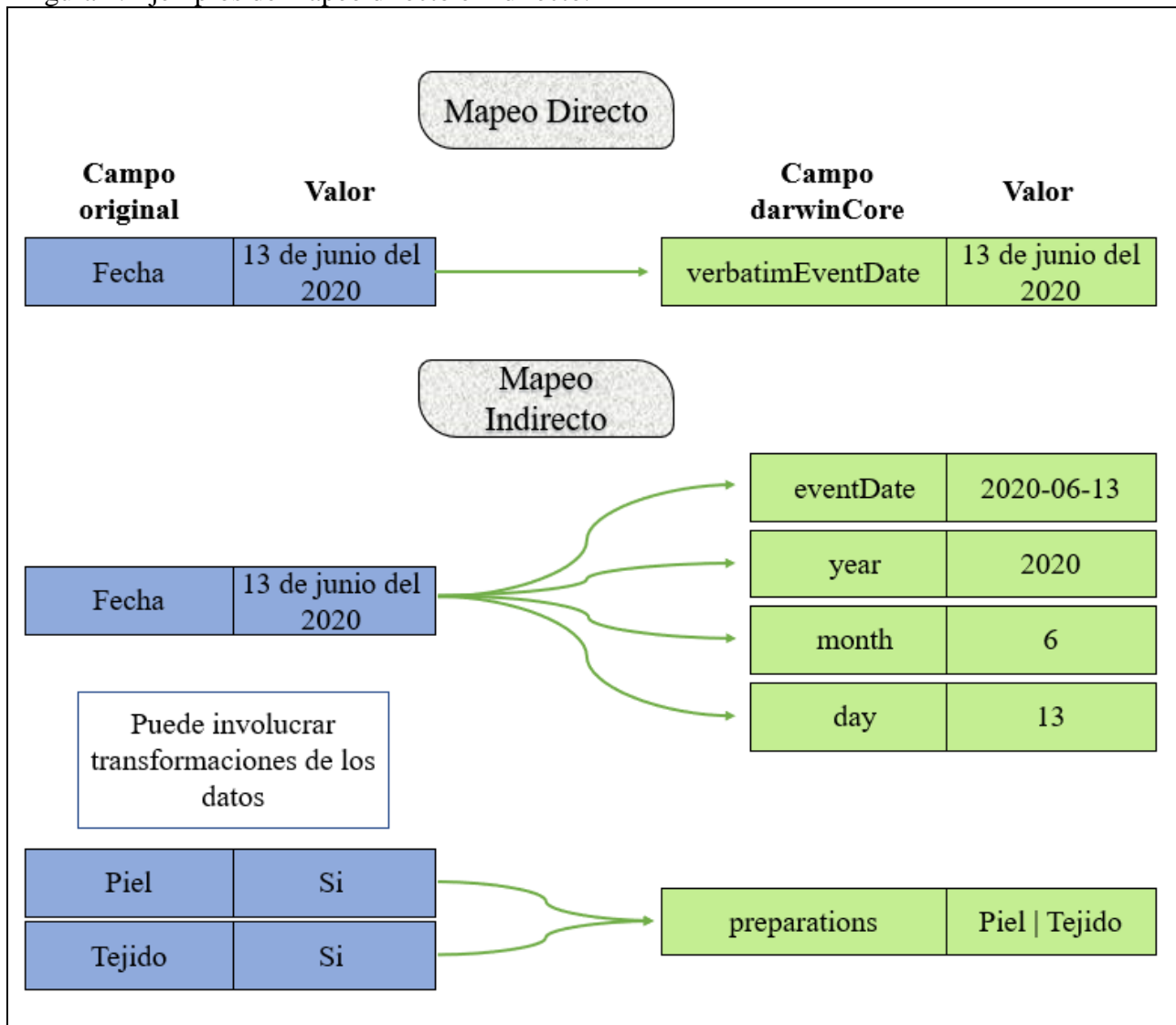
El mapeo indirecto es aquel en que se crean campos distintos a los obtenidos por el mapeo directo, ya que se puede deducir la información requerida, a través del trabajo con valores de algún campo de mapeo directo para completarlos, por ejemplo:

- ✓ Campo *verbatimEventDate* (Fecha original): 13 de junio del 2020, de este campo se pueden extraer indirectamente los campos, *eventDate*, *year*, *month* y *day* (Figura 4).

También puede involucrar transformaciones de los datos, como por ejemplo:

- ✓ Campo *preparations* (Preparación) consiste en una lista de preparaciones y métodos de conservación para un espécimen, por lo que es posible unirse en un solo campo en el estándar dwC (Figura 4).

Figura 4. Ejemplos de mapeo directo e indirecto.



4 HERRAMIENTAS DE GESTIÓN PARA EL MANEJO Y CALIDAD DE BASES DE DATOS DE BIODIVERSIDAD

4.1 Herramientas para el tratamiento de nombres científicos

4.1.1 Herramientas de normalización de nombres científicos

Estas herramientas permiten validar de manera automática y masiva los nombres científicos en un conjunto de datos. Se comparan los nombres científicos del conjunto, contra el catálogo taxonómico a partir del cual recupera la jerarquía taxonómica de cada taxón (reino, filo, clase, orden, familia, género) y su estado taxonómico (sinónimo, aceptado, ambiguo).

4.1.1.1 Species matching - GBIF

La herramienta compara los nombres científicos de especímenes de cualquier reino que contenga el conjunto de datos contra el árbol taxonómico de GBIF. Esta herramienta solo permite archivos delimitados por comas (.csv) o documento de texto que tengan como mínimo una columna llamada “*scientificName*”, opcionalmente el archivo puede contener las columnas “*kingdom*” y “*id*”.

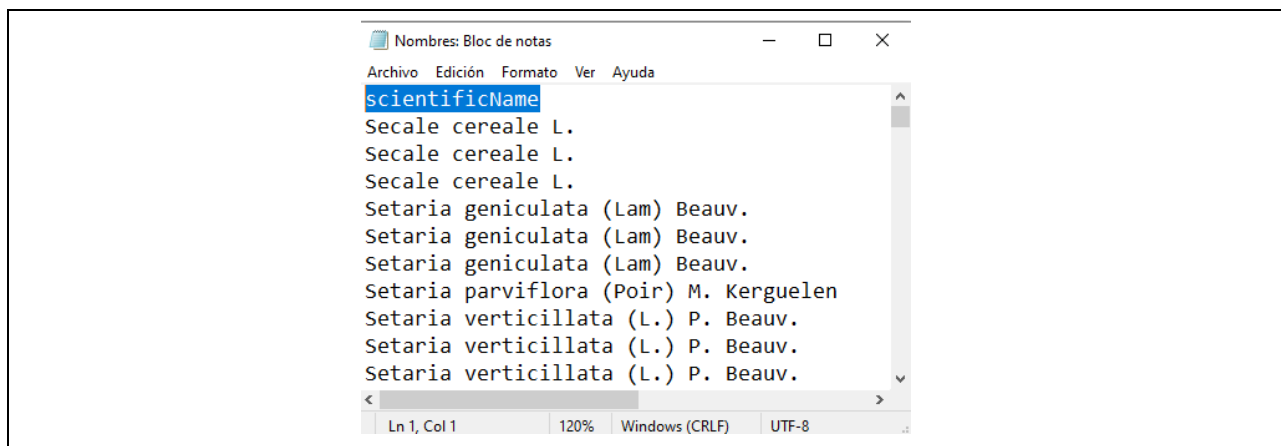
Enlace Web: [Species matching](#)

Paso 1

Ingresa a la herramienta en línea [Species Matching](#) y revise las opciones que brinda la herramienta.

Paso 2

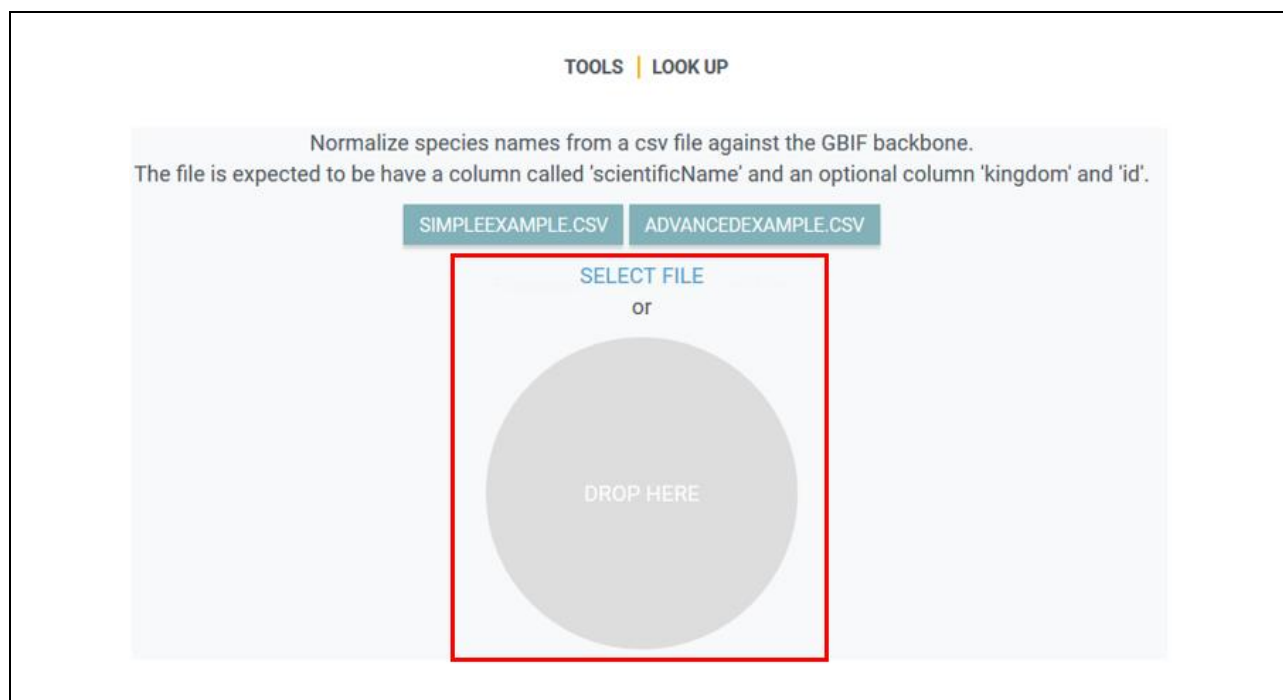
1. Abra el documento de Excel que contenga el campo nombres científicos.
2. Ubique el campo que contiene los **nombres científicos**, copie la columna completa (Ctrl+C) y péguela (Ctrl+V) en un documento en blanco de texto (bloc de notas). Asegúrese de que la lista de los nombres científicos esté encabezada con el campo “*scientificName*”.
3. Guarde el documento como “nombres_originales.txt”.



Paso 3

Desde la herramienta Species Matching ya puede integrar los nombres para normalizar. Puede realizarlo de dos maneras:

1. Pulsando el botón **SELECT FILE** para navegar y encontrar en su equipo el archivo a validar (nombres_originales.txt) o arrastrando con el ratón el archivo al globo donde se indica **DROP HERE**.










2. Automáticamente se previsualizará el listado de nombres y se dará la opción a elegir un reino (Plantae, Animalia, Fungi, Cromista, Bacteria, Protozoo, Virus, Arqueas).
3. Seleccione el reino de acuerdo con el caso de uso o puede no seleccionar y GBIF asignará un reino de acuerdo con su *Backbone Taxonomy*. Haga clic en **MATCH TO GBIF BACKBONE**.



Paso 4

1. Como resultado aparecerá una página en la que se muestran los siguientes campos:

TOOLS LOOK UP				
verbatimScientificName	preferredKingdom	matchType	confidence	scientificName (editable)
Isolepis cernuua (Vahl) Roem. & Schult.	any	FUZZY	100	 Isolepis cernua (Vahl) Roem. & Schult.
Isolepis cernuua (Vahl) Roem. & Schult.	any	FUZZY	100	 Isolepis cernua (Vahl) Roem. & Schult.
Isolepis cernuua (Vahl) Roem. & Schult.	any	FUZZY	100	 Isolepis cernua (Vahl) Roem. & Schult.
Isolepis cernuua (Vahl) Roem. & Schult.	any	FUZZY	100	 Isolepis cernua (Vahl) Roem. & Schult.
Isolepis nigricans Kunth	any	EXACT	100	 Isolepis nigricans Kunth
Isolepis nigricans Kunth	any	EXACT	100	 Isolepis nigricans Kunth
Scirpus sp.	any	HIGHERRANK	94	 Scirpus L.

El campo **matchType** indica con texto y en clave de colores la coincidencia del nombre proporcionado en origen con el *Backbone* de GBIF. Haga clic sobre este campo para ordenar la información que contiene y localiza aquellos nombres calificados como **FUZZY** (dudoso), **NONE** o **HIGHERRANK** (no se han encontrado coincidencias al mismo nivel taxonómico y se propone uno superior).

2. Utilice la herramienta de edición para comprobar manualmente los nombres cuyo **matchType** no es **EXACT**, en caso de que existan, sustituya la propuesta por el valor adecuado si considera que hay algún nombre con una coincidencia mejor que la propuesta.
3. Dirigirse a la parte inferior derecha “Generate CSV”, Guarde el archivo “Normalized.csv” resultante en su equipo.

Previous1...2324252627Next

Exclude unmatched

GENERATE CSV

Paso 6

1. Para visualizar los anteriores resultados importe los datos “Normalized.csv” en un archivo Excel siguiendo las indicaciones del subcapítulo “4.5.2. Importar archivos de texto en Excel”.
2. Renombre el nuevo archivo como “nombres_normalizados.xlsx”.
3. A continuación, vamos a incorporar los nombres normalizados en nuestra tabla original. Para ello, seleccione desde “Nombres_normalizados.xlsx” los valores de los campos *scientificName*, *kingdom*, *phylum*, *class*, *order*, *family*, *genero*, *specificEphitet* y *rank*, copie (Ctrl+C) y pegue (Ctrl+V) en la base de datos donde mapeo sus campos al estándar DwC, según corresponda.

4.1.1.2 I Plant Collaborative

Esta herramienta realiza la comparación de los nombres científicos del reino **Plantae** de su base de datos con diversos catálogos taxonómicos de plantas (TPL, GCC, ILDIS, TROPICOS, USDA), los cuales se pueden seleccionar todos o solo dejar los catálogos que considere más adecuados.

iPlant Collaborative Taxonomic Name Resolution Service v4.0

Home | TNRS Application | Instructions | Sources | About | Collaborators | Known Issues | Optimize Your !

Enter List Upload and Submit List Retrieve Results

Enter scientific names to check

Enter up to 5000 names.

1

Click here for support Clear Submit List 2

Name processing settings

Processing Mode: Edit ?
Selected mode: Perform Name Resolution

Match Accuracy: Edit ?
Allow partial matches, Selected minimum threshold: 0.05

Sources: Edit ? 3
[TPL, GCC, ILDIS, TROPICOS, USDA]

Family Classification: Edit ?
Selected classification source: TROPICOS

1: Cuadro de trabajo, lugar donde se pegan nombres científicos de nuestra BD.

2: Submit list, botón para ejecutar el proceso y donde la aplicación entrega el resultado.

3: Fuentes, se puede dejar la forma predeterminada con todas las fuentes o si selecciona “Edit” puede seleccionar fuentes de catálogos taxonómicos de su preferencia.

Paso 1

Se deben ingresar los nombres científicos (copiándolos desde su base de datos estandarizada de herbario) en la herramienta “[Taxonomic Name Resolution Service](#)” (TNRS) de la plataforma IPlantCollaborative.

Nota: Varios de estos nombres en su digitalización o previo a ésta presentan errores de tipeo o en las terminaciones. Esta herramienta permite la comparación del nombre científico de la planilla original con los obtenidos de distintas fuentes, sugiriendo por similitud un nombre científico sin estos errores.

Paso 2

Luego de ingresados los nombres científicos a la herramienta, haga clic en la parte inferior derecha del cuadro principal de trabajo en “*Submit List*”. Entrega los resultados taxonómicos para las especies ingresadas, propuestas por la plataforma TNRS.

Enter List Upload and Submit List Retrieve Results

Enter scientific names to check

Zannichellia palustris L.
Deyeuxia nardifolia (Griseb.) P.
Deyeuxia arundinacea Phil.
Calamagrostis tenuifolia (Phil.)
Bromus unioloides Kth.
Elymus repens (L.) Gould

[Click here for support](#)
Close
Submit List

Name processing settings

Processing Mode: [Edit](#)
Selected mode: Perform Name Resolution

Match Accuracy: [Edit](#)
Allow partial matches, Selected minimum threshold: 0.05

Sources: [Edit](#)
[TPL, GCC, ILDIS, TROPICOS, USDA]

Family Classification: [Edit](#)
Selected classification source: TROPICOS

Best match settings Download settings Download results

Name Submitted	Name Matched	Name Source	Overall Score	Taxonomic Status	Accepted Name	Details
Zannichellia palustris L.	Zannichellia palustris L.	TPL TROPICOS USDA	1.00	Accepted	Zannichellia palustris L.	Details
Deyeuxia nardifolia (Griseb.) P.	Deyeuxia nardifolia (Griseb.) Phil. (+1 more)	TPL	.96	Synonym	Calamagrostis nardifolia (Griseb.) Hi.	Details
Deyeuxia arundinacea Phil.	Deyeuxia arundinacea Phil. (+5 more)	TPL	1.00	Synonym	Calamagrostis eminens (J.Presl) Ste.	Details
Calamagrostis tenuifolia (Phil.)	Calamagrostis tenuifolia (Phil.) R.E.Fr. (+2 more)	TPL	.89	Synonym	Calamagrostis curvula (Wedd.) Pilg.	Details
Bromus unioloides Kth.	Bromus unioloides Kunth (+2 more)	TPL TROPICOS USDA	.86	Synonym	Bromus catharticus Vahl	Details
Elymus repens (L.) Gould	Elymus repens (L.) Gould	TPL TROPICOS USDA	1.00	Accepted	Elymus repens (L.) Gould	Details

Page 1 of 1

Displaying 1 - 6 of 6

Paso 3

Se abrirá el siguiente cuadro para realizar la descarga de los resultados. Haga clic en el cuadro rojo donde dice “Download results”. Luego configure la descarga como aparece en el cuadro verde y haga clic en “Ok”.

Best match settings Download settings **Download results**

Name Submitted	Name Matched
Zannichellia palustris L.	Zannichellia palustris L.
Deyeuxia nardifolia (Griseb.) P.	Deyeuxia nardifolia (Griseb.) Phil. (+1 more)
Deyeuxia arundinacea Phil.	Deyeuxia arundinacea Phil. (+5 more)
Calamagrostis tenuifolia (Phil.)	Calamagrostis tenuifolia (Phil.) R.E.Fr. (+2 more)
Bromus unioloides Kth.	Bromus unioloides Kunth (+2 more)
Elymus repens (L.) Gould	Elymus repens (L.) Gould

Page 1 of 1

Download Options

Results to download

☒ Best matches only ?
☐ All matches ?

Download format

☐ Simple ?
☒ Detailed ?

Download as:
 ?
Encoding:

OK Cancel

Paso 4

1. Para visualizar los resultados anteriores, importe los datos “tnrs_result.txt” en un archivo Excel siguiendo las indicaciones del subcapítulo “4.5.2. Importar archivos de texto en Excel”.
2. Después de realizada la descarga de la herramienta se debe crear una columna y concatenar los valores de los campos “Name_matched” y “Name_matched_author”. Para así obtener el nombre científico completo según el estándar DwC y llevarlo a la base de datos estandarizada.

	A	B	C	D
1				
2	Name_submitted	scientificName	Name_matched	Author_matched
3	Zannichellia palustris L.	Zannichellia palustris L.	Zannichellia palustris	L.
4	Deyeuxia nardifolia (Griseb.) P.	Deyeuxia nardifolia (Griseb.) Phil.	Deyeuxia nardifolia	(Griseb.) Phil.
5	Deyeuxia arundinacea Phil.	Deyeuxia arundinacea Phil.	Deyeuxia arundinacea	Phil.
6	Calamagrostis tenuifolia (Phil)	Calamagrostis tenuifolia (Phil.) R.E.Fr.	Calamagrostis tenuifolia	(Phil.) R.E.Fr.
7	Bromus unioloides Kth.	Bromus unioloides Kunth	Bromus unioloides	Kunth
8	Elymus repens (L.) Gould	Elymus repens (L.) Gould	Elymus repens	(L.) Gould
9				

3. Ir al campo “Warnings”, las celdas que no presenten valor o presenten el valor “[Ambiguous match]”, deben copiarse los valores de la columna creada, en la planilla darwinCore en el campo “*scientificName*”. Si aparece el valor [Partial_match], el “Name_submitted” debe ser verificado, porque significa que no existe coincidencia entre el nombre científico ingresado y los nombres de los catálogos taxonómicos de la herramienta, por lo tanto se debe hacer una revisión individual de los nombres.
4. Verificado el campo del nombre científico de la herramienta, debe ser copiado en su planilla darwinCore en el campo “*scientificName*”.
5. En el campo “*specificEpithet*” (Epíteto específico) de DwC de sus bases de datos estandarizada: Se copian los valores del campo “Specific_epithet_matched” de la base de datos de salida de TNRS.

4.1.2 Catálogos taxonómicos globales en línea para el chequeo taxonómico

4.1.2.1 Instituto de Botánica Darwinion: Catálogo de las Plantas Vasculares del Conosur



INSTITUTO DE BOTÁNICA
DARWINION

El "Catálogo de las Plantas Vasculares del Conosur" es un proyecto organizado en colaboración entre diversas instituciones de la Argentina y de los Estados Unidos de América. Su objetivo es completar un inventario actualizado y computarizado de la flora de Argentina. En 1994 se publicó la lista crítica de las Poaceae de Argentina, publicación a la que siguieron el resto de las Monocotiledóneas, junto con Pteridophyta, Gimnospermas y las Dicotiledóneas. La información florística existente en la base de datos se mantiene actualizada a medida que nuevos cambios taxonómicos o nuevos registros de plantas del cono sur son publicados (Instituto de Botánica Darwinion, 2020).

4.1.2.2 Tropicos



La plataforma Tropicos pertenece al Jardín botánico de Missouri. Fue originalmente creado para la investigación interna, pero desde entonces ha sido puesto a la disposición de la comunidad científica del mundo. Todos los datos de la nomenclatura, bibliográficos, y especímenes han sido acumulados en las bases de datos electrónicas durante los últimos 25 años y están a disposición del público. Tropicos trabaja con los siguientes grupos taxonómicos: Dicotiledóneas, Monocotiledóneas, Helechos, Gimnospermas, Briofitos, Hongos y Algas (Tropicos, 2020).

4.1.2.3 The Plant List



The Plant List es una lista de trabajo de especies de plantas conocidas, cuyo objetivo es tener una cobertura completa a nivel de especies para todos los nombres de musgos y hepáticas y sus aliados (Briófitos) y de plantas vasculares que incluyen las plantas con flores (Angiospermas), coníferas, cícadas y sus aliados (Gimnospermas) y los helechos y sus aliados, incluidas las colas de caballo y los musgos de club (Pteridofitas) (The Plant List, 2013).

4.1.2.4 AlgaeBase



AlgaeBase es una base de datos de información sobre algas que incluye organismos terrestres, marinos y de agua dulce. En la actualidad, es uno de los repositorios más completos sobre datos de algas marinas. También se han incluido los pastos marinos, aunque son plantas con flores (Guiry y Guiry, 2020).

4.2 Herramientas para el tratamiento de coordenadas

4.2.1 Conversión de coordenadas

4.2.1.1 Canadensys Coordinate Conversion

La herramienta permite convertir coordenadas geográficas de DDMMSS (grados, minutos, segundos) a grados decimales. Para ello se escriben pares de coordenadas en líneas separadas o pegar columnas de latitud y longitud desde la base de datos.

Ejemplos de valores de entrada

45° 32' 25" N, 129° 40' 31" W

Paso 1

Abra la base de datos donde se encuentran las coordenadas del registro biológico.

Paso 2

Ingresa a la aplicación en línea [Coordinate conversion](#) y revise las opciones que le brinda la herramienta.

Paso 3

1. En su archivo con el caso de uso, ubique los elementos que describen la latitud y longitud originales (*verbatimLatitude* y *verbatimLongitude*) y realice el procedimiento de concatenar estos campos para formar el campo "*verbatimCoordinates*".
2. Verifique que los separadores decimales se encuentren todos como **punto (.)** y **NO** como **coma (,)** o la herramienta no los transformará.

Para cambiar el separador decimal, seleccione las columnas de latitud y longitud, ubique en Excel la herramienta de "buscar y reemplazar" (Ctrl+L) y llene la ventana como se muestra en la siguiente imagen. Luego haga clic en "Reemplazar todos", y Excel le informará cuántos registros fueron cambiados.

Paso 4

Copie las celdas de los elementos “*verbatimCoordinates*” (sin incluir el encabezado) y péguelos directamente en el cuadro de texto de Canadensys.

5° 42' 27.5" N	76° 0' 6.2" W
6° 20' 3.120" N	75° 36' 18.650" W
6° 20' 3.120" N	75° 36' 18.650" W
6° 20' 3.120" N	75° 36' 18.650" W
6° 20' 3.120" N	75° 36' 18.650" W
6° 37' 1.7" N	75° 39' 6.8" W
6° 37' 1.7" N	75° 39' 6.8" W
6° 37' 1.7" N	75° 39' 6.8" W
6° 37' 1.7" N	75° 39' 6.8" W
6° 37' 1.7" N	75° 39' 6.8" W
6° 37' 1.7" N	75° 39' 6.8" W
6° 37' 1.7" N	75° 39' 6.8" W
6° 37' 1.7" N	75° 39' 6.8" W
6° 37' 1.7" N	75° 39' 6.8" W
6° 37' 1.7" N	75° 39' 6.8" W

Paso 5

Haga clic en el botón “*Submit*” o “Enviar” (ubicado en la esquina inferior izquierda). Se obtendrá una tabla como se muestra en la siguiente figura.

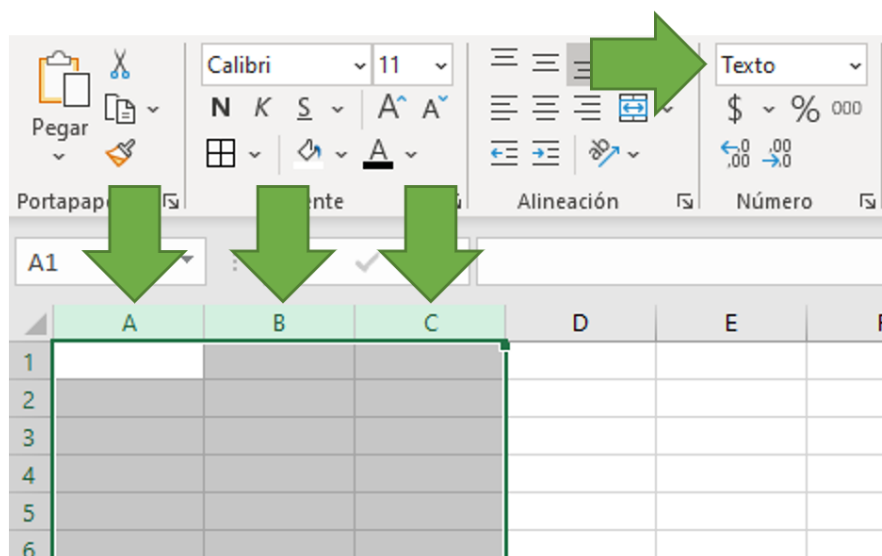
Coordinate conversion results

original	decimalLatitude	decimalLongitude
6° 12' 26.210" N 75° 29' 24.110" W	6.2072806	-75.4900306
6° 20' 3.120" N 75° 36' 18.650" W	6.3342	-75.6051806
6° 20' 3.120" N 75° 36' 18.650" W	6.3342	-75.6051806
6° 20' 3.120" N 75° 36' 18.650" W	6.3342	-75.6051806
6° 20' 3.120" N 75° 36' 18.650" W	6.3342	-75.6051806
6° 12' 26.210" N 75° 29' 24.110" W	6.2072806	-75.4900306
6° 12' 26.210" N 75° 29' 24.110" W	6.2072806	-75.4900306
6° 12' 26.210" N 75° 29' 24.110" W	6.2072806	-75.4900306
6° 12' 26.210" N 75° 29' 24.110" W	6.2072806	-75.4900306
6° 12' 26.210" N 75° 29' 24.110" W	6.2072806	-75.4900306
6° 12' 26.210" N 75° 29' 24.110" W	6.2072806	-75.4900306
6° 12' 26.210" N 75° 29' 24.110" W	6.2072806	-75.4900306
6° 12' 26.210" N 75° 29' 24.110" W	6.2072806	-75.4900306
6° 12' 26.210" N 75° 29' 24.110" W	6.2072806	-75.4900306
6° 12' 26.210" N 75° 29' 24.110" W	6.2072806	-75.4900306
6° 12' 26.210" N 75° 29' 24.110" W	6.2072806	-75.4900306
6° 12' 26.210" N 75° 29' 24.110" W	6.2072806	-75.4900306
6° 12' 26.210" N 75° 29' 24.110" W	6.2072806	-75.4900306

Nota: Los resultados se encuentran en el costado derecho, como *decimalLatitude* y *decimalLongitude*. Si ingresa coordenadas que ya se encuentran en coordenadas decimales, Canadensys no cambiará la información original, excepto que reemplazará las orientaciones cardinales con el respectivo signo negativo para Sur y Oeste (S-W).

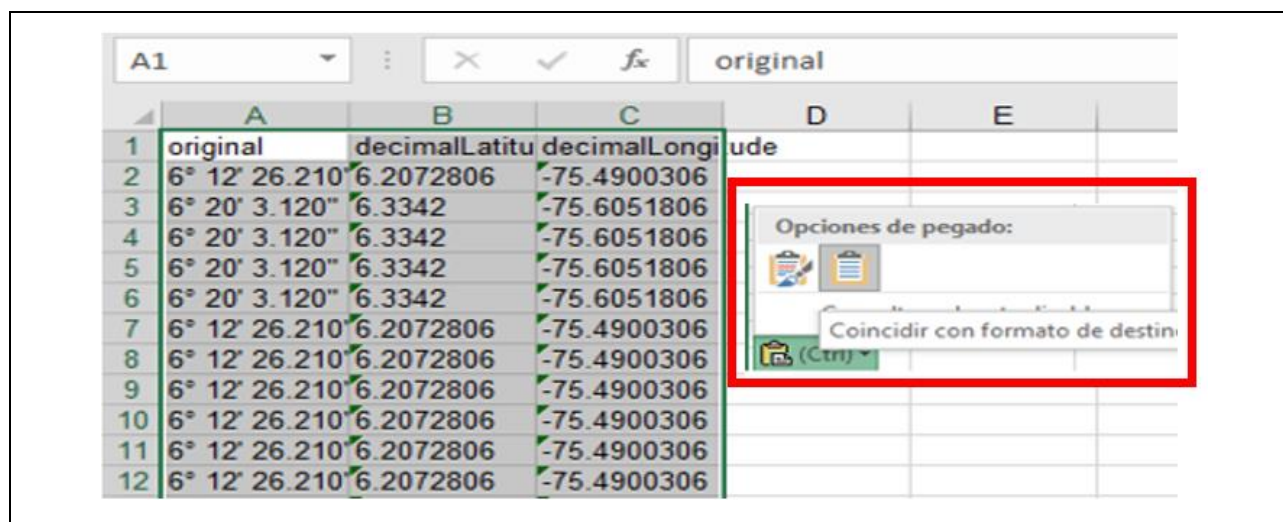
Paso 6

Abra un nuevo documento Excel y seleccione las 3 primeras Columnas (A, B, C) y asígneles formato de “texto” en la pestaña de inicio, como se ve a continuación.



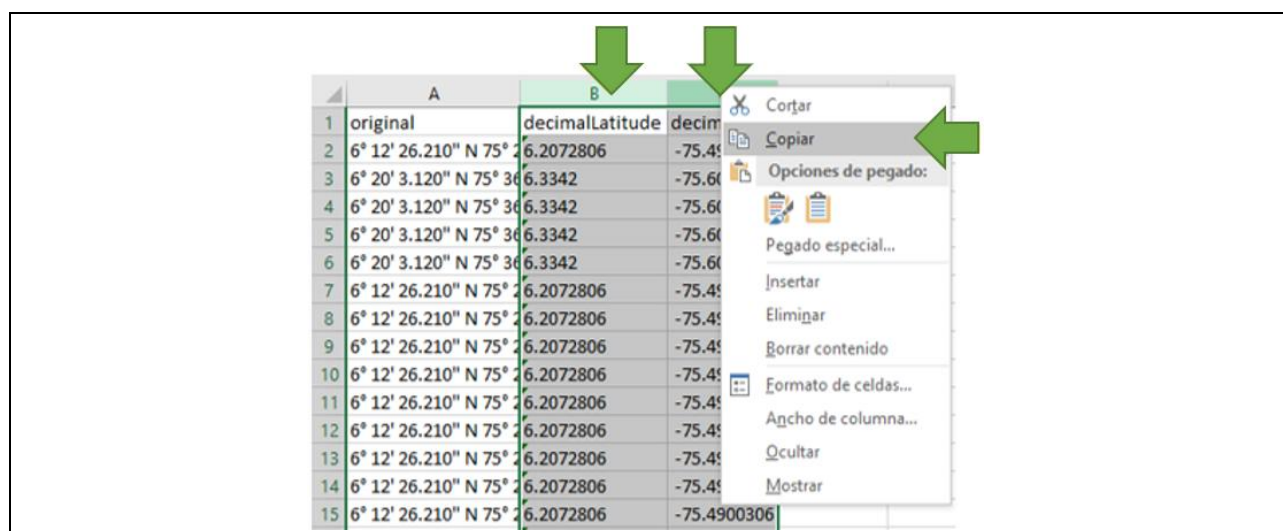
Paso 7

Copie los resultados de Canadensys y péguelos en el nuevo libro Excel, seleccionando la opción “Coincidir con formato de destino”.



Paso 8

Copie estos valores de las columnas “*decimalLatitude*” y “*decimalLongitude*”. Haga clic derecho sobre las columnas y seleccione copiar (Ctrl+C).



Paso 9

Pegue los valores copiados en las columnas *decimalLatitude* y *decimalLongitude* en su base de datos estandarizada.

4.2.1.2 ArcGeek

La planilla “Conversor de Coordenadas de ArcGeek” permite realizar la conversión de las coordenadas geográficas UTM a grados decimales, permitiendo trabajar por lotes de datos.

Los datos de entrada de la planilla son el Huso y el hemisferio (N-S), luego se ingresan los registros que se encuentran en UTM a la planilla y automáticamente entrega los resultados en grados decimales.

Conversión de Coordenadas UTM a Grados Decimales				
Ingrese sus coordenadas				
UTM		Hem. Lat.	Grados decimales	
Este	Norte		Latitud	Longitud
641225	4849046	G	-46.49734666	-73.15957396
639402	4849700	G	-46.49184362	-73.18351477
641225	4849046	G	-46.49734666	-73.15957396
677964	4834071	G	-46.62330058	-72.67540808
640911	4849138	G	-46.49658488	-73.16369169

4.3 Herramientas para el tratamiento de fechas

4.3.1 Canadensys Date Parsing

Date Parsing es una herramienta en línea, desarrollada por Canadensys (nodo GBIF Canadá), que permite atomizar y estandarizar fechas de forma masiva, de acuerdo con el estándar ISO 8601 con formato AAAA-MM-DD; siendo el empleado para el estándar darwinCore.

Paso 1

Abra la base de datos estandarizada, diríjase al campo de “*verbatimEventDate*” del registro biológico (este procedimiento puede aplicarse en otros campos que contengan fechas).

Paso 2

Seleccione los valores de la columna “*verbatimEventDate*” y cópiela (Ctrl + C).

Paso 3

Ingresa a la aplicación Date Parsing

Date parsing

Use this tool to parse dates into their component parts. Type or paste dates on separate lines, optionally preceded by your own identifier followed by a tab or a pipe.

Jun 13, 2008

Submit

Coordinate conversion

Date parsing

Tools API

About

Example input

Jun 13, 2008
15 Jan 2011
2009 IV 02
2 VII 1986

1 | 1999/02/24
2 | 02/17/1921

Paso 4

Pegue los datos (Ctrl+V), previamente copiados en la aplicación. Luego haga clic en “Submit” para iniciar la conversión de fechas.

Date parsing

Use this tool to parse dates into their component parts. Type or paste dates on separate lines, optionally preceded by your own identifier followed by a tab or a pipe.

2001/03/19
19/03/2001
Feb 10, 2001
2001
19/03/2001
20/03/2001
Mar 2001
2001/03/19
Feb 2001
May 2001
Mar 19, 2001
2001
19/03/2001
19/03/2001
Feb 10, 2001
2001

Submit

Paso 5

Tras haberle dado clic al botón “Submit”, obtendrá el siguiente resultado. Nótese que donde dice “ISO 8601” se encuentra el resultado de la estandarización de las fechas para el campo “eventDate”.

Date parsing results				
original	year	month	day	ISO 8601
2001/03/19	2001	3	19	2001-03-19
19/03/2001	2001	3	19	2001-03-19
Feb 10, 2001	2001	2	10	2001-02-10
2001	2001			2001
19/03/2001	2001	3	19	2001-03-19
20/03/2001	2001	3	20	2001-03-20
Mar 2001	2001	3		2001-03
2001/03/19	2001	3	19	2001-03-19

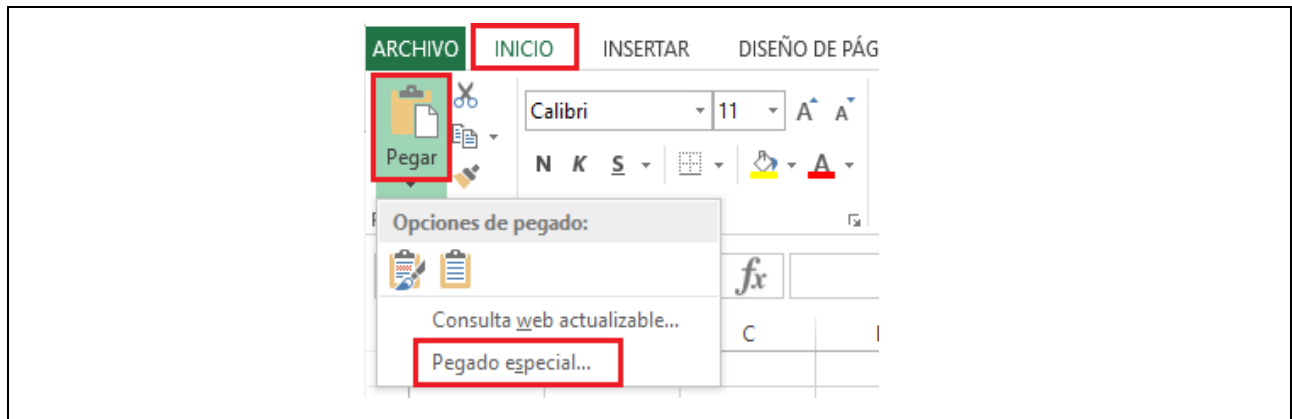
Paso 6

Seleccione toda la tabla arrojada como resultado y cópiela (Ctrl + C).

Date parsing results				
original	year	month	day	ISO 8601
2001/03/19	2001	3	19	2001-03-19
19/03/2001	2001	3	19	2001-03-19
Feb 10, 2001	2001	2	10	2001-02-10
2001	2001			2001
19/03/2001	2001	3	19	2001-03-19
20/03/2001	2001	3	20	2001-03-20
Mar 2001	2001	3		2001-03
2001/03/19	2001	3	19	2001-03-19

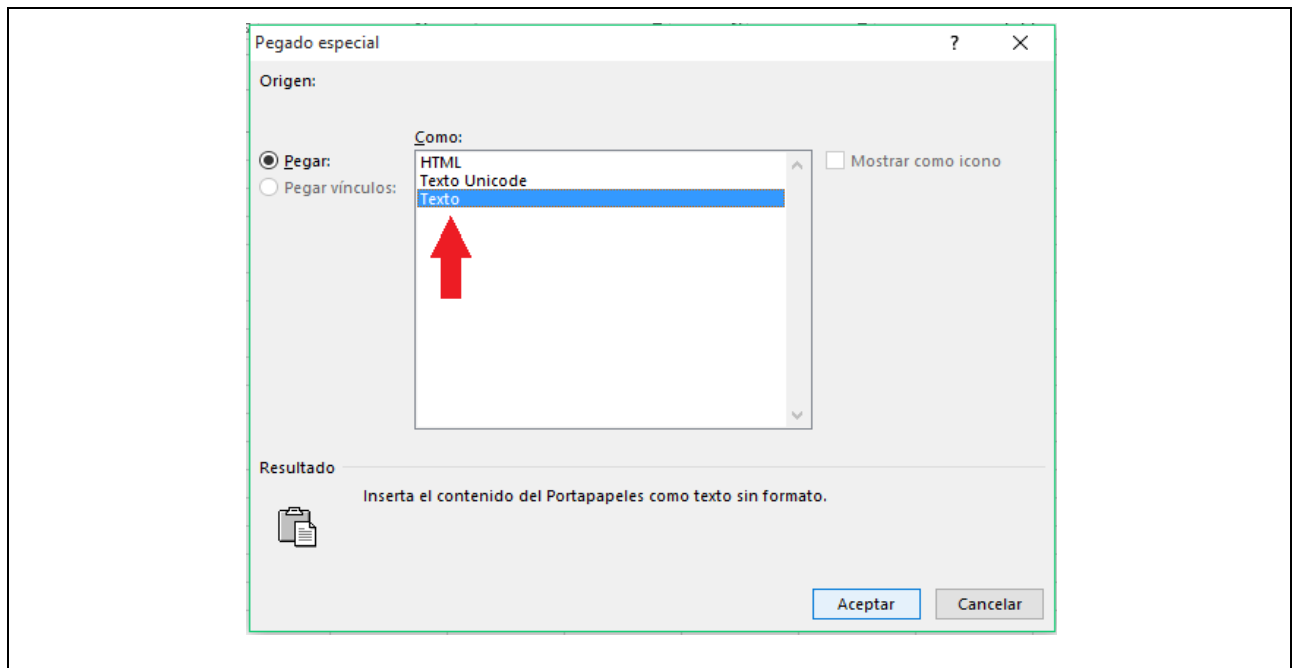
Paso 7

Pegue la tabla copiada resultante del paso previo en un nuevo archivo u hoja de Excel. Para hacerlo ingrese a la pestaña de “Inicio” en el menú superior. Allí seleccione “Pegar” y “Pegado especial”.



Paso 8

En el menú de pegado especial seleccione “Texto” y de clic en “Aceptar”.



Paso 9

Copie los valores de las fechas estandarizadas en sus respectivos campos: *eventDate*, *year*, *month* y *day*) en su base de datos estandarizada.

4.4 Herramientas para corrección de formato

4.4.1 Open Refine



Open Refine es una herramienta que dispone de un conjunto de características para trabajar con datos tabulares que mejoran la calidad general de las bases de datos. Se trata de una aplicación que se ejecuta fuera de la computadora como un pequeño servidor web, al que se accede desde un navegador web.

La carga de datos se puede hacer desde diversas fuentes de datos: TSV, CSV, SV, Excel (.xls y .xlsx), JSON, XML, RDF as XML y datos de Google Docs. La carga de datos implica dos etapas, la primera es la creación del proyecto y la segunda es el análisis de la fuente.

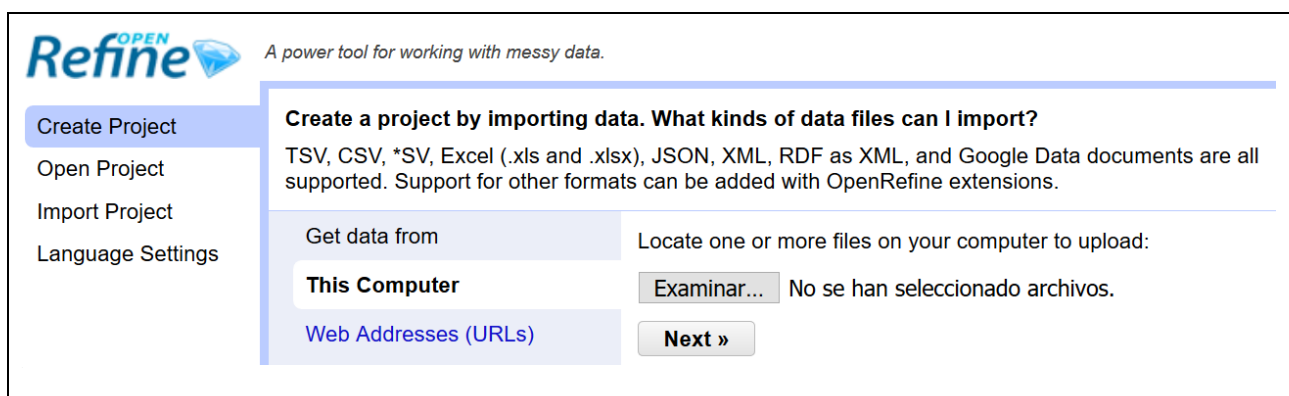
A continuación se detallarán los pasos a seguir para mejorar la calidad de los datos con Open Refine, extraídos de “Open Refine-Guía básica” (SiB Colombia, 2019).

4.4.1.1 Crear un proyecto

1. Tenga presente el lugar donde almacenó la base de datos estandarizada.
2. Abra *OpenRefine* y diríjase a la pestaña “*Create Project*”, para cargar su base de datos siga la ruta:

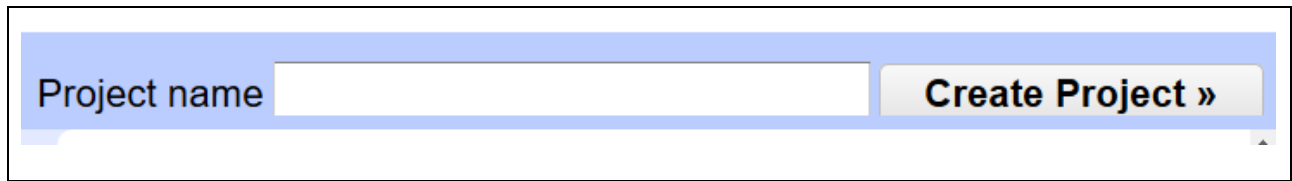
Get data from> This Computer, y haga clic en “*Examinar*” (Choose Files)

3. Seleccione el archivo de su base de datos estandarizada y haga clic en “*Next*”.



4. Un panel de selección aparecerá, este le permite especificar el tipo de datos que se cargan y configurar la manera en la que los datos son leídos. Si los caracteres de tildes y ñ's muestran caracteres raros cambie el *Character encoding* por UTF-8.

5. En la esquina superior derecha verá un cuadro de texto en el que puede cambiar el nombre del proyecto; nómbrelo y haga clic en el botón “*Create Project*”:




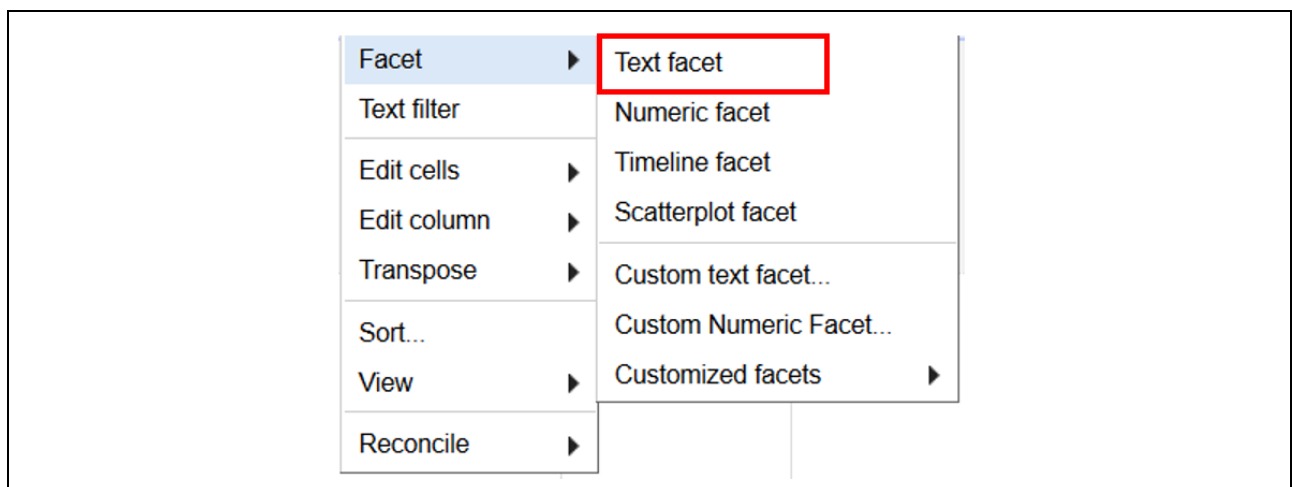
The image shows a user interface element with a light blue header. On the left, the text 'Project name' is followed by a white rectangular input field. To the right of the input field is a grey button with the text 'Create Project »' in bold black font.

6. Espere a que cargue el archivo, esto puede tomar un tiempo dependiendo del tamaño de éste.

4.4.1.2 Faceting

Es un método para filtrar los datos en conjuntos más pequeños para facilitar el uso y análisis, puede hacerse para el texto, los números y las fechas.

1. Diríjase al campo de interés donde desea realizar un filtro de sus datos, haga clic en el Menú Columna  y siga la ruta que se muestra en la imagen para hacer un *Text Facet*:



2. A su izquierda aparecerá una ventana con el nombre de la columna y el Facet (cuadro rojo) que se realizó. Haga clic en *count* (cuadro verde) para organizar las clases de la más a la menos abundante y en *name* (cuadro amarillo) para organizarlas en orden alfabético.

The screenshot shows the Refine interface for the dataset 'BDBrifofitasCSV csv'. The 'municipality' facet is active, showing 27 choices sorted by name. The choices are: Comuna de Alhué (11), Comuna de Alto del Carman (2), Comuna de Alto del Carmen (2), Comuna de Chanco (1), and Comuna de Chonchi (2). The 'name' and 'count' columns are highlighted. The main table shows 319 rows with columns: untryCode, country, stateProvince, county, and municipality. The first row shows data for Chile, Región Metropolitana de Santiago, Provincia de Santiago, and Comuna de La Pintana.

- Corrija las inconsistencias en los nombres si las hubiese. Para esto acerque el cursor al valor que desea corregir y haga clic en *Edit*, luego en el cuadro de texto que aparece corrija el error y haga clic en *Apply*. Verá que todos los valores serán corregidos de manera automática y las celdas se transformarán de forma masiva.

The screenshot illustrates the process of editing a municipality name. The 'municipality' facet is shown with 27 choices. The 'Comuna de Alto del Carman' entry is selected, and the 'edit' button is clicked. A text input field appears with the name 'Comuna de Alto del Carman' and a red squiggly line indicating a spelling correction. The 'Apply' button is highlighted. Below, the facet is shown again with 26 choices, and the 'Comuna de Alto del Carmen' entry is now highlighted, indicating the successful correction.

- Este procedimiento también lo puede realizar en las distintas columnas que presenten inconsistencias. Especialmente en columnas con vocabulario controlado como *“basisOfRecord”* (Base del registro) para que se ajusten según el estándar DwC (Ver Guía rápida de DwC).




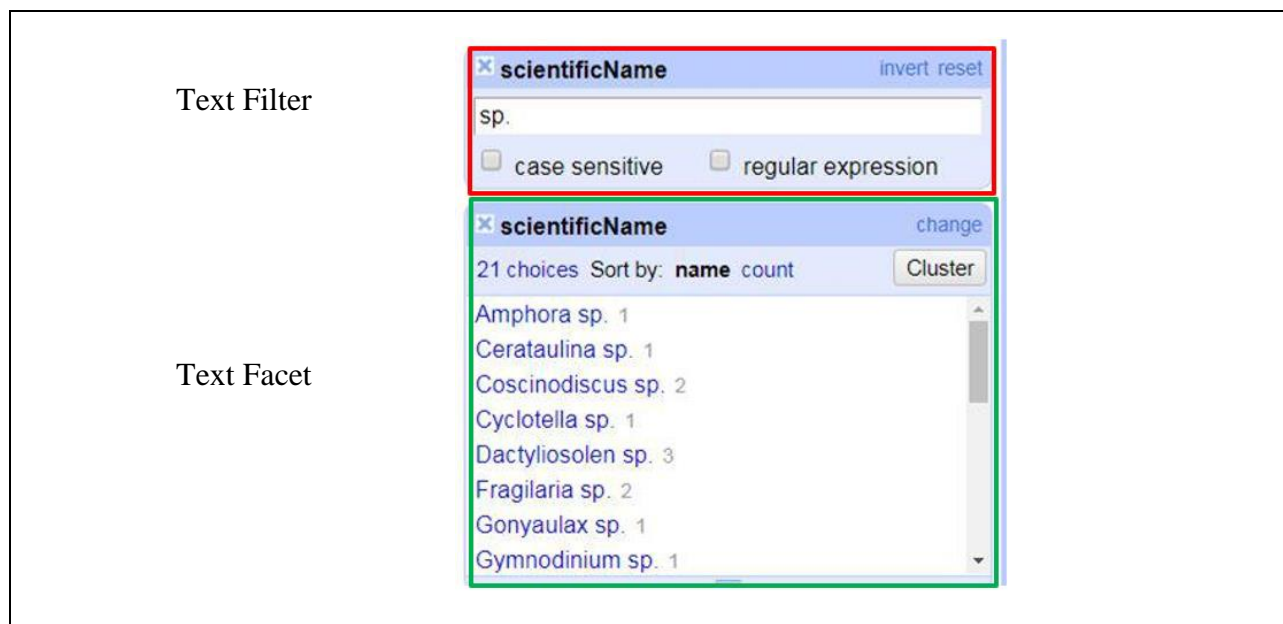
5. Al finalizar este ejercicio diríjase en el menú lateral y seleccione la opción “*Remove All*”. Así removerá todos los *Facets* y Filtros que tenga en uso.



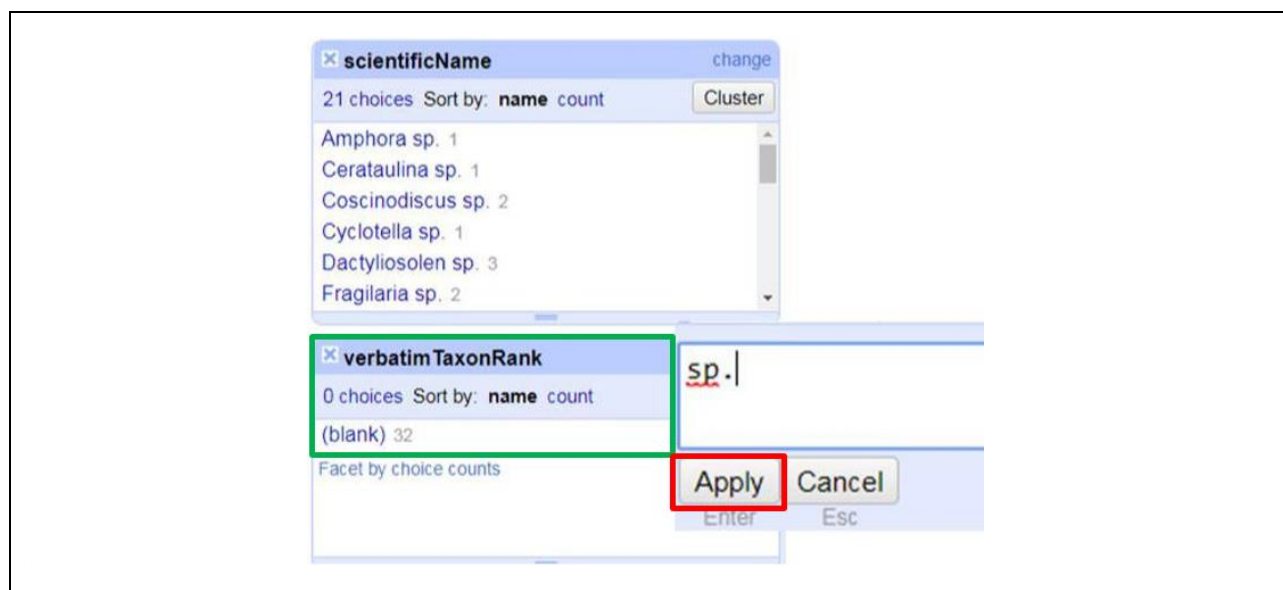
4.4.1.3 Filtering


Esto permite realizar filtros básicos y reemplazo de valores.

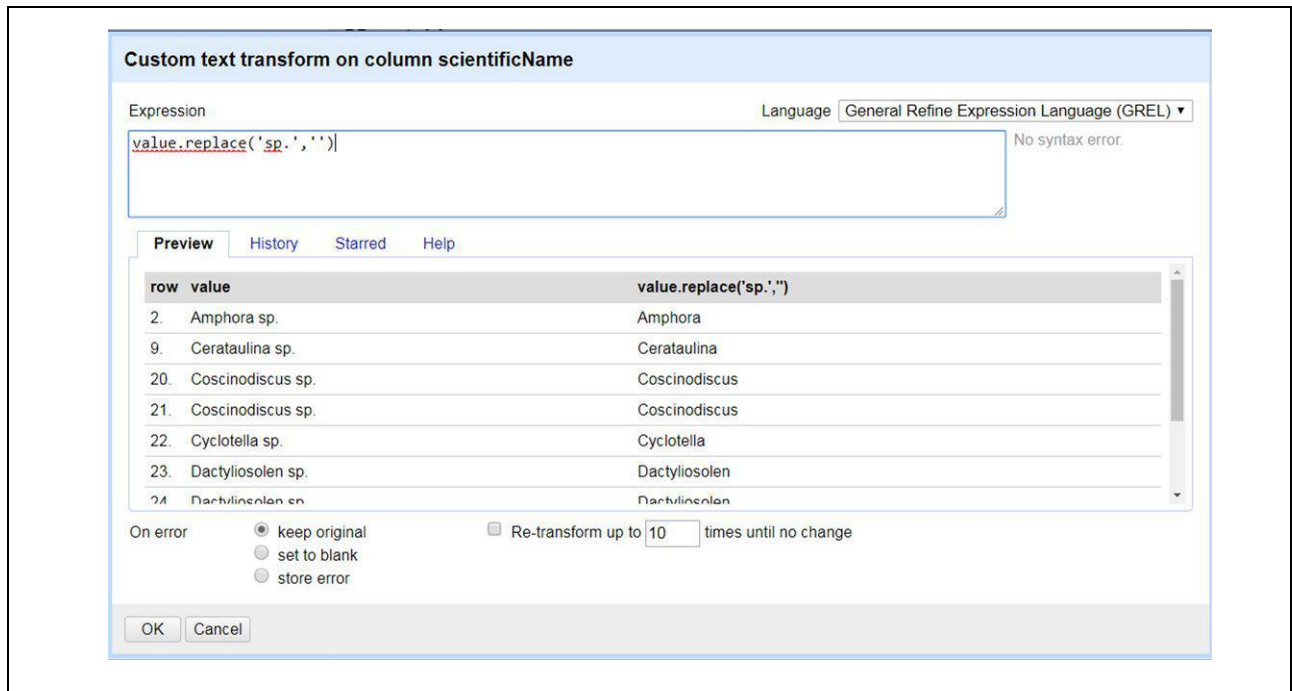
1. Diríjase a la columna a tratar (en este ejemplo *scientificName*), haga clic en el Menú Columna  y luego en *Text filter*, aparecerá la ventana del Filtro.
2. Escriba en el campo de texto sp. y realice un *Text Facet* en *scientificName* para visualizar los registros con este valor. Este tipo de clasificaciones no determinadas (ejemplo: sp.) no deben documentarse en el campo *scientificName*, para ello se emplea el campo *verbatimTaxonRank*.



3. Realice un *Text Facet* en *verbatimTaxonRank* y edite masivamente reemplazando las celdas vacías (*blank*) con *sp.*, haga clic en *Apply*.



4. Diríjase nuevamente al menú Columna  de *scientificName* y siga la ruta *Edit cells* → *Transform*, luego ingrese la fórmula → *value.replace* (“ sp.”,””) tal y como se muestra a continuación:



5. Haga clic en OK y verá el mensaje de confirmación de que los cambios se han realizado.


Nota: Empleando este comando *value.replace*, podemos sustituir cualquier valor de una columna poniendo dentro de un paréntesis inicialmente el valor a buscar (ej. "sp"), entre comillas ["X"] y luego separado por una coma [,] el valor de reemplazo (en este caso ninguno por lo cual se ponen unas comillas vacías [""]).

6. Al finalizar este ejercicio diríjase en el menú lateral y seleccione la opción *Remove All*. Así removerá todos los *Facets* y Filtros que tenga en uso.

4.4.1.4 Clustering

Es una herramienta que permite agrupar en un campo los valores que presentan similitud, dando la opción de cambiar algunos de los valores para que estos queden de forma homogénea.

El ejemplo a continuación muestra un *Clustering* realizado en el campo *recordedBy*:

1. Diríjase a la columna *recordedBy*, haga clic en el Menú Columna  y luego en *Text facet*, aparecerá la ventana del *Facet* con las diferentes entradas de datos (*choices*):



2. En la parte superior derecha verá el botón *Cluster* haga clic, aparecerá la ventana de *Cluster Edit* para la columna *recordedBy*.
3. Podrá ver la siguiente información:

Contenido del <i>Cluster & Edit</i>	
<i>Cluster size</i>	La cantidad versiones que el algoritmo muestra como similares.
<i>Row count</i>	El número de registros por cluster.
<i>Values in cluster</i>	Los valores seleccionados por el algoritmo para esa agrupación y el número de registros por valor.
<i>Merge</i>	En este cuadro se selecciona si los valores se fusionan en el valor que propone el algoritmo por defecto.
<i>New cell value</i>	En este campo de texto se puede escribir un valor completamente nuevo para el clúster. También se puede hacer clic en cualquier valor para asignarlo como valor por defecto.

Cluster & Edit column "recordedBy"

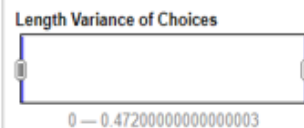
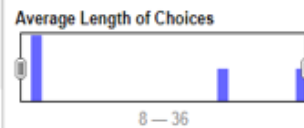
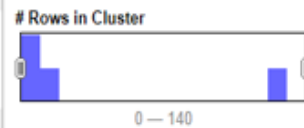
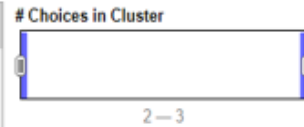
This feature helps you find groups of different cell values that might be alternative representations of the same thing. For example, the two strings "New York" and "new york" are very likely to refer to the same concept and just have capitalization differences, and "Gödel" and "Godel" probably refer to the same person. [Find out more ...](#)

Method key collision

Keying Function fingerprint

4 clusters found

Cluster Size	Row Count	Values in Cluster	Merge?	New Cell Value
3	4	<ul style="list-style-type: none"> David H; Arango A; Bedoya J (2 rows) David H; Arango A; Bedoya J (1 rows) David H; Arango A; Bedoya J (1 rows) 	<input type="checkbox"/>	David H; Arango A; Bedoya J
3	132	<ul style="list-style-type: none"> Vargas I (130 rows) Vargas I (1 rows) Vargas I (1 rows) 	<input type="checkbox"/>	Vargas I
2	3	<ul style="list-style-type: none"> Rodríguez Wilson; Giraldo F; Marín M (2 rows) Wilson Rodríguez; Giraldo F; Marín M (1 rows) 	<input type="checkbox"/>	Rodríguez Wilson; Giraldo F; Marín
2	12	<ul style="list-style-type: none"> Giraldo D (11 rows) D Giraldo (1 rows) 	<input type="checkbox"/>	Giraldo D



Select All Unselect All

Export Clusters

Merge Selected & Re-Cluster

Merge Selected & Close

Close

4. Vaya a *Keying Function*, seleccione *ngram-fingerprint* y en *Ngram Size* escriba 1.

Cluster & Edit column "recordedBy"

This feature helps you find groups of different cell values that might be alternative representations of the same thing. For example, the two strings "New York" and "new york" are very likely to refer to the same concept and just have capitalization differences, and "Gödel" and "Godel" probably refer to the same person. [Find out more ...](#)

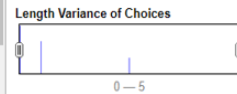
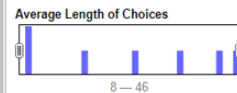
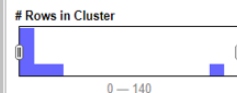
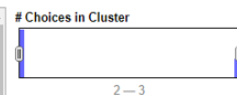
Method key collision

Keying Function ngram-fingerprint

Ngram Size 1

7 clusters found

Cluster Size	Row Count	Values in Cluster	Merge?	New Cell Value
3	4	<ul style="list-style-type: none"> David H; Arango A; Bedoya J (2 rows) David H; Arango A; Bedoya J (1 rows) David H; Arango A; Bedoya J (1 rows) 	<input type="checkbox"/>	David H; Arango A; Bedoya J
3	132	<ul style="list-style-type: none"> Vargas I (130 rows) Vargas I (1 rows) Vargas I (1 rows) 	<input type="checkbox"/>	Vargas I
2	3	<ul style="list-style-type: none"> Rodríguez Wilson; Giraldo F; Marín M (2 rows) Wilson Rodríguez; Giraldo F; Marín M (1 rows) 	<input type="checkbox"/>	Rodríguez Wilson; Giraldo F; Marín
2	15	<ul style="list-style-type: none"> López J; Idárraga A; Correa D; Sánchez L (13 rows) López J; Idárraga A; Correa D; Sánchez Lorena (2 rows) 	<input type="checkbox"/>	López J; Idárraga A; Correa D; Sán
2	12	<ul style="list-style-type: none"> Giraldo D (11 rows) D Giraldo (1 rows) 	<input type="checkbox"/>	Giraldo D
2	4	<ul style="list-style-type: none"> Molina L; Agudelo A; González A; Sierra S; Zúñiga F (3 rows) 	<input type="checkbox"/>	Molina L; Agudelo A; González A; S



Select All Unselect All

Export Clusters

Merge Selected & Re-Cluster

Merge Selected & Close

Close

- Tras realizar el *Cluster* si encuentra que se agrupo de forma correcta y equivalen al mismo valor y solo tienen distinto formato seleccione el cuadrado en *Merge*, eligiendo algunos de los valores o si ninguno de ellos es correcto modifíquelo en *New Cell Value*.
- Una vez escoja las entradas que desee fusionar o modificar y vaya a *Merge Selected & Close* para agrupar los valores y volver a la ventana principal. El resultado del proceso debería verse así:

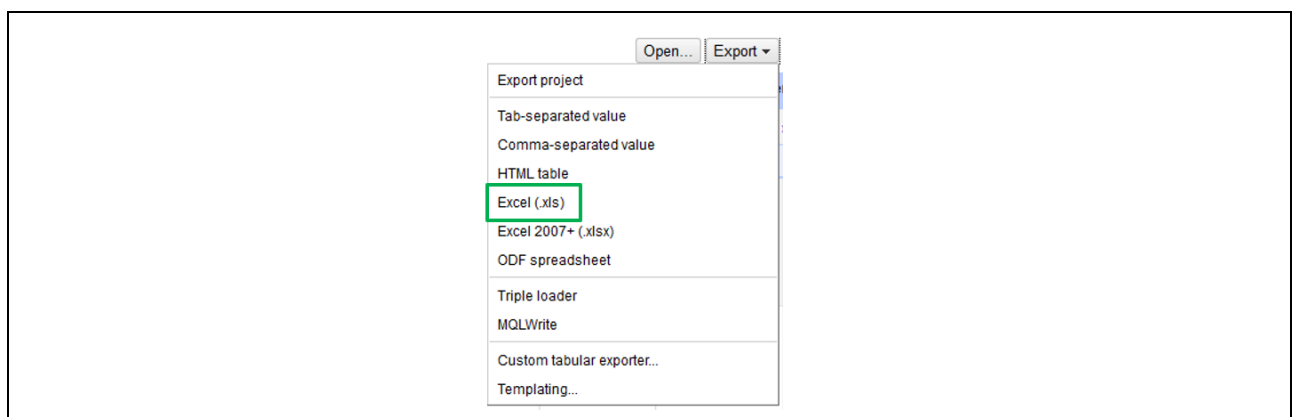


- Al finalizar este ejercicio diríjase en el menú lateral y seleccione la opción *Remove All*. Así removerá todos los *Facets* y Filtros que tenga en uso.

4.4.1.5 Exportación del archivo

Existen múltiples maneras de exportar los archivos en OpenRefine, la siguiente es la que ha mostrado funcionar en todos los casos.

- En la esquina superior derecha haga clic en el botón *Export* y seleccione la opción Excel. Se descargará automáticamente el archivo.



4.5 Configuración de Microsoft Excel para mejorar la gestión de datos

4.5.1 Funciones de Excel para la gestión de datos

4.5.1.1 Función Concatenar

Existen campos en el estándar DwC que se forman a partir de la unión de dos o más campos, ejemplo: Campo ***occurrenceID***, para construirlo son necesarios los elementos ***institutionCode***, ***collectionCode*** y ***catalogNumber***, que deben estar documentados en el estándar. Para realizar la construcción de este identificador de manera masiva en el conjunto de datos, puede utilizar una función de combinación en Excel como:

= concatenar(Valor instituteCode;";";Valor collectionCode;";";Valor catalogNumber)

Arrastre esta función para que se ejecute en todos los registros. Una vez asignados los identificadores en el campo ***occurrenceID*** selecciónelos en su totalidad, copie y pegue sobre las mismas celdas como formato “solo valores” para garantizar que las celdas guarden los datos como texto y no como fórmula.

4.5.1.2 Función duplicados del campo catalogNumber

Es necesario realizar la función duplicados, para la verificación de los valores duplicados que se encuentran en el ***catalogNumber***, ya que este número debiese ser único e irrepetible de cada muestra.

Para ello se utilizó la función formato condicional, realizando los siguientes pasos:

- Seleccionar la columna catalogNumber
- Diríjase a Inicio
- Formato Condicional
- Reglas para resaltar celdas
- Valores duplicados
- Realizar filtro en la columna de acuerdo con el color de celda, y ordenados de menor a mayor, así se podrán visualizar de mejor manera los valores duplicados de las celdas.

1	catalogNumber	scientificName	acceptedNameUsage
2	1	Lophosoria quadripinnata	Lophosoria quadripinnata
3	1	Acacia caven	Vachellia caven
4	2	Acacia caven	Vachellia caven
5	2	Lophosoria quadripinnata	Lophosoria quadripinnata
6	3	Acacia caven	Vachellia caven
7	3	Polystichum chilense	Polystichum chilense var. dusenii

4.5.2 Importar archivos de texto en Excel

Existen tres formatos para el almacenamiento de archivos de texto con potencial implementación en forma de tabla, y que **no** son los formatos tradicionales de *Microsoft Office Excel*® (.xls, .xlsx, etc), estos archivos son los **delimitados por tabulaciones** y los **separados por comas**.

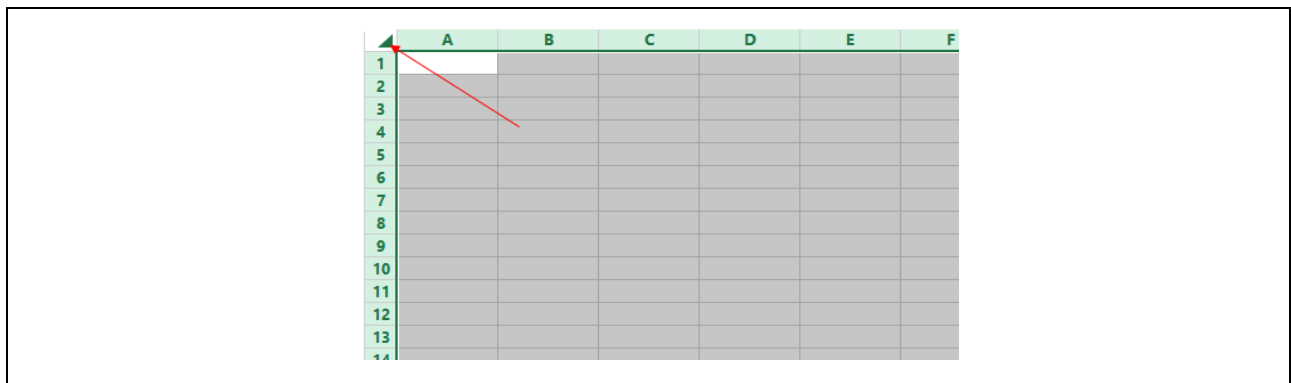
- a) **Archivos de valores separados por tabulaciones (.tsv):** en este formato el carácter de tabulación (tab) separa cada campo de contenido almacenado.
- b) **Archivos de texto delimitados por tabulaciones (.txt):** en este formato el carácter de tabulación (tab) delimita cada campo de contenido almacenado en un archivo de texto.
- c) **Archivos de valores separados por comas (.csv):** en este formato el carácter de coma (,) o de punto y coma (;) separa cada campo de contenido almacenado.

4.5.2.1 Adecuación de la hoja de cálculo

Este procedimiento se realiza previamente en la tabla ajustándola a un formato único con el fin de evitar errores ligados principalmente a números y/o fechas.

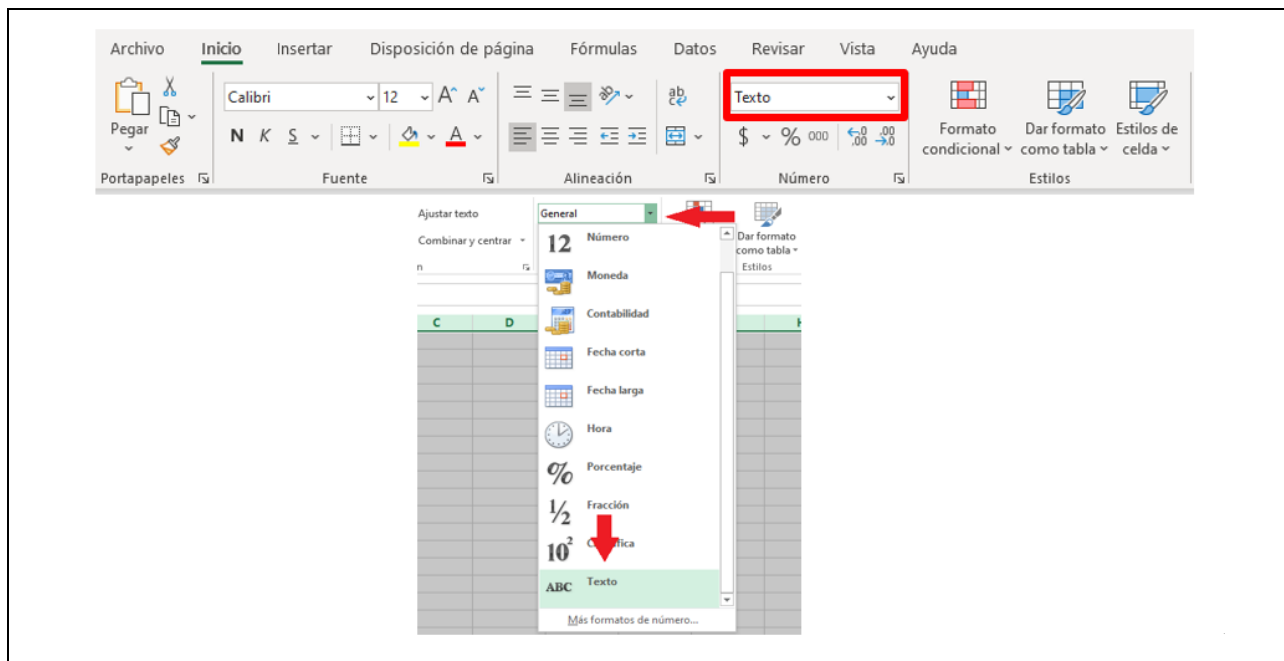
Paso 1

De clic en la esquina superior izquierda de la hoja de cálculo (espacio entre la columna A y la fila 1), de esta manera seleccionará la hoja de cálculo en su totalidad.



Paso 2

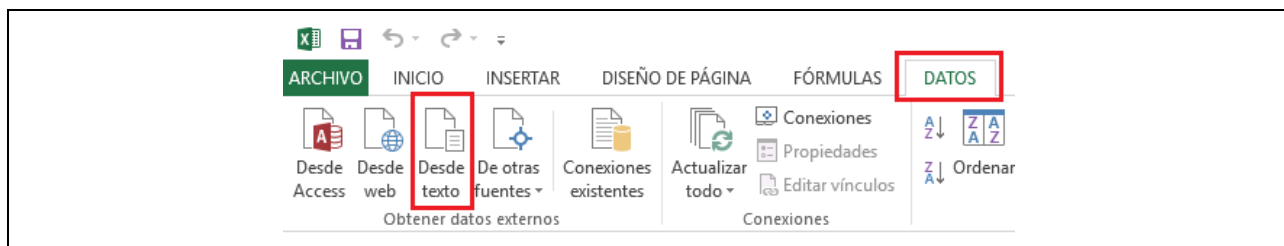
Al abrir un archivo en *Excel* siempre aparecerá seleccionado el formato “**General**” de forma predeterminado. De clic en el desplegable del formato de la hoja de cálculo, y al final de la lista seleccione “**Texto**”.



4.5.2.2 Importar archivo .txt

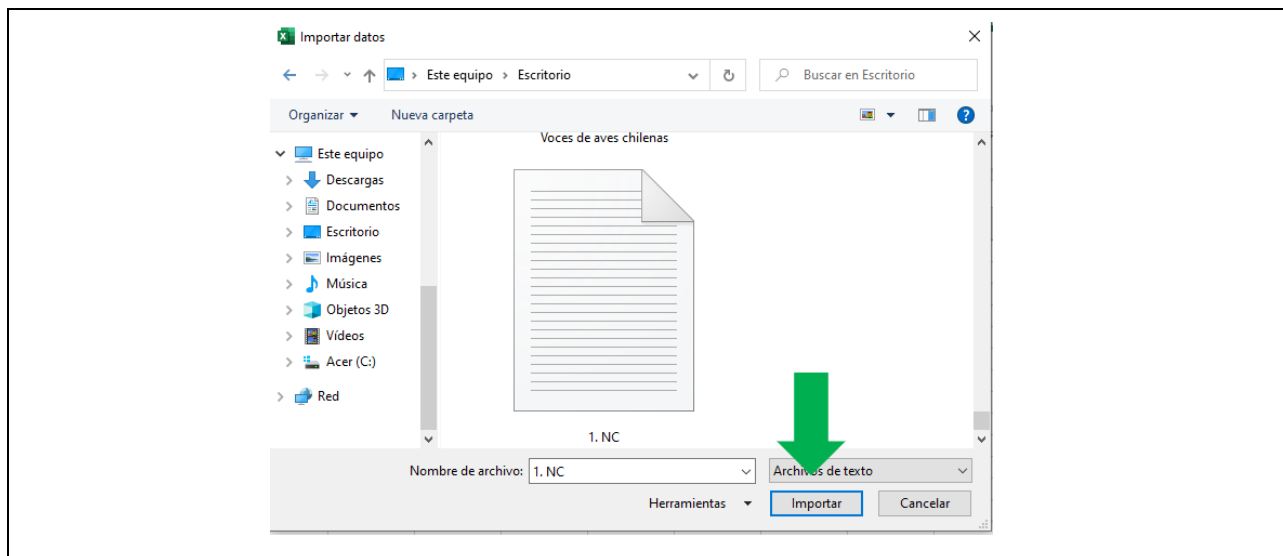
Paso 1

Seleccione la pestaña “**Datos**” del menú superior. Luego escoja obtener datos externos “**Desde texto**”.



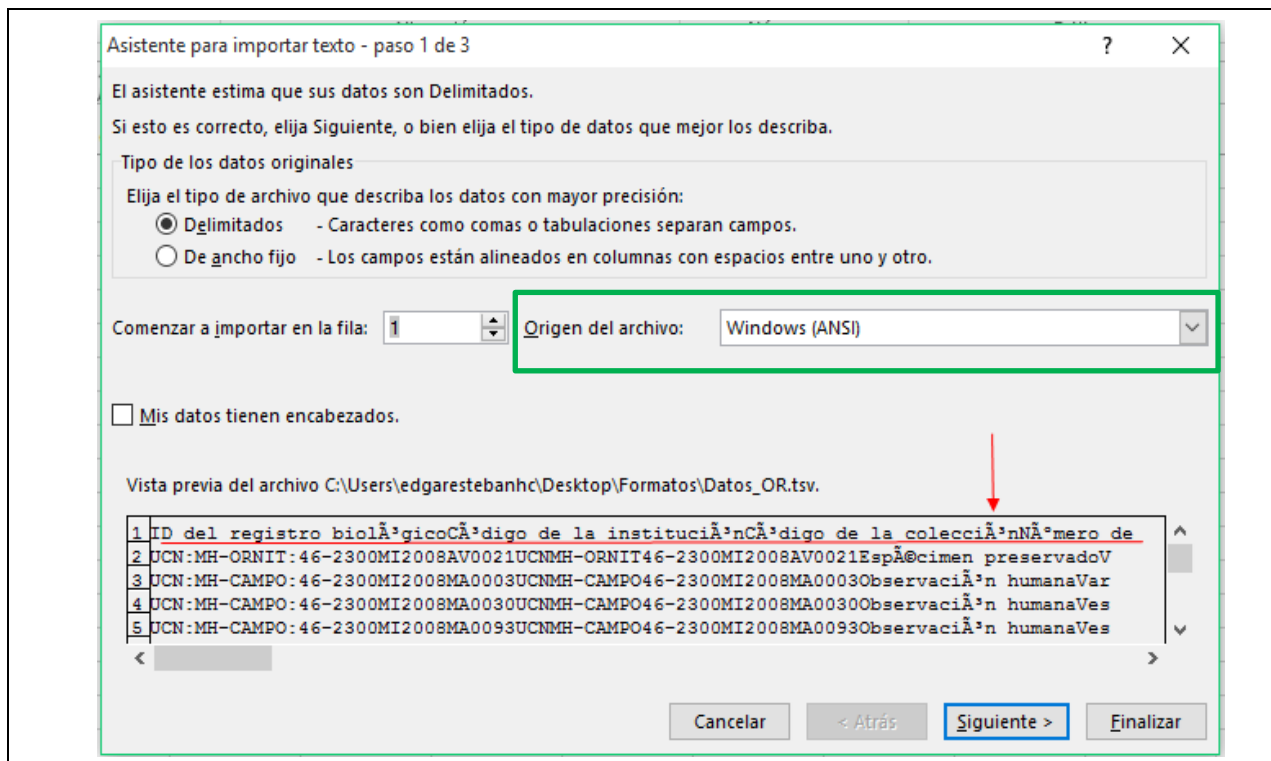
Paso 2

Diríjase a la carpeta donde tiene su archivo de texto de interés y haga clic en **Importar**.



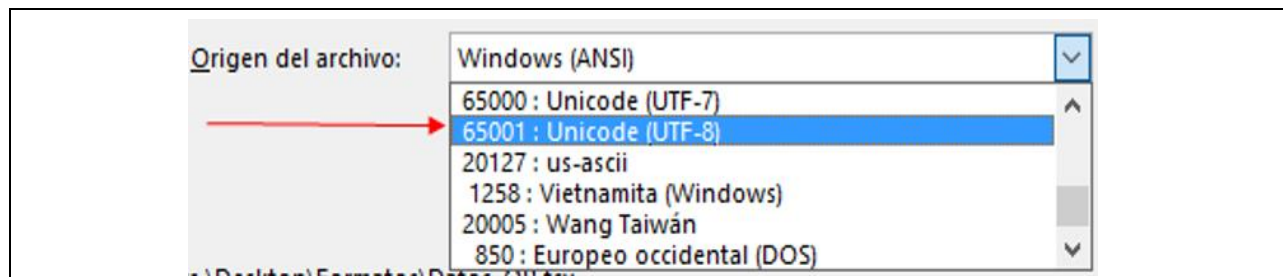
Paso 3

Excel abrirá el “Asistente para importar texto”, el cual cuenta con tres pasos. Nótese que la primera fila del archivo (encabezado) aparece con caracteres raros (ver la flecha roja).



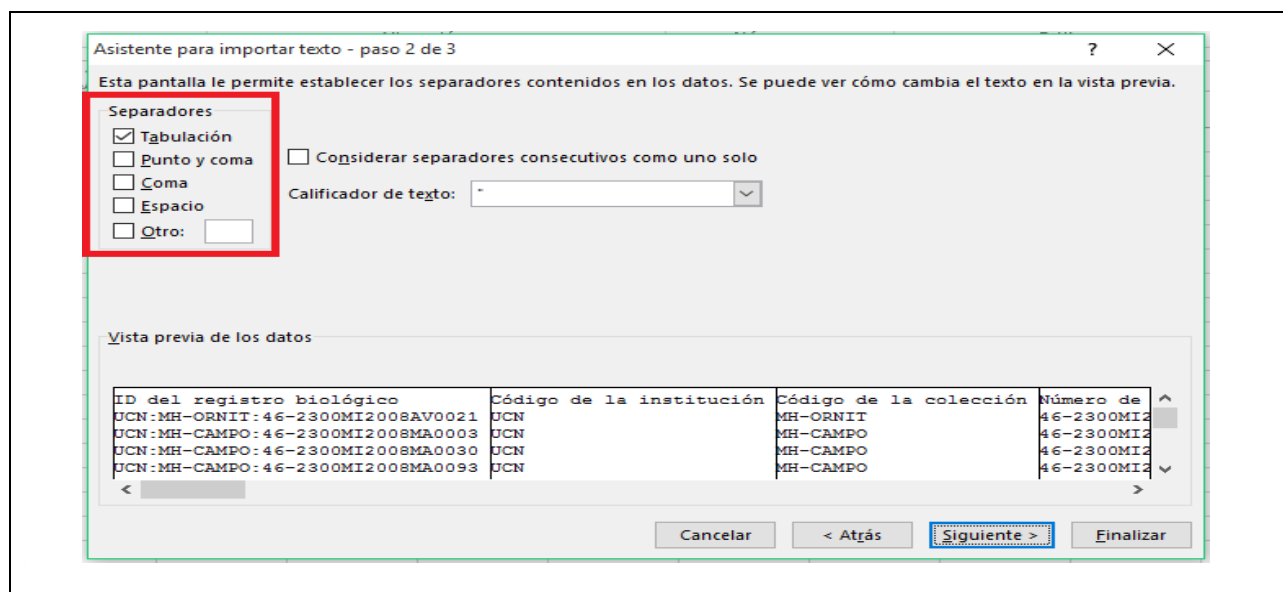
Paso 4

Para solucionar el problema de los “caracteres raros”, busque en **Origen del archivo** la codificación “**Unicode (UTF-8)**”, que se encuentra cerca del final de la lista desplegable del lado derecho del “**Asistente para importar texto**” y luego haga clic en **Siguiente >**





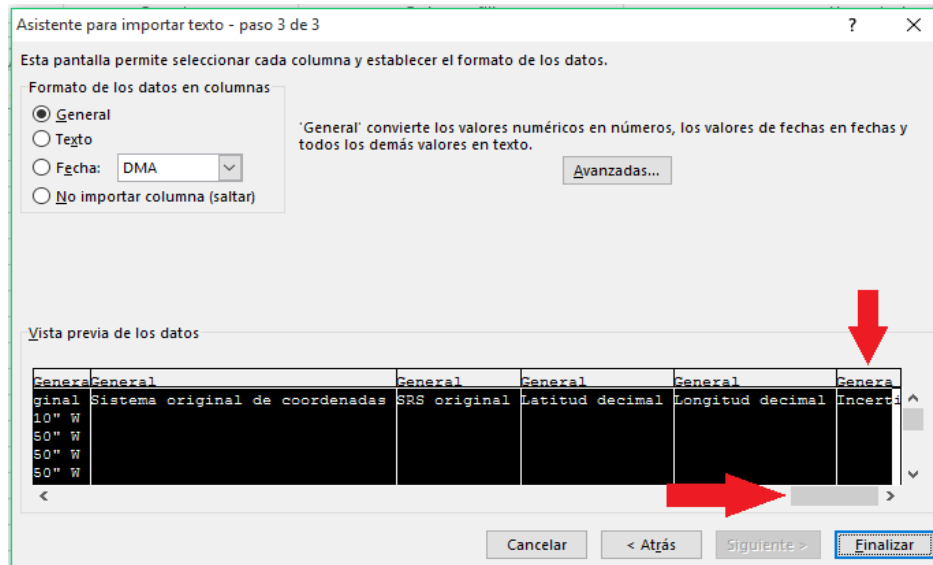
Paso 5

En el segundo paso del “Asistente para importar texto” debe escoger el tipo de separador que corresponda, y luego de clic en **Siguiente >**



Paso 6

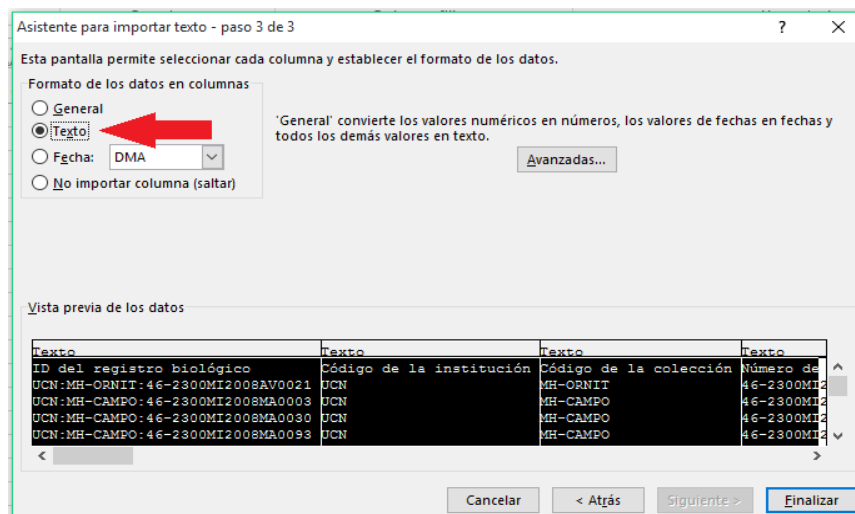
En el tercer paso del “**Asistente para importar texto**”, debe hacer clic sobre la primera columna del conjunto de datos, luego presione la tecla “**Shift**”  y sin soltarla arrastre la barra de deslizamiento hasta el límite derecho donde observará la última columna del conjunto de datos, de clic sobre esta columna y suelte la tecla “**Shift**” . De esta manera usted habrá seleccionado todas las columnas del conjunto de datos.



Paso 7

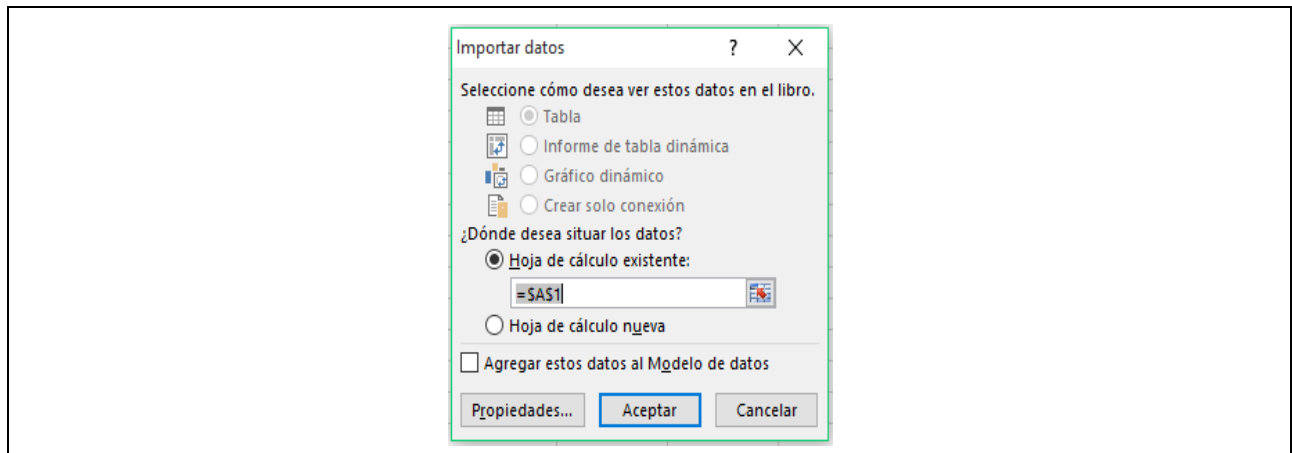
Luego seleccione **"Texto"** en el **"Formato de los datos en columnas"** y de clic en

Finalizar



Paso 8

En la ventana de Importar datos haga clic en aceptar.



Paso 9

A continuación, tendrá el archivo correctamente importado en Excel.

5 HERRAMIENTAS DE PUBLICACION VÍA GBIF: INTEGRATED PUBLISHING TOOLKIT

5.1 Proceso de publicación

La herramienta *Integrated Publishing Toolkit* (IPT) permite compartir el conjunto de datos que se ha tratado a través de la creación de un nuevo recurso, en el cual se cargará el archivo reconociendo los campos que estén adecuados al estándar darwinCore, posteriormente se desplegará la opción de documentar los metadatos. Tras completar los metadatos, el recurso se encontrará en condiciones de ser publicado y visible en GBIF.

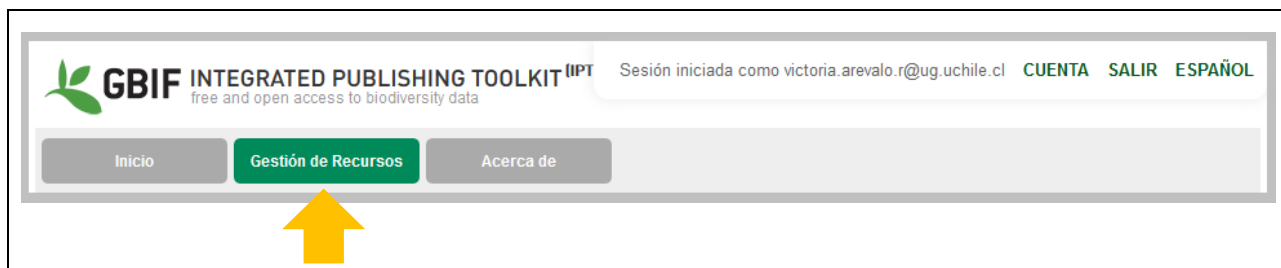
Previo a la carga de datos se deben contactar con el Nodo GBIF correspondiente a su país. En el caso de Chile con el Nodo GBIF-Chile a cargo del Ministerio del Medio Ambiente para realizar el registro de usuario en la plataforma IPT.

A continuación, se detallan los pasos para la carga del conjunto de datos y documentación de metadatos de la “[Guía básica del IPT](#)” (Beltrán *et al.*, 2016).

5.2 Carga de datos

Paso 1

Ingresa a la [Plataforma IPT](#) con el usuario y contraseña asignados, diríjase al botón del menú “Gestión de Recursos” (1), podrá ver en la parte inferior de la página una sección que le permite crear un nuevo recurso (2) o editar los existentes.



Paso 2

Al final de la página de Gestión de recursos diríjase a la sección “Crear nuevo recurso”. Asigne un “Nombre corto” para identificar de manera única el recurso en la instancia del IPT, el cual será utilizado también en la URL para acceder al recurso. Este nombre no es el que se verá en la publicación final, más bien es para que el usuario identifique el nombre del recurso de manera simple.

El nombre debe contener al menos tres caracteres, estos podrán ser caracteres alfanuméricos, pero no deberá incluir espacios en blanco o puntuación, excepto guiones o guiones bajos.

Crear un Nuevo Recurso

Puede crear un nuevo recurso en blanco, subir un recurso existente guardado como un archivo Darwin Core comprimido, o subir un recurso existente en un IPT, utilizando su carpeta de configuración de recursos en formato .zip. Por favor, consulte el Manual del Usuario para obtener instrucciones más específicas. Se requiere un nombre corto.

Nombre corto

Tipo

Importar desde archivo ☐

Crear

Paso 3

Seleccione un Tipo de recurso de la lista desplegable, dependiendo la naturaleza de los datos que piensa publicar. Para conjuntos de datos provenientes de herbarios este correspondería a “Ocurrencia”. En la siguiente imagen el recuadro rojo indica la opción que se debe seleccionar en la lista desplegable del “Tipo” de recurso.

Crear un Nuevo Recurso

Puede crear un nuevo recurso en blanco, subir un recurso existente guardado como un archivo Darwin Core comprimido, o subir un recurso existente en un IPT, utilizando su carpeta de configuración de recursos en formato .zip. Por favor, consulte el Manual del Usuario para obtener instrucciones más específicas. Se requiere un nombre corto.

Nombre corto

Tipo

- Registros biológicos
- Registros biológicos
- Lista de chequeo
- Evento de muestreo
- Solamente metadatos
- Otro

IPT Version 2.4.0-r6f4f375

[Acerca del IPT](#)

[Manual del](#)

Paso 4

Luego de asignar el nombre y el tipo solo debe hacer clic en el botón "Crear". De esta manera será dirigido a la página “Vista general del recurso”.

Vista General: [sqf](#)

Esta es la página para la vista general del recurso *sqf*.

Conjunto de Datos

[Examinar...](#) No se ha seleccionado ningún archivo.

[Conectar a la base de datos](#) [Limpiar](#)

Sus archivos del conjunto de datos y fuentes SQL para generar un Archivo Darwin Core.

Mapeo Darwin Core

Su mapeo entre el conjunto de datos y los elementos del Darwin Core.

Metadatos

[Editar](#)

Los metadatos del recurso.

Versiones publicadas

[Publicar](#)

Autopublicación

[Seleccionar la frecue](#)

Una vista previa de la versión pendiente a ser publicada en comparación con la versión actual, si existe.

	Versión pendiente
Versión	1.0 Vista previa
Visibilidad	Private
Licencia de los datos	-
Publicado en	-

Visibilidad **Privado**

[Público](#)

Este recurso es privado para los administradores. Para hacer este recurso disponible a todo el mundo, hágalo público.

Administradores de Recursos

[Agregar](#)

[Eliminar](#)

El Administrador ha concedido permiso para modificar este recurso.

Creador Victoria Arevalo, victoria.arevalo.r@ug.uchile.cl

Paso 5

En la sección “Conjunto se Datos” se debe escoger la opción “Examinar”. Con el fin de evitar errores en la lectura de los formatos, se recomienda cargar el archivo en formato de “texto delimitado por comas” (.CSV). Al seleccionar el archivo el nombre aparecerá al lado derecho del botón “Examinar”. En caso de que desee cambiar el archivo escogido, haga clic en limpiar y vuelva a seleccionar nuevamente el archivo.

Conjunto de Datos

[Examinar...](#) EIF -Mono&Rosid.xlsx

[Agregar](#) [Limpiar](#)

Paso 6

Una vez haya asegurado que el archivo seleccionado es el correcto, haga clic en el botón “Agregar” para asociar este archivo al recurso. Esto lo llevará a la página de información del conjunto de datos, allí podrá ver si su conjunto fue añadido correctamente y encontrará información detallada de la fuente de datos.

Título del recurso **PBUCH_SQF**

Datos fuente

Edite su formato de datos fuente

Nombre de la fuente

dwc-sqf

Legible

☒

Columnas

49

Expediente

/opt/ipt/resources/pbuch_sqf/sources/dwc-sqf.txt

tamaño

1.5 MB

filas

2,096

Modificado

28 de febrero de 2019 6:51:55 PM

Registro de origen

[Descargar](#)

Analizar

Número de filas de encabezado

1

Cotizaciones de campo

"

Codificación de caracteres

ISO-8859-1

Guardar

Cancelar

Eliminar archivo fuente

Delimitador de campo

,

Delimitador multivalor

"

Formato de fecha

YYYY-MM-DD

Paso 7

Verifique que el número de Columnas y Filas de su conjunto de datos coincida con lo reportado en esta página.

En caso de que haya diferencias es posible que IPT esté interpretando filas o columnas vacías como si tuvieran contenido, para solucionar este error debe dar clic en el botón “Eliminar”, luego corregir desde el archivo original eliminando todas las filas y columnas vacías y finalmente volver a seleccionar y cargar el archivo en su recurso. Otro error común es que aparezcan menos campos (Columnas), si el nombre de la columna está escrito de forma distinta a los que se encuentran en el estándar DwC, no los reconocerá. Para ello es necesario verificar en su base de datos estandarizada que estén bien escritos los campos.

Después de revisar todos los campos y realizar cambios pertinentes, haga clic en el botón guardar para almacenar la configuración. Si el almacenamiento se realizó de forma correcta, aparecerá la página de la vista general del recurso y en la parte derecha del área “Datos Fuente” la

información resumida del archivo, en la siguiente imagen se puede distinguir la vista general del recurso, en recuadro rojo se encuentra información resumida del archivo.

Vista General: [Plantas Vasculares del Herbario EIF](#)

Esta es la página para la vista general del recurso *Plantas Vasculares del Herbario EIF*.

Conjunto de Datos

Examinar...
No se ha seleccionado ningún archivo

Conectar a la base de datos
Limpiar

Sus archivos del conjunto de datos y fuentes SQL para generar un Archivo Darwin Core.

Última modificación Jun 2, 2020

plantas-vasculares-eif_2 [archivo] 3.1 MB, 5,238 filas, 59 columnas. Jun 24, 2019 Editar

Mapeo Darwin Core

Darwin Core Occurrence
Agregar

Su mapeo entre el conjunto de datos y los elementos del Darwin Core.

No ha sido modificado desde la última publicación

Core

Darwin Core Occurrence 59 elementos mapeados en plantas-vasculares-eif_2. Jun 24, 2019 Editar

Metadatos

Editar

Los metadatos del recurso.

Última modificación May 12, 2020

Paso 8

La sección Mapeo darwinCore (“*Darwin Core Mappings*” o Asignaciones del núcleo de Darwin) permite mapear los datos cargados anteriormente, adicional al nombre del campo, se verifican campos con formatos específicos o vocabulario controlado. Ésta posee una lista desplegable donde se tendrá que seleccionar la opción más adecuada a su recurso. En este caso, de herbario o para otra colección biológica se seleccionó “Darwin Core Ocurrencia”, para ello se debe hacer clic en el botón “Añadir” como se distingue a continuación donde (1) opción de la lista desplegable a seleccionar y (2) corresponde al botón “añadir”.

Tras esto será dirigido a la página de selección del conjunto de datos.

Asignaciones del núcleo de Darwin

Darwin Core ocurrencia
Core

1
Darwin Core ocurrencia
Darwin Core Taxon
Darwin Core Event

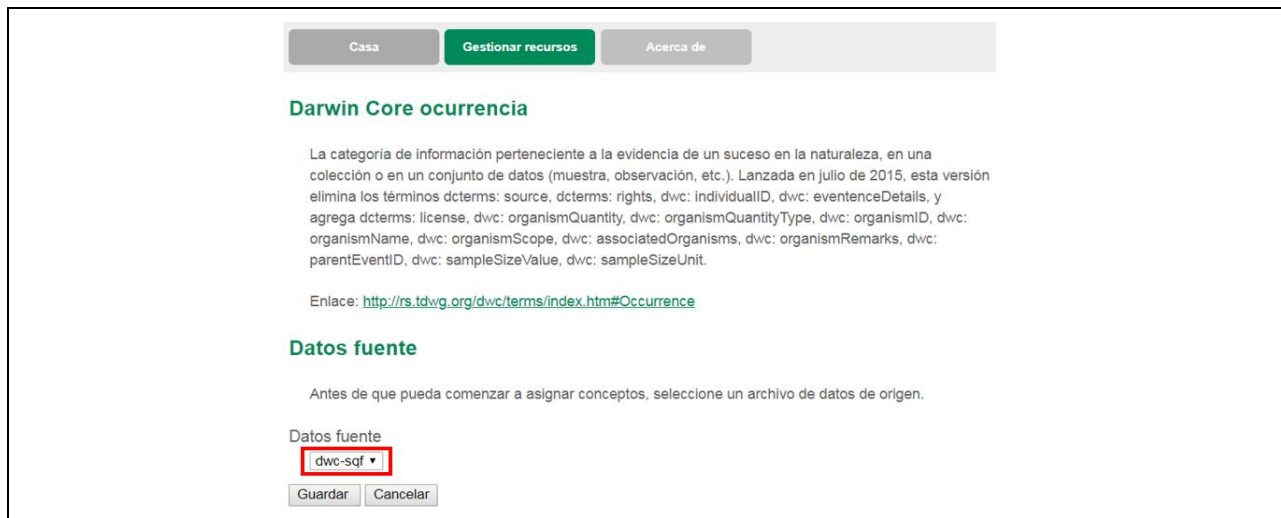
Asignaciones del núcleo de Darwin

Darwin Core ocurrencia

2
Añadir

Paso 9

En esta página en la parte superior aparece una breve descripción del tipo de Core seleccionado (*Occurrence*, *Checklist*, *Sampling event*, *Metadata*) y en la parte inferior la opción para escoger el archivo que se desea mapear. Seleccione el archivo a mapear (su nombre debe coincidir con el archivo seleccionado en el “Paso 5”) y haga clic en el botón “Guardar”.



Casa Gestionar recursos Acerca de

Darwin Core ocurrencia

La categoría de información perteneciente a la evidencia de un suceso en la naturaleza, en una colección o en un conjunto de datos (muestra, observación, etc.). Lanzada en julio de 2015, esta versión elimina los términos dcterms: source, dcterms: rights, dvc: IndividualID, dvc: eventenceDetails, y agrega dcterms: license, dvc: organismQuantity, dvc: organismQuantityType, dvc: organismID, dvc: organismName, dvc: organismScope, dvc: associatedOrganisms, dvc: organismRemarks, dvc: parentEventID, dvc: sampleSizeValue, dvc: sampleSizeUnit.

Enlace: <http://rs.tdwg.org/dwc/terms/index.htm#Occurrence>

Datos fuente

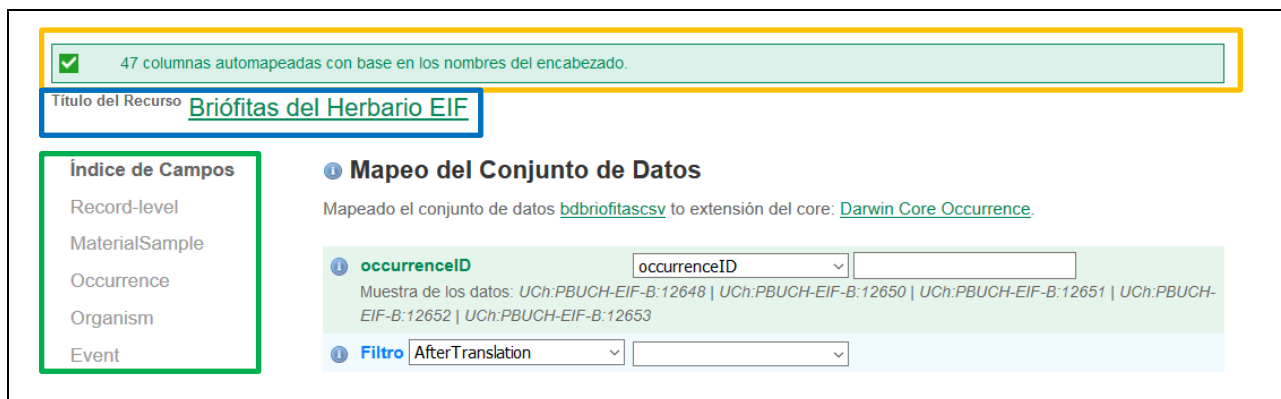
Antes de que pueda comenzar a asignar conceptos, seleccione un archivo de datos de origen.

Datos fuente

dwc-sqf

Guardar Cancelar

Posteriormente será dirigido a la página detallada del mapeo del conjunto de datos encontrará en la parte superior un mensaje con el estado del mapeo, que indica el número de columnas que fueron mapeadas automáticamente, debajo de este mensaje se encuentra el título del recurso y en el panel izquierdo el “Índice de Campos” con las secciones del estándar DwC, navegue a través de las secciones para observar los elementos mapeados.



✓ 47 columnas automapeadas con base en los nombres del encabezado.

Título del Recurso **Briófitas del Herbario EIF**

Índice de Campos

- Record-level
- MaterialSample
- Occurrence
- Organism
- Event

Mapeo del Conjunto de Datos

Mapeado el conjunto de datos [bdbriofitascsv](#) to extensión del core: [Darwin Core Occurrence](#).

occurrenceID occurrenceID

Muestra de los datos: UCh:PBUCH-EIF-B:12648 | UCh:PBUCH-EIF-B:12650 | UCh:PBUCH-EIF-B:12651 | UCh:PBUCH-EIF-B:12652 | UCh:PBUCH-EIF-B:12653

Filtro AfterTranslation

Paso 10

Revise que la cantidad de campos mapeados automáticamente coincida con el total de sus campos en estándar DwC, navegue a través de las secciones para observar los elementos mapeados. Una vez finalice exitosamente el mapeo de elementos haga clic en el botón “Guardar”

que se encuentra en la parte izquierda de la pantalla al final del Índice de Campos y luego regrese a la vista general del recurso.

Donde (1) Mensaje que muestra cuántas columnas fueron mapeadas automáticamente de acuerdo con los nombres de los encabezados. (2) Índice de campos del estándar DwC, puede consultar sección por sección cuáles elementos fueron mapeados. (3) Cuando un elemento ha sido mapeado aparece su nombre equivalente en la lista desplegable y una muestra de los datos que contiene el elemento. (4) Si al lado del elemento aparece un ícono de libro, significa que el campo cuenta con vocabulario controlado, al hacer clic en este ícono lo dirigirá a la página detallada del vocabulario.

The screenshot displays the 'Gestionar recursos' (Manage resources) page. At the top, there are navigation buttons: 'Casa' (Home), 'Gestionar recursos' (Manage resources), and 'Acerca de' (About). Below the navigation bar, a green message box (1) states: 'Automaticé 49 columnas basadas en nombres de encabezados.' (Automatically mapped 49 columns based on header names). The resource title is 'PBUCH SQF'. On the left, a sidebar (2) titled 'Índice de campo' (Field index) lists various categories: 'Nivel de registro', 'MaterialSample', 'Ocurrencia', 'Organismo', 'Evento', 'Ubicación', 'Contexto geológico', and 'Identificación'. The main area is titled 'Mapeo de datos de origen' (Data source mapping) and shows the assignment of data from 'dwc-sqf' to the 'Ocurrencia de Darwin Core' extension. A table (3) displays mapped fields, with 'ocurrenceID' highlighted. Below the table, a 'Filtrar' (Filter) section is visible. At the bottom, the 'basisOfRecord' field is shown with a dropdown menu and a 'Traducción' (Translation) section with an 'Agregar' (Add) button. A red box (4) highlights a book icon next to the 'basisOfRecord' field, indicating a controlled vocabulary.


5.3 Elaboración de metadatos

Los metadatos se definen como “Los datos acerca de los datos” y describen el conjunto de datos en aspectos como “quién, qué, dónde, cómo, cuándo”; son de gran utilidad para que el usuario evalúe qué tan adecuados son los datos que consultará para su uso y permiten tener una visión general del contenido del conjunto de datos.

5.3.1 Procedimiento para la elaboración y carga de metadatos

Paso 1

Haga clic en el botón “Editar” en la sección metadatos.


Metadatos

[Editar](#)

Sus metadatos de recursos.

Última modificación el 21 de marzo de 2019

Al editar los metadatos se abrirá la página "Metadatos Básicos", la primera de una serie de páginas de metadatos. Al lado derecho de la página encontrará las secciones que se deben completar (señalado en recuadro rojo).

Título del recurso **PBUCH SQF**

Metadatos basicos

Ingrese todas las propiedades obligatorias en la página Metadatos básicos y luego continúe ingresando los metadatos en las otras páginas que sean aplicables a su recurso. Cuantos más metadatos proporcione, mayor será la posibilidad de que otros investigadores encuentren su recurso y sean citados.

Título*

PBUCH_SQF

Organización editorial *

Ministerio del Medio Ambio

Tipo*

Ocurrencia

Lenguaje de metadatos *

Spanish

Frecuencia de actualización*

Desconocido

Subtipo

Muestra

Idioma de los datos *

Spanish

Licencia de datos *

Creative Commons Reconocimiento No Comercial (CC-BY-NC) 4.0

Sección

- [Metadatos basicos](#)
- [Cobertura geografica](#)
- [Cobertura taxonomica](#)
- [Cobertura temporal](#)
- [Palabras clave](#)
- [Fiestas Asociadas](#)
- [Datos del proyecto](#)
- [Métodos de muestreo](#)
- [Citas](#)
- [Datos de recopilación](#)
- [enlaces externos](#)
- [Metadatos adicionales](#)

Paso 2

Complete la mayor cantidad de campos de las distintas secciones de metadatos, con la información disponible.

Es muy importante que al terminar la documentación de cada sección de clic en el botón "Guardar" presente al final de cada sección, en caso contrario los cambios no serán tenidos en cuenta y los metadatos quedarán incompletos.

5.3.2 Tipos de metadatos

5.3.2.1 Metadatos básicos

En esta sección se documentan campos como el título, la descripción, el tipo de licencia y los tres principales contactos del recurso (contacto, creador y proveedor de los metadatos). Esta sección cuenta con campos obligatorios.

CAMPO	DESCRIPCIÓN
Título	Es el título del recurso
Organización editorial	Organización responsable de la publicación (producción, liberación) del recurso.
Frecuencia de actualización	Frecuencia con la que se realizan cambios en el recurso después de que el recurso inicial ha sido publicado.
Tipo	Lista desplegable en la cual se debe seleccionar el tipo del conjunto de datos a compartir. Para herbarios corresponde Reg. Biológico
Subtipo	Depende del tipo escogido en el punto anterior. Para herbarios es Espécimen .
Idioma de los metadatos	El idioma en el cual están escritos los metadatos.
Idioma de los datos	El idioma en el cual están escritos los datos del recurso.
Licencia de los datos	Forma estandarizada de definir los usos apropiados de su recurso (Dominio público CC 1.0, Atribución CC BY 4.0 y Atribución-No comercial CC BY NC 4.0)
Descripción	Resumen de los aspectos más importantes que contiene el recurso
Contactos del recurso	Personas u organizaciones que deben ser contactadas para obtener más información acerca del recurso. Por lo menos debe contener: *APELLIDO; *POSICIÓN; *ORGANIZACIÓN
Creadores del recurso	Personas u organizaciones responsables de la creación original del contenido del recurso. Por lo menos debe contener: *APELLIDO; *POSICIÓN; *ORGANIZACIÓN
Proveedor de los metadatos	Personas u organizaciones responsables de producir los metadatos del recurso. Por lo menos debe contener: *APELLIDO; *POSICIÓN; *ORGANIZACIÓN

5.3.2.2 Cobertura geográfica

En esta sección se documenta la información del área geográfica que abarca el recurso. Si su conjunto de datos cuenta con coordenadas, pueden extraerse los datos de la latitud y longitud decimal e incluir una descripción acerca de la cobertura geográfica.

CAMPO	DESCRIPCIÓN
Establecer cobertura global	Marca la Delimitación geográfica del área cubierta por el recurso es a nivel global
Longitud mín. / máx.	Valores de la longitud decimal mínima y máxima
Latitud mín. / máx.	Valores de la latitud decimal mínima y máxima
Descripción	Descripción de la cobertura geográfica

5.3.2.3 Cobertura taxonómica

Esta sección permite al usuario incluir la información taxonómica del recurso. En la descripción se documenta de forma clara y sencilla los datos taxonómicos más relevantes del recurso y se listan los nombres científicos, comunes y el rango de éstos.

CAMPO	DESCRIPCIÓN
Descripción	Descripción del rango de taxones representados en el recurso
Añadir varios taxa	Este enlace abre un cuadro de texto llamado “lista de taxa”, en el cual se puede ingresar un taxón por línea, usando la tecla “Enter” para separar los taxones, éstos serán tratados como nombres científicos. Al terminar la lista debe dar clic en el botón agregar

5.3.2.4 Cobertura temporal

En esta sección se incluye información sobre el periodo de tiempo que abarca el recurso.

CAMPO	DESCRIPCIÓN
Tipo de cobertura temporal	Lista desplegable que permite escoger el tipo de cobertura temporal; Para el caso de herbarios esta corresponde a Rango de fechas.
Fecha de inicio	La fecha en que comenzó la cobertura
Fecha final	La fecha en que terminó la cobertura

5.3.2.5 Palabras Clave

Esta sección permite al usuario incluir palabras clave que pueden ayudar a dar una idea breve acerca del contenido del recurso. Las palabras clave pueden o no estar asociadas a un tesoro.

CAMPO	DESCRIPCIÓN
Tesoro/Vocabulario	Nombre de un tesoro o vocabulario controlado, del cual se derivó el conjunto de palabras clave. Si las palabras clave no están regidas por un tesoro o vocabulario, se ingresa “n/a”, que indica, no aplica.
Lista de palabras clave	Lista de palabras clave que describen o están relacionadas con el recurso, las palabras deben estar separadas por comas.

5.3.2.6 Partes asociadas

Esta sección contiene información sobre las personas u organizaciones que están asociadas con el recurso además de las que se incluyen en la sección de metadatos básicos. Si los datos de las partes asociadas son los mismos que los del contacto del recurso, éstos pueden copiarse haciendo clic en el enlace “Copiar datos del contacto del recurso”.

Partes Asociadas

Partes Asociadas de este recurso (p. ej. la institución propietaria).

[\[Copiar los datos del contacto del recurso \]](#)

[\[Eliminar esta\(e\) parte asociada \]](#)

Nombre

Apellido*

Posición*

Organización*

Dirección

Ciudad

Departamento/Estado/Provincia

País

Código Postal

Teléfono

Correo Electrónico

Página Web

Perfil en Línea

Identificador Personal

Rol

[Añadir nuevo\(a\) parte asociada](#)

Guardar

Cancelar

5.3.2.7 Datos del Proyecto

Esta sección documenta en caso de que los datos hayan sido producidos en el marco de un proyecto, contiene información acerca del nombre del proyecto, fuentes de financiación, descripción y personas asociadas al proyecto.

CAMPO	DESCRIPCIÓN
Título	Permite el ingreso a información acerca del proyecto bajo el cual se produjo el recurso. Esto es solo si los datos fueron producidos bajo un proyecto
Identificador	Identificador único para el proyecto de investigación
Descripción	Resumen del proyecto de investigación
Fondos	Información sobre las fuentes de financiamiento para el proyecto
Descripción del Área de Estudio	Documenta el área física asociada al proyecto de investigación. Puede incluir descripciones de la cobertura geográfica, temporal y taxonómica de la ubicación de la investigación

Descripción del diseño	Una descripción textual general del diseño de investigación. Puede incluir informes detallados de objetivos, motivaciones, teoría, hipótesis, estrategia, diseño estadístico y trabajo real.
Personal del proyecto	Personas involucradas en el proyecto debe contener: *APELLIDO *POSICIÓN *ORGANIZACIÓN

5.3.2.8 Métodos de Muestreo

En esta sección se incluye información acerca de los métodos de muestreo utilizados de forma general y los pasos específicos de éstos.

CAMPO	DESCRIPCIÓN
Área de Estudio	Representa tanto un área de muestreo específica como la frecuencia de muestreo. La extensión del estudio geográfico suele ser un sustituto para el área más amplia documentada en el campo "Descripción del área de estudio" de la página de metadatos del proyecto
Descripción del Muestreo	Este campo permite una descripción basada en texto de los procedimientos de muestreo utilizados en el proyecto de investigación.
Control de Calidad	representa una descripción de las acciones tomadas para controlar o evaluar la calidad de los datos resultantes de los pasos del método asociado
Descripción de etapas	Repetición de conjunto de elementos que documentan la serie de métodos, procedimientos utilizados y pasos conducentes a datos resultantes

5.3.2.9 Citas

En el elemento referencias de esta sección se explica cómo se debe citar el recurso e incluye el identificador de éste, adicionalmente, en el elemento bibliografía pueden documentarse los recursos bibliográficos relacionados con el conjunto de datos, tales como publicaciones que se usaron para generar los datos o que fueron resultado de la producción de éstos.

CAMPO	DESCRIPCIÓN
Cómo citar este recurso	Cita del recurso, se puede generar de manera automática dando clic al enlace superior "[Generación automática Desactivada - Activar]". Además, se puede crear manualmente la cita deseada.
Identificador de la citación para este Recurso (<i>URL o identificador único</i>)	URL persistente utilizado para referenciar el recurso. Se recomienda que el identificador esté incluido en la referencia del recurso.

Citas bibliográficas	Bibliografía que se utilizó para generar el recurso
----------------------	---

5.3.2.10 Datos de la colección

Esta sección debe documentarse sólo si los datos pertenecen a una colección biológica y contiene información acerca del nombre de las colecciones, el identificador, los métodos de preservación de los especímenes y las unidades curatoriales.

CAMPO	DESCRIPCIÓN
Colecciones	Nombre, Identificador de la colección de la colección parental.
Métodos de preservación de especímenes	Lista desplegable que permite escoger entre diferentes métodos de preservación.
Unidades Curatoriales	Tipo de método: Rango de Conteo o Conteo con Incertidumbre.

5.3.2.11 Enlaces Externos

En esta sección se incluyen los enlaces a la página web del recurso u otros enlaces alternos.

CAMPO	DESCRIPCIÓN
Página web del recurso	Contiene enlaces a la página web del recurso, así como enlaces alternos al recurso (archivo de bases de datos, hojas de cálculo, datos relacionados, etc.)
Otros formatos de datos	Nombre, conjunto de caracteres, URL del archivo, formato del archivo y versión del formato del archivo.

5.3.2.12 Metadatos Adicionales

Esta sección final sirve para documentar información no contemplada en ninguna de las secciones anteriores, se puede cargar el logo del recurso e incluir el propósito e información adicional del recurso, así como identificadores alternativos.

CAMPO	DESCRIPCIÓN
Fecha de Creación	Fecha en la cual la primera versión del recurso fue publicada.
Fecha de última Publicación	Fecha en la que el recurso fue publicado por última vez.
URL del Logo del Recurso	Logo representativo del recurso.
Propósito	Resumen del fin o los fines por los cuales se desarrolló el conjunto de datos Pueden incluirse los objetivos para crear el conjunto de datos y lo que busca soportar o apoyar éste.
Descripción de Mantenimiento	Descripción de la frecuencia de mantenimiento del recurso.

Información Adicional	Cualquier información de posible interés que no ha sido registrada en alguno de los otros campos para metadatos
Identificadores alternativos	Lista de identificadores adicionales o alternativos para el recurso.

5.4 Creación del Archivo Darwin Core

Esta sección crea el Archivo Darwin Core y se procede a la publicación del recurso. Cuando se desea publicar se debe notificar al equipo coordinador de GBIF-Chile sobre la generación de este archivo para que los administradores finalicen el proceso.

5.4.1 Versiones publicadas

Esta sección es la encargada de publicar una versión del recurso, generando el Archivo Darwin Core (A-DwC), el EML (formato .xml para metadatos) y el RTF (formato de metadatos en Word); el A-DwC compila el conjunto de datos mapeado y los metadatos en un archivo comprimido (.zip).

Teniendo listos los procesos anteriores y sin errores se procede a hacer clic en el botón “Publicar”, luego se despliega una ventana donde puede resumir los cambios que ha realizado a la versión, para que quede como respaldo de los cambios en las distintas versiones.

La autopublicación puede ser configurada para ser publicado automáticamente en una fecha programada, seleccionando algunas de las frecuencias de publicación. Para desactivar la autopublicación, seleccione “Desactivar” y presione “Publicar”.

The screenshot shows the 'Versiones publicadas' (Published Versions) interface. It includes a 'Publicar' button with a warning icon (labeled 1), an 'Autopublicación' section with a dropdown menu for frequency selection (labeled 2), and a comparison table between the 'Versión actual' (1.8) and 'Versión pendiente' (1.9) (labeled 3). The table shows publication dates and download links.

Versión actual		Versión pendiente	
Versión	1.8 Ver ✓	Versión	1.9 Vista previa
Publicado en	Jun 25, 2019	Publicado en	-
Reporte de publicación	Descargar	Reporte de publicación	-

(1) Botón de publicar la versión (2) Lista desplegable de autopublicación (C) Sección informativa del recurso publicado (vista previa y descarga del recurso).

5.4.2 Estado de la publicación

Después de hacer clic en el botón “Publicar” se abre la página de “Estado de Publicación”, en ella aparecerán mensajes de estado destacando el éxito o falla de cada etapa de la publicación. La publicación de una nueva versión es un evento de todo o nada, lo que significa que todos los pasos deben terminar con éxito o de lo contrario la versión no se publica y se revierte a la anterior.

Cuando la publicación no contiene ningún error, aparece un aviso en un recuadro verde donde se indica que la versión publicada finalizó con **éxito**.

✓ Inició la publicación del recurso prueba1, versión #1.0

Conjunto de datos prueba

Estado de la Publicación

05-oct-2015 19:19:02

✓ Publishing version #1.0 of resource prueba1 finished successfully

Continúe a la [vista general del recurso](#).

Descarga completa [Reporte de publicación](#)

Mensajes del Sistema

- Archive generation started for version #1.0 19:18:42
- Start writing data file for Darwin Core Occurrence 19:18:42
- Data file written for Darwin Core Occurrence with 15 records and 19 columns 19:18:42
- Start writing data file for Darwin Core Measurement Or Facts 19:18:42
- Data file written for Darwin Core Measurement Or Facts with 15 records and 4 columns 19:18:42

Existen algunas publicaciones que poseen errores en las bases de datos, teniendo como resultado una publicación **no exitosa**. Por lo que el IPT reconoce e indica el error generado a través de un cuadro rojo. Además, muestra donde se genera el error, en este caso el campo “basisOfRecord” debe coincidir con el vocabulario controlado del estándar DwC como: *PreservedSpecimen*, *LivingSpecimen*, *FossilSpecimen*, *HumanObservation* y *MachineObservation*. Así que es necesario realizar esa modificación en la base de datos que fue cargada al IPT y luego subirla nuevamente.

✓ Inició la publicación del recurso prueba1, versión #1.0

Conjunto de datos de prueba

Estado de la Publicación

05-oct-2015 18:48:41

Falló la publicación de la versión #1.0 del recurso prueba1: Falló la generación del archivo para el recurso prueba1: Can't validate DwC-A for resource prueba1. Each row in the occurrence file(s) must have a basisOfRecord, and each basisOfRecord must match the Darwin Core Type Vocabulary (please note comparisons are case insensitive)

Continúe a la [vista general del recurso](#)

Descarga completa [Reporte de publicación](#)

- 15 line(s) have basisOfRecord that does not match the Darwin Core Type Vocabulary (please note comparisons are case insensitive) 18:44:39
- Archive validation failed, because not every row in the occurrence file(s) has a valid basisOfRecord (please note all basisOfRecord must match Darwin Core Type Vocabulary, and comparisons are case insensitive) 18:44:39

6 REFERENCIAS BIBLIOGRAFÍA

AMAYA, L. 2018. Taller CESP: Transferencia de Capacidades Técnicas Asociadas a la Publicación de Datos de Biodiversidad a través de GBIF. Presentación del Nodo GBIF Chile. Hotel Fundador Santiago de Chile. Patrocinantes: GBIF, GBIF-Chile, GBIF-Es, SIB-Colombia, Sistema Nacional de Datos Biológicos, Ministerio del Medio Ambiente (Chile) y Ministerio del Ambiente (Colombia). 41 p. [En línea] <<https://www.gbif.es/wp-content/uploads/2018/05/2.-Presentacion-del-Nodo-GBIF-Chile.pdf>>

BELTRÁN N., BUITRAGO L., DEAN V., DELGADO M. 2016. Guía básica del IPT, Herramienta de publicación de información sobre biodiversidad. [En Línea] <https://drive.google.com/file/d/1jhGH1L04fiQwY2ZJvvZU_wPhAjIF-LHy/view>

BUITRAGO, L. 2018. Taller CESP: Transferencia de Capacidades Técnicas Asociadas a la Publicación de Datos de Biodiversidad a través de GBIF. Conceptos Básicos Sobre Calidad y Curación de Datos. Hotel Fundador Santiago de Chile. Patrocinantes: GBIF, GBIF-Chile, GBIF-Es, SIB-Colombia, Sistema Nacional de Datos Biológicos, Ministerio del Medio Ambiente (Chile) y Ministerio del Ambiente (Colombia).

CERDA, C. y BIDEgain, I. 2018 Environ Monit Assess 190: 207. [En línea] <<https://doi.org/10.1007/s10661-018-6588-4>>

CEZÓN, K. 2018. Taller CESP: Transferencia de Capacidades Técnicas Asociadas a la Publicación de Datos de Biodiversidad a través de GBIF. Flujo de publicación de datos de bioiversidad. Hotel Fundador Santiago de Chile. Patrocinantes: GBIF, GBIF-Chile, GBIF-Es, SIB-Colombia, Sistema Nacional de Datos Biológicos, Ministerio del Medio Ambiente (Chile) y Ministerio del Ambiente (Colombia). 41 p. [En línea] <<https://www.gbif.es/wp-content/uploads/2018/05/4.-Flujo-geneeral-de-publicacion-de-datos-de-biodiversidad.pdf>>

COMISIÓN NACIONAL DEL MEDIO AMBIENTE (CONAMA). 2008. Biodiversidad de Chile, Patrimonio y Desafíos, Ocho Libros Editores (Santiago de Chile), 640 pp

GBIF. 2018 (a). Regional capacity enhancement to Latin America by establishing Chile's node MID-TERM ACTIVITY REPORT [En Línea] <https://assets.ctfassets.net/uo17ejk9rkwj/CXzqXAXQOcMOcSMIgs6eA/fe2f95e0f67f37f6d1ac529019701f2c/CESP2017-0007_Mid-term_report_template_final.pdf>

GBIF. 2018 (b). What is GBIF?. [En línea] <<https://www.gbif.org/what-is-gbif>>.

GBIF. 2018 (c) The GBIF Network. [en línea] <<https://www.gbif.org/the-gbif-network>>

GONZÁLEZ, A. 2006. Obtención de Datos Sobre Organismos: GBIF y Otras Bases de Datos Distribuidas. Réplica del taller: “GBIF ECOLOGICAL NICHE MODELLING” Madrid, mayo del 2006.

GUIRY, M.D. Y GUIRY, G.M. 2020. AlgaeBase. World-wide electronic publication, National University of Ireland, Galway. [En línea] <<https://www.algaebase.org>>.

INSTITUTO DE BOTÁNICA DARWINION. 2020. Catálogo de Plantas Vasculares de Flora del Cono Sur. [Consultado: 10 de junio 2020] [En línea] <<http://www.darwin.edu.ar/>>.

JAMES, S.; SOLTIS, P.; BELBIN, L.; CHAPMAN, A.; NELSON, G. y COLLINS, M. 2018. Herbarium data: Global biodiversity and societal botanical needs for novel research. *Applications in Plant Sciences* 6(2): e1024. <https://doi.org/10.1002/aps3.1024>.

MARTIN, B. 2016. Base de datos de biodiversidad de la Comunidad de Madrid y su aplicación a estudios de flora y vegetación. Tesis Doctoral. Facultad de Ciencias Biológicas, Departamento de Ecología. Universidad Complutense de Madrid.

MMA. 2017. Estrategia Nacional de Biodiversidad (2017-2030), desarrollada en el contexto del proyecto “Planificación Nacional de la Biodiversidad para apoyar la implementación del Plan Estratégico de la CBD 2011-2020”, ejecutado por el MMA de Chile y el Programa de las Naciones Unidas para el Desarrollo (PNUD), y financiado por el Fondo para el Medio Ambiente Mundial entre los años 2012 y 2015. 102 p. [En Línea] <http://portal.mma.gob.cl/wp-content/uploads/2018/03/Estrategia_Nac_Biodiv_2017_30.pdf>

NOSS, R. 1990. Indicators for Monitoring Biodiversity: A Hierarchical Approach. *Conservation Biology*, 4(4). [En Línea] <<http://www.jstor.org/stable/2385928>>. pp. 355-364.

ORTIGOSA, D. 2017. La taxonomía en la era de la biodiversidad. Universidad Nacional Autónoma de México (UNAM). Facultad de Ciencias, Unidad Multidisciplinaria de Docencia e Investigación, Sisal.

SECRETARÍA DEL CONVENIO SOBRE LA DIVERSIDAD BIOLÓGICA (SCDB). 2006. Perspectiva Mundial sobre Diversidad Biológica 2. Montreal, 81 + vii páginas [En Línea] <<https://www.cbd.int/doc/gbo/gbo2/cbd-gbo2-es.pdf>>

SECRETARÍA DEL CONVENIO SOBRE LA DIVERSIDAD BIOLÓGICA (SCDB). 2014. Perspectiva Mundial sobre la Diversidad Biológica 4. Montreal, 155 páginas [En Línea] <https://www.cbd.int/gbo/gbo4/publication/gbo4-es-hr.pdf>

SIB COLOMBIA. 2018. OpenRefine - Guía básica. Limpieza de datos sobre biodiversidad. Sistema de Información sobre Biodiversidad de Colombia, Bogotá D.C., Colombia, 22 pp.

THE PLANT LIST. 2013. Versión 1.1. Published on the Internet. [En línea] <<http://www.theplantlist.org/>>.

TROPICOS. 2020. Missouri Botanical Garden. [En línea] <<http://www.tropicos.org>>.

UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION (UNESCO). 2017. Temáticas MAB en ALC, BIODIVERSIDAD. Oficina de la UNESCO en MONTEVIDEO, Oficina Regional de Ciencia para América Latina y el Caribe [En Línea] <<http://www.unesco.org/new/es/office-in-montevideo/natural-sciences/ecological-sciences/mab-lac-themes/biodiversidad/>>.

VILLASEÑOR, JL. 2015. ¿La crisis de la biodiversidad es la crisis de la taxonomía? *Botanical Sciences*, 93(1), pp. 03-14.

WIECZOREK, J; BLOOM, D; GURALNICK, R; BLUM, S; DÖRING, M; GIOVANNI, R; ROBERTSON, T. y VIEGLAIS, D. 2012. Darwin Core: An Evolving Community-Developed Biodiversity Data Standard. *PLoS ONE* 7(1): e29715. <https://doi.org/10.1371/journal.pone.0029715>

WILLIS, C.; ELLWOOD, E.; PRIMACK, R; DAVIS, C; PEARSON, K; GALLINAT, A; YOST, J; NELSON, G. MAZER, J; ROSSINGTON, N; SPARKS, T. y SOLTIS, P. 2017. Old plants, new tricks: Phenological research using herbarium specimens. *Trends in Ecology and Evolution* 32: 531–546. DOI: <https://doi.org/10.1016/j.tree.2017.03.015>

YOST, J., SWEENEY, P., GILBERT, E., NELSON, G., GURALNICK, R., GALLINAT, A., ELLWOOD, E., ROSSINGTON, N., WILLIS, CH., BLUM, S., WALLS, R., HASTON, E., DENSLOW, M., ZOHNER, C., MORRIS, A., STUCKY, B., CARTER, J., BAXTER, D., BOLMGREN, K., DENNY, E., DEAN, E., PERSON, K., DAVIS, CH., MISHLER, B., SOLTIS, P. y MAZER, S. 2018. Digitization protocol for scoring reproductive phenology from herbarium specimens of seed plants. *Applications in Plant Sciences* 6(2): e1022.